



A User's Guide to

WINSTEPS[®]

MINISTEP

**Rasch-Model
Computer Programs**

Program Manual 4.5.2 by

John M. Linacre

winsteps.com

Copyright © 2020 by John M. Linacre. All rights reserved.

Permission to copy is granted.

ISBN 0-941938-03-4

Winsteps[®] is a registered trademark

Table of Contents

1	Control Variables Index	17
2	Control Variables by function	22
3	Input Files Index	27
4	Output Tables Index	28
5	Output Files Index	33
6	Output Graphs and Plots Index	34
7	Introduction	35
7.1	Rasch analysis and WINSTEPS.....	35
7.2	References.....	37
7.3	About the Program Manual and Software.....	38
7.4	Getting further help.....	38
8	Installation and Execution	39
8.1	What is supplied.....	39
8.2	Installation instructions for WINSTEPS.....	39
8.3	Silent Installation of Winsteps from the Command Line.....	40
8.4	Starting WINSTEPS in Windows.....	41
8.5	Using Winsteps under Windows.....	43
8.6	Uninstalling WINSTEPS.....	44
8.7	Stopping WINSTEPS.....	44
9	Menus	45
9.1	Menu bar.....	45
9.2	Batch menu.....	46
9.3	Data Setup menu.....	46
9.4	Diagnosis menu.....	46
9.5	Edit menu.....	47
9.6	Excel/RSSST menu.....	48
9.7	File menu.....	54
9.8	Graphs menu.....	55
9.9	Help menu.....	57
9.10	Output Files menu.....	59
9.11	Output Tables menu.....	59
9.12	Plots menu.....	61
9.13	Right-click menus.....	61

9.14	Specification menu.....	62
10	Control and Data Files	64
10.1	Control and Data File Setup Window.....	64
	Control and data file setup window	64
	Reading in a pre-existent control file	65
	Data display	66
	Item labels	67
	Category labels	67
10.2	Do-it-yourself control and data file construction.....	68
10.3	Control file and template.txt.....	70
10.4	Data file.....	72
10.5	Data from Excel files.....	72
10.6	Data from R Statistics files.....	73
10.7	Data from SAS files.....	73
10.8	Data from SPSS files.....	74
10.9	Data from STATA files (version 13 and earlier).....	75
10.10	Data from Text files (with Tabs, Commas, Semicolons).....	76
10.11	Data from Excel and other spreadsheets.....	76
10.12	Data file with other delimiters.....	78
11	* Control Variables	78
11.1	AGREEFILE= paired-person agreement counts.....	78
11.2	AGREEGROUP= compare with other groups in Table 35 = Yes.....	80
11.3	ALPHANUM= alphabetic numbering.....	80
11.4	ANCESTIM= alternative method for anchored analyses.....	80
11.5	ASCII= characters.....	80
11.6	ASYMPTOTE= item upper and lower asymptotes.....	82
11.7	BATCH= Batch/script mode analysis.....	83
11.8	BOXSHOW= draw boxes around Output Tables.....	88
11.9	BYITEM= display graphs for items.....	89
11.10	CATREF= reference category for Table 2.....	89
11.11	CFILE= scored category label file.....	90
11.12	CHART= graphical plots in Tables 10, 13-15.....	91
11.13	CHISQUARE= in IFILE= and PFILE=.....	92
11.14	CLFILE= codes label file.....	92
11.15	CMATRIX= category matrix = Yes.....	94
11.16	CODERANGE = numerical range of data codes.....	95
11.17	CODES= valid data codes.....	96
11.18	CONVERGE= select convergence criteria.....	98
11.19	CSV= comma-separated values in output files.....	100
11.20	CURVES= probability curves for Table 21.....	100
11.21	CUTHI= cut off responses with high expectations.....	100

11.22	CUTLO= cut off responses with low expectations.....	101
11.23	DATA= name of data file.....	102
11.24	DATESHOW= show date in Table headings.....	103
11.25	DELIMITER= or SEPARATOR= data field delimiters.....	103
11.26	DIF= columns within person label for Table 30.....	105
11.27	DISCRIMINATION= item discrimination.....	107
11.28	DISFILE= category/distractor/option count file.....	108
11.29	DISOPTION= distractor file option = C.....	109
11.30	DISTRACTOR= output option counts in Tables 10, 13-15.....	109
11.31	DPF= columns within item label for Table 31.....	110
11.32	EDFILE= edit data file.....	110
11.33	EFILE= expected scores on items.....	113
11.34	END LABELS or END NAMES.....	114
11.35	EQFILE= code equivalences.....	114
11.36	ETYPE= option for EFILE=.....	115
11.37	EXCAT= extreme category adjustment.....	115
11.38	EXCELNUMBER= Excel converts numeric labels into numbers.....	115
11.39	EXTRSC= extreme score correction for extreme measures.....	116
11.40	FITHIGH= higher bar in charts.....	117
11.41	FITI= item misfit criterion.....	117
11.42	FITLOW= lower bar in charts.....	117
11.43	FITP= person misfit criterion.....	118
11.44	FORMAT= reformat data.....	118
11.45	FORMFD= the form feed character.....	124
11.46	FRANGE= half-range of fit statistics on plots.....	124
11.47	FREQUENT= heading in Tables 1, 12, 16.....	124
11.48	FSHOW= show files created from control file.....	124
11.49	GRFILE= probability curve coordinate output file.....	124
11.50	GROUPS= or ISGROUPS= assigns items to rating scale groupings.....	126
11.51	GRPFROM= location of ISGROUPS.....	126
11.52	GUFIL= (G0ZONE=, G1ZONE=) Guttmanized response file.....	126
11.53	HEADER= display or suppress subtable headings.....	127
11.54	HIADJ= correction for top rating scale categories.....	127
11.55	HLINES= heading lines in output files.....	128
11.56	IAFILE= item anchor file.....	128
11.57	IANCHQU= anchor items interactively.....	131
11.58	ICORFILE= item residual correlation file.....	132
11.59	ICOLORFILE= colors for item labels in PKMAPs.....	133
11.60	IDDELETE= item one-line item deletion.....	134
11.61	IDELQU= delete items interactively.....	135
11.62	IDFILE= item deletion file.....	136

11.63	IDROPEXTREME= drop items with extreme scores.....	138
11.64	IFILE= item output file.....	139
11.65	ILFILE= item label file.....	142
11.66	IMAP= item label on item maps Tables 1, 12.....	143
11.67	INUMB= label items by sequence numbers.....	143
11.68	IOFSFIELDS= field selection for IFILE=.....	144
11.69	IPEXTREME= placement of extreme scores on maps.....	145
11.70	IPMATRIX= response-level matrix.....	146
11.71	IPRFILE= change codes for blocks of persons and items.....	147
11.72	IREFER= identifying items for recoding.....	149
11.73	ISELECT= item selection criterion.....	151
11.74	ISFILE= item structure output file.....	153
11.75	ISGROUPS= or GROUPS= assigns items to rating scale groupings.....	157
11.76	ISORT= column within item name for alphabetical sort in Table 15.....	161
11.77	ISRANGE= category range of a rating scale.....	162
11.78	ISUBTOTAL= columns within item label for Table 27.....	162
11.79	ITEM= title for item labels.....	163
11.80	ITEM1= column number of first response.....	163
11.81	ITLEN= maximum length of item label.....	164
11.82	IVALUEX= recoding of data.....	165
11.83	IWEIGHT= item (variable) weighting.....	166
11.84	KEYFROM= location of KEYn.....	168
11.85	KEYn= scoring key.....	169
11.86	KEYSCR= reassign scoring keys.....	170
11.87	LCONV= logit change at convergence.....	172
11.88	LESS= heading in Tables 1, 12, 16.....	172
11.89	LINELength= length of printed lines.....	172
11.90	LOCAL= locally restandardize fit statistics.....	172
11.91	LOGFILE= accumulates control files.....	174
11.92	LOWADJ= correction for bottom rating scale categories.....	174
11.93	MAKEKEY= construct MCQ key.....	174
11.94	MATRIX= correlation output format.....	174
11.95	MAXPAGE= the maximum number of lines per page.....	175
11.96	MFORMS= reformat input data.....	175
11.97	MHSLICE= Mantel-Haenszel slice width.....	181
11.98	MHZERO= adjustment to zero cells in Mantel cross-tabulation.....	181
11.99	MISSCORE= scoring of missing data codes.....	182
11.100	MJMLE= or MUCON= maximum number of JMLE iterations.....	184
11.101	MNSQ= show mean-square or standardized fit statistics.....	184
11.102	MODELS= assigns model types to items.....	185
11.103	MODFROM= location of MODELS.....	186

11.104	MORE= heading in Tables 1, 12, 16.....	187
11.105	MPROX= maximum number of PROX iterations.....	189
11.106	NAME1= first column of person label.....	189
11.107	MRANGE= half-range of measures on plots.....	190
11.108	MTICK= measure distance between tick marks on plots = 0, auto-size.....	190
11.109	MUCON= or MJMLE= maximum number of JMLE iterations.....	190
11.110	NAMLEN= length of person label.....	190
11.111	NAMLMP= name length on map for Tables 1, 12, 16.....	191
11.112	NEWSCORE= recoding values.....	191
11.113	NI= number of items.....	193
11.114	NORMAL= normal distribution for standardizing fit.....	193
11.115	OFILE= report output file.....	193
11.116	OSORT= option/distractor sort.....	194
11.117	OUTFIT= sort misfits on infit or outfit.....	194
11.118	PAFILE= person anchor file.....	195
11.119	PAIRED= correction for paired comparison data.....	197
11.120	PANCHQU= anchor persons interactively.....	198
11.121	PCORFIL= person residual correlation file.....	198
11.122	PDELETE= person one-line item deletion.....	199
11.123	PDELQU= delete persons interactively.....	199
11.124	PDFILE= person deletion file.....	200
11.125	PDROPEXTREME= drop persons with extreme scores.....	201
11.126	PERSON= title for person labels.....	201
11.127	PFILE= person output file.....	201
11.128	PKMAP= customize Table 36 diagnostic maps.....	204
11.129	PLFILE= person label file.....	211
11.130	PMAP= person label on person map: Tables 1, 16.....	211
11.131	POFSFIELDS= field selection for PFILE=.....	212
11.132	PRCOMP= residual type for principal components analyses in Tables 23, 24.....	212
11.133	PSELECT= person selection criterion.....	213
11.134	PSORT= column within person label for alphabetical sort in Table 19.....	215
11.135	PSUBTOTAL= columns within person label for Tables 28, 35.....	215
11.136	PTBISERIAL= point-biserial-type correlation coefficient.....	217
11.137	PVALUE= proportion correct or average rating.....	219
11.138	PWEIGHT= person (case) weighting.....	219
11.139	QCMLE = Quasi-CMLE estimates.....	220
11.140	QUOTED= quote-marks around labels.....	220
11.141	RARE= heading in Tables 1, 12, 16.....	221
11.142	RCONV= score residual at convergence.....	221
11.143	REALSE= inflate S.E. for misfit.....	221
11.144	RESCORE= response recoding.....	222

11.145	RESFROM= location of RESCORE.....	223
11.146	RFILE= scored response file.....	224
11.147	RMSR= report root-mean-square residuals in measure tables.....	224
11.148	ROW1HEADING = heading in first row of output file.....	225
11.149	SAFILE= structure-threshold input anchor file.....	225
11.150	SAITEM= item numbers in SAFILE with one grouping.....	235
11.151	SANCHQ= anchor category structure interactively.....	235
11.152	SCOREFILE= person score file.....	236
11.153	SCREEN= screen log file.....	238
11.154	SDELQU= delete category structure interactively.....	238
11.155	SDFILE= category structure deletion file.....	238
11.156	SEPARATOR= or DELIMITER= data field delimiters.....	238
11.157	SFILE= category structure-threshold output file.....	238
11.158	SFUNCTION= function to model Andrich thresholds with polynomials.....	239
11.159	SICOMPLETE= simulate complete data.....	243
11.160	SIEXTREME= simulate extreme scores.....	244
11.161	SIFILE= simulated data file.....	244
11.162	SIMEASURE= measure or data.....	244
11.163	SINUMBER= number of simulated data files.....	244
11.164	SIRESAMPLE= number of persons resampled.....	245
11.165	SISEED= simulated data seed.....	245
11.166	SPFILE= supplementary control file.....	245
11.167	STBIAS= correct for estimation bias.....	246
11.168	STEPT3= include category structure summary in Table 3.2 or 21.....	248
11.169	STKEEP= keep unobserved intermediate categories in structure.....	248
11.170	SUBSETS= perform subset detection.....	250
11.171	SVDEPOCHS= maximum epochs for SVD estimation.....	251
11.172	SVDFACTORS= singular-value decomposition factors.....	251
11.173	SVDFILE= singular-value decomposition file.....	251
11.174	SVDMIN= singular-value decomposition minimum improvement.....	252
11.175	SVDTYPE= singular-value decomposition residual type.....	252
11.176	T1CAT= show category number.....	252
11.177	T1I#= number of items summarized by "#" symbol in Table 1.....	252
11.178	T1P#= number of persons summarized by "#" symbol in Table 1.....	253
11.179	T1WEIGHT= show by weight in Table 1.....	253
11.180	T2SELECT= selects Table 2 subtables.....	253
11.181	T45OPTIONS= field choices for Table 45.....	253
11.182	T7OPTIONS = selects Table 7.1 and 11.1 detail lines.....	254
11.183	TABLES= output tables.....	254
11.184	TARGET= estimate using information-weighting.....	255
11.185	TCCFILE= Test Characteristic Curve and Test Information Function.....	256

11.186	TFILE= input file listing tables to be output.....	258
11.187	TITLE= title for output listing.....	261
11.188	TOTALSCORE= show total scores with extreme observations.....	261
11.189	TRPROFILE= transposing the data matrix.....	261
11.190	TRPOTYPE= transposing specifications.....	262
11.191	UAMOVE= move all anchor values a pre-set amount.....	262
11.192	UANCHOR= anchor values supplied in user-scaled units.....	263
11.193	UASCALE= anchor user-scale value of 1 logit.....	263
11.194	UCOUNT= number of unexpected responses: Tables 6, 10.....	264
11.195	UDECIMALS= number of decimal places reported.....	264
11.196	UEXTREME= include extreme scores for UIMEAN= or UPMEAN=.....	264
11.197	UIMEAN=or UMEAN= the mean or center of the item difficulties.....	265
11.198	UMEAN=or UIMEAN= the mean or center of the item difficulties.....	266
11.199	UPMEAN= the mean or center of the person abilities.....	266
11.200	USCALE= the user-scaled value of 1 logit.....	267
11.201	UTF8ENC= UTF-8 encoding.....	268
11.202	UTF8SUB= substitute character for multibyte UTF-8.....	268
11.203	W300= Output files in Winsteps 3.00 format.....	269
11.204	WEBFONT= font of webpages.....	269
11.205	WHEXACT= Wilson-Hilferty exact normalization.....	269
11.206	WMLE = Warm's Mean Likelihood Estimation.....	270
11.207	XFILE= analyzed response file.....	270
11.208	XWIDE= columns per response.....	272
11.209	#1#= user-defined replaceable token.....	273
11.210	&END end of control variables.....	273
11.211	&INST start of control instructions.....	273
11.212	; starts a comment.....	273
11.213	@Field= name for location in label.....	274
12	Examples of Control and Data Files	274
12.1	Example 0: Rating scale data: The Liking for Science data.....	274
12.2	Example 1: Dichotomous data: Simple control file with data included.....	277
12.3	Example 2: Control and anchor files.....	278
12.4	Example 3: Item recoding and item deletion.....	278
12.5	Example 4: Selective item recoding.....	279
12.6	Example 5: Scoring key for items, also CAT responses.....	280
12.7	Example 6: Keys in data record FORMAT.....	280
12.8	Example 7: A partial credit analysis.....	282
12.9	Example 8: Items with various rating scale models.....	282
12.10	Example 9: Grouping and modeling items.....	283
12.11	Example 10: Combining tests with common items.....	284

12.12	Example 11: Item responses two characters wide.....	287
12.13	Example 12: Comparing high and low samples with rating scales.....	287
12.14	Example 13: Paired comparisons as the basis for measurement.....	290
12.15	Example 14: Multiple rescorings, response structures and widths.....	291
12.16	Example 15: Figure skating: Multidimensionality, DIF or Bias.....	295
12.17	Example 16: Disordered categories - thresholds and person anchoring.....	297
12.18	Example 17: Rack, stack, principal components.....	300
12.19	Example of subset reporting.....	302
12.20	Examples on www.winsteps.com	307
13	Analysis Window	307
13.1	The Analysis Window.....	307
14	Output Tables	309
14.1	Table 0 Analysis Report.....	309
	Table 0.1 Analysis identification	309
	Table 0.2 Convergence report	310
	Table 0.3 Control file	311
	Table 0.4 Subset details	312
14.2	Table 1 Wright item-person maps of the latent variable	312
14.3	Table 2 Category plots.....	326
	Table 2 Multiple-choice distractor plot	326
	Table 2 Polytomies: Most probable, expected, cumulative, structure, average measures	328
	Table 2.1 Modal observed categories (most probable categories)	331
	Table 2.2 Mean observed categories (average or expected categories: Rasch-half-point thresholds)	333
	Table 2.3 Median observed categories (cumulative category probabilities: Rasch-Thurstone thresholds)	333
	Table 2.4 Andrich thresholds (structure calibrations)	336
	Table 2.5 Observed average measures of persons by scored category (empirical averages)	337
	Table 2.6 Observed average measures of persons (empirical averages)	338
	Table 2.7 Expected average measures of persons	340
	Table 2.11 Modal categories by category score (most probable)	341
	Table 2.12 Mean categories by category score (average or expected: Rasch-half-point thresholds)	342
	Table 2.13 Median categories by category score (cumulative probabilities: Rasch-Thurstone thresholds)	342
	Table 2.14 Andrich thresholds (structure calibrations)	345
	Table 2.15 Observed average measures of persons for scored categories (empirical averages)	346
	Table 2.16 Observed average measures of persons (empirical averages)	347
	Table 2.17 Expected average measures of persons	348
14.4	Table 3.1 Summaries of persons and items.....	349
14.5	Table 3.2+ Summary of dichotomous, rating scale or partial credit structures.....	353
14.6	Table 4.1 Person infit fit plot.....	362
14.7	Table 5.1 Person outfit fit plot.....	362
14.8	Table 5.2 Person infit-outfit fit plot.....	363
14.9	Table 6 Person report in misfit order.....	364
	Table 6.1 Person statistics in misfit order	364
	Table 6.2 Person statistics graphically in misfit order	364
	Table 6.4 Person most-misfitting response strings	364
	Table 6.5 Person most-unexpected observations	365
	Table 6.6 Person most-unexpected response list	365

14.10	Table 7.1 Person misfitting responses.....	366
14.11	Table 8.1 Item infit fit plots.....	366
14.12	Table 9.1 Item outfit fit plot.....	367
14.13	Table 9.2 Item infit-outfit fit plot.....	368
14.14	Table 10 Item report in misfit order.....	368
	Table 10.1 Item statistics in misfit order	368
	Table 10.2 Item statistics graphically in misfit order	372
	Table 10.3 Item option & distractor frequencies in misfit order	373
	Table 10.4 Item most-misfitting response strings	374
	Table 10.5 Item most-unexpected observations	375
	Table 10.6 Item most-unexpected response list	375
14.15	Table 11.1 Item misfitting responses.....	376
14.16	Table 12 Item maps with full item labels.....	377
14.17	Table 13 Item report in measure order.....	383
	Table 13.1 Item statistics in measure order	383
	Table 13.2 Item statistics graphically in measure order	383
	Table 13.3 Item option & distractor frequencies in measure order	383
14.18	Table 14 Item report in entry order.....	384
	Table 14.1 Item statistics in entry order	384
	Table 14.2 Item statistics graphically in entry order	385
	Table 14.3 Item option & distractor frequencies in entry order	385
14.19	Table 15 Item report in label order.....	386
	Table 15.1 Item statistics in label order	386
	Table 15.2 Item statistics graphically in label order	386
	Table 15.3 Item option & distractor frequencies in label order	386
14.20	Table 16 Person maps with full person labels.....	388
14.21	Table 17.1 Person statistics in measure order.....	390
14.22	Table 17.2 Person statistics graphically in measure order.....	390
14.23	Table 18.1 Person statistics in entry order.....	390
14.24	Table 18.2 Person statistics graphically in entry order.....	390
14.25	Table 19.1 Person statistics in label order.....	390
14.26	Table 19.2 Person statistics graphically in label order.....	390
14.27	Table 20.1 Complete score-to-measure table on test of all items.....	391
14.28	Table 20.2 Person score and measure distribution.....	394
14.29	Table 20.3 Complete score-to-calibration table for tests based on whole sample.....	396
14.30	Table 21 Category Probability curves and Expected score ogive.....	397
14.31	Table 22.1, 22.2, 22.3 Sorted response matrix (Guttman scalogram).....	399
14.32	Table 23 Item multidimensionality.....	402
	Table 23.0 Variance components scree plot for items	402
	Table 23.1, 23.11 Principal components plots of item loadings	410
	Table 23.2, 23.12 Item Principal components analysis/contrast of residuals	413
	Table 23.3, 23.13 Item contrast by persons	414
	Table 23.4, 23.14 Item contrast loadings sorted by measure	415
	Table 23.5, 23.15 Item contrast loadings sorted by entry number	416
	Table 23.6, 23.16 Person measures for item clusters in contrast	416
	Table 23.99 Largest residual correlations for items	417
14.33	Table 24 Person multidimensionality.....	420
	Table 24.0 Variance components scree plot for persons	420

	Table 24.1, 24.11 Principal components plots of person loadings	421
	Table 24.2, 24.12 Person Principal components analysis/contrast of residuals	422
	Table 24.3, 24.13 Person contrast by items	423
	Table 24.4, 24.14 Person contrast loadings sorted by measure	425
	Table 24.5, 24.15 Person contrast loadings sorted by entry number	425
	Table 24.99 Largest residual correlations for persons	426
14.34	Table 25 Item report in displacement order.....	426
	Table 25.1 Item statistics in displacement order	426
	Table 25.2 Item statistics graphically	426
	Table 25.3 Item option & distractor frequencies	426
14.35	Table 26 Item report in correlation order.....	428
	Table 26.1 Item statistics in correlation order	428
	Table 26.2 Item statistics graphically	428
	Table 26.3 Item option & distractor frequencies	428
14.36	Table 27 Item subtotals.....	430
	Table 27.1 Item subtotal summaries on one line	430
	Table 27.2 Item subtotal measure bar charts	432
	Table 27.3+ Item subtotal detailed summary statistics	433
14.37	Table 28 Person subtotals.....	433
	Table 28.1 Person subtotal summaries on one line	433
	Table 28.2 Person subtotal measure bar charts	435
	Table 28.3+ Person subtotal detailed summary statistics	436
14.38	Table 29 Empirical ICCs and option frequencies.....	437
14.39	Table 30 DIF Differential Item Functioning.....	438
	Table 30.1 Differential item functioning DIF pairwise	438
	Table 30.2, 30.3 Differential item functioning DIF list	441
	Table 30.4 Differential item functioning DIF fit summary	443
	Table 30.5, 30.6 Within-class fit report	444
	Table 30.7 Item measure profiles for classes of persons	445
14.40	Table 31 DPF Differential Person Functioning.....	446
	Table 31.1 Differential person functioning DPF pairwise	446
	Table 31.2, 31.3 Differential person functioning DPF list	447
	Table 31.4 Differential person functioning DPF fit summary	448
	Table 31.5, 31.6 Within-class fit report	449
	Table 31.7 Person measure profiles for classes of items	449
14.41	Table 32 Control specifications.....	451
14.42	Table 33 DGF Differential Group Functioning.....	451
	Table 33.1, 33.2 Differential group functioning DGF pairwise	451
	Table 33.3, 33.4 Differential group functioning DGF list	453
	Table 33.7, 33.8 Item group-Person group profiles	454
14.43	Table 34.1 Columnar statistical comparison and scatterplot.....	454
14.44	Table 35 Paired-person (cheating) agreement.....	455
14.45	Table 36 PKMAP person diagnostic maps.....	457
14.46	Tables 37-41 Person Keyforms.....	461
	Table 37 Person KeyForms - Measure order	461
	Table 38 Person KeyForms - Entry order	463
	Table 39 Person KeyForms - Alphabetical order	463
	Table 40 Person KeyForms - Misfit order	463
	Table 41 Person KeyForms - Misfit order, only unexpected responses	463
14.47	Table 42.1 Person statistics in displacement order.....	463
14.48	Table 42.2 Person statistics graphically in displacement order.....	463

14.49	Table 43.1 Person statistics in correlation order.....	463
14.50	Table 43.2 Person statistics graphically in correlation order_2.....	463
14.51	Table 44.1 Global statistics.....	463
14.52	Table 45 Person measures after each item.....	465
14.53	Table heading.....	468
15	Charts	469
15.1	Person-item histograms or barcharts.....	469
16	Graphs	473
16.1	Graphs window.....	473
16.2	Graphs window - bottom right.....	476
16.3	Category information function.....	478
16.4	Category probability curves (Rasch-Andrich thresholds).....	478
16.5	Conditional probability curves.....	480
16.6	Cumulative probabilities.....	481
16.7	Cumulative thresholds (Rasch-Thurstone thresholds).....	482
16.8	Empirical category curves.....	483
16.9	Empirical ICC, IRF.....	483
16.10	Empirical option curves.....	484
16.11	Empirical randomness.....	485
16.12	Expected score ICC, IRF (Rasch-half-point thresholds).....	485
16.13	Item information function.....	486
16.14	Model and empirical ICCs.....	487
16.15	Multiple item ICCs.....	487
16.16	Non-uniform DIF ICCs.....	490
16.17	Points and lines.....	492
16.18	Category probability and empirical curves.....	493
16.19	Test Characteristic Curve TCC.....	493
16.20	Test Information Function TIF.....	494
16.21	Test Empirical Randomness.....	495
17	Plots	495
17.1	Bubble charts: Developmental pathways.....	495
17.2	Cluster Measure Plot - Table 23.6.....	499
17.3	Datapoint labels.....	501
17.4	DGF DIF-DPF Plot - Table 33.3.....	502
17.5	DIF Plot - Table 30.2.....	502
17.6	DIF/DPF/DTF Scatterplots of paired measures - Table.....	504
17.7	DPF Plot - Table 31.2.....	505
17.8	Incremental person measures - Table 45.....	506
17.9	Keyform plot.....	506

17.10	Scatterplot: Compare statistics.....	508
18	Output dialogs	514
18.1	AGREEFILE and Table 35 Specifications.....	514
18.2	Compare statistics.....	516
18.3	ConstructMap (GradeMap) interface.....	516
18.4	DGF specifications.....	517
18.5	DIF specifications.....	518
18.6	DPF specifications.....	519
18.7	Edit Initial Settings (Winsteps.ini) file.....	519
18.8	Facets interface: this requires Facets software, available separately.....	522
18.9	Output file formats.....	525
18.10	Output file specifications and Field selection.....	526
18.11	Simulated file specifications.....	528
18.12	SVD specifications.....	533
18.13	Table/Plot 45 specifications.....	533
18.14	Transpose file specifications.....	533
18.15	XFILE= field selection.....	535
19	Special Topics	536
19.1	Advice to novice analysts.....	536
19.2	Anchored estimation.....	538
19.3	Andrich Thresholds and disordered rating categories.....	539
19.4	Arithmetical expressions.....	543
19.5	Automating file selection.....	544
19.6	Average measures, distractors and rating scales.....	544
19.7	Batch mode example: Score Tables.....	545
19.8	Bayesian prior distributions.....	546
19.9	Between-group and Within-group fit statistics.....	548
19.10	Biserial correlation.....	548
19.11	Bonferroni - Multiple t-tests.....	549
19.12	Bradley-Terry model and paired comparisons.....	551
19.13	Category boundaries and thresholds.....	551
19.14	Category mean-square fit statistics.....	552
19.15	Chi-squares, d.f., probabilities and Mean-squares.....	553
19.16	CMLE, MMLE, PMLE: Using estimates from other software in Winsteps.....	554
19.17	Column and classification selection and transformation.....	557
19.18	Combining items into superitems and testlets.....	558
19.19	Comparing estimates with other Rasch software.....	558
19.20	Comparing tests or instruments.....	559
19.21	Computer-Adaptive Testing (CAT).....	559
19.22	Convergence considerations.....	559

19.23	Correlations: point-biserial, point-measure, residual.....	561
19.24	Decimal, percentage and continuous data.....	567
19.25	Demographic codes - changing.....	569
19.26	Dependency and unidimensionality - Repeated measures - Longitudinal studies.....	569
19.27	Dichotomous mean-square fit statistics.....	572
19.28	DIF - DPF - bias - interactions concepts.....	573
19.29	DIF for two person groups.....	580
19.30	Dimensionality.....	580
19.31	Dimensionality investigation - an example.....	580
19.32	Dimensionality: contrasts & variances.....	583
19.33	Dimensionality: when is a test multidimensional?.....	592
19.34	Disjoint strings of responses.....	594
19.35	Displacement measures.....	594
19.36	Edit taskbar caption.....	596
19.37	Equating and linking tests.....	597
19.38	Estimation bias correction - warnings.....	604
19.39	Estimation methods: JMLE, PROX, WMLE, CMLE.....	608
19.40	Exact Match: OBS% and EXP%.....	612
19.41	Excel plotting.....	612
19.42	Excel-formatted output Tables.....	613
19.43	Extra Specifications prompt.....	614
19.44	Extreme scores: what happens.....	615
19.45	Fit diagnosis: infit outfit mean-square standardized.....	615
19.46	Global fit statistics.....	622
19.47	Glossary.....	623
19.48	Guttman parameterization of rating scales.....	630
19.49	Half-rounding.....	631
19.50	How big an analysis can I do?.....	632
19.51	How long will an analysis take?.....	632
19.52	Inestimable measures.....	632
19.53	Information - item and test.....	632
19.54	Installing multiple versions of Winsteps.....	633
19.55	IRT to Rasch conversion.....	633
19.56	Item difficulty: definition.....	634
19.57	Item difficulties from score-to-measure table.....	634
19.58	Item discrimination or slope estimation.....	636
19.59	Item map in Excel or Word.....	638
19.60	Item Maps - Combining.....	638
19.61	Iterations - PROX & JMLE.....	639
19.62	JMLE is inconsistent!.....	642
19.63	Local Dependence.....	642

19.64	Logit and probit.....	643
19.65	Mantel and Mantel-Haenszel DIF statistics.....	645
19.66	Misfit diagnosis: infit outfit mean-square standardized.....	648
19.67	Missing data.....	654
19.68	Mixed-Mixture Models and Saltus models.....	656
19.69	Müller's continuous rating scale model CRSM.....	657
19.70	Multidimensional Model - Multidimensionality.....	657
19.71	Multiple copy-and-paste or other repetitive actions.....	658
19.72	Multiple Response items - MR questions.....	658
19.73	Non-uniform DIF tables.....	658
19.74	Null or unobserved categories: structural and incidental zeroes.....	661
19.75	One item test - One observation per respondent.....	662
19.76	Order of elements in Control file.....	664
19.77	Partial Credit model.....	664
19.78	Person Response Functions (PRF).....	665
19.79	Person-free and Item-free measurement.....	665
19.80	Pivot anchoring.....	666
19.81	Plausible values.....	668
19.82	Poisson counts.....	668
19.83	Polytomous mean-square fit statistics.....	670
19.84	Printing pretty tables and graphs.....	671
19.85	Prior distributions.....	671
19.86	Probabilities from measures.....	671
19.87	Quality-control misfit selection criteria.....	672
19.88	R Statistics.....	673
19.89	Random number generator.....	674
19.90	Rank order data.....	675
19.91	Rasch - Why use Rasch methodology?.....	675
19.92	Rasch Model.....	676
19.93	Rasch Software.....	676
19.94	Rating scale conceptualization: Andrich, Thurstone, half-point thresholds.....	676
19.95	Ratio scale of measures: twice as much.....	678
19.96	Rectangular copying.....	678
19.97	Reinstalling Winsteps.....	680
19.98	Reliability and separation of measures.....	681
19.99	Response-pattern frequency.....	685
19.100	Rules for assigning values to control variables (key-words).....	686
19.101	Scoring items with multiple responses.....	687
19.102	Shortcut Keys.....	687
19.103	Simulating data.....	687
19.104	Smartphone - running Winsteps on your PC.....	687

19.105	Specifying how data are to be recoded.....	688
19.106	Speeded tests.....	688
19.107	Split items - procedure.....	688
19.108	Standard errors: model and real.....	691
19.109	Standard setting.....	693
19.110	Starting Winsteps from the DOS prompt.....	693
19.111	Subsets and connection ambiguities.....	693
19.112	Subtest scoring and weighting.....	699
19.113	TAM R Statistics and Winsteps.....	700
19.114	Teaching Rasch measurement, Rasch analysis_2.....	700
19.115	Testlets.....	701
19.116	Theta or B? parameterizing ability.....	706
19.117	t-statistics.....	706
19.118	Unidimensionality.....	707
19.119	Unobserved and dropped categories.....	707
19.120	User-friendly rescaling: zero point and unit.....	708
19.121	Using a word processor or text editor.....	711
19.122	UTF-8 Chinese character set.....	712
19.123	Validity investigation.....	713
19.124	Videos and Tutorials.....	714
19.125	Weighting items and persons.....	714
19.126	Winsteps: history and steps.....	716
19.127	Wright maps in Word, Excel, R, Notepad++.....	716
Index of Specifications and Files		730

1 Control Variables Index

Before your analysis, these control variables can be set in your [Control file](#) and at the [Extra specifications](#) prompt. After your analysis, these control variables can be set in the [Specification-menu dialog box](#) and in other [pull-down menus](#). The current settings of these variables can be seen from the [Output Files menu](#). Click on [Control variables file](#).

AGREEFILE=	paired-person agreement file
AGREEGROUP=	compare with other groups in Table 35
ALPHANUM=	alphabetic numbering to exceed 9 with XWIDE=1
ANCESTIM=	Special alternative estimation method for anchored analyses
ASCII=	output only ASCII characters
ASYMPTOTE=	report estimates of upper and lower asymptotes
BATCH=	set batch mode
BOXSHOW=	draw boxes around Output Tables
BYITEM=	show empirical curves for items
CATREF=	reference category: Table 2
CFILE=	scored category label file
CHART=	graphical plots of measures
CHISQUARE=	output Chi-square instead of Mean-square in IFILE= and PFILE=
CLFILE=	codes label file
CMATRIX=	category matrix in Table 3.2, etc.
CODERANGE=	numerical range of data codes
CODES=	valid & missing data codes
CONVERGE=	select convergence criteria
CSV=	comma-separated values in output files
CURVES=	probability curve selection: Tables 21, 2
CUTHI=	cut off responses with high expectations
CUTLO=	cut off responses with low expectations
DATA=	name of data file
DATESHOW=	display analysis date and Winsteps version number in Table headings
DELIMITER=	delimiter of input response data fields
DIF=	person label columns for DIF
DISCRIM=	display item discrimination estimates
DISFILE=	category/distractor/response option count file
DISOPTION=	distractor file reporting option
DISTRACTOR=	output category/distractor/option counts
DPF=	item label columns for DPF
EDFILE=	edit data file
EFILE=	expected scores on the items
EQFILE=	code equivalences
EXCAT=	extreme-category adjustment
EXCELNUMBER =	Excel converts numeric labels into numbers = No
EXTRSCORE=	extreme score adjustment
FITHIGH=	lower bar in charts
FITI=	item misfit criterion

FITLOW=	lower bar in charts
FITP=	person misfit criterion
FORMAT=	reformat data
FORMFD=	form-feed character
FRANGE=	half-range of fit statistics on plots
FREQUENT=	column heading in Tables 1, 12, 16
FSHOW=	show files created from control file
GOZONE=	percent of 0's within 0-zone among which all 1's are turned to 0's
G1ZONE=	percent of 1's within 1-zone among which all 0's are turned to 1's
GRFILE=	graphing file for probability curves
GROUPS=	assigns items to rating scale or partial credit groupings (same as ISGROUPS=)
GRPFROM=	location of ISGROUPS=
GUFIL=	Guttmanized file of responses
HEADER=	display or suppress Sub-Table Headings after the first
HIADJ=	correction for top categories
HLINES=	heading lines in output files
IAFILE=	item anchor file
IANCHQU=	anchor items interactively
ICOLORFILE=	colors for item labels in PKMAPs
ICORFILE=	inter-item residual correlations
IDELETE=	one-line item deletion list
IDELQU=	delete items interactively
IDFILE=	item deletion file
IDROPEXTR=	remove items with extreme scores
IFILE=	item output file
ILFILE=	item label file
IMAP=	item label for item maps
INUMB=	label items by sequence number
IOFSFIELDS=	field selection for IFILE=
IPEXTREME=	placement of extreme scores on maps
IPMATRIX=	response matrix (Output Files menu only)
IPRFILE=	change codes for blocks of persons and items
IREFER=	identifying items for recoding
ISELECT=	item selection criterion
ISFILE=	item-structure output file
ISGROUPS=	assigns items to rating scale or partial credit groupings (same as GROUPS=)
ISORT=	sort column in item label
ISRANGE=	category range of an item group
ISUBTOTAL=	subtotal items by specified columns
ITEM=	title for item labels
ITEM1=	column number of first response
ITLEN=	maximum length of item label
IVALUEx=	recoding for items
IWEIGHT=	item (variable) weighting

KEYFORM=	keyform output file (backwards compatibility only)
KEYFROM=	location of KEYn=
KEYn=	scoring key
KEYSCR=	reassign scoring keys
LCONV=	logit change at convergence
LESS=	column heading in Tables 1, 12, 16
LINELENGTH=	length of printed lines
LOCAL=	locally restandardize fit statistics
LOGFILE=	accumulates control files
LOWADJ=	correction for bottom categories
MAKEKEY=	construct an MCQ scoring key
MATRIX=	correlation output format
MAXPAGE=	maximum number of lines per page
MFORMS=	reformatting input records & multiple forms
MHSlice=	size of Mantel or Mantel-Haenszel slice
MHZERO=	adjustment to zero cells in Mantel cross-tabulation
MISSCORE=	scored value of missing data: not your missing-data code
MJMLE=	maximum number of JMLE iterations
MNSQ=	show mean-square or t standardized fit
MODELS=	assigns model types to items
MODFROM=	location of MODELS=
MORE=	column heading in Tables 1, 12, 16
MPROX=	maximum number of PROX iterations
MRANGE=	half-range of measures on plots in Tables
MTICK=	measure distance between tick-marks on plots in Tables
MUCON=	maximum number of JMLE iterations
NAME1=	first column of person label
NAMLEN=	length of person label
NAMLMP=	label length on maps
NEWSCORE=	recoded data values
NI=	number of items
NORMAL=	normal distribution for standardizing fit
OFILE=	report output file
OSORT=	category/option/distractor sorted by score, count or measure
OUTFIT=	sort misfits on outfit or infit
PAFILE=	person anchor file
PAIRED=	correction for paired comparison data
PANCHQU=	anchor persons interactively
PCORFIL=	inter-person residual correlations
PDELETE=	one-line person deletion list
PDELQU=	delete persons interactively
PDFILE=	person deletion file
PDROPEXTR=	remove persons with extreme scores
PERSON=	title for person labels

<u>PFILE=</u>	person output file
<u>PIVOT=</u>	structure anchoring (backwards compatibility only)
<u>PKMAP=</u>	customize Table 36 diagnostic maps
<u>PLFILE=</u>	person label file
<u>PMAP=</u>	person label for person maps
<u>POFSFIELDS=</u>	field selection for PFILE=
<u>PRCOMP=</u>	residual type for principal components/contrast analysis
<u>PSELECT=</u>	person selection criterion
<u>PSORT=</u>	sort column in person label
<u>PSUBTOTAL=</u>	subtotal persons by specified columns
<u>PTBISERIAL=</u>	report point-biserial correlations
<u>PVALUE=</u>	report item proportion-correct-values or average ratings
<u>PWEIGHT=</u>	person (case) weighting
<u>QCMLE=</u>	Quasi-Conditional Maximum Likelihood Estimates
<u>QUOTED=</u>	quote-marks around labels
<u>RARE=</u>	column heading in Tables 1, 12, 16
<u>RCONV=</u>	score residual at convergence
<u>REALSE=</u>	inflate S.E. of measures for misfit
<u>RESCORE=</u>	controls response recoding
<u>RESFRM=</u>	location of RESCORE=
<u>RFILE=</u>	scored response output file
<u>RMSR=</u>	report root-mean-square residual in measure tables
<u>ROW1HEADING=</u>	heading in first row of output file
<u>SAFILE=</u>	structure anchor file
<u>SAITEM=</u>	multiple ISGROUPS= format in SFILE= & SAFILE=
<u>SANCHQU=</u>	anchor structure interactively
<u>SCOREFILE=</u>	score-to-measure file
<u>SDELQU=</u>	delete structure interactively
<u>SDFILE=</u>	structure deletion file
<u>SEPARATOR=</u>	delimiter of input response data fields
<u>SFILE=</u>	structure output file
<u>SFUNCTION=</u>	function to model Andrich thresholds with polynomials
<u>SICOMPLETE=</u>	simulate complete data
<u>SIEXTREME=</u>	simulate extreme scores = Yes
<u>SIFILE=</u>	simulated data file
<u>SIMEASURE=</u>	measure or data = Yes
<u>SINUMBER=</u>	number of simulated data files = 1
<u>SIRESAMPLE=</u>	number of persons resampled
<u>SISEED=</u>	simulated data seed = 0 (clock)
<u>SPFILE=</u>	supplemental control file
<u>STBIAS=</u>	correct for estimation bias
<u>STEPT3=</u>	structure summary in Table 3.2 or 21
<u>STKEEP=</u>	keep non-observed intermediate categories in structure
<u>SUBSETS=</u>	perform subset detection

SVDEPOCHS=	maximum epochs for SVD estimation
SVDFACTORS=	singular-value decomposition factors
SVDFILE=	singular-value decomposition file
SVDMIN=	singular-value decomposition minimum improvement
SVDMETHOD=	singular-value decomposition residual type
T1I#=	items per # in Table 1
T1P#=	persons per # in Table 1
T1WEIGHTED=	weighted items and persons shown once or by weight in Table 1
T2SELECT=	selects Table 2 subtables
T45OPTIONS=	field choices for Table 45
T7OPTIONS=	selects Table 7.1 and 11.1 detail lines
TABLES=	output table selection
TARGET=	information-weighted estimation
TCCFILE=	Test Characteristic Curve (TCC) and Test Information Function (TIF)
TFILE=	input file of table numbers to be output
TITLE=	title for output listing
TOTALSCORE=	show total observed score and count
TRPROFILE=	transposed file
TRPOTYPE=	transposed file type
UAMOVE=	amount to increase anchor values
UANCHOR=	user anchors (backward compatibility only)
UASCALE=	anchor user-scale value of 1 logit
UCOUNT=	most unexpected responses in Tables 6 and 10.
UDECIMALS=	number of decimal places reported
UEXTREME=	include extreme scores for UIMEAN= or UPMEAN= = No
UIMEAN=	reported user-set mean of item measures
UMEAN=	reported user-set mean of item measures
UPMEAN=	reported user-set mean of person measures
USCALE=	reported user-scaled value for 1 logit
UTF8ENC=	file is or is not encoded in UTF-8
UTF8SUB=	substitute character for multibyte UTF-8
W300=	produce IFILE= and PFILE= in 3.00 format
WEBFONT=	font to display using ASCII=Webpage
WHEXACT=	Wilson-Hilferty exact normalization
WMLE=	Warm's Mean Likelihood Estimates
XFILE=	analyzed response file
XWIDE=	columns per response
@FIELD=	user-defined field locations
#1#=	user-defined replaceable tokens
&INST	start of control variables (optional)
&END	end of control variables, before item labels
END LABELS END NAMES	end of item labels or names, before data

2 Control Variables by function

Data file layout:	
DATA=	name of data file
DELIMITER=	delimiter of input response data fields
FORMAT=	reformat data
ILFILE=	item label file
INUMB=	label items by sequence number
ITEM1=	column number of first response
ITLEN=	maximum length of item label
MFORMS=	reformatting input records & multiple forms
NAME1=	first column of person label
NAMLEN=	length of person label
NI=	number of items
PLFILE=	person label file
SEPARATOR=	delimiter of input response data fields
XWIDE=	columns per response
@FIELD=	user-defined field locations
Data selection and recoding:	
ALPHANUM=	alphabetic numbering to exceed 9 with XWIDE=1
CODES=	valid & missing data codes
CUTHI=	cut off responses with high expectations
CUTLO=	cut off responses with low expectations
EDFILE=	edit data file
IREFER=	identifying items for recoding
IVALUEx=	recoding for items
IWEIGHT=	item (variable) weighting
KEYn=	scoring key
KEYFROM=	location of KEYn=
KEYSCR=	reassign scoring keys
MAKEKEY=	construct an MCQ scoring key
MISSCORE=	scored value of missing data: not your missing-data code
NEWSCORE=	recoding values
PWEIGHT=	person (case) weighting

RESCORE=	response recoding
RESFRM=	location of RESCORE=
Items: deleting, anchoring and selecting:	
IAFILE=	item anchor file
IANCHQU=	anchor items interactively
IDDELETE=	one-line item deletion list
IDELQU=	delete items interactively
IDFILE=	item deletion file
IDROPEXTR=	remove items with extreme scores
ISELECT=	item selection criterion
Person: deleting, anchoring and selecting:	
PAFILE=	person anchor file
PANCHQU=	anchor persons interactively
PDELETE=	one-line person deletion list
PDELQU=	delete persons interactively
PDFILE=	person deletion file
PDROPEXTR=	remove persons with extreme scores
PSELECT=	person selection criterion
Rating scales, partial credit items and polytomous response structures:	
GROUPS=	assigns items to rating scale or partial credit groupings (same as ISGROUPS=)
GRPFROM=	location of ISGROUPS=
ISGROUPS=	assigns items to rating scale or partial credit groupings (same as GROUPS=)
MODELS=	assigns model types to items
MODFROM=	location of MODELS=
STKEEP=	keep non-observed categories in structure
Category structure: anchoring, labeling, deleting:	
CFILE=	scored category label file
CLFILE=	codes label file
SAFILE=	structure anchor file
SAITEM=	multiple ISGROUPS= format in SFILE= & SAFILE=

SANCHQU=	anchor structure interactively
SDELQU=	delete structure interactively
SDFILE=	structure deletion file
Measure origin, anchoring and user-scaling:	
UAMOVE=	amount to increase anchor values
UANCHOR=	anchor values in user-scaled units (only for backward compatibility)
UASCALE=	anchor user-scaled value for 1 logit
UDECIMALS=	number of decimal places reported
UEXTREME=	include extreme scores for UIMEAN= or UPMEAN= = No
UIMEAN=	reported user-set mean of item measures
UMEAN=	reported user-set mean of item measures
UPMEAN=	reported user-set mean of person measures
USCALE=	reported user-scaled value for 1 logit
Output table selection and format:	
ASCII=	output only ASCII characters
FORMFD=	form-feed character
HEADER=	display or suppress Sub-Table Headings after the first
ITEM=	title for item labels
LINELENGTH=	length of printed lines
MAXPAGE=	maximum number of lines per page
PERSON=	title for person labels
TABLES=	output table selection
TFILE=	input file of table numbers to be output
TITLE=	title for output listing
TOTALSCORE=	show total observed score and count
WEBFONT=	font to display using ASCII=Webpage
Output tables, files and graphs: specific controls	
ASYMPTOTE=	report estimates of upper and lower asymptotes
BOXSHOW=	display border boxes around columns of numbers in tables
BYITEM=	show empirical curves for items
CATREF=	reference category: Table 2

<u>CHART=</u>	graphical plots of measures
<u>CMATRIX=</u>	category matrix in Table 3.2 , etc.
<u>CURVES=</u>	probability curve selection: Tables 21, 2
<u>DISCRIM=</u>	display item discrimination estimates
<u>DISTR=</u>	output category/distractor/option counts
<u>EQFILE=</u>	code equivalences
<u>FITHIGH=</u>	lower bar in charts
<u>FITI=</u>	item misfit criterion
<u>FITLOW=</u>	lower bar in charts
<u>FITP=</u>	person misfit criterion
<u>FRANGE=</u>	half-range of fit statistics on plots
<u>HIADJ=</u>	correction for top categories
<u>IMAP=</u>	item label for item maps
<u>IPEXTREME=</u>	placement of extreme scores on maps
<u>ISORT=</u>	sort column in item label
<u>ISUBTOTAL=</u>	subtotal items by specified columns
<u>LOWADJ=</u>	correction for bottom categories
<u>MNSQ=</u>	show mean-square or t standardized fit
<u>MATRIX=</u>	correlation output format
<u>MRANGE=</u>	half-range of measures on plots
<u>MTICK=</u>	measure distance between tick-marks on plots in Tables
<u>NAMLMP=</u>	label length on maps
<u>OSORT=</u>	category/option/distractor sorted by score, count or measure
<u>OUTFIT=</u>	sort misfits on outfit or infit
<u>PMAP=</u>	person label for person maps
<u>PSORT=</u>	sort column in person label
<u>PSUBTOTAL=</u>	subtotal persons by specified columns
<u>PVALUE=</u>	report item proportion-correct-values or average ratings
<u>STEPT3=</u>	structure summary in Table 3 or 21
<u>T1#=</u>	items per # in Table 1
<u>T1P#=</u>	persons per # in Table 1
<u>T7OPTIONS=</u>	selects Table 7.1 and 11.1 detail lines
<u>TRPOTYPE=</u>	transposed file type
<u>UCOUNT=</u>	most unexpected responses in Tables 6 and 10.

W300=	produce IFILE= and PFILE= in 3.00 format
Output file format control:	
CSV=	comma-separated values in output files
HLINES=	heading lines in output files
QUOTED=	quote-marks around labels
Estimation, operation and convergence control:	
ANCESTIM=	Special alternative estimation method for anchored analyses
CONVERGE=	select convergence criteria
EXTRSCORE=	extreme score adjustment
LCONV=	logit change at convergence
LOCAL=	locally restandardize fit statistics
MJMLE=	maximum number of JMLE iterations
MPROX=	maximum number of PROX iterations
MUCON=	maximum number of JMLE iterations
NORMAL=	normal distribution for standardizing fit
PAIRED=	correction for paired comparison data
PTBISERIAL=	compute point-biserial correlations
RCONV=	score residual at convergence
REALSE=	inflate S.E. of measures for misfit
STBIAS=	correct for estimation bias
TARGET=	information-weighted estimation
WHEXACT=	Wilson-Hilferty exact normalization
Program operation:	
BATCH=	set batch mode
FSHOW=	show files created from control file
MAKEKEY=	construct an MCQ scoring key
SPFILE=	supplemental control file
&END	end of control variable list
&INST	start of control variable list (ignored)
END LABELS	end of item label list

Secondary (post-hoc) processing	
DIF=	person label columns for DIF
DPF=	item label columns for DPF
G0ZONE=	percent of 0's within 0-zone among which all 1's are turned to 0's
G1ZONE=	percent of 1's within 1-zone among which all 0's are turned to 1's
PRCOMP=	residual type for principal components/contrast analysis
SICOMPLETE=	simulate complete data
SIEXTREME=	simulate extreme scores = Yes
SINUMBER=	number of simulated data files = 1
SVDDEPTH=	singular-value decomposition levels
SVDMAX=	singular-value decomposition maximum change
SVDTYPE=	singular-value decomposition residual type

For the Output File control variables, please see [Output Files Index](#)

3 Input Files Index

These can be accessed from the [Edit Menu](#). They are specified in your Control File or at [Extra Specifications](#).

CFILE=	scored category label file
CLFILE=	codes label file
DATA=	name of data file
DIF=	person label columns for DIF
DPF=	item label columns for DPF
EDFILE=	edit data file
EQFILE=	code equivalences
GROUPS=	assigns items to rating scale or partial credit groupings (same as ISGROUPS=)
IAFILE=	item anchor file
ICOLORFILE=	colors for item labels in PKMAPs
IDFILE=	item deletion file
ILFILE=	item label file
IPRFILE=	change codes for blocks of persons and items
IREFER=	identifying items for recoding
ISRANGE=	category range of a rating scale
ISUBTOTAL=	subtotal items by specified columns
IWEIGHT=	item (variable) weighting
MFORMS=	reformatting input records & multiple forms
MODELS=	assigns model types to items
PAFILE=	person anchor file
PDFILE=	person deletion file
PIVOT=	structure anchoring (backwards compatibility only)
PLFILE=	person label file

PSUBTOTAL=	subtotal persons by specified columns
PWEIGHT=	person (case) weighting
RESCORE=	controls response recoding
SAFILE=	structure-threshold anchor file
SDFILE=	structure deletion file
SPFILE=	supplemental control file
TFILE=	input file of table numbers to be output

4 Output Tables Index

Click on the Table name to go to the Table description:

Output Tables	Output Files	Batch	Help	Specification	Plots	Excel/RSSST	Graphs	Data Setup
Request Subtables				1. Variable (Wright) maps				20. KID Score table
3.2+ Rating (partial credit) scale				2.2 General Keyform				21. Probability curves
2. Measure forms (all)				2.5 Category Averages				29. Empirical curves
				3.1 Summary statistics				22. Scalograms
10. TAP (column): fit order				6. KID (row): fit order				40. KID Keyforms: fit order
13. TAP: measure				17. KID: measure				37. KID Keyforms: measure
14. TAP: entry				18. KID: entry				38. KID Keyforms: entry
15. TAP: alphabetical				19. KID: alphabetical				39. KID Keyforms: alphabetical
25. TAP: displacement				42. KID: displacement				41. KID Keyforms: unexpected
26. TAP: correlation				43. KID: correlation				36. KID diagnostic PKMAPs
11. TAP: responses				7. KID: responses				35. KID Paired Agreement
9. TAP: outfit plot				5. KID: outfit plot				45. KID Incremental Measures
8. TAP: infit plot				4. KID: infit plot				44. Global fit statistics
12. TAP: Wright map				16. KID: Wright map				34. Comparison of two statistics
23. TAP: dimensionality				24. KID: dimensionality				32. Control variable list
27. TAP: subtotals				28. KID: subtotals				20.3 TAP Score table
30. TAP: DIF, between/within				31. KID: DPF, between/within				33. KID-TAP: DGF: DIF & DPF

Table	Description
1	<p>Maps of person and item measures. Show Rasch measures.</p> <p>1.0 One page map with item and person labels.</p> <p>1.1 Map of distributions - persons and items</p> <p>1.2 Item labels with person distribution (squeezed onto one page)</p> <p>1.3 Person labels with item distribution (squeezed onto one page)</p> <p>1.4 (Polytomies) Rating scale or partial credit map of distributions: persons with items at high, mean, low</p> <p>1.5 (Polytomies) Item map with expected score zones (Rasch-half-point thresholds)</p> <p>1.6 (Polytomies) Item map with 50% cumulative probabilities (Rasch-Thurstone thresholds)</p> <p>1.7 (Polytomies) Item map with Andrich thresholds (modal categories if ordered)</p> <p>1.8 (Polytomies) Item map with measures at category scores (maximum probability of categories)</p> <p>1.9 (Polytomies) Item map of average person measure for each category score</p> <p>1.10 One page map with person names by measure, item names by easiness.</p> <p>1.11 Map of distributions - persons by ability and items by easiness</p> <p>1.12 Map of item labels, by easiness, with person distribution</p> <p>1.13 Map of person labels with item distribution by easiness</p>
2	<p>Measures and responses plots. Response categories for each item, listed in measure order, plotted against person measures, shown as modal categories, expected values and cumulative probabilities.</p> <p>Table 2 for dichotomous and multiple-choice items.</p> <p>Table 2 for polytomous, rating-scale, partial credit, items.</p>

	<p><i>By scored categories (illustrated by an observed category code for each score)</i></p> <p>2.1 Modal categories (most probable) 2.2 Mean categories (average or expected: Rasch-half-point thresholds) 2.3 (Polytomies) Median categories (cumulative probabilities: Rasch-Thurstone thresholds) 2.4 (Polytomies) Structure calibrations (Rasch model parameters: rating scale, partial credit, "restricted", "unrestricted": Rasch-Andrich thresholds) 2.5 Observed average measures of persons for scored categories (empirical averages) 2.7 Expected average measures of persons</p> <p><i>By observed categories</i></p> <p>2.6 Observed average measures of persons (empirical averages)</p> <p><i>By category scores</i></p> <p>2.11 Modal categories (most probable) 2.12 Mean categories (average or expected: Rasch-half-point thresholds) 2.13 (Polytomies) Median categories (cumulative probabilities: Rasch-Thurstone thresholds) 2.14 (Polytomies) Structure calibrations (Rasch model parameters: rating scale, partial credit, "restricted", "unrestricted") 2.15 Observed average measures of persons for scored categories (empirical averages) 2.16 Observed average measures of persons for observed categories (empirical averages) 2.17 Expected average measures of persons</p>
3	<p>Summary statistics. Person, item, and category measures and fit statistics. 3.1 Summaries of person and items: means, S.D.s, separation, reliability. 3.2 Summary of rating categories and probability curves. (STEPT3=Yes) 3.3-... Summary of rating categories and probability curves with ISGROUPS= (STEPT3=Yes)</p>
4	<p>Person infit plot. Person infit statistics plotted against person measures. 4.1 Person infit vs. person measure plot.</p>
5	<p>Person outfit plot. Person outfit statistics plotted against person measures. 5.1 Person outfit vs. person measure plot. 5.2 Person infit vs. person outfit plot.</p>
6	<p>Person statistics - fit order. Misfitting person list. Scalogram of unexpected responses. 6.1 Table of person measures in descending order of misfit. (Specify FITP=0 to list all persons) 6.2 Chart of person measures, infit mean-squares and outfit mean-squares. (Chart=Yes) 6.4 Scalogram of most misfitting person response strings. 6.5 Scalogram of most unexpected responses. 6.6 Most unexpected responses list.</p>
7	<p>Misfitting Persons. Lists response details for persons with t standardized fit greater than FITP=. 7.1 Response strings for most misfitting persons.</p>
8	<p>Item infit plot. Item infit plotted against item calibrations. 8.1 Item infit vs. item measure plot.</p>
9	<p>Item outfit plot. Item outfits plotted against item calibrations. 9.1 Item outfit vs. item measure plot. 9.2 Item infit vs. item outfit plot</p>
10	<p>Item statistics - fit order. Misfitting item list with option counts. Scalogram of unexpected responses. 10.1 Table of item measures in descending order of misfit. (Specify FITI=0 to list all persons) 10.2 Chart of item measures, infit mean-squares and outfit mean-squares. (Chart=Yes) 10.3 Item response-structure categories/options/distractors: counts and average abilities. (Distractors=Yes) 10.4 Scalogram of most misfitting item response strings. 10.5 Scalogram of most unexpected responses. 10.6 Most unexpected responses list.</p>
11	<p>Misfitting Items. Response details for items with t standardized fit greater than FITI=. 11.1 Response strings for most misfitting items.</p>
12	<p>Item distribution map with full labels for items and an person distribution. 12.2 Item labels with person distribution (same as 1.2) 12.5 (Polytomies) Item map with expected score zones (Rasch-half-point thresholds) (same as 1.5)</p>

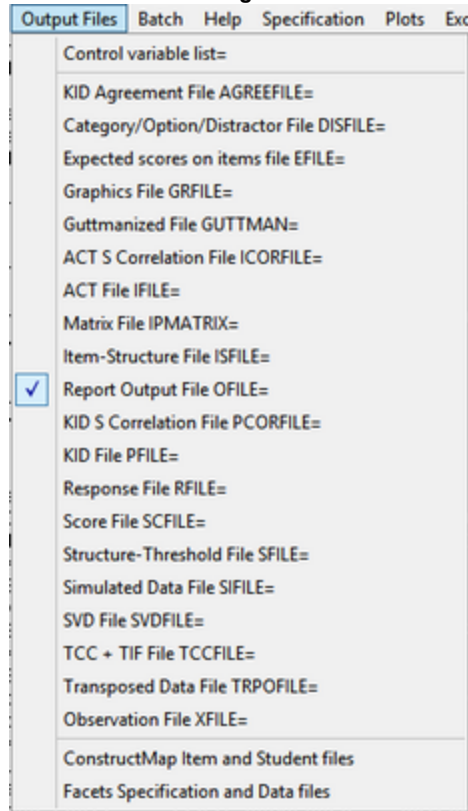
	<p>12.6 (Polytomies) Item map with 50% cumulative probabilities (Rasch-Thurstone thresholds) (same as 1.6)</p> <p>12.7 (Polytomies) Item map with Andrich thresholds (modal categories if ordered) (same as 1.7)</p> <p>12.8 (Polytomies) Item map with measures at category scores (maximum probability of categories) (same as 1.8)</p> <p>12.9 (Polytomies) Item map of average person measure for each category score (same as 1.9)</p> <p>12.12 Item labels, by easiness, with person distribution (same as 1.12)</p>
13	<p>Item statistics - measure order list and graph with category/option/distractor counts.</p> <p>13.1 Table of items in descending measure order.</p> <p>13.2 Chart of item measures, infit mean-squares and outfit mean-squares. (Chart=Yes)</p> <p>13.3 Item response-structure categories/options/distractors: counts and average abilities. (Distractors=Yes)</p>
14	<p>Item statistics - entry order list and graph with category/option/distractor counts.</p> <p>14.1 Table of items in entry number (sequence) order.</p> <p>14.2 Chart of item measures, infit mean-squares and outfit mean-squares. (Chart=Yes)</p> <p>14.3 Item response-structure categories/options/distractors: counts and average abilities. (Distractors=Yes)</p>
15	<p>Item statistics - alphabetical order list and graph with category/option/distractor counts.</p> <p>15.1 Table of item measures in alphabetical order by label. (Specify ISORT= for sort column)</p> <p>15.2 Chart of item measures, infit mean-squares and outfit mean-squares. (Chart=Yes)</p> <p>15.3 Item response-structure categories/options/distractors: counts and average abilities. (Distractors=Yes)</p>
16	<p>Person distribution map with full labels for persons and an item distribution.</p> <p>16.3 Person labels with item distribution (same as 1.3)</p> <p>16.13 Person labels with item distribution by easiness (same as 1.13)</p>
17	<p>Person statistics - measure order list and chart.</p> <p>17.1 Table of persons in descending measure order.</p> <p>17.2 Chart of person measures, infit mean-squares and outfit mean-squares. (Chart=Yes)</p>
18	<p>Person statistics - entry order list and chart.</p> <p>18.1 Table of persons in entry number (sequence) order.</p> <p>18.2 Chart of person measures, infit mean-squares and outfit mean-squares. (Chart=Yes)</p>
19	<p>Person statistics - alphabetical order list and chart.</p> <p>19.1 Table of person measures in alphabetical order by label. (Specify PSORT= for sort column)</p> <p>19.2 Chart of person measures, infit mean-squares and outfit mean-squares. (Chart=Yes)</p>
20	<p>Measures for all scores on a test of all calibrated items, with percentiles.</p> <p>20.1 Table of person measures for every score on complete test. (Specify ISELECT= for subtests).</p> <p>20.2 Table of measures for every score, with sample percentiles and norm-referenced measures.</p> <p>20.3 Table of item difficulty measures (calibrations) for every score (proportion-correct-values or average ratings) by complete sample.</p>
21	<p>Category probability curves. Category probabilities plotted against the difference between person and item measures, then the expected score and cumulative probability and expected score ogives. See also Graphs menu</p> <p>21.1 Category probability curves (shows Rasch-Andrich thresholds). CURVES=1xx</p> <p>21.2 Expected score ogive (model Item Characteristic Curve, model Item Response Function). CURVES=x1x</p> <p>21.3 Cumulative category probability curves (shows Rasch-Thurstone thresholds). CURVES=xx1</p>
22	<p>Sorted observations. Data sorted by person and item measures into scalogram patterns.</p> <p>22.1 Guttman scalogram of sorted scored responses.</p> <p>22.2 Guttman scalogram showing out-of-place responses.</p> <p>22.3 Guttman scalogram showing original responses.</p>
23	<p>Item principal components/contrasts. Identifies structure and dimensionality in response residuals (BIGSTEPS Table: 10.3)</p> <p>23.0 Scree plot of variance components</p> <p>23.1. 23.11.... Plot of loadings on first contrast in residuals vs. item measures.</p> <p>23.2. 23.12.... Items in contrast loading order.</p> <p>23.3. 23.13.... Persons exhibiting contrast.</p> <p>23.4. 23.14.... Items in measure order.</p> <p>23.5. 23.15.... Items in entry order.</p> <p>23.6. 23.16.... Person measures for item clusters. Matching scatterplot.</p>

	23.99 Tables of items with highly correlated residuals.
24	Person principal components/contrasts. Identifies structure in residuals (not in BIGSTEPS) 24.0 Scree plot of variance components. 24.1. 24.11.... Plot of loadings on first contrast in residuals vs. person measures. 24.2. 24.12.... Persons in contrast loading order. 24.3. 24.13.... Items exhibiting contrast. 24.4. 24.14.... Persons in measure order. 24.5. 24.15.... Persons in entry order. 24.99 Tables of persons with highly correlated residuals.
25	Item statistics - displacement order list and graph with category/option/distractor counts. 25.1 Table of items in descending displacement order. 25.2 Chart of item measures, infit mean-squares and outfit mean-squares. (Chart=Yes) 25.3 Item response-structure categories/options/distractors: counts and average abilities. (Distractors=Yes)
26	Item statistics - correlation order list and graph with category/option/distractor counts. 26.1 Table of items in ascending correlation order (Point-biserial, if PTBIS=Yes , else Point-measure). 26.2 Chart of item measures, infit mean-squares and outfit mean-squares. (Chart=Yes) 26.3 Item response-structure categories/options/distractors: counts and average abilities. (Distractors=Yes)
27	Item subtotals. 27.1 Measure sub-totals, controlled by ISUBTOT= 27.2 Measure sub-totals bar charts, controlled by ISUBTOT= 27.3.... Measure sub-totals summary statistics, controlled by ISUBTOT=
28	Person subtotals. 28.1 Measure sub-totals, controlled by PSUBTOT= 28.2 Measure sub-totals bar charts, controlled by PSUBTOT= 28.3.... Measure sub-totals summary statistics, controlled by PSUBTOT=
29	Empirical item character curves and response frequency plots. 29.1.... Expected and Empirical ICCs (see also Graph Menu) Empirical category code frequencies
30	Differential Item Function across Person classifications 30.1 DIF report (paired), controlled by DIF= 30.2 DIF report (measure list: person class within item) 30.3 DIF report (measure list: item within person class) 30.4 DIF report (item-by-person class chi-squares, between-class fit) 30.5 Within-class fit report (person class within item) 30.6 Within-class fit (report item within person class) 30.7 Item measure profiles for classes of persons
31	Differential Person Function across Item classifications 31.1 DPF report (paired), controlled by DPF= 31.2 DPF report (measure list: item class within person) 31.3 DPF report (measure list: person within item class) 31.4 DPF report (person by item-class chi-squares, between-class fit) 31.5 Within-class fit report (item class within person) 31.6 Within-class fit report person class within item) 31.7 Person measure profiles for classes of items
32	Control Variable Listing of the current settings of all Winsteps control variables - appears on the Output Files pull-down menu.
33	Differential Group Function: Item Class vs. Person Class interaction/bias 33.1 DGF report (paired person classes on each item class), controlled by DIF= and DPF= 33.2 DGF report (paired item classes on each person class) 33.3 DGF report (list of person classes within item class) 33.4 DPF report (list of item classes within person class) 33.7 Item group-Person group profiles 33.8 Item group-Person group profiles
34	File Comparison (Plot menu) 34.1 Column comparison of two statistics (and Excel scatterplot)

35	Paired Comparison of Person Response Strings 35.1.35.11.... % Same (paired student) observed responses 35.2.35.12.... % Same (paired student) scored responses 35.3.35.13.... % Same (paired student) observed highest (right) responses 35.4.35.14.... % Same (paired student) observed lowest (wrong) responses 35.5.35.15.... % Same (paired student) missing responses
36	Person PKMAP Plots 36.1.... PKMAPS of persons
37	Person KeyForms - Measure order. 37.1-... Keyforms of responses of persons.
38	KeyForms of Persons - Entry order. 38.1-... Keyforms of responses of persons.
39	KeyForms of Persons - Alphabetical order. 39.1-... Keyforms of responses of persons.
40	KeyForms of Persons - Misfit order. 40.1-... Keyforms of responses of persons.
41	KeyForms of Persons - Misfit order - only Unexpected Responses. 41.1-... Keyforms of unexpected responses.
42	Person statistics - displacement order list and chart. 42.1 Table of persons in descending displacement order. 42.2 Chart of person measures, infit mean-squares and outfit mean-squares. (Chart=Yes)
43	Person statistics - Correlation order list and chart. 43.1 Table of persons in ascending correlation order (Point-biserial, if PTBIS=Yes , else Point-measure). 43.2 Chart of person measures, infit mean-squares and outfit mean-squares. (Chart=Yes)
44	Global statistics 44.1 Table of global counts and fit statistics
45	Table 45 45.1 Person measures after each item
0	Control Variables and Convergence report. Lists the control variables and shows estimation convergence. (Only appears at end of Output Report File). 0.1 Title page and Analysis identification 0.2 Convergence table 0.3 Control file 0.4 Subset details

5 Output Files Index

Click on the Table name to go to the Table description:

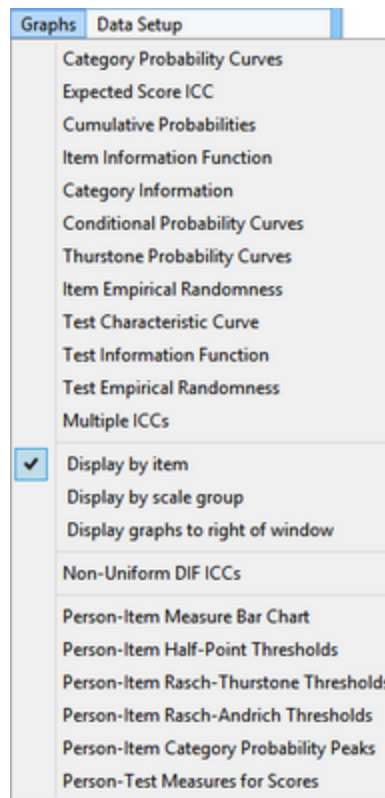


These can be accessed from the [Output Files](#) pull-down menu. Once created, they can be accessed from the [Edit Menu](#). The output files specified in the control file can be displayed immediately they are created with [FSHOW=Yes](#).

Output files	
Control variable file	control variable listing (same as Table 32)
AGREEFILE=	paired-person agreement counts
DISFILE=	category/distractor/response option count file
EFILE=	expected scores on the items
GRFILE=	graphing file for probability curves
GUTTMAN=	Guttmanized file of responses
ICORFILE=	inter-item residual correlations
IFILE=	item output file
IPMATRIX=	response matrix - only from Output Files menu
ISFILE=	item-structure output file (Do not use for anchoring)
MFORMS=	shows input responses as reformatted by MFORMS=
OFILE=	change report output file
PCORFILE=	inter-person residual correlations
PFILE=	person output file
RFILE=	scored response output file
SCOREFILE=	score-to-measure file
SFILE=	structure-threshold output file (Use for anchoring)
SIFILE=	simulated data file

SVDFILE=	singular-value decomposition file
TCCFILE=	Test Characteristic Curve and Test Information Function
TRPOFILE=	transposed file
XFILE=	analyzed response file
ConstructMap	ConstructMap (formerly GradeMap) files - only from Output Files menu
Facets	create Facets Specification and Data files - only from Output Files menu
KEYFORM=	Keyform skeleton for Excel plot - see Plots menu
LOGFILE=	accumulates control files - only in the Control file

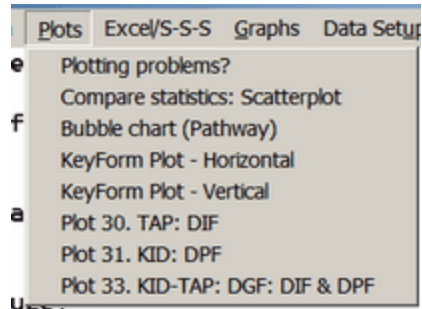
6 Output Graphs and Plots Index



Graphs - from the [Graphs](#) pull-down menu

- [Category information function](#)
- [Conditional probability curves](#)
- [Cumulative probabilities](#)
- [Empirical category curves](#)
- [Empirical ICC \(Item Characteristic Curve\)](#)
- [Empirical randomness](#)
- [Expected score ICC](#)
- [Item information function](#)
- [Multiple item ICCs](#)
- [Non-uniform DIF ICCs](#)
- [Points and lines](#)
- [Probability category curves](#)
- [Test characteristic curve](#)
- [Test information function](#)
- [Test randomness](#)

[Person-item Histogram](#)



Plots - from the [Plots](#) pull-down menu

[Compare statistics](#) generates an Excel-based cross-plot of the values in Table 34.

[Bubble chart](#) generates a Bond & Fox-style bubble chart.

[Keyform](#) plot - Horizontal generates a horizontal Keyform layout.

[Keyform](#) plot - Vertical generates a vertical Keyform layout.

[DIF plot](#) - the DIF values in Table 30.

[DPF plot](#) - the DPF values in Table 31.

[DIF+DPF plot](#) - the DIF+DPF values in Table 33.

[Person measures after items](#) plot in Table 45

7 Introduction

7.1 Rasch analysis and WINSTEPS

Winsteps is Windows-based software which assists with many applications of the Rasch model, particularly in the areas of educational testing, attitude surveys and rating scale analysis. There is more information at: www.winsteps.com

Winsteps started from "Rating Scale Analysis" (Wright & Masters, 1982), available by free download at www.rasch.org (green book).

Rasch analysis is a method for obtaining objective, fundamental, additive measures (qualified by [standard errors](#) and quality-control fit statistics) from stochastic observations of ordered category responses. Georg Rasch, a Danish mathematician, formulated this approach in 1953 to analyze responses to a series of reading tests (Rasch G, *Probabilistic Models for Some Intelligence and Attainment Tests*, Chicago: MESA Press, 1992, with instructive Foreword and Afterword by B.D. Wright). Rasch is pronounced like the English word *rash* in Danish, and like the English sound *raa-sch* in German. The German pronunciation, *raa-sch*, is used to avoid misunderstandings.

The person and item total raw scores are used to estimate additive measures. Under Rasch model conditions, these measures are [item-free](#) (item-distribution-free) and [person-free](#) (person-distribution-free). So that the measures are statistically equivalent for the items regardless of which persons (from the same population) are analyzed, and for the persons regardless of which items (from the same population) are analyzed. Analysis of the data at the response-level indicates to what extent these ideals are realized within any particular data set. Rasch analysis is "conjoint measurement". The person abilities and item difficulties are measured on the same scale. If you add something to the item difficulties then you add the same amount to the person abilities (thetas) in order to keep the relationship between the person and items the same.

The Rasch models implemented in Winsteps include the Georg Rasch dichotomous, Andrich "rating scale", Masters "partial credit", Bradley-Terry "paired comparison", Glas "success model", Linacre "failure model" and most combinations of these models. Other models such as binomial trials and Poisson can also be analyzed by [anchoring](#) (fixing) the response structure to accord with the response model. (If you have a particular need, please let us know as Winsteps is continually being enhanced.)

The estimation method is [JMLE](#), "Joint Maximum Likelihood Estimation", with initial starting values provided by [PROX](#), "Normal Approximation Algorithm".

The Rasch Family of Models

The necessary and sufficient transformation of ordered qualitative observations into additive measures is a Rasch model. Rasch models are logit-linear models, which can also be expressed as log-linear models. Typical Rasch models operationalized with Winsteps are:

The dichotomous model:
 $\log_e(P_{ni1} / P_{ni0}) = B_n - D_i$

The polytomous "Rating Scale" model:
 $\log(P_{nij} / P_{ni(j-1)}) = B_n - D_i - F_j$

The polytomous "Partial Credit" model:
 $\log(P_{nij} / P_{ni(j-1)}) = B_n - D_i - F_{ij} = B_n - D_{ij}$

The polytomous "Grouped response-structure" model:
 $\log(P_{nij} / P_{ni(j-1)}) = B_n - D_{ig} - F_{gj}$

where

P_{nij} is the probability that person n encountering item i is observed in category j ,

B_n is the "ability" (θ) measure of person n ,

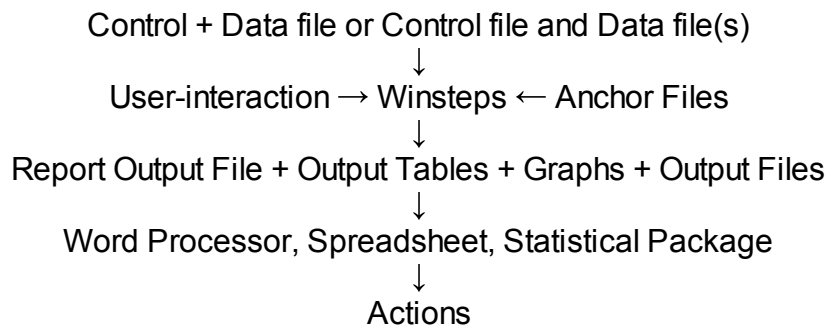
D_i is the "difficulty" (δ) measure of item i , the point where the highest and lowest categories of the item are equally probable.

F_j is the "calibration" measure of category j relative to category $j-1$, the point where categories $j-1$ and j are equally probable relative to the measure of the item. No constraints are placed on the possible values of F_j .

Also models with the form of "[Continuation Ratio](#)" models, such as the "Success" model and the "Failure" model.

For methods of estimation, see RSA, pp. 72-77.

Work-flow with Winsteps



Winsteps is designed to construct Rasch measurement from the responses of a set of persons to a set of items. Responses may be recorded as letters or integers and each recorded response may be of one or two characters. Alphanumeric characters, not designated as legitimate responses, are treated as missing data. This causes these observations, but not the corresponding persons or items, to be omitted from the analysis. The responses to an item may be dichotomous ("right"/"wrong", "yes"/"no"), or may be on a rating scale ("good"/"better"/"best", "disagree"/"neutral"/"agree"), or may have "partial credit" or other hierarchical structures. The items may all be grouped together as sharing the one response structure, or may be sub-groups of one or more items which share the same response structure.

Winsteps begins with a central estimate for each person measure, item calibration and response-structure calibration, unless pre-determined, "anchor" values are provided by the analyst. An iterative version of the [PROX](#) algorithm is used reach a rough convergence to the observed data pattern. The [JMLE](#) method is then iterated to obtain more exact estimates, standard errors and fit statistics.

Output consists of a variety of useful plots, graphs and tables suitable for import into written reports. The statistics can also be written to data files for import into other software. Measures are reported in [Logits](#) (log-odds units) unless user-rescaled. Fit statistics are reported as mean-square residuals, which have approximate chi-square distributions. These are also reported t standardized, $N(0,1)$.

7.2 References

Please cite the current Winsteps computer program as:

Linacre, J. M. (2020). Winsteps® Rasch measurement computer program. Beaverton, Oregon: Winsteps.com

For this User's Guide, the citation is:

Linacre, J. M. (2020). Winsteps® Rasch measurement computer program User's Guide. Beaverton, Oregon: Winsteps.com

APA Style:

Linacre, J.M. (2020). Winsteps® (Version 4.5.2) [Computer Software]. Beaverton, Oregon: Winsteps.com. Retrieved January 1, 2020. Available from <https://www.winsteps.com/>

For a webpage in Winsteps Help:

Linacre, J. M. (2020, April) Web page title, Retrieved from www.winsteps.com/winman/webpage.htm

RSA means Wright B.D. & Masters G.N. Rating Scale Analysis, Chicago: MESA Press, 1982, especially p. 100:

www.rasch.org/books.htm

BTD means Wright B.D. & Stone M.H. Best Test Design, Chicago: MESA Press, 1979: www.rasch.org/books.htm

Other recommended sources:

Rasch Measurement Transactions: www.rasch.org/rmt/

Journal of Applied Measurement: www.jampress.org

"Applying the Rasch Model: Fundamental Measurement in the Human Sciences", by Trevor G. Bond & Christine M. Fox, 2001. Mahwah NJ: Lawrence Erlbaum Assoc. 0-8058-3476-1. [Authors' website](#). Using Winsteps: [Bond & Fox examples](#)

"Introduction to Rasch Measurement", Everett V. Smith, Jr. & Richard M. Smith (Eds.) JAM Press, 2004 www.jampress.org

Rasch, G. (1960). Probabilistic Models for Some Intelligence and Attainment Tests (Reprint, with Foreword and Afterword by B. D. Wright, Chicago: University of Chicago Press, 1980). Copenhagen, Denmark: Danmarks Paedagogiske Institut. www.rasch.org/books.htm

Some Rasch papers available online:

For quick access to online papers, [Google](#) "+Winsteps +Rasch +Linacre filetype:pdf" - as of November 2009, this produced 2,290 hits.

<https://www.science.gov/topicpages/m/many-facet+rasch+model.html#>

<https://worldwidescience.org/topicpages/m/many-faceted+rasch+model.html#>

www.uky.edu/~kdbrad2/Rasch_Symposium.pdf Constructing and Evaluating Measures: Applications of the Rasch Measurement Model

pareonline.net/pdf/v10n4.pdf Test Equating by Common Items and Common Subjects: Concepts and Applications

www.jalt.org/pansig/2005/HTML/Weaver.htm How entrance examination scores can inform more than just admission decisions (skip the short section in Japanese).

Other Winsteps references (please submit yours if you want it listed here):

Peeters M.J., Stone G.E. (2009). An instrument to objectively measure pharmacist professionalism as an outcome: a pilot study. Canadian Journal of Hospital Pharmacy, 62(3), 209-16.

7.3 About the Program Manual and Software

You don't need to know about every Winsteps option in order to use the program successfully. Glance through the examples and find one similar to yours. Adapt the example to match your requirements. Then "fine tune" your analysis as you become familiar with further options.

Most of this Guide is in proportionately-spaced type.

```
When it is important to be precise about blanks or
spaces, or about
column alignment,
fixed-space type is used.
```

When it is important to show everything that appears on a long line, small type is used.

Suggestions that we have found helpful are shown like this in italics.

Authors: please cite the current Winsteps computer program as shown in [References](#).

Winsteps software is written in [Absoft Pro Fortran](#) and Microsoft Visual Basic 6.0. The software installer is constructed by [Setup Factory](#). Help files and PDFs are written using [Help & Manual](#).

We acknowledge the kind permission granted by Chris Hanscom of [Veign](#) for the use of their **Jeweled Style Command Button**.

This product includes software developed by vbAccelerator (<http://vbaccelerator.com/>).

Thank you to the many Winsteps users who have reported bugs and sent us suggestions for improvements. Please keep up the good work!

An Ode from a user:

Winsteps

This is a program that's much alive,
With scores and maps, and curves Ogive,
Persons and items to separate,
Giving results that do relate.

Although we love what Winsteps does,
Some problems make us Yelp!
But if we give John Michael a buzz,
He is always there to help!

Jim Houston, Nov. 30, 2001

7.4 Getting further help

Common installation problems are solved at: www.winsteps.com/problems.htm

Winsteps is a powerful weapon in the struggle to wrest meaning from the chaos of empirical data. As you become skilled in using Winsteps, you will find that it helps you to conceptualize what you are measuring, and to diagnose measurement aberrations. The [Special Topics](#) section of this User's Guide contains a wealth of information and [advice](#).

[Rasch Measurement Transactions](#), contains instructive articles on the fundamentals of Rasch analysis as well as the latest ideas in theory and practice. There are other useful books and journals, including: Journal of Applied Measurement, Trevor Bond & Christine Fox: "Applying the Rasch Model", Lawrence Erlbaum Assoc.

You may also find that you can use a more personal word of advice on occasion. The author of Winsteps, Mike Linacre, is happy to answer e-mailed questions to do with the operation of Winsteps or the nature of Rasch analysis. More prolonged consultations can also be arranged.

8 Installation and Execution

8.1 What is supplied

Winsteps is supplied in several forms:

1)	MinistepInstall.exe	To install Ministep, the student/evaluation version with maximum dataset size of 75 persons (rows) by 25 items (columns). This is freeware.
2)	WinstepsTimeLimitedInstall.exe	To install Winsteps. This full version of Winsteps expires in a few weeks. It is used for training.
3)	WinstepsPasswordInstall.exe	To install Winsteps with password-protected installation. This is the full licensed version.
4)	Bond&FoxStepsInstall.exe	To install Bond&FoxSteps, the reduced-capability version for use with "Applying the Rasch Model" (Bond & Fox, 2015).

These create directory, C:\Winsteps, and install Winsteps or Ministep in it.

Sample control and data (.TXT) files are also installed in c:\Winsteps\EXAMPLES to help you get started:

KCT.TXT is the Knox Cube Test data (BTD p.31 - see Section 1.1) The results in BTD were obtained with more approximate algorithms and do not agree exactly with Winsteps results.

SF.TXT is the Liking For Science data (RSA p.18)

There are many more [EXAMPLE](#) files described later in this manual.

8.2 Installation instructions for WINSTEPS

Under Windows XP/98/NT/ME/2000/...:

Run WinstepsInstall.exe from the downloaded file or from the CD-ROM.

If program hangs during "Constructing Winsteps.ini ..." then see Initialization Fails

To Run Winsteps:

Click on *Start* button (or use desktop Winsteps icon)

Point to *Programs*

Point to *Winsteps*

Click on *Winsteps* icon

Click No to Welcome dialog box

Type **SF.TXT**

Press Enter key

Press Enter key

Press Enter key

Winsteps will run. Examine its output with the pull-down menus, particularly [Output Tables](#)

Additional Notes:

a) Winsteps is usually installed in the C:\Winsteps directory.

b) When Winsteps ends, pull down the "**Edit**" menu to edit control and output files.

c) All information on the screen is also in the Report output file.

d) Files in the C:\TEMP directory and with the suffix .TMP may be deleted.

Macintosh Computer: Winsteps users report successful operation under Fusion, Virtual PC and Soft PC.

8.3 Silent Installation of Winsteps from the Command Line

1. Download WinstepsPasswordInstall(version number).exe or MinistepInstall.exe

Example: WinstepsPasswordInstall371.exe

2. Obtain your installation password.
3. Installing in folder c:\Winsteps

In the folder which contains WinstepsPasswordInstall(version number).exe

At a DOS prompt or in a batch file, type in:

WinstepsPasswordInstall(version number).exe /serial:"Installation Password"

Example: WinstepsPasswordInstall371.exe /serial:"abcd1234"

Press enter at the DOS prompt, or save and launch the batch file.

Winsteps or Ministep should install

or 3. Installing in a different folder:

Silent Install of Winsteps

1. Download WinstepsPasswordInstallxxx.exe to your work folder
2. In your work folder, create text file: setupvars.ini
3. The contents of setupvars.ini are:
[SetupValues]
%AppFolder%=C:\Winsteps (or your target folder)
%AppShortcutFolderName%=Winsteps
- 4.
5. In your work folder, create text file: install.bat
6. The contents of install.bat are:
WinstepsPasswordInstallxxx.exe "/S:c:\(work folder path)\setupvars.ini" /serial:"123456789" ← your installation password goes here

Example:

My installation password is: *my%password*

WinstepsPasswordInstall3902.exe "/S:c:\myworkfolder\setupvars.ini" /serial:"my%%password"

Notice that % has become%%

7. To install Winsteps silently, double-click on install.bat in the Explorer view of your work folder.
8. View c:\Winsteps (or your target folder) to verify that Winsteps.exe has installed correctly
9. Back up these files in your work folder, then delete them: WinstepsPasswordInstallxxx.exe, setupvars.ini, install.bat

Silent Install of Ministep

1. Download Ministep.exe to your work folder
2. In your work folder, create text file: setupvars.ini
3. The contents of setupvars.ini are:
[SetupValues]
%AppFolder%=C:\Winsteps (or your target folder)
%AppShortcutFolderName%=Winsteps
4. In your work folder, create text file: install.bat
5. The contents of install.bat are:
MinistepInstall.exe "/S:c:\(full path name)\setupvars.ini"
6. To install Ministep silently, double-click on install.bat in the Explorer view of your work folder.
7. View c:\Winsteps (or your target folder) to verify that Ministep.exe has installed correctly
8. Back up these files in your work folder, then delete them: Ministep.exe, setupvars.ini, install.bat

8.4 Starting WINSTEPS in Windows

A typical analysis requires two components: control information and data. These can be in separate computer files or can be combined in one file. The results of the analysis are written to output files.

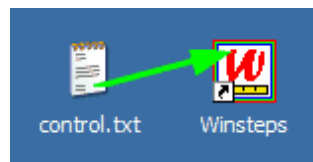
To launch your Winsteps analysis, here are several methods:

1) Double-click the Winsteps Icon on the Desktop. Winsteps launches. File menu: Open File.

To change the Winsteps starting folder,
right-click on the icon
highlight and click on properties
click the shortcut tab
change the "Start in" path to the path to the desired folder.

or 2) Click on Start button
Point to Programs
Point to Winsteps
Click on Winsteps icon

or 3) Drag your control file onto your Winsteps icon.



or 4) Add Winsteps to the Windows Sendto menu, and right-click on your control file.

To add Winsteps to Sendto:

1. Open an Explorer window. Enter the following into the address bar and press the Enter key:
shell:sendto
2. Copy the desktop shortcuts for Winsteps into the Send To folder.
3. To run an analysis, right click on the control file to open the Send To menu, select Winsteps.

If program hangs during "Constructing Winsteps.ini ..." then see Initialization Fails



Welcome to Winsteps!

Control + Data Setup Procedure: If you need help to set up a new control file, click this and go to [control and data file setup](#).

Import from Excel, R, SAS, SPSS, STATA, Tabbed Text: takes you to the [Excel/RSSST](#) dialog

Text-File Instructions: takes you to the [do-it-yourself instructions](#)

No: Winsteps asks you for the names of your input and report output files. There are example files already in the Winsteps\examples directory.

Don't ask again: makes "No" the standard option here. You can reset this using [Edit Initial Settings](#)

Help displays this page.

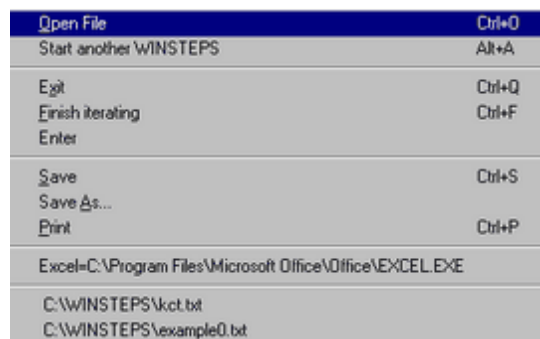
To select your control file:

Winsteps asks: **Please enter name of Winsteps control file:**

(a) You can type it in ...

Please enter name of Winsteps control file: KCT.TXT(Enter)

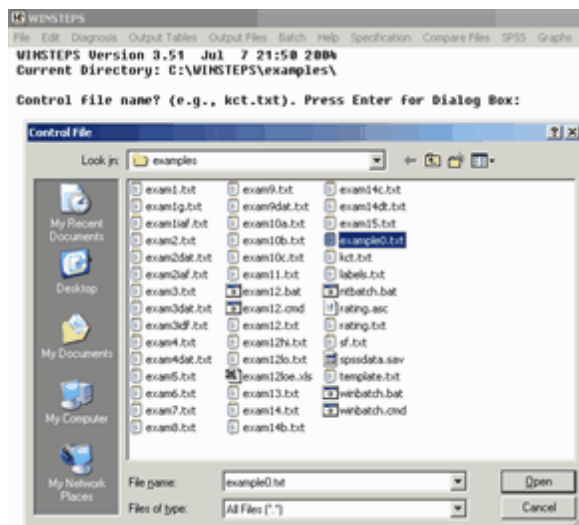
or (b) Click on the [Files](#) pull-down menu, "Open File", to get a file dialog box.



or (c) Click-on a file name from the bottom of the Files menu list.

or (d) Press your **Enter** key

This displays a standard Windows open dialog box - try it.



You can also edit files directly from the file dialog box by right-clicking on them and selecting "Edit".

Example Analysis:

Control file name? (e.g., kct.txt). Press Enter for Dialog Box: Example0.txt(Enter)

Please enter name of report output file: (Enter)

If you only press Enter, a temporary output file will be used.

If you type in a file name, that file will be created and used.

If you want the file dialog box, use 'Open File' on the [File](#) pull-down menu.

Extra specifications (if any). Press Enter to analyze):? (e.g., [MJMLE=1](#)), or press Enter: (Enter)

Usually you have none, so merely press Enter.

This is used for making [just-in-time changes](#) to your control file instructions, for this run only.

Winsteps will now construct measures (i.e., analyze) the Liking for Science data from RSA.

Use the [Edit pull-down menu](#) to simplify inspection and editing of the control, input and output files. This is done with Notepad or [your own text editor](#).

At the end of the run, using the [Output Tables](#) pull-down menu you can request further output Tables to be produced interactively and displayed on the screen. If you want to save any of these, use the Save As option when they are displayed. If you omit to save them, they are written as "ws.txt" files, and can be recovered from the Recycle Bin.

At the end of the run, you can also use the [Output Files](#) pull down menu, to write out person and item measures to computer-readable [PFILE=](#) and [IFILE=](#) files.

If the Edit menu and Output Tables menu don't work properly, then see [Changing](#) your Word Processor setting.

8.5 Using Winsteps under Windows

```
example0.txt
File Edit Diagnosis Output Tables Output Files Batch Help Specification Plots Excel/RSSST Graphs Data Setup
Control file name? (e.g., exam1.txt). Press Enter for Dialog Box:
C:\Winsteps37011\examples\example0.txt
Current Directory: C:\Winsteps37011\examples\

Report output file name (or press Enter for temporary file, Ctrl+O for Dialog Box):

Extra specifications (if any). Press Enter to analyze:

Temporary Workfile Directory: C:\DOCUME~1\Mike\LOCALS~1\Temp\
Reading Control Variables ..
Input in process:
Input Data Record:
1211102012222021122021020 M Rossner, Marc Daniel
^I ^N ^P
75 KID Records Input.

CONVERGENCE TABLE
-Control: \examples\example0.txt Output: \examples\Z0U044WS.TXT
| PROX ACTIVE COUNT EXTREME 5 RANGE MAX LOGIT CHANGE |
| ITERATION KID ACT CATS KID ACT MEASURES STRUCTURE|
>=====<
| 1 75 25 3 3.78 3.20 3.8918 |
```

Winsteps provides a familiar "pull-down" user interface, intended to provide the user with maximum speed and flexibility. There are three main ways that you direct Winsteps:

(a) **You respond to prompts on the screen.**

There are two frequent ones:

Name of control file:

This is the ASCII file where your control specifications reside.
You can press **Enter** to browse for it.

Report output file name:

Press Enter for a temporary output file, or
Type in an output file name or use the pull-down file menu
The report output file is always in text format.

Extra specifications (if any). Press Enter to analyze):

Press Enter
This is used for making [just-in-time changes](#) to your control file instructions, for this run only.

If these prompts do not appear, see [Edit Initial Settings](#)

(b) You use the pull-down menus.

A frequent choice is the first choice on both the [File](#) and [Edit](#) menus:

Edit Control File=

Select this to edit your Winsteps control file.

(c) You use the Notepad text editor.

All editing and display of Winsteps output is done using text files and Notepad or [your own text editor](#).

This gives you great flexibility to:

- modify control and anchor files
- view, copy and paste output into Word (or other) files

8.6 Uninstalling WINSTEPS

Depending on the installation procedure:

- (a) Select "Uninstall" from the **Programs** menu
or
- (b) Go to "Settings", "Control Panel", "Add/Remove Programs" and double-click on "Winsteps Uninstall"
- (c) Delete Winsteps directory, and delete the Winsteps entries from "Windows\Start Menu\Programs"
- (d) Use a Windows clean-up utility to tidy up loose ends.
- (e) Delete files in C:\TEMP and C:\WINDOWS\TEMP (or your Windows temporary file) and files ending *.....ws.txt*

8.7 Stopping WINSTEPS

The Winsteps program ceases execution when

1) The program stops itself:

The estimation procedure has reached an acceptable level of convergence and all pre-specified output has been produced. This happens when:

- a) The estimates are within the convergence criteria ([LCONV=](#) and [RCONV=](#) as controlled by [CONVERGE=](#))
- b) The maximum number of iterations has been reached ([MPROX=](#) and then [MJMLE=](#))

To instruct Winsteps to run indefinitely (up to 2,000,000,000 iterations), set

```
MJMLE=0  
LCONV=0  
RCONV=0  
CONVERGE=F
```

- c) The estimates are not improving. This can occur when the limits of the computational precision of your computer have been reached.

2) You stop the iterative process:

- a) If you press Ctrl with F (or use [File](#) menu) during [PROX](#) iterations:

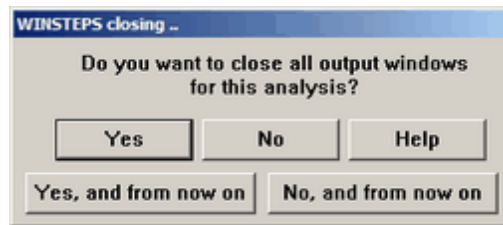
PROX iteration will cease as soon extreme scores have been identified and point-biserial correlations have been calculated. JMLE iterations then start.

- b) If you press Ctrl with F during [JMLE](#) iterations:
JMLE iteration will cease at the end of this iteration. Fit statistics will then be calculated and output tables written to disk.
- c) If you press Ctrl with F during the output phase:
Output will cease at the end of the current output operation.
Acknowledgment of your Ctrl with F instruction is shown by the replacement of = by # in the horizontal bar drawn across your screen which indicates progress through the current phase of analysis.

3) You cancel Winsteps execution immediately:
From the Winsteps File menu, choose Exit.
No more analysis or output is performed.

When Winsteps exits ...

It deletes all temporary files it has created and releases memory. You may have output Tables, files or graphs open on your screen. Winsteps asks if you want these closed.



Yes: close all open windows. If some windows have been modified, but not saved, you may be asked if you want to save those.

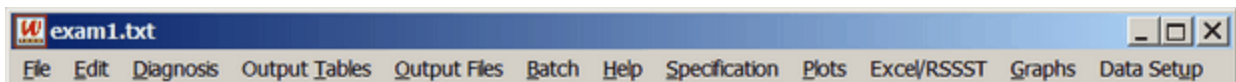
No: leave all windows as they are, but close the Winsteps analysis window.

To make Yes or No the automatic standard, click "**and from now on**". This choice may be reset in [Edit Initial Settings](#)

9 Menus

9.1 Menu bar

Winsteps analysis window has a useful set of pull-down menus:

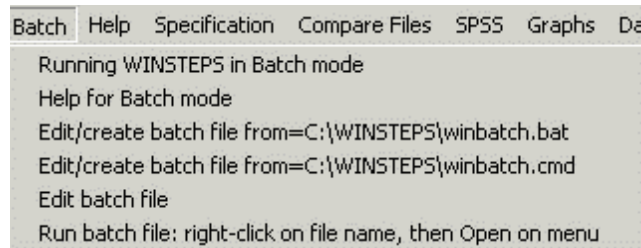


The top line shows the name of the file currently being analyzed. The name can be changed with [Edit Task Bar caption](#). This is useful when you are copying the window (using, for instance, your PrintScreen key) to include in a document.

File	overall control of the analysis
Edit	display and editing of input and output files and Tables
Diagnosis	Tables for understanding, evaluating and improving your measurement system
Output Tables	produces all output Tables produced by Winsteps
Output Files	produces output primarily intended for input into other software
Batch	facilitates running Winsteps in batch mode
Help	displays Help file
Specification	allows entry of specifications after the analysis, one at a time, in the form of specification=value
Plots	uses Excel to display and compare analyses
Excel/RSSST	reformat Excel .xls, R .rda, SAS .sas7bdat, SPSS .sav, STATA .dta, and Tabbed-text .txt files into Winsteps control and data files
Graphs	computer-graphics for test, item and category display
Data Setup	provides a immediate means for setting up control and data files

9.2 Batch menu

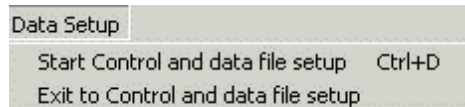
This facilitates running Winsteps in batch mode.



Running Winsteps in Batch mode: Summary instructions for running Winsteps in batch mode.

Help	Displays help information for batch mode
Edit	Edit batch file
Run	Run batch file: done by right-clicking on batch file name (.bat or .cmd), then clicking on open on the Windows right-click menu

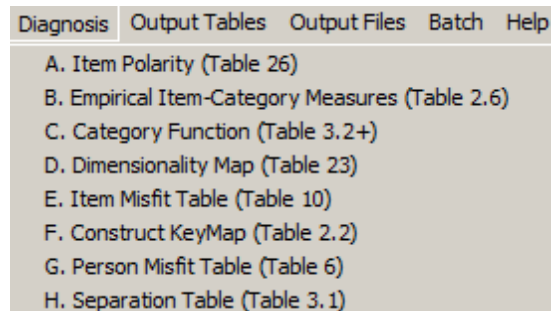
9.3 Data Setup menu



This is described at [Control and data file setup](#).

9.4 Diagnosis menu

The Diagnosis pull-down menu suggests a step-by-step procedure for investigating the results of your analysis.



A. Item Polarity: check that all items are aligned in the same direction on the latent variable, same as [Table 26](#). Check that all items have positive correlations. Use [IREFER=](#) and [IVALUE=](#) to point all items in the same direction, or [KEY1=](#) to correct a multiple-choice key error. [IDFILE=](#) to delete (for the moment) uncooperative items.

B. Empirical Item-Category Measures: check that all categories for all items are aligned in the same direction, same as [Table 2.6](#). For multiple-choice items, see [Table 2](#) for MCQ. Check that correct answers, and higher category values corresponding to "more" of the variable, are to the right.

C. Category Function: check that all categorization functioned as intended, same as [Table 3.2](#) onwards. Check that the "[average measures](#)" for the categories advance, and that no category is especially noisy. Use [IREFER=](#) and [IVALUE=](#) to collapse or remove discordant categories. Use [ISGROUPS=](#) to identify category functioning. If more details are required, look at the [option/distractor](#) analysis of the Item Tables.

D. Dimensionality: check that all items share the same dimension, same as [Table 23](#). This identifies sub-structures, "secondary dimensions", in the data by performing a [principal components/contrast](#) decomposition of the observation residuals. If there are large sub-structures, then it may be wiser to divide the data into two measurement instruments.

E. Item Misfit: check that items cooperate to measure, same as [Table 10](#). Are there misbehaving items? Look for large mean-squares, and also for contradictory use of responses in the [option/distractor](#) listing.

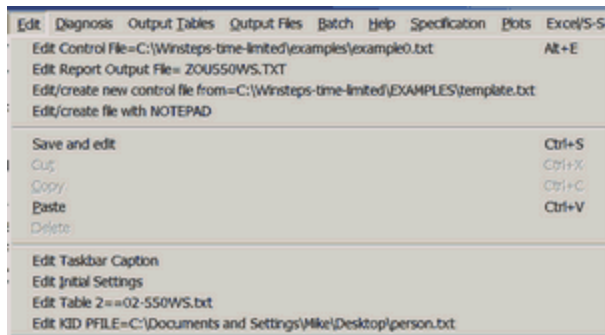
F. Construct KeyMap: check that the item hierarchy is as intended (construct validity), same as [Table 2.2](#). This locates items, response categories and your sample in one picture. Does your item measure hierarchy make sense? What is the typical person in your sample saying?

G. Person Misfit: check that persons cooperate to measure, same as [Table 6](#). Are there misbehaving persons? Look for large mean-squares, and also look at the unexpected observations in the Most Unexpected Responses subtable.

H. Separation: check that the items discriminate different levels of person performance ("test" [reliability](#)), same as [Table 3.1](#). Also that persons are able to discriminate differences in item calibration.

9.5 Edit menu

Display and editing of input and output files and tables.

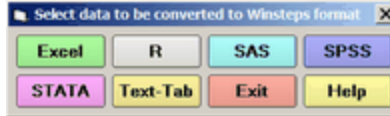


Edit Control File=	display and edit the current control file. Alters this analysis if no computation has been done, otherwise the next analysis done with this control file
Edit Report Output File=	display and edit the report output file written during the main analysis phase. This contains Table 0 and output specified with TABLES= and TFILE=
Edit/create new control file from=\template.txt	template.txt is a generic control file which can be edited and saved under another name to give you a flying start at setting up your own analysis. There is a control and data file setup procedure. It is easier!
Edit/create file with Notepad	launches Notepad or your own text editor
Save and edit	save the Winsteps analysis window to a disk file and open for editing
Cut	copy characters from an analysis window line to the Windows clipboard and delete them from the screen
Copy	copy characters from an analysis window line to the Windows clipboard
Paste	paste characters from the Windows clipboard to an analysis window line
Delete	delete character from an analysis window line
<i>for more substantial editing, save the analysis screen using the File pull-down menu</i>	

Edit Taskbar Caption	change the task name displayed on the Windows taskbar
Edit initial settings	change standard files and settings in Winsteps.ini
Edit Table ...	display and edit a Table produced from the Diagnosis, Output Tables or other pull-down menu
Edit ... File	display and edit a file produced from the Output Files or in other ways
Formatted data==	Formatted data after FORMAT= .

9.6 Excel/RSSST menu

Convert data files from proprietary formats into Winsteps files.



Excel: This is described in [Data from Excel files](#)

R: This is described in [Data from R Statistics files](#)

SAS: This is described in [Data from SAS files](#)

SPSS: This is described in [Data from SPSS files](#)

STATA (version 13 and earlier): This is described in [Data from STATA files](#)

Text-Tab: This is described in [Data from Text files \(with Tabs\)](#)

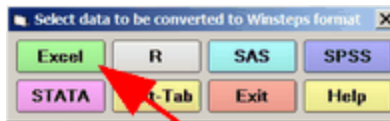
Winsteps does not function with every version of these programs.

Suggestion: Save your input file as a tab-separated DOS text file.

Then select Text-Tab input

The same procedure applies to all these file types. We will use Excel as the example.

0. Use the Winsteps Excel/RSSST menu: Excel option:

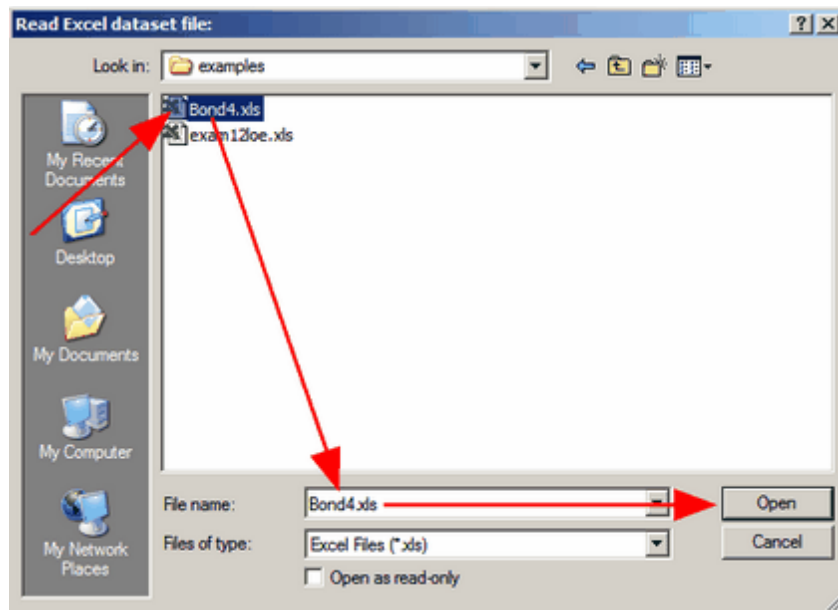


1. Select the Input file to be converted:



and more

2. Then select the desired input file and click "Open"

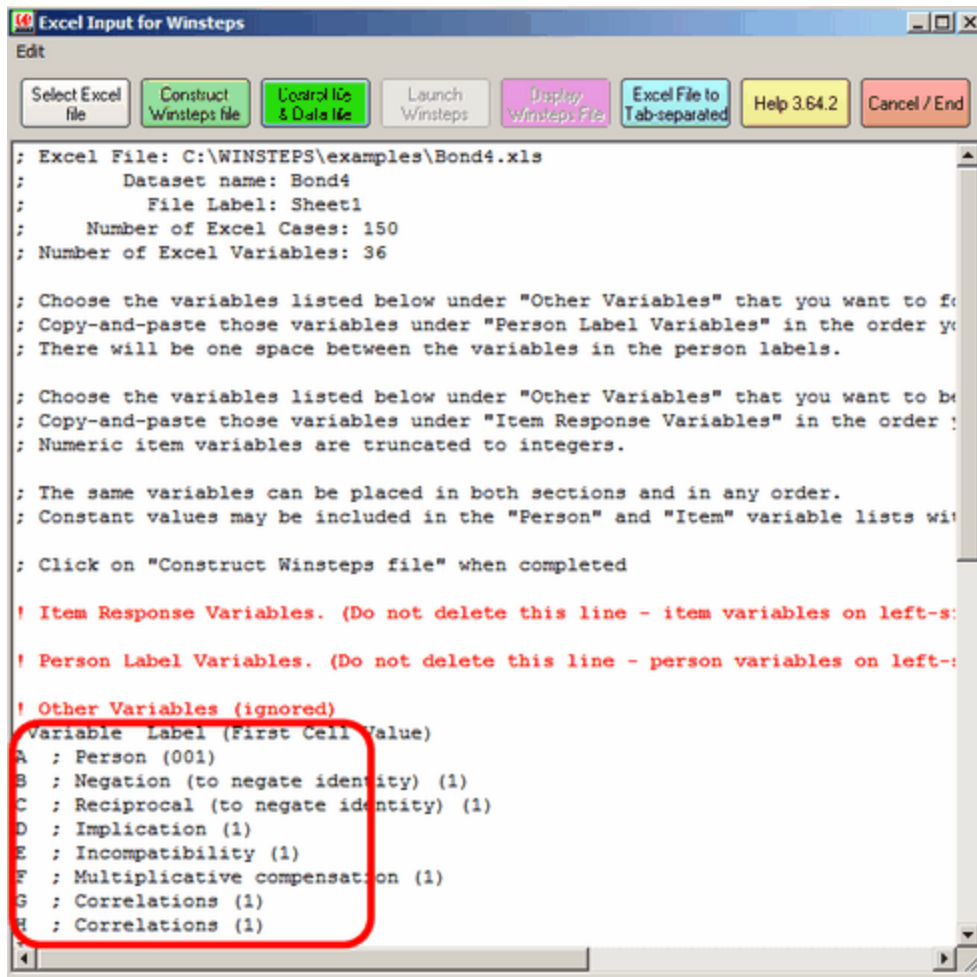


3. Look at your input variables.

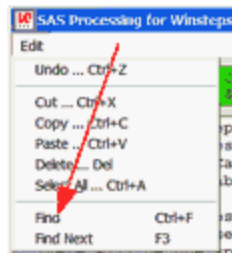
In the Excel spreadsheet, the first row is the variable names. Then each row is one person (subject, case). Each column contains one variable: person name, item response, demographic variable.

A summary report about your input file, and some brief instructions are shown.

The input variables are listed in the selection window under "Other Variables". They are identified by their column letters. You can also select "Case number" and "Row number" when available.



If your variable list is long, use "Edit", "Find" to locate the variables:



4. Copy-and-paste the Item Response Variables

Which are the variables that contain the responses, observations, ratings, etc. that are to be analyzed by Winsteps? Copy-and-Paste the variable names under "! Item Response Variables". You can change the order if you want to. In this example, I have alphabetized the item variables by what will be the Winsteps item labels.

```

; Click on "Construct Winsteps file" when c
! Item Response Variables. (Do not delete t
M ; Affirmation of p (1)
J ; Conjunction (1)
G ; Correlations (1)
H ; Correlations (1)
I ; Correlations (1)
K ; Disjunction (1)
D ; Implication (1)
E ; Incompatibility (1)
F ; Multiplicative compensation (1)
B ; Negation (to negate identity) ; Item 1
C ; Reciprocal (to negate identity) (1)
! Person Label Variables. (Do not delete th

```

How to copy-and-paste:

- Drag your mouse pointer over an SPSS variable name in the "Other Variables" list to highlight it.
- Press your Ctrl key. Hold it down and press the C key. Release the keys. The variable name is copied to the Windows clipboard.
- Position the mouse pointer on the line under "Item Response Variables"
- Press your Ctrl key. Hold it down and press the V key. Release the keys. The variable name is pasted into the window.
- Repeat the process for all the other variable names. You can also highlight a block of variable names at the same time.

5. Copy-and-paste the Person Label Variables

Which are the variables that contain the person identification, demographics, etc. that are to be reported by Winsteps? Copy-and-Paste the variable names under "! Person Label Variables". You can change the order if you want to. You can also reuse an Item Variable.

In this example, the person label includes, the "Person" column, one of the items "Disjunction" because I want Winsteps to report person sub-totals based on responses to that item, and also the Excel row number, in case I need to refer back to the original Excel spreadsheet.

```

B ; Negation (to negate identity) ;
C ; Reciprocal (to negate identity)
! Person Label Variables. (Do not delete)
A ; Person (001)
K ; Disjunction (1)
@Row number ; in Excel spreadsheet
! Other Variables (ignored)
;Variable Label (First Cell Value)
A ; Person (001)

```

6. Optional extras

A. In each Winsteps data line, the item responses are to the left and the person labels are to the right:

```
111010011100111 Jose M
```

If you prefer the person labels to the left, and the item responses to the right:

```
Jose M 111010011100111
```

then cut-and-paste "!Person Label Variables" and its variables above "! Item Response Variables"

B. You can include constant values in the Item or Person sections. For instance, perhaps this dataset is from Region "B23" and you want this in every person label.

C. Item responses then person labels, or Person labels then item responses.

Item responses, then person labels:

```

; Click on "Construct Winsteps file" when completed
! Item Response Variables. (Do not delete this line -
wr1_writing_1 ; F8.0 First items on the test
wr2           ; F8.0
wr3           ; F8.0
! Person Label Variables. (Do not delete this line -
name         ; F8.0

```

Person labels, then item responses:

```

; Click on "Construct Winsteps file" when completed
! Person Label Variables. (Do not delete this line -
name         ; F8.0
! Item Response Variables. (Do not delete this line
wr1_writing_1 ; F8.0 First items on the test
wr2           ; F8.0
wr3           ; F8.0

```

In this example, the Person Variables are moved before the Item Variables, so they will appear to the left in the Winsteps file. Constant "B23" is included.

```

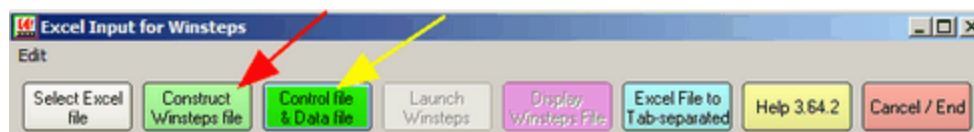
; Click on "Construct Winsteps file" when comp:
! Person Label Variables. (Do not delete this :
A ; Person (001)
K ; Disjunction (1)
"B23"
@ROW number ; in Excel spreadsheet
! Item Response Variables. (Do not delete this
M ; Affirmation of p (1)
J ; Conjunction (1)
G ; Correlations (1)
H ; Correlations (1)
I ; Correlations (1)
K ; Disjunction (1)
D ; Implication (1)
E ; Incompatibility (1)
F ; Multiplicative compensation (1)
B ; Negation (to negate identity) ; Item label:
C ; Reciprocal (to negate identity) (1)
! Other Variables (ignored)
;Variable Label (First Cell Value)
A ; Person (001)

```

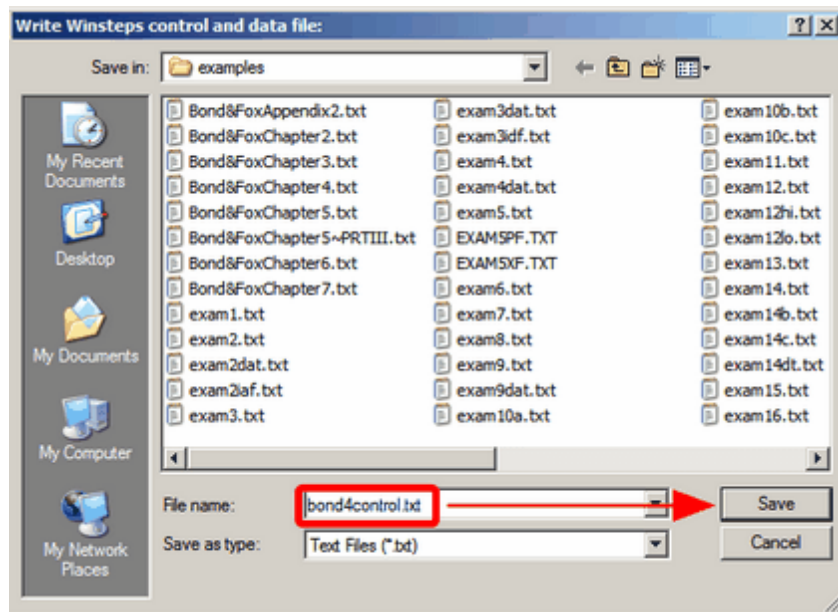
To save this information for a later file conversion, Select All (ctrl+A), copy (ctrl+C) and paste (ctrl+V) into a convenient document.

7. Construct the Output file, a Winsteps control and data file.

Click on "Construct Winsteps File":



Type in the name of the new Winsteps control and data file. "bond4control.txt" in this example. Click on "Save".



If you want a separate Winsteps control file and a Winsteps data file, then click on "Control file & Data file", and enter the name of the control file into the first file dialog box, and the name of the data file into the second file dialog box.

The Winsteps control and data file displays in a text-editor window.

```

bond4control.txt - WordPad
File Edit View Insert Format Help

&INST
Title= "C:\WINSTEPS\examples\Bond4.xls"
; Excel file created or last modified: 10/28/2007 2:41:57 AM
; Bond4
;   Excel Cases processed = 150
; Excel Variables processed = 36
ITEM1= 15 ; Starting column of item responses
NI= 11 ; Number of items
NAME1 = 1 ; Starting column for person label in data record
NAMLEN = 14 ; Length of person label
XWIDE = 1 ; this matches the biggest data value observed
CODES= 01 ; matches the data
TOTALSCORE = Yes ; Include extreme responses in reported score
; Person Label variables: columns in label: columns in line
@Person = 1E3
@Disjunction = 5E5
; @B23 = 7E9
@Rot = 11E13
&END ; Item labels follow: columns in label
Affirmation of p ; Item 1 : 1-1
Conjunction ; Item 2 : 2-2
Correlations ; Item 3 : 3-3
Correlations ; Item 4 : 4-4
Correlations ; Item 5 : 5-5
Disjunction ; Item 6 : 6-6
Implication ; Item 7 : 7-7
Incompatibility ; Item 8 : 8-8
Multiplicative compensation ; Item 9 : 9-9
Negation (to negate identity) ; Item labels courtesy of Trevor
Reciprocal (to negate identity) ; Item 11 : 11-11
END NAMES
001 1 B23 2 111111111111
002 1 B23 3 111111111111
003 1 B23 4 111111010111
004 1 B23 5 111111111111
005 1 B23 6 111111111111
006 1 B23 7 111111111111
007 1 B23 8 111111111111
008 1 B23 9 111111111111

```

Winsteps includes `ISGROUPS=0` in the control file because Winsteps does not know the structure of your data. Please delete `ISGROUPS=0` if your data are dichotomous or use the Andrich Rating-Scale model. If `ISGROUPS=0` is in the Winsteps control file, the partial-credit model is used.

If one of your data lines is an answer-key then:
 Cut-and-paste the answer-key data line.
 Remove it from the data section.
 Paste it before `&END`
 Edit it to start:
`KEY1= (answer-key)`
 Save the control and data file.

8. Click on "Launch Winsteps" to analyze this control and data file.



9. Click on "Display Winsteps File" to redisplay the Winsteps control and data file, if you need to see them again.

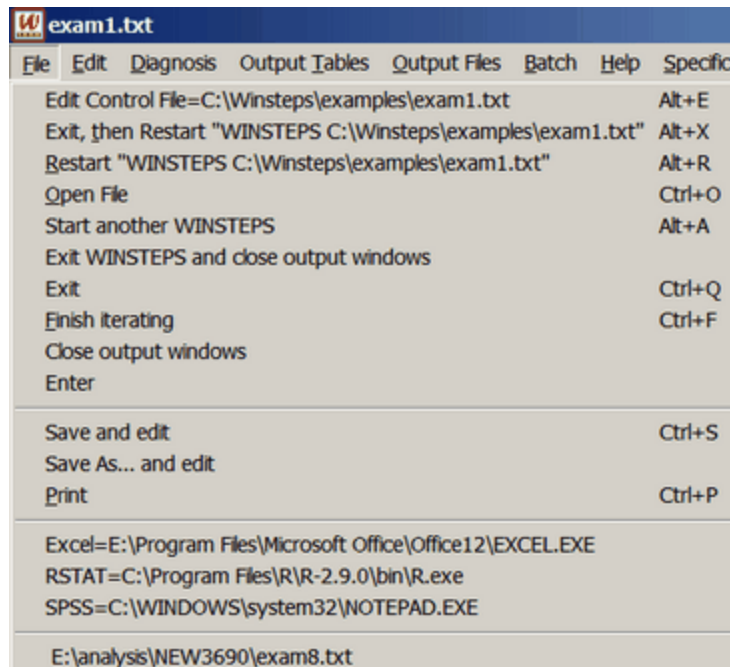
10. Click on "File to Tab-Separated" to display the contents of an input file in tab-separated format.

Tab-separated files can be conveniently pasted into Excel. The tab-separated values display in a temporary file. Copy and paste the contents, or "Save as" for a permanent copy.

11. Click on "Cancel / End" to conclude this dialog.

9.7 File menu

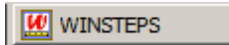
This menu launches and terminates Winsteps analysis.



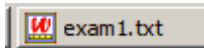
Edit Control File=	Edit the current control file. Alters this analysis if no computation has been done, otherwise the next analysis done with this control file.
Exit, then Restart "WINSTEPS ..."	Stop and then restart this analysis, usually after editing the control file.
Restart "WINSTEPS ..."	Restart this analysis, leaving the current one running.

Open File	Select control file for this analysis
Start another Winsteps	Launch a new copy of Winsteps. More than one copy of Winsteps can run at the same time.
Exit WINSTEPS and close output windows	Exit from Winsteps and close output windows without asking.
Exit	Exit from Winsteps
Finish iterating	Finish the current phase as quickly as possible.
Close open output windows	close any open windows for output tables, files or plots.
Enter	Acts as the Enter key
Save and edit	Save the information displayed on the processing screen to disk.
Save As... and edit	Save the screen output to a named disk file.
Print	Print the screen output
Excel=	Location of the Excel program (if installed on your computer): can be change in Edit Initial Settings
RSTAT=	Location of the R Statistics program (if installed on your computer): can be change in Edit Initial Settings
SPSS=	Location of the SPSS program (if installed on your computer): can be change in Edit Initial Settings
(file name)	Up to 30 must recently used Winsteps control files. Click on one of these to analyze it

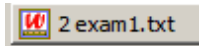
When you launch Winsteps, the Windows task bar shows:



After a control file is selected, the task bar indicates it:



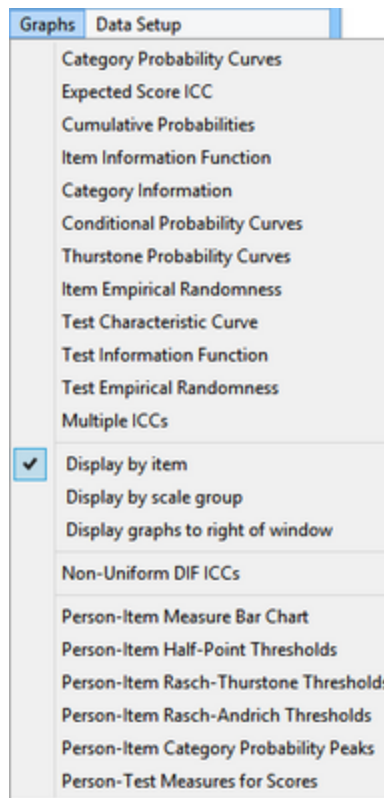
When a second run accesses the same control file:



and the number increments for further uses of the same control file.

9.8 Graphs menu

Winsteps produces bit-mapped images, using the **Graphs** menu. Winsteps produces character-based graphs in [Table 21](#)

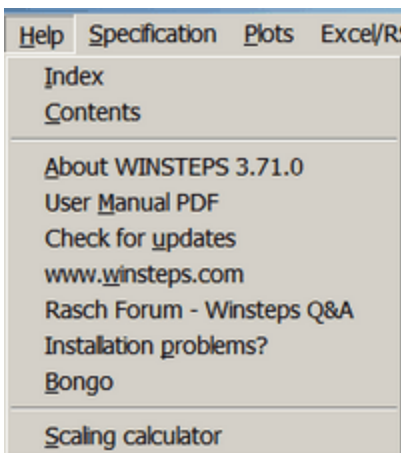


Initially, select which type of curves you want to see. When the graphs display, you can select the other options. You can look at the others later without going back to this menu. Graphs are plotted relative to the central [difficulty](#) of each item or response structure unless "Absolute Scaling" is clicked in the Graphs window. Model-based curves (such as probability and information functions) are the same for all items which share the same model definition in [ISGROUPS=](#). Empirical curves differ across items and rating scales.

Category Probability Curves	model-based probability of observing each category of the response structure at each point on the latent variable (relative to the item difficulty)
Expected and Empirical Score ICC	model-based Item Characteristic Curve (or Item Response Function IRF) for the item or response structure. This is controlled BYITEM= or the last two entries in this menu.
Cumulative Probabilities	model-based sum of category probabilities. The Thurstone thresholds (category median boundaries) are the points at which the probability is .5. Click on a line to obtain the category accumulation.
Item Information Function	model-based Fisher statistical information for the item. This is also the model variance of the responses, see RSA p. 100.
Category Information	model-based item information partitioned according to the probability of observing the category. Click on a line to obtain the category number.
Conditional Probability Curves	model-based relationship between probabilities of adjacent categories. These follow dichotomous logistic ogives. The Andrich thresholds (points of equal probability of adjacent thresholds) are the points at which the probability is .5. Click on a line to obtain the category pairing.
Thurstone Probability Curves	model-based sum of category probabilities. The Thurstone thresholds (category median boundaries) are the points at which the probability is .5. Click on a line to obtain the category accumulation.
Item Mean-Square Fit	observed randomness (mean-square fit) in each interval on the variable with logarithmic scaling. The model expectation is 1.0
Test Characteristic Curve	model-based test score-to-measure characteristic curve.
Test Information Function	model-based test information function, the sum of the item information functions.

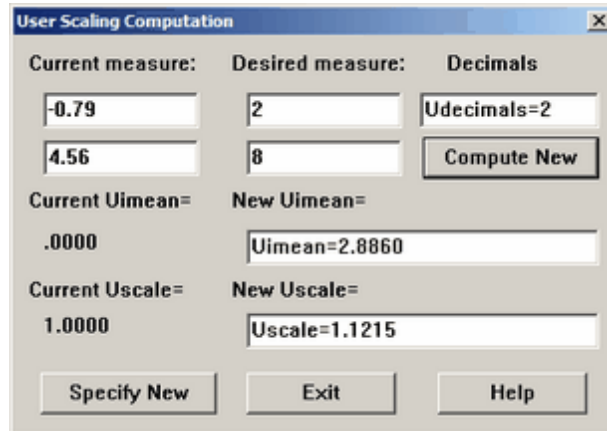
Test Mean-Square Fit	the observed randomness (mean-square fit) in each interval on the variable with logarithmic scaling. The model expectation is 1.0
Multiple Item ICCs	displays several model and empirical Item Characteristic Curves / Item Response Functions simultaneously. Click on MultiICC in Graph window
Display by item	shows these curves for individual items, also controlled by BYITEM= . Model-based output is the same for all items with the same ISGROUPS= designation.
Display by scale group	for each ISGROUPS= code, a set of curves is shown. An example item number is also shown - all other items in the grouping are included in the one set of grouping plots. Also controlled by BYITEM= .
Display graphs to right of window	The graphs are displayed in the bottom right of the Graphs window so that they can be pulled larger. Clicking this alternates between the standard Graph window and the bottom-right window .
Non-Uniform DIF ICCs	the empirical item characteristic curve for each DIF= person-classification-group.
Vertical histograms showing the the distributions of the person abilities and ...	
Person-Item Difficulty Histograms	item difficulties
Person-Item Half-Point Thresholds	rating-scale half-point thresholds between categories: similar to Table 1.5
Person-Item Rasch-Thurstone Thresholds	rating-scale Rasch-Thurstone thresholds between categories: similar to Table 1.6
Person-Item Rasch-Andrich Thresholds	rating-scale Andrich thresholds between categories: similar to Table 1.7
Person-Item Category Peaks	rating-scale category centers: similar to Table 1.8 , where expected score on the item is the category number, and also the category has the highest probability of being observed.
Person-Test Measures for Scores	locations of the measures corresponding to scores on the currently active set of items

9.9 Help menu



Index	displays the Index of control variables in the Help file
Contents	displays the Table of Contents of the Help file.
About WINSTEPS	displays Winsteps version number and compile date. Please mention these when reporting problems.
User Manual PDF	displays the Winsteps Help file as a printable User Manual PDF.
Check for updates	shows webpage of Winsteps updates on www.winsteps.com/wingood.htm

www.winsteps.com	takes you to our website: www.winsteps.com	
Installation problems?	shows webpage of Winsteps installation problems and solutions on www.winsteps.com/problems.htm	
Bongo	is the Winsteps "Adjutant's Bugle Call" - play this when you are summoning the data in preparation for constructing measures!	梵語
Scaling calculator	is designed to help you linearly rescale your measures in the way most meaningful for your audience:	



Under **Current measure**: enter two measures from your current analysis.

Under **Desired measure**: enter the values with which you want them to be reported.

Under **Decimals**: enter the number of decimal places for the measure Tables, [Udecimals=](#).

Press **Compute New** to calculate the revised values of [Uimean=](#) and [Uscale=](#).

The current values of Uimean= and Uscale= are displayed and also the revised **New** values. The New values can be altered if you wish.

Press **Specify New** to action the **New** values. Or the values can be copied (Ctrl+c) and pasted into your Winsteps control file.

Also clicking on "[Output Files](#)", "Control Variable List" produces the full list of control variables including the new values of UIMEAN= and USCALE=

Example: We want the mean and S.D. of our sample to be 0 and 1.

Current Table 3.1

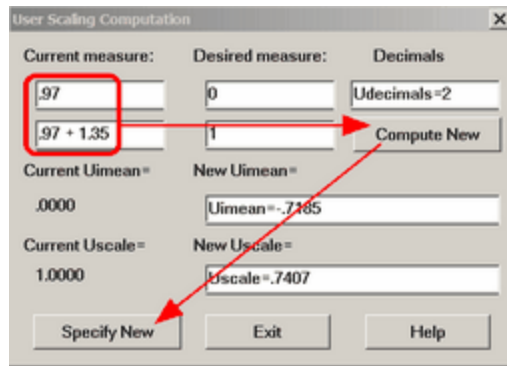
	RAW		
	SCORE	COUNT	MEASURE
MEAN	31.7	25.0	.97
S.D.	8.7	.0	1.35

The current mean = .97

The current mean + 1 S.D. = .97 + 1.35

The desired mean = 0

The desired mean + 1 S.D. = 0+1



New Table 3.1

	RAW	COUNT	MEASURE
	SCORE		
MEAN	31.7	25.0	.00
S.D.	8.7	.0	1.00

Example: If I use [IAFILE=](#) and [SAFILE=](#) for some items, how do I rescale the analysis to set the reported person measure range as 0-100?

1. Do your analysis with [IAFILE=](#) and [SAFILE=](#). [USCALE=](#) in this analysis is the value of [USCALE=](#) in the analysis that output the [SAFILE=](#)
2. [Output Tables menu: Table 20.1](#). Identify the measure you want to be 0, and the measure you want to be 100
3. Help menu: Scaling calculator
 "Current Measure:" are the two measures from Table 20.1
 "Desired Measure:" are the two measures you want
 Click on "Compute New:" write down the value of [Uscale=](#) for your next analysis
 Click on "Specify New"
4. [Output Tables menu: Table 20.1](#). The range should be 0-100.
5. Output [IFILE=if.txt](#), [SFILE=sf.txt](#)
6. Do the analysis again: [IAFILE=if.txt](#), [SAFILE=sf.txt](#) and [USCALE=](#) the value of [Uscale=](#) from 3.
7. [Output Tables menu: Table 20.1](#). The range should be 0-100.

9.10 Output Files menu

The Output Files menu produces output primarily intended for input into other software. When a selection is clicked on, the [output file specifications](#) dialog box is usually shown so that you can choose the output file format.

For details, please see [Output Files Index](#)

9.11 Output Tables menu

Output Tables are listed in the [Output Table Index](#). They are written into temporary files if selected from the Output Tables menu. Output Tables are written into the Report Output File if specified using [TABLES=](#) or [TFILE=](#) in the control file. [Table 0](#) is always written to the Report Output File.

Output Tables are always in text format (.txt). To export a Table in Excel format see [Excel-formatted Output Tables](#).

For details of the Output Tables menu, please see [Output Tables Index](#)

3.2 Rating (partial credit) response-structure and most Tables shown

Click on the Table to write it to a file and show it on the screen. Here is "3.2 Rating response-structure Structure". It is written into temporary file 03-859ws.txt. "03" refers to Table number 3. "859" is a unique number for this analysis. "ws.txt" means "Winsteps text file".

TABLE 3.2 LIKING FOR SCIENCE (Wright & Masters p. ZOU859ws.txt Oct 9 10:54 2002
 INPUT: 76 PUPILS, 25 ACTS REPORTED: 75 PUPILS, 12 ACTS, 3 CATS WINSTEPS 3.36

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY	OBSERVED	OBSVD	SAMPLE	INFINIT	OUTFIT	ANDRICH	CATEGORY			
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	THRESHOLD	MEASURE	
0	0	667	33	-1.30	-1.30	.96	.95	NONE	(-2.04)	00 dislike
1	1	757	37	-.08	-.09	.90	.78	-.82	.00	01 neutral
2	2	609	30	1.40	1.41	1.09	1.33	.82	(2.04)	02 like

AVERAGE MEASURE is mean of measures in category.

Tables 27, 28, 30, 31, 33

These all allow the user to change the relevant control command on execution. ISUBTOTAL= controls the sub-total segments for Table 27 with a selection command so you are asked to confirm or change this value, before the Table is produced.

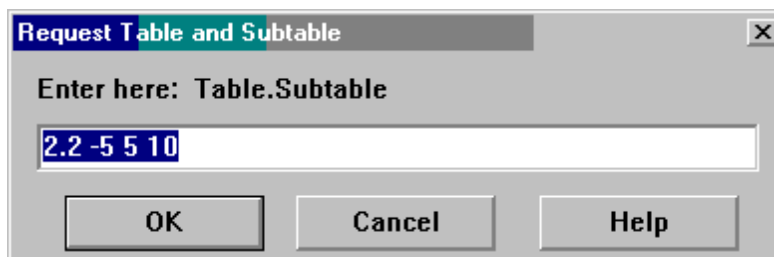


Tables 23, 30, 31, 33

These allow the user to produce Excel plots of parts of the Table.

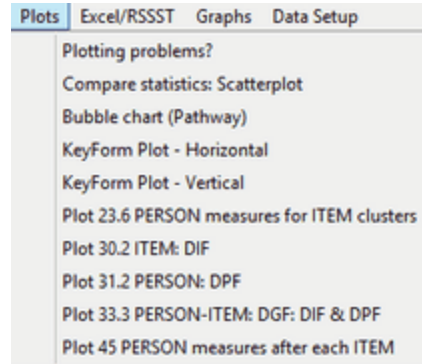
Request Subtables

Any Table (except Table 0) can be displayed using this command. It also accepts the special fields available with TFILE=



Weight Selection. See weighting. When IWEIGHT= or PWEIGHT= are used in estimation, reports can be adjusted to reflect those weights or not. Weights of zero are useful for pilot items, variant items or persons with unusual characteristics. These can be reported exclusively or excluded from reports.

9.12 Plots menu



Here is the [Plots menu](#).

Plotting problems?	these are usually due to the Winsteps-Excel interface. See www.winsteps.com/problems.htm
Compare statistics: Scatterplot	enables you to draw scatter-plots (xy-plots) of Winsteps statistics within or between analyses. It also produces the tabular output of Table 34 .
Bubble chart (pathway)	generates a Bond & Fox-style bubble chart.
Keyform Plot - Horizontal	generates a horizontal Keyform layout.
Keyform Plot - Vertical	generates a vertical Keyform layout.
Plot 23.6 Person measures for Item clusters	scatterplots the person measures in Table 23.6 on the 3 clusters of items in Table 23.1
Plot 30.2. (Item): DIF	DIF plots the DIF values in Table 30.2 .
Plot 31.2. (Person): DPF	DPF plots the DPF values in Table 31.2 .
Plot 33.3. (Person-Item): DGF: DIF& DPF	DIF & DPF plots the DIF+DPF values in Table 33.3 .
Plot 45. Person measures after each item	plots the values in Table 45

9.13 Right-click menus

Mouse Right-Click on the Winsteps screen

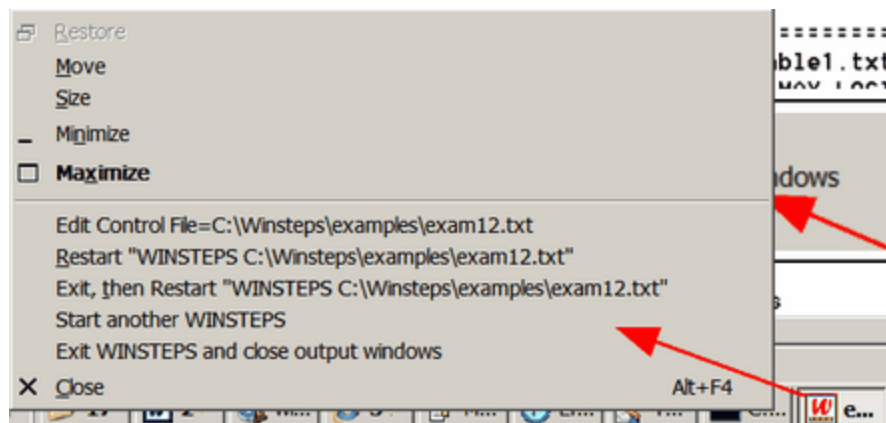
Right-click anywhere on the Winsteps screen, and a menu will be displayed.

During estimation: It is the [File menu](#).

After estimation: It is the [Output Tables menu](#).

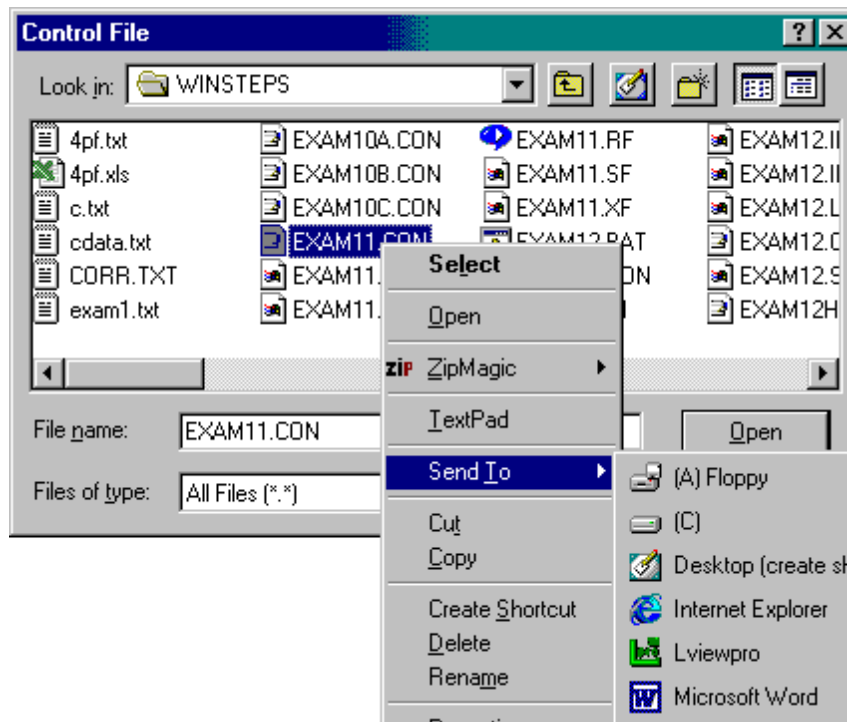
Mouse Right-Click on the Task Bar

Right-click on the task bar, to obtain the Winsteps system menu. This has much of the functionality of the [File menu](#).



Mouse Right-Click on File Dialog Box:

Right-click a file name, and get the **Send To** menu.



Add functionality to the **Send To** menu by copying short-cuts into `c:\windows\SendTo` (or the equivalent SendTo folder in your version of Windows) - a useful program to add is WordPad or [your own text editor](#). To do this:

- Start
- Find
- Files or Folders
- Named: WordPad in C:
- when Wordpad.exe appears, Right-click on it.
- Send To: Desktop: Create shortcut
- Exit from Find
- On Desktop:
- My Computer
- C:
- Windows (if this does not appear: then View: Folder Option: View: Show all files)
- Send To
- Drag WordPad shortcut from Desktop into SendTo folder.
- WordPad is now in the Send To menu.

9.14 Specification menu

This allows entry of [most specifications](#) after the analysis, one at a time, in the form of *specification=value*. Click on "OK" to action the specification and return to the standard screen. "OK and again" to action the specification and redisplay this entry box.

In general:

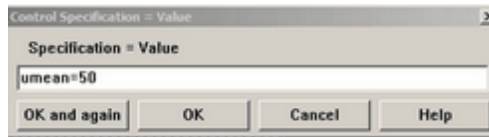
Control variables in the control file or at the "[Extra specification](#)" prompts are applied during measure estimation.

Control variables from the pull-down variables only affect the output.

For instance, [DELETE=](#)

In the control file or at the extra specifications prompt: these deleted items do not take part in the measure estimation.

From the "Specification" pull-down menu: these deleted items are not reported, but have taken part in measure estimation, fit statistics, etc.



[Most specifications](#) can be entered after the analysis has completed. They do not change the analysis but do alter the output. They are useful for making selections (e.g., [PSELECT=](#) and [ISELECT=](#)), setting output Table control values (e.g., [MRANGE=](#)) and changing user-scaling (e.g., [USCALE=10](#)).

"not done" in the Analysis window indicates that

1. a command that should be implemented before estimation (usually by inclusion in the control file) is being done after estimation (usually from the Specification dialog box). For instance, since anchor values (IAFILE=) is implemented before estimation, then UASCALE= (which tells Winsteps the scaling of the anchor values) must also be implemented before estimation.
2. a command has no effect or removes all the data

Multi-line specifications and the Specification pull-down menu:

Specifications with "file name only" do not support "***".

- [CFILE=](#) scored category label file: file name only, blank deletes labels
- [CLFILE=](#) codes label file: file name only, blank deletes labels
- [IDFILE=](#) item deletion file: file name only: blank resets temporary deletions
- [ILFILE=](#) item label file: file name only: blank not allowed
- [PDFILE=](#) person deletion file: file name only: blank resets temporary deletions

So, instead of

- ```
CLFILE=*
1 strongly disagree
2 disagree
3 agree
4 strongly agree
*
```

use the [Edit menu](#), "Create/Edit with Notepad", then, in Notepad, type

- ```
1 strongly disagree
2 disagree
3 agree
4 strongly agree
```

save as "clfile.txt"

and in the Specification dialog box, enter:

```
CLFILE = clfile.txt
```

Example: We want score Table 20 for items 1-12, 14-20, 31

Specification dialog box: [IDELETE=](#)+1-20,13,+31

[Output Tables menu](#): Table 20

Incremental commands

These commands add to the effect of previous commands: IDELETE=, PDELETE=, ISELECT=, PSELECT=

Example 1: In the Specification menu dialog box:

- PDELETE= ; reinstates everyone who is temporarily deleted
- PDELETE= 23 ; temporarily deletes person 23 and reinstates no one
- PDELETE= +23 ; temporarily deletes everyone, then reinstates person 23 if temporarily deleted
- PDELETE= 23 +23 ; temporarily deletes person 23 then reinstates person 23 and deletes no one
- PDELETE= 23 +23 +5 ; temporarily deletes person 23 then reinstates persons 23 and 5 and deletes no one

Example 2: In the Specification menu dialog box:

ISELECT= ; reinstates everyone who is temporarily deselected

ISELECT= A ; deselects everyone whose person label does not start "A" and reinstates no one

ISELECT= ?B ; deselects everyone whose person label does not have "B" in its second column and reinstates no one

ISELECT= A then ISELECT= ?B is the same as ISELECT= AB

10 Control and Data Files

10.1 Control and Data File Setup Window

10.1.1 Control and data file setup window

This interface simplifies setting up Winsteps control and data files. It can be used for entering data with matching specifications or for constructing the specifications that match existing data.

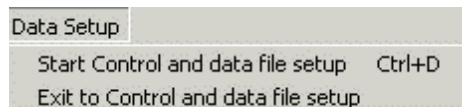
When you launch Winsteps, the Welcome dialog box displays, or use the [Data Setup pull-down menu](#):



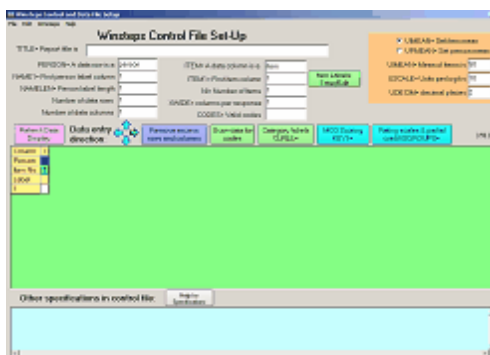
If this does not display, then go to [Edit Initial Settings](#) and click on the "Welcome" option and then OK. Restart Winsteps.

Control + Data Setup Procedure

If you are already in Winsteps, then on the menu bar:



This displays the Control File Set-Up screen:



A multiple-choice test key, KEY1=, can be specified, if desired.

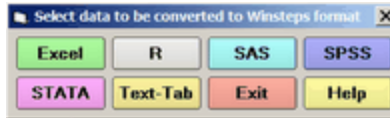
Items can be clustered into similar response-structure groupings using ISGROUPS=, using a one character code for each grouping.

Use the Files menu to read in pre-existing control or data files. Uses the boxes and the data grid to enter new control and data information. Use the Files or Winsteps menu to Save what you have done.

After performing the set-up, save the file and return to Winsteps using the Winsteps pull-down menu.

Import from Excel, R, SAS, SPSS, STATA, Tabbed Text

Click on this for the Excel/RSSST dialog box:



This constructs a Winsteps control and data file from a pre-existing Excel, R, SAS, SPSS, STATA, or Tabbed-Text data file. This has the same functionality as the [Excel/RSSST menu](#)

Text-file Instructions.

Shows the Help page: [Do-it-yourself control and data file construction](#)

All Winsteps control and data files are in text format. These can be edited with Notepad and most other word processing and text editing software.

No.

Closes this dialog box. Use "[Open File](#)" to select a control file.

Don't ask again.

Click on this to make "No." the continuing option. To reinstate this dialog box, go to [Edit Initial Settings](#) and click on the "Welcome" option and then OK. Restart Winsteps.

Help

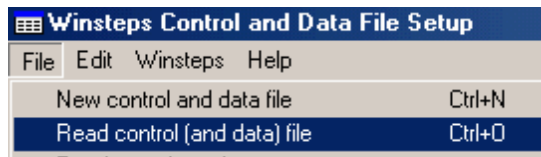
Click on this for this Help page.



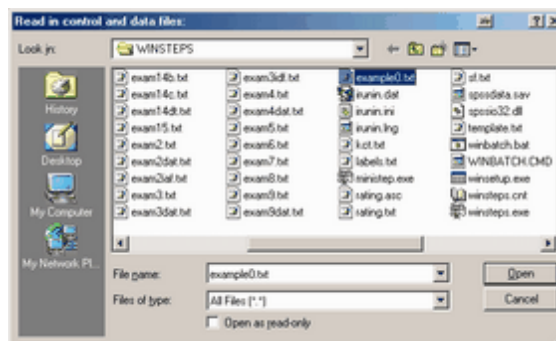
10.1.2 Reading in a pre-existent control file

Reading in a pre-existing control file is easy. You can add or change control specifications, and add or change the data.

From the [Setup](#) screen, use the **File** pull-down menu:



Select your desired pre-existing control or data file:



This fills in the Setup screen:

The control specifications and data are those in the Control and Data file. Control values are filled in as far as possible. The data are filled in with one row per data record and one character per data column.

To see the item labels (between &End and END LABELS in the Control file) either drag the column wider or click on "Item Labels Enter/Edit"

Item deletion: type the specification:

`DELETE=(item number)`

into the box at the bottom of the data setup screen.

This omits the item from your analysis



10.1.3 Data display

This grid displays the data file with one character per column.

During data entry, more columns are automatically added to the right, as needed.

Double-click the extreme row or column for an extra row or column.

Click on "Refresh Data Display" if the display does not show the current specification settings.

Press shift + left-click on the Person or Item No. row to dynamically select person label or item response columns.

Column	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
Person:	R	i	c	h	a	r	d																									
Item No.:																																
Label:																																
ISGROUPS=																																
1	R	i	c	h	a	r	d																									
2	T	r	a	c	i	e																										
3	W	a	i	t	e	r																										
4	B	l	a	i	s	e																										
5	R	o	n																													
6	W	i	l	l	i	a	m																									
7	S	u	s	a	n																											
8	L	i	n	d	a																											
9	K	i	m																													

You can copy-and-paste data into the data-display box. In this box there is one character per cell.

Note: `FORMAT=` is ignored in this display



10.1.4 Item labels

Output reports and displays are much more useful and informative when the items are identified with short, clear identifying labels.

These are usually entered in the specification control file after `&END` and before `END LABELS` (or `END NAMES`). There is one item identifying label per line, so there should be as many lines of item identification as there are items on the test or instrument.

Blank lines and lines containing only comments starting with `;` are ignored between `&END` and `END LABELS`. Items without labels are labeled with their item entry number.

In the Setup routine, they are entered in a special screen.



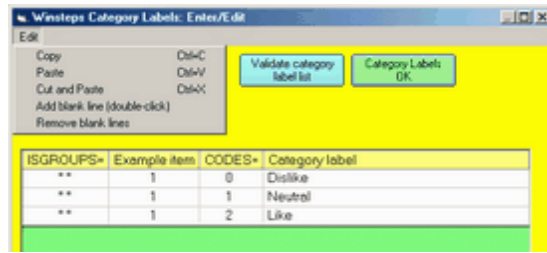
10.1.5 Category labels

Category labels describe the categories in a response structure, levels in a partial credit item, or such like.

Categories are identified in the data codes (`CODES=`). If there are different categories for different items, then use item grouping (`ISGROUPS=`) to identifier clusters of items which share the same category structures. Both of these can be entered on the main set-up screen.

Example grouping " " means that this is the standard common grouping.

Double-click on the bottom line for another blank line.



10.2 Do-it-yourself control and data file construction

There is a [control and data file setup](#) or [Tabbed Text](#) procedure. They are easier!

Here is a step-by-step guide to setting up and running Winsteps. It is a little tricky the first time, but you'll soon find it's a breeze!

The first stage is to set up your data in a rectangular data file in "MS-DOS text with line breaks" format.

1. Obtain your data

You'll need to be a little organized. Think of your data as a wall of pigeon-holes.

(a) Each column corresponds to one item, probe, prompt, task, agent,

For each column, you will need an item name or label. Make these short, only one or two words long. Make a list of these in a document file. Put the label of the first item on the first line, etc.

Put END LABELS on the line after the last item.

Your list will look like this:

```
Eating
Dressing
....
Walking
Stair climbing
END LABELS
```

You can use Notepad or [your own text editor](#) or pull-down the Winsteps "Edit" menu, and select "Create/Edit file with Notepad"

(b) Each row of pigeon-holes corresponds to one person, subject, case, object, ...

You will need some useful identifying codes for each row such as age, gender, demographics, diagnosis, time-point.

Winsteps doesn't require these. but its is much more useful when they appear. Give each of these identifiers one or two letter codes, e.g., F=Female, M=Male, and give each identifier a column of pigeon-holes.

(c) The Data must be carefully lined up.

It is simpler if each data point, observation, response, rating can be squeezed into one character - numeric or alphabetic.

Now create the data file. It will look like something this:

```
M 29 B 001 210212110200102
F 27 W 002 122121010201020
F 32 H 003 222210102112100
M 31 W 004 002010021000210
```

or, less conveniently,

```
M29B001210212110200102
F27W002122121010201020
```


F32H003222210102112100
M31W004002010021000210

After the END LABELS line, or in a separate file,
on each line enter the person identifying codes. Line them up so that each column has a meaning. This is easier if you set
the font to Courier.

Then enter the responses, starting with the first item and continuing to the last.
Do not place spaces or tabs between the responses.
If the lines start to wrap-around, reduce the font size, or increase the page size.

Excel, SPSS, SAS, ACCESS

Your data may already be entered in a spread-sheet, statistics program or database.

"Copy and Paste", Save As, Export or Print to disk the data from that program into "DOS-text with line breaks" or ASCII file.

If the program puts in extra blanks or separators (e.g., Tabs or commas), remove them with a "global replace" in your text
editor or word processor.

To replace a Tab with nothing, highlight the space where a Tab is. Then Ctrl+c to copy. Global replace. Ctrl+V put a Tab into
"From". Put nothing in "To". Action Global Replace.

In Excel, reduce the column width to one column, then
"Save As" Formatted Text (Spaced delimited) (*.prn)

In SPSS, see [SPSS pull-down menu](#).

2. Set up your Winsteps Control and Data file

(a) Edit Template.txt

Pull-down the [Edit menu](#), select "Create new control file from=\Template.txt"

The Template.txt will be displayed on your screen by Notepad or [your own text editor](#).

(b) Template.txt is a Winsteps Control and Data file

Find the three sections:

top: down to &END are Control Specifications
we will edit this in a moment

middle: between &END and END LABELS are the Item Labels
copy and paste the item labels from your list into this area.
one item label per line, in the order of the labels.

bottom: below END LABELS are the data
copy and paste the person labels and responses into this area.
one person per line (row).

(c) Edit the Control Specifications

Find the line "[Title=](#)"

Replace Put your page heading here with your own page heading.

Look at a data line, and count across:

In which column does the person identifying label start, the first position of the person name?	
This is the Name1= value	e.g., if it is column 4, then Name1=-4
How long is the person identifying label, the name length?	

This is the Namlen= value	e.g., if it is 10 columns, then Namlen=10
In which column is the response to the first item?	
This is the Item1= value	e.g., if the first response is in column 12, then Item1=12
How many items are there, the number of items?	
This is the NI= value	e.g., if the number of items is 50, then NI=50
What are the valid item response codes in the data file?	
This is the Codes= value	e.g., if the codes are 1,2,3,4, then Codes=1234

If your codes are not numeric, then you will need to rescore them.
See [Data recoding](#)

This is usually enough for your first Winsteps run.

(d) "Save As" the Template.txt file with your own file name.
Winsteps accepts any valid file name.

3. Run Winsteps

To the prompt:
Control file name? (e.g., KCT.txt). Press Enter for Dialog Box
Press the Enter key
Select your control file from the dialog box
Press the Enter key

Report output file name (or press Enter for temporary file):
Press the Enter key

Extra specifications (if any). Press Enter to analyze):
Press the Enter key
This is used for making [just-in-time changes](#) to your control file instructions, for this run only.

4. Your analysis commences

5. Your analysis concludes.

If there is an error message:
select "Edit Control File=" from the Winsteps Edit menu
correct the control file
save it
select "Exit, then Restart Winsteps" from the [File menu](#)

If "Measures constructed" -
use the [Output Tables](#) pull-down menus to look at the Output Tables
here is the [list of output tables](#).

6. Exit Winsteps using the X in the top right corner.

10.3 Control file and template.txt

The control file tells what analysis you want to do. The template file, TEMPLATE.TXT. gives you an outline to start from. The easiest way to start is to look at one of the examples in the next section of this manual, or on the program disk. The control file contains control variables. These are listed in the index of this manual. Only two control variables must have values assigned for every analysis: [NI=](#) and [ITEM1=](#). Almost all others can be left at their automatic standard values, which means that you can defer learning how to use most of the control variables until you know you need to use them.

When in doubt, don't specify control variables, then they keep their standard values.

Here is a version of TEMPLATE.TXT. Copy and paste this, if your TEMPLATE.TXT is corrupted.

```
; this is a WINSTEPS specification control file template.
; Save it with your own name, e.g., control.txt

; a semi-colon means a comment: remove semi-colons as needed.

&INST ; optional

TITLE = "Put your page heading here"

;Input Data Format
NAME1 = 1 ; column of start of person information
NAMLEN = 30 ; maximum length of person information
ITEM1 = ? ; column of first item-level response
NI = ?? ; number of items = test length
XWIDE = 1 ; number of columns per response
PERSON = Person ; Persons are called ...
ITEM = Item ; Items are called ...

; DATA = ; data after control specifications

; For rescoring
; 0 1 2 3 4 5 6 7
; 123456789012345678901234567890123456789012345678901234567890
;ISGROUPS=0 ; specify that each item has its own response structure (partial credit)
;IREFER=AABBCC....

; Data Scoring
CODES = "01" ; valid response codes
;IVALUEA= "01" ; for rescoring for item type A
;IVALUEB= "10" ; for rescoring for item type B
;IVALUEC= " " ; for rescoring for item type C
; ; Codes in IREFER with no IVALUE are not changed

CLFILE = * ; label the categories in Table 3.2
0 Strongly Disagree ; 0 in the data means "Strongly Disagree"
1 Strongly Agree ; 1 in the data means "Strongly Agree"
*

;NEWSCORE = "10" ; use to rescore all items
; KEY1 = ; key for MCQ items

; XWIDE = 2 ; for all codes 00 to 99
; CODES = "000102030405060708091011121314151617181920212223242526272829+
; +303132333435363738394041424344454647484950515253545556575859+
; +606162636465666768697071727374757677787980818283848586878889+
; +90919293949596979899"

; codes reversed, in case needed
; NEWSOCR= "999897969594939291908988878685848382818079787776757473727170+
; +696867666564636261605958575655545352515049484746454443424140+
; +393837363534333231302928272625242322212019181716151413121110+
; +09080706050403020100"

; MISSCORE = -1 ; -1 = missing data treated as not administered

;User Scaling
UMEAN = 50 ; user-set item mean - standard is 0.00
```

```
USCALE = 10           ; user-scaled measure units - standard is 1.00
UDECIIM = 1          ; reported decimal places - standard is 2
MRANGE = 50          ; half-range on maps - standard is 0 (auto-scaled)
```

&END

;Put item labels here for NI= lines

END LABELS

;Put data here - and delete this comment to prevent it being processed as a data line.

10.4 Data file

If your data file is small, it is easiest merely to have it at the end of your control file. If your data is extensive, keep it in a separate data file.

Your data file is expected to contain a record for each person containing a person-id field and a string of responses to some items. Your data can be placed either at the end of your control file or in a separate disk file.

Winsteps reads up to 30 columns of person-id information as standard. Normally the person-id is assumed to end when the response data begin or when the end of your data record is reached. However, an explicit length of up to 300 characters can be given using the [NAMLEN=](#) control variable.

By the term "response" is meant a data value which can be a category label or value, score on an item or a multiple-choice option code. The responses can be one or two characters wide. Every record must contain responses (or missing data codes) to the same items. The response (or missing data code) for a particular item must be in the same position in the same format in every record. If every person was not administered every item then mark the missing responses blank or make them some otherwise unused code, so that the alignment of item responses from record to record is maintained.

A table of valid responses is entered using the [CODES=](#) character string. Any other response found in your data is treated as missing. By using the [CODES=](#), [KEYn=](#), [NEWSCORE=](#) and [VALUE=](#) options, virtually any type of response, e.g. "01", "1234", " 1 2 3 4", "abcd", " a b c d", can be scored and analyzed. Missing responses are usually ignored, but the [MISSCORE=](#) control variable allows such responses to be treated as, say, "wrong".

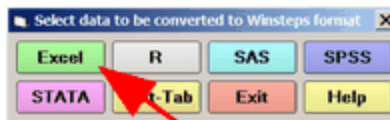
When writing a file from **SPSS**, the syntax is:

FORMATS ITEM1 ITEM2 ITEM3 (F1). i.e., FORMATS varlist (format) [varlist.]

The procedure is FORMATS and then the variable list. Enclosed in parentheses is the format type. F signifies numeric while 1 signifies the width. (F2) would signify a numeric with a width of 2 columns for [XWIDE=2](#). See pages 216 and 217 of the SPSS Reference Guide (1990). See also the [SPSS pull-down menu](#).

10.5 Data from Excel files

Winsteps control and data files can easily be constructed from Excel .xls files. Use the [Excel/RSSST menu](#): Excel option:



In the Excel spreadsheet, the first row is the variable names. Then each row is one person (subject, case). Each column contains one variable: person name, item response, demographic variable.

Here is Bond4.xls - the example from Bond & Fox, "Applying the Rasch Model", chapter 4:

	A	B	C	D	E	F
1	Person	Negation	Reciprocal	Implication	Incompatibility	Mul
2	001	1	1	1	1	1
3	002	1	1	1	1	1
4	003	1	1	0	1	0
5	004	1	1	1	1	1
6	005	1	1	1	1	1
7	006	1	1	1	1	1

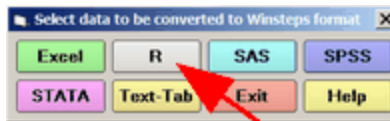
Each Excel column is an "Excel Variable". The first row contains the variable name. "Person" is the person identification - we want this to go into the Winsteps Person Label. "Negation", "Reciprocal", ... are items on the test instrument. The responses, 1 or 0, are in each column for each person - we want these to be the Winsteps Items.

Now follow the procedure at [Excel/RSSST menu](#)

For another conversion technique see [Data from Excel and other spreadsheets](#)

10.6 Data from R Statistics files

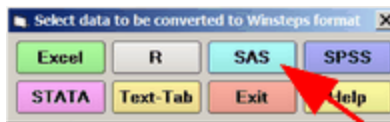
Winsteps control and data files can easily be constructed from [R Statistics](#) *.rda* and *.rdata* files. Use the [Excel/RSSST menu](#): R option:



Now follow the procedure at [Excel/RSSST menu](#).

10.7 Data from SAS files

Winsteps control and data files can easily be constructed from SAS *.sas7bdat* files. Use the [Excel/RSSST menu](#): SAS option:



SAS file conversion requires the "SAS Local Provider", downloadable free (after free registration) from <http://www.sas.com/> - more precisely from <https://support.sas.com/downloads/package.htm?pid=648>

.sas7bdat files from SAS 9.4 or later may not be compatible with the SAS Local Provider. Please create SAS files with the SAS option:

```
options ExtendObsCounter=no;
```

see [Usage Note 49496: The option EXTENDOBSCOUNTER=YES might cause the error](#)

An alternative is to export your SAS file in *.CSV* format or to Excel.

SAS conversion problems may also be due to lack of Microsoft Data Access Components (MDAC) support in Windows, available by download free from Microsoft.

Now follow the procedure at [Excel/RSSST menu](#)

2. SAS provides an environment within which Winsteps can run. Kazuaki Uekawa suggests:

Sample instructions:

```
/*type the location where Winsteps is installed*/
%let win= C:\Winsteps\Winsteps;
....
option xwait xsync;
/*This run uses the whole sample*/
x "start &win &WD&scale..con &WD&scale._whole.out ifile=&WD&scale._whole.ifile pfile=&WD&scale._whole.pfile ";
....
/*item files produced by Winsteps are now read by SAS*/
....
```

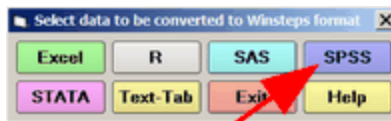
10.8 Data from SPSS files

Winsteps control and data files can easily be constructed from SPSS **.sav** files (but not from PASW .spv files or other SPSS formats). SPSS file conversion requires spssio32.dll which is installed with your copy of SPSS, or is in your SPSS software package (often in an obscure folder). If this is not present, Winsteps installs a minimal spssio32.dll

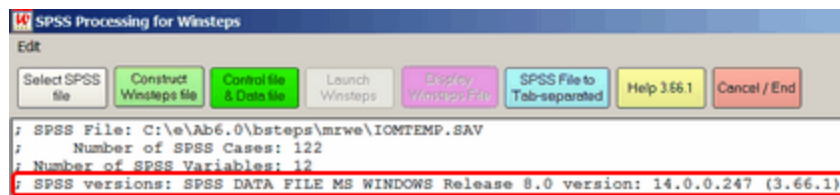
The Winsteps version of spssio32.dll only extracts 8-character variable names. If your variable names are longer, then convert your SPSS file into an Excel file, and use the [Excel routine](#) to create your Winsteps file

Here is a summary of the procedure to convert an SPSS file into a Winsteps file:

1. Launch Winsteps
2. If you see a button label "Import from ... SPSS", please click on it with your mouse. If you don't, please click on "Excel/RSSST" on the menu bar at the top of the Ministep window.
3. The "Select data" buttons display.



4. Click on the SPSS button.
5. A window opens "Select SPSS processing for Winsteps"
6. Click on "Select SPSS file"
7. A window opens "Read SPSS dataset file"
8. Navigate to your .sav file, click on its name, then click on the OPEN button.
9. The SPSS variable list displays.



SPSS File:	Name of the SPSS .sav data file
------------	---------------------------------

Number of SPSS Cases:	Number of cases (person records) found in the SPSS file.
Number of SPSS Variables:	Number of SPSS variables (items, person labels, demographic indicators, etc.)
SPSS versions:	
SPSS DATA FILE	SPSS version which created the SPSS .sav file
version:	spssio32.dll version number (10.0 is the minimal interface supplied with Winsteps). Your version of SPSS has a copy of spssio32.dll with a higher version number. If it is not active, please copy it into the Windows System folder, and delete spssio32.dll in the Winsteps folder.
(4.5.2)	Winsteps interface routine (wininput.exe) performing the conversion. This is usually the same as your Winsteps version number.

10. Now please follow the file-conversion procedure from Step 3 in [Excel/RSSST menu](#)

Example dataset spssdata.sav is courtesy of Parisa Daftarifard.

10.9 Data from STATA files (version 13 and earlier)

STATA have changed the internal format of their .dta files in STATA version 14. To read a STATA 14 or later file with Winsteps, first in STATA,

`. saveold filename13, version(13)`

Then filename13.dta can be processed by Winsteps.

Winsteps control and data files can easily be constructed from STATA .dta files. Use the [Excel/RSSST menu](#): STATA option:



Now follow the procedure at [Excel/RSSST menu](#)

You can simplify some of the variable selection if you construct your person variable in STATA using STATA instructions similar to:

`Gen pname newvar = string(var1,"%05.0f") + string(var2,"%01.0f") + string(var3,"%01.0f") + string(var4,"%01.0f")`

This becomes the person-name variable for Winsteps.

Alan Acock writes: Fred Wolfe has provided a Stata module which produces Winsteps control and data files. It is a Stata command that does two things:

1. It converts the data you want to analyze using Winsteps to a dataset that Winsteps can read.
2. It generates the Winsteps command file to do the analysis in Winsteps.

To install this command go to the command window in Stata and enter the following command:

```
ssc install raschcvt
```

This will find the command on the internet and install the command. You only need to run this once. Here is an example of how you would use the command. From an open session of Stata enter the following command in your command window:

```
raschcvt var1 var2 var3 var4 id , outfile(c:\Winsteps\testme) id(id) max(4)
```

This will generate a dataset for Winsteps and put it in the directory where Winsteps is located. The file will be called

`testme.dat`

This will also generate a command file called

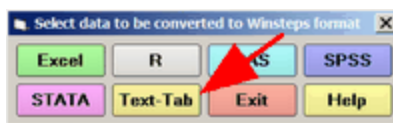
testme.con

and put it in the same directory as the data. This file may have a couple lines at the top and bottom that need to be edited out. The Rasch analysis will be performed on the variables that you list. In this example the variables are var1, var2, var3, and var4. The identification variable should be at the end of this list and happens to be called id in this example. It is important that each variable has a variable label in the Stata dataset as Wolfe's command assumes this when generating the command file. The outfile(c:\Winsteps\testme) after the comma determines the name and location of the two files created, the id(id) gives the name you select for the identification variable, and the max(4) gives the maximum score, in this case the variables have a maximum value of 4.

Use your editor to remove any lines at the top or bottom in the testme.con file that are superfluous. When you run Winsteps simply click on enter or file open, and then browse to find *testme.con*. When prompted for the output file name enter an appropriate name. When asked for special features simply press enter. This will run your data in Winsteps.

10.10 Data from Text files (with Tabs, Commas, Semicolons)

Winsteps control and data files can easily be constructed from Text files with Tabs, commas or semicolons, usually .txt files. Use the [Excel/RSSST menu](#): Text-Tab option:



Now follow the procedure at [Excel/RSSST menu](#).

The Text files (with Tabs) must be in standard Windows .txt file format, with one row for each person and one column for each item (variable). The columns are separated by tab characters. Additional columns can contain person identification and demographics.

The first row contains the names (labels) for the columns, also tab-separated.

Here is a typical file. The white spaces are tab characters.

Person	Item1	Item2	Item3	Item4	Item5
001	1	1	1	1	1
002	1	1	1	1	1
003	1	1	0	1	0
004	1	1	1	1	1
005	1	1	1	1	1

10.11 Data from Excel and other spreadsheets

When possible, use the technique described at [Data from Excel files](#). But it is straight-forward to copy data from an Excel spreadsheet into a Winsteps data file.

(i) Organize your data.

Transform all item responses into columns one or two columns wide, e.g., "1" or "23"

Transform all demographics into columns, one column wide, e.g., "M" and "F" for male and female.

(ii) Organize your Excel spread sheet.

Put all item responses (one item per column) into one block to the left of your spreadsheet.

Put all person identifiers (one item per column) into one block, immediately to the right of the last item column.

(iii) Organize your column widths.

Make all item column widths the same (usually one or two columns).

Person identifier widths can match the identifiers, but these are best at one column wide.

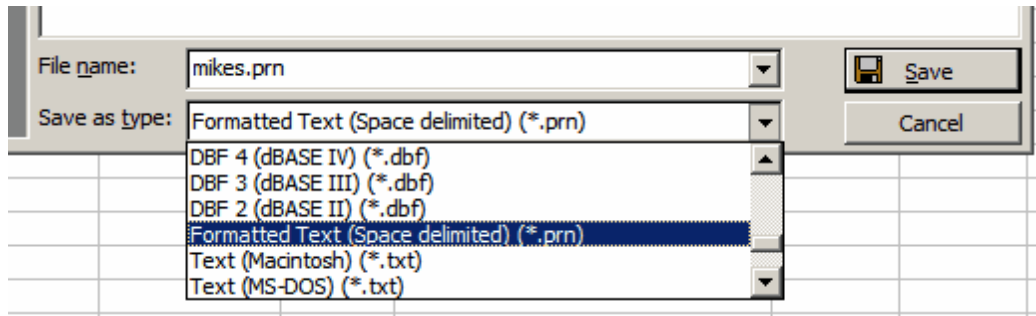
(iv) Replace missing data with "*" or "." or "-"

Global replace nothing in a cell with a convenient clear missing data indicator, which is **not a number**.

(v) Use the Excel format function to inset leading zeroes etc.
Select the item columns, then
Format - Cells - Custom
and enter 0 for 1 character wide columns, 00 for 2 character-wide columns, etc.

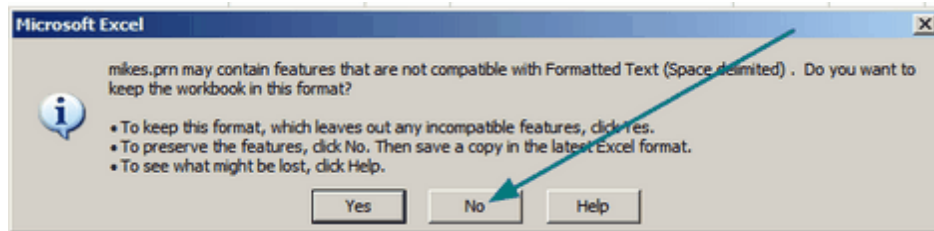
(vi) Select all cells.

(vii) Copy into clipboard (Ctrl+C), or write to a tab-delimited file
or **"Save as" a "Formatted Text (space delimited) (*.prn)" file (respond No to next dialog box)**

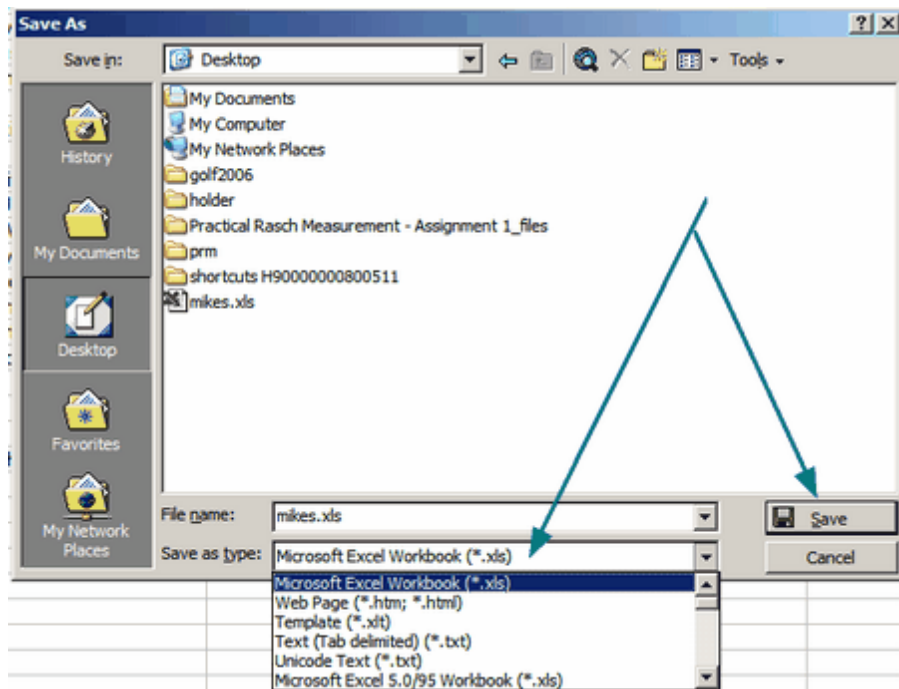


The data has been saved in the .prn format. This is a txt file.

To the "Keep" question: respond No:



To the "Save As" question, respond "Microsoft Excel Workbook", "Save"



- (viii) Open Notepad or [your own text editor](#).
Paste (Ctrl+V) or open the tab-delimited file.
- (ix) Removing tabs
Highlight a tab (area between two columns)
Copy (Ctrl+C)
Replace all tabs: (Ctrl+V) tab if necessary with nothing.
- (x) The file should now look like a standard Winsteps rectangular data file.
Save as a text file.

To obtain a data matrix like this from Excel:

```
110101010
001011001
```

1. Set the Excel column widths so that the Excel matrix looks like this.
2. "Save As" the matrix as "Formatted Text (Space delimited) (*.prn)"

10.12 Data file with other delimiters

Global replace the delimiters with tabs and use the [Excel/RSSST menu](#): Data as Text with Tabs option.

It is often convenient to organize your data with delimiters, such as commas, semi-colons or spaces, rather than in fixed column positions. However, often the delimiter (a Tab, space or comma) only takes one column position. In which case, it may be easier to include it in the [CODES=](#) or use [MFORMS=](#) or [FORMAT=](#). See also [DELIMITER=](#).

11 * Control Variables

11.1 AGREEFILE= paired-person agreement counts

(controlled by [PSELECT=](#), [ISELECT=](#), [PSUBTOTAL=](#))

AGREEFILE= reports the paired-person information underlying [Table 35](#). Use [PSELECT=](#) and [ISELECT=](#) to choose subsets to compare for agreement. Sub-samples (groups of persons) can be specified with [PSUBTOTAL=](#). Only pairs in the same groups are shown if [AGREEGROUP=](#)Yes.

AGREEFILE=? opens a Browse window.

When AGREEFILE= is launched from the [Output Files](#) menu, the [AGREEFILE= dialog box](#) displays.

Contents of AGREEFILE=

```
; AGREEMENT FILE FOR An MCQ Test: administration was Computer-Adaptive ISELECT=* PSELECT=* AGREEGROUP=Y
TWOOBS TWOSAME TWOSCOR ONEHIGH TWOHIGH ONELOW TWOLOW ONEMISS TWOMISS MEASURE MEASURE AVERAGE T35.1 T35.2 T35.3 T35.4 T35.5 STUDEN1
STUDEN2 P-LABEL P-LABEL2
  9      7      8      5      4      5      3      60      33      .40      .44      .42  77.8  88.9  80.0  60.0  55.0      1
      7 IM CAT IH JAN
  9      7      7      6      4      5      3      60      24      .40     -.96     -.28  77.8  77.8  66.7  60.0  40.0      1
      8 IM CAT IL ROB
  7      3      4      5      2      5      1      62      35      .40     -.05     .18  42.9  57.1  40.0  20.0  56.5      1
      14 IM CAT IL TAK
```

Field	Columns	Description
TWOOBS	1-8	count of items for which both persons have non-missing observations
TWOSAME	9-16	count of items for which both persons have the same non-missing observations

TWOSCORE	17-24	count of items for which both persons have non-missing observations scored the same
ONEHIGH	25-32	count of items for which one or both persons have non-missing observations scored in the highest (correct) category
TWOHIGH	33-40	count of items for which both persons have the same non-missing observations scored in the highest (correct) category
ONELOW	41-48	count of items for which one or both persons have non-missing observations scored in the lowest (wrong) category
TWOLOW	49-56	count of items for which both persons have the same non-missing observations scored in the lowest (wrong) category
ONEMISS	57-64	count of items for which one or both persons have a missing observation
TWOMISS	65-72	count of items for which both persons have missing observations
MEASURE(1)	73-80	person ability measure for first person
MEASURE(2)	81-88	person ability measure for second person
AVERAGE	89-96	average person ability of the two persons: Table 35 x-axis measure
T35.1	97-102	Table 35.1 y-axis percent = % Same (Paired Kid) Observed Responses
T35.2	103-108	Table 35.2 y-axis percent = % Same (Paired Kid) Scored Responses
T35.3	109-114	Table 35.3 y-axis percent = % Same (Paired Kid) Observed Highest (Right) Responses
T35.4	115-120	Table 35.4 y-axis percent = % Same (Paired Kid) Observed Lowest (Wrong) Responses
T35.5	121-126	Table 35.5 y-axis percent = % Same (Paired Kid) Missing Responses when either person has a Missing Response
PERSON(1)	127-134	person entry number for first person
PERSON(2)	135-142	person entry number for second person
P-LABEL(1)	143-....(length of longest person label)	person label for first person
P-LABEL(2)	(one blank, then) ...-....	person label for second person

Example 1: an MCQ test with 60 questions, scored 1, 0. One person answered questions 1-40 and the other answered 10-50. They were scored 1-1 on 10 items, 1-0 on 8 items, 0-1 on 5 items, 0-0 on 8 items = 31 items.

TWOOBS: both students must be observed on the item. They were both observed on items 10-40, so that 31 will appear here.

TWOSAME: counts of items to which both students made the same response. There was only one correct response to each item, so SAME = 10 (1-1) correct answers + 2 (out of 8 0-0) incorrect same choice of distractors = 12.

TWOSCORE: the count of items on which they scored the same (for a dichotomy, both correct or both incorrect). They made the SAME score on 10 (1-1) + 8 (0-0) = 18.

ONEHIGH: one or both persons scored in the highest category, "1", in 10+8+5 = 23 items. The "highest" category is the highest score possible on each item, for dichotomies usually scored 1. For Likert rating scales, the high category is usually scored 5.

TWOHIGH: 10 pairs of responses are in the highest category (1-1).

ONELOW: one or both persons score in the lowest category, "0", in 8+5+8 = 21 items. The "lowest" category is the lowest score possible on each item, for dichotomies usually scored 0. For Likert rating scales, the low category is usually scored 1.

TWOLOW: 8 pairs of responses are in the lowest category (0-0).

ONEMISS: there are 60 questions, the both responded to 31, so one or other or both students had missing data on ONEMISS = 60 - 31 = 29 items.

TWOMISS: items 51-60 were missing for both students. TWOMISS = 10.

11.2 AGREEGROUP= compare with other groups in Table 35 = Yes

AGREEGROUP = Yes

For the overall plots in [Table 35](#), and [AGREEFILE=](#), AGREEGROUP=Yes if persons are to be compared only with others in their own classification groups.

AGREEGROUP = Yes

For the overall plots in [Table 35](#), and [AGREEFILE=](#), AGREEGROUP=No if persons are to be compared with all other persons.

11.3 ALPHANUM= alphabetic numbering

Normally [XWIDE=1](#) limits the numerical score range to 0-9, but with ALPHANUM= this can be extended much further.

ALPHANUM= is followed by a string of characters that represent the cardinal numbers in order starting with 0, 1, 2, 3, ...

Example: Represent the numbers 0-20 using XWIDE=1

```
ALPHANUM = 0123456789ABCDEFGHIJK ; then A=10, B=11, . . ., K=20
XWIDE = 1
CODES = 0123456789ABCDEFGHIJK
NI = 5
ITEM1 = 1
&END
21BF3 Mary ; Mary's responses are 2, 1, 11, 15, 3
K432A Mark ; Mark's responses are 20, 4, 3, 2, 10
```

11.4 ANCESTIM= alternative method for anchored analyses

ANCESTIM = No (the default)	Not done. Estimation with anchors is done using standard procedure.
ANCESTIM = YES	When there are IAFLE= or PAFILE= anchor values, MJMLE is done with an alternative estimation method.

The standard Winsteps anchoring procedure, ANCESTIM=No, is usually successful, but sometimes reports convergence when the [Displacements](#) for the anchor values (set by [IAFILE=](#), [PAFILE=](#)) and/or freely estimated persons and items are unnecessarily large.

The special alternative anchoring estimation method, ANCESTIM=YES, attempts to minimize the Displacements from the anchor values by global shifting of the estimates of the unanchored persons (PAFILE=) or items (IAFILE=) or both (PAFILE= and IAFLE= together).

Advice: if ANCESTIM=No (the default) produces unsatisfactory Displacements, please try ANCESTIM=YES, and *vice-versa*.

11.5 ASCII= characters

Encoded drawing characters can produce prettier-looking [output tables](#).

ASCII=Yes	use ASCII characters (the standard) for drawing table boxes.
-----------	--

ASCII=Webpage	use HTML characters for drawing table boxes, output to web browser.
ASCII=Doc	use HTML characters for drawing table boxes, output to Microsoft Word (or equivalent).
ASCII=No	use MS-DOS box characters for drawing table boxes.

[BOXSHOW=](#)No removes the lines around the table boxes.

ASCII=Yes (displays in your text editor)

```

ASCII=Yes ; Letter Gothic Line
-----
|CATEGORY|OBSERVED|OBSVD|SAMPLE|INFIT|OUTFIT|COHERENCE|ESTIM| | | |
|LABEL|SCORE|COUNT|AVRGE|EXPECT|MNSQ|MNSQ|M->C|C->M|DISCR|
|0|0|203|50|-3.29|-3.27|.98|.76|90%|90%|
|1|1|203|50|2.99|2.97|.94|.70|90%|90%|1.04|
-----
0 Wrong
1 Right

ASCII=Yes ; Lucida Console
-----
|CATEGORY|OBSERVED|OBSVD|SAMPLE|INFIT|OUTFIT|COHERENCE|ESTIM| | | |
|LABEL|SCORE|COUNT|AVRGE|EXPECT|MNSQ|MNSQ|M->C|C->M|DISCR|
|0|0|203|50|-3.29|-3.27|.98|.76|90%|90%|
|1|1|203|50|2.99|2.97|.94|.70|90%|90%|1.04|
-----
0 Wrong
1 Right

```

Fonts: Letter Gothic Line, Lucida Console, Consolas, Courier New, SImPL, SimSUN, Bitstream Vera Sans Mono, Crystal, Onuava, Andale Mono.

ASCII=Webpage (displays in your web browser)

```

ASCII=Webpage ; Courier New
-----
|CATEGORY|OBSERVED|OBSVD|SAMPLE|INFIT|OUTFIT|COHERENCE|ESTIM| | | |
|LABEL|SCORE|COUNT|AVRGE|EXPECT|MNSQ|MNSQ|M->C|C->M|DISCR|
|0|0|203|50|-3.29|-3.27|.98|.76|90%|90%|
|1|1|203|50|2.99|2.97|.94|.70|90%|90%|1.04|
-----
0 Wrong
1 Right

ASCII=Webpage ; Lucida Console
-----
|CATEGORY|OBSERVED|OBSVD|SAMPLE|INFIT|OUTFIT|COHERENCE|ESTIM| | | |
|LABEL|SCORE|COUNT|AVRGE|EXPECT|MNSQ|MNSQ|M->C|C->M|DISCR|
|0|0|203|50|-3.29|-3.27|.98|.76|90%|90%|
|1|1|203|50|2.99|2.97|.94|.70|90%|90%|1.04|
-----
0 Wrong
1 Right

```

Fonts: Courier New, Lucida Console, Andale Mono, DejaVu Sans Mono

[WEBFONT=](#) sets the font priority.

ASCII=Doc (displays in Microsoft Word or equivalent)

```

ASCII=Webpage ; Courier New
-----
|CATEGORY|OBSERVED|OBSVD|SAMPLE|INFIT|OUTFIT|COHERENCE|ESTIM| | | |
|LABEL|SCORE|COUNT|AVRGE|EXPECT|MNSQ|MNSQ|M->C|C->M|DISCR|
|0|0|203|50|-3.29|-3.27|.98|.76|90%|90%|
|1|1|203|50|2.99|2.97|.94|.70|90%|90%|1.04|
-----
0 Wrong
1 Right

ASCII=Webpage ; Lucida Console
-----
|CATEGORY|OBSERVED|OBSVD|SAMPLE|INFIT|OUTFIT|COHERENCE|ESTIM| | | |
|LABEL|SCORE|COUNT|AVRGE|EXPECT|MNSQ|MNSQ|M->C|C->M|DISCR|
|0|0|203|50|-3.29|-3.27|.98|.76|90%|90%|
|1|1|203|50|2.99|2.97|.94|.70|90%|90%|1.04|
-----
0 Wrong
1 Right

```

Fonts: Courier New, Lucida Console, Andale Mono, DejaVu Sans Mono

[WEBFONT=](#) sets the font priority.

Example: Output a Table to Microsoft Word:

1. Do the Winsteps analysis
2. Specification Menu box: ASCII=Doc
3. Output Tables menu: select Table or enter Sub-Table
4. Table is output to Microsoft Word
5. In Winsteps, Specification Menu box: ASCII=Yes

ASCII=No (displays in your text editor)

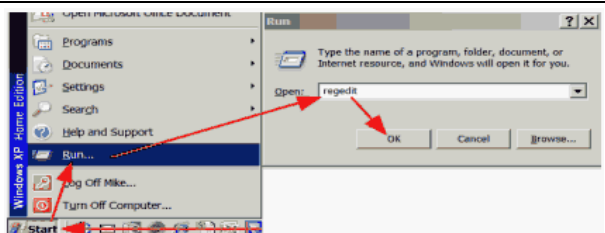
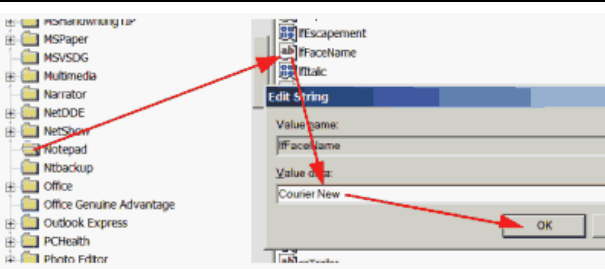

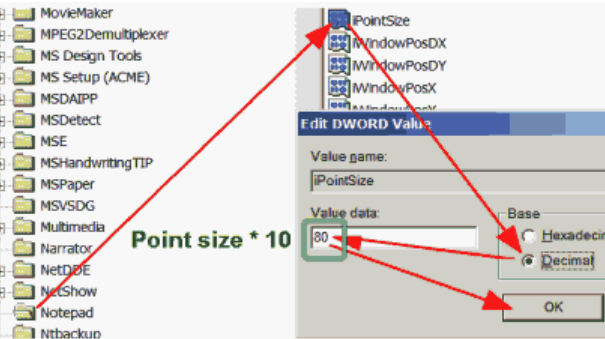
```

ASCII=No ; Letter Gothic Line
-----
|CATEGORY|OBSERVED|OBSVD|SAMPLE|INFIT|OUTFIT|COHERENCE|ESTIM| | | |
|LABEL|SCORE|COUNT|AVRGE|EXPECT|MNSQ|MNSQ|M->C|C->M|DISCR|
|0|0|203|50|-3.29|-3.27|.98|.76|90%|90%|
|1|1|203|50|2.99|2.97|.94|.70|90%|90%|1.04|
-----
0 Wrong
1 Right

```

Fonts: Letter Gothic Line, MS Line Draw.

To make permanent (default) changes in Notepad font face and/or size:

<p>Windows "Start" Click on "Run" Type in "regedit" Click on "OK"</p>	
<p>Registry Editor: Click on the + in front of "HKEY_CURRENT_USER" Click on the + in front of "Software" Click on the + in front of "Microsoft" Click on "Notepad" For the type face: Double-click on "IfFaceName" Type in "Courier New" (or "Letter Gothic Line") Click on "OK"</p>	
<p>For the font size: Double-click on "iPointSize" Click on "Decimal" Type in 80 (for point-size 8 multiplied by 10) Click on "OK"</p> <p>Close registry Click on top right </p>	 <p style="color: green; font-weight: bold;">Point size * 10</p>

11.6 ASYMPTOTE= item upper and lower asymptotes

Persons responding to multiple-choice questions (MCQ) can exhibit guessing and carelessness. In the three-parameter IRT model (3-PL), guessing is parameterized as a lower asymptote to the item's logistic ogive of the probability of a correct answer. In the four-parameter IRT model (4-PL), carelessness is parameterized as an upper asymptote. Winsteps reports a first approximation to these parameter values, but does not use the estimates to alter the Rasch measures. The literature suggests that when the lower asymptote is .10 or greater, it is "substantial" (How Many IRT Parameters Does It Take to Model Psychopathology Items? Steven P. Reise, Niels G. Waller, Psychological Methods, 2003, 8, 2, 164-184).

ASYMPTOTE=Y	report the values of the Upper and Lower asymptotes in the Item Tables and FILE=
ASYMPTOTE=N	do not report values for the Upper and Lower asymptotes.

Example 1: Estimate the 4-PL IRT parameters for the Knox Cube Test data:

Run Exam1.txt

After the analysis completes, use the "Specification" pull-down menu:

Enter: DISCRIM = Yes to report the Item Discrimination

Enter: ASYMP = Yes to report the asymptotes

On the "Output Tables" menu, select an item table, e.g., Table 14.

```

-----+-----
|ENTRY  RAW          | INFIT | OUTFIT | PTMEA|ESTIM| ASYMPOTE |
|NUMBER SCORE COUNT MEASURE ERROR|MNSQ  ZSTD|MNSQ  ZSTD|CORR.|DISCR|LOWER UPPER| TAP
-----+-----
| 4    32    34   -4.40   .81|.90   .0|.35   .8|.55|1.09|.00 1.00|1-3-4
| 6    30    34   -3.38   .64|1.17  .6|.96   .6|.53|.87|.10 1.00|3-4-1
| 7    31    34   -3.83   .70|1.33  .9|2.21 1.2|.40|.54|.09 .98|1-4-3-2
-----+-----

```

Example 2: Polytomous data: Liking for Science

Run Example0.txt

After the analysis completes, use the "Specification" pull-down menu:

Enter: DISCRIM = Yes to report the Item Discrimination

Enter: ASYMP = Yes to report the asymptotes

On the "Output Tables" menu, select an item table, e.g., Table 14.

ENTRY	TOTAL	TOTAL		INFIT	OUTFIT	PTMEASURE-A	ESTIM	ASYMPTOTE						
NUMBER	SCORE	COUNT	MEASURE	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	DISCR	LOWER	UPPER	ACT	
1	109	75	-.40	.55	-3.5	.49	-2.5	.64	.49	1.52	.00	2.00	WATCH BIRDS	
5	37	75	2.42	2.30	5.6	3.62	7.3	.05	.61	-.54	.49	2.00	FIND BOTTLES AND CANS	
23	42	75	2.18	2.41	6.3	4.11	9.0	.00	.61	-.90	.56	1.64	WATCH A RAT	

Estimation

Item Response Theory (IRT) three-parameter and four-parameter (3-PL, 4-PL) models estimate lower-asymptote parameters ("guessability", "pseudo-guessing") and upper-asymptote parameters ("mistake-ability") and use these estimates to modify the item difficulty and person ability estimates. Rasch measurement models guessability and mistake-ability as misfit, and does not attempt to make adjustments for item difficulties and person abilities. But initial approximations for the values of the asymptotes can be made, and output by Winsteps with ASYMPOTOTE=Yes.

The algebraic representation of the discrimination and lower asymptote estimate by Winsteps are similar to 3-PL IRT, but the estimation method is different, because Winsteps does not change the difficulties and abilities from their 1-PL values. Consequently, in Winsteps, discrimination and asymptotes are indexes, not parameters as they are in 3-PL.

A lower-asymptote model for dichotomies or polytomies is:

$$T_{ni} = c_i + (m_i - c_i) (E_{ni}/m_i)$$

where T_{ni} is the expected observation for person n on item i , c_i is the lower asymptote for item i , m_i is the highest category for item i (counting up from 0), and E_{ni} is the Rasch expected value (without asymptotes). Rewriting:

$$c_i = m_i (T_{ni} - E_{ni}) / (m_i - E_{ni})$$

This provides the basis for a model for estimating c_i . Since we are concerned about the lower asymptote, let us construct a weight, W_{ni} ,

$B_i = B(E_{ni}=0.5)$ as the ability of a person who scores 0.5 on the item,

then $B_{ni} = B_n - D_i$ and $W_{ni} = B_{ni} - B_i$ for all $B_n < B_i$ otherwise $W_{ni} = 0$, for each observation X_{ni} with expectation E_{ni} ,

$$c_i \approx \Sigma(W_{ni} m_i (X_{ni} - E_{ni})) / \Sigma(W_{ni} (m_i - E_{ni}))$$

Similarly, for d_i , the upper asymptote,

$$d_i \approx \Sigma(W_{ni} m_i X_{ni}) / \Sigma(W_{ni} E_{ni}) \text{ for } B_{ni} > B(E_{ni}=m_i-0.5)$$

The lower asymptote is the lower of c_i or the item p-value. The upper asymptote is the higher of d_i or the item p-value. If the data are sparse in the asymptotic region, the estimates may not be good. This is a known problem in 3-PL estimation, leading many analysts to impute, rather than estimate, asymptotic values.

Birnbaum A. (1968) Some latent trait models and their uses in inferring an examinee's ability. In F.M. Lord & M.R. Novick, Statistical theories of mental test scores (pp. 395-479). Reading, MA: Addison-Wesley.

Barton M.A. & Lord F.M. (1981) An upper asymptote for the three-parameter logistic item-response model. Princeton, N.J.: Educational Testing Service.

11.7 BATCH= Batch/script mode analysis

1. In standard mode: If you want Winsteps to close itself after performing an analysis and writing out any output specified in the control file, e.g., by [TABLES=](#), then specify BATCH=YES in the control file. You can launch batch files from the [Batch menu](#).

If you want to run Winsteps so that it is visible, but automatically closes then:

Winsteps.exe inputfile.txt outputfile.txt
and include the command "batch=yes" in the inputfile.txt

2. In Batch/Script mode: If you want Winsteps to run in "background" with the minimum user interaction, then specify BATCH=YES in the Shortcut, DOS or Shell command which invokes Winsteps.

If you want to run Winsteps "hidden" then call it from the MS-DOS prompt:
Winsteps.exe batch=yes inputfile.txt outputfile.txt

If you want to run many Winsteps analyses consecutively, then use BATCH=Yes. [Here](#) is an example using Winsteps to simulate multiple datasets and analyze them.

Running Winsteps in Batch mode: *If this won't work for you, see **Difficulty** below.*

In most versions of Windows, .bat and .cmd function in the same way.

Please use whatever works in your version of Windows. First, test what works on your computer:

Create a simple text file on your Desktop with Notepad
contents are:

```
START /WAIT NOTEPAD  
START /WAIT NOTEPAD
```

Name it "test.bat"

Double-click on it. Does Notepad launch? Close Notepad. Does another Notepad launch. If so, all is OK. Use a .bat batch file with START

If not, rename "test.bat" to "test.cmd".

Double-click on it. Does Notepad launch? Close Notepad. Does another Notepad launch. If so, all is OK. Use a .cmd batch file with START

If not,
rename "test.cmd" to "test.bat"
change the contents of "test.bat" to

```
NOTEPAD  
NOTEPAD
```

Double-click on it. Does Notepad launch? Close Notepad. Does another Notepad launch. If so, all is OK. Use a .bat batch file without START.

If none of these work for you, then your version of Windows may not support batch files. Please contact www.winsteps.com.

Example: Blanks act like new lines in a control file. Commas act like blanks:

```
START /WAIT ..\Winsteps BATCH=YES SF.txt SF.OUT TFILE=* 1 * PERSON=CASE IDELETE=3,4,24
```

Under Windows-2000, -XP and later Windows-NT

It is often useful to run multiple Winsteps tasks, one after the other, without keyboard intervention. This can be accomplished by running Winsteps in **CMD batch mode**.

- i) On the main Winsteps screen, click on the "Batch" menu item.
- ii) On the pull-down menu, select "Edit batch file".
- iii) In the dialog box, select **Winbatchcmd.cmd** and click on "Open"
- iv) The following batch file is available to edit:

```
echo This is the version for Windows-NT, 2000  
echo This is a batch file to run Winsteps in batch mode
```



```

echo Edit the next lines and add more.
echo Format of lines is:
echo START /WAIT c:\Winsteps\Winsteps BATCH=YES c:\folder\Control-file c:\folder\Output-file
Extra=specifications
START /WAIT ..\Winsteps BATCH=YES EXAMPLE0.txt EXAMPLE0.OUT TABLES=111
START /WAIT ..\Winsteps BATCH=YES SF.txt SF.OUT TFILE=* 1 * PERSON=CASE
START /WAIT ..\Winsteps BATCH=YES KCT.txt KCT.OUT TFILE=* 3 20 * MRANGE=4

```

These characters have special meanings in batch files: @ & ^ ()

- v) The lines starting with "echo" are comments.
- v) Lines starting "START /WAIT c:\Winsteps\Winsteps BATCH=YES" execute Winsteps from the Winsteps folder
- vi) The format is START /WAIT Winsteps BATCH=YES control-file output-file extra-specifications
- vii) Each new Winsteps line is an additional run of the Winsteps program
- viii) Edit and save this file. You can save it with any name ending ".cmd"
- ix) From the "Batch" pull-down menu, select "Run batch file".
- x) Right-click on the desired batch file
- xi) **In the right-click menu, left-click on "open"**
- x) The batch file will run - if nothing happens, the batch file is incorrect.
- xi) Exit from the Winsteps dialog by clicking on "Cancel".
- xii) You can minimize the batch screen by clicking on the underline in the top right corner.
- xiii) You can cancel the batch run by right clicking on the Batch icon in the Task bar, usually at the bottom of the screen.

Example: I want to automatically run multiple DIF reports for the same set of data.

Since Winsteps can only perform one DIF analysis at a time in batch mode, you can use anchor files:

First line in batch file, produce measure files

```
Winsteps BATCH=YES infile outfile dif=$s1w1 ifile=ifile.txt pfile=pfile.txt sfile=sfile.txt
```

Later lines in batch file, use measure files as anchor files

```
Winsteps BATCH=YES infile outfile2 dif=$s2w1 ifile=iafile.txt pfile=pafile.txt safile=sfile.txt
tfile=* 30 *
```

```
Winsteps BATCH=YES infile outfile3 dif=$s3w1 ifile=iafile.txt pfile=pafile.txt safile=sfile.txt
tfile=* 30 *
```

A Winsteps batch processor for Windows

Batch files under Windows are used to test out new features in Winsteps. Here is what is done:

- a) Create a new subfolder of c:\Winsteps, called c:\Winsteps\test
- b) Copy into folder "test" all the control and data files to be analyzed. For instance all the Winsteps example control and data files, which are found in c:\Winsteps\examples
- c) Use Notepad to create a file in c:\Winsteps\test to do the analysis. This file is "saved as" test.bat
This file contains, for instance:

```

start /w c:\Winsteps\Winsteps batch=yes exam1.txt exam1.out DISC=YES TABLES=111
start /w c:\Winsteps\Winsteps batch=yes exam9.txt exam9.out DISC=YES TABLES=111
start /w c:\Winsteps\Winsteps batch=yes sf.txt sf.out DISC=YES TABLES=111

```

You can replace ..\Winsteps with the pathname to your copy of Winsteps.exe

- d) double-click on test.bat in c:\Winsteps\test to run this batch file.
- e) Winsteps "flashes" on the task bar several times, and progress through the batch file is shown in a DOS-style window.
- e) the .out files are written into c:\Winsteps\test

Example 1. Windows XP file: test.cmd to do 10 analyses from the same control and data files

```

rem - change the directory to the control and data files
chdir c:\Winsteps\examples
chdir

```

```

echo check that the change directory worked
pause
START /Wait c:\Winsteps\Winsteps.exe exam1.txt SCReadDIFg2.out.txt pselect=$2*
Title=READING_DIF_grade_2
echo check that Winsteps was called as expected: add BATCH=Y to the line above when it works
correctly
pause
START /Wait c:\Winsteps\Winsteps.exe BATCH=Y exam1.txt exam1g3.out.txt pselect=$3*
Title=READING_DIF_grade_3
START /Wait c:\Winsteps\Winsteps.exe BATCH=Y exam1.txt exam1g4.out.txt pselect=$4*
Title=READING_DIF_grade_4
START /Wait c:\Winsteps\Winsteps.exe BATCH=Y exam1.txt exam1g5.out.txt pselect=$5*
Title=READING_DIF_grade_5
START /Wait c:\Winsteps\Winsteps.exe BATCH=Y exam1.txt exam1g6.out.txt pselect=$6*
Title=READING_DIF_grade_6
START /Wait c:\Winsteps\Winsteps.exe BATCH=Y exam1.txt exam1g7.out.txt pselect=$7*
Title=READING_DIF_grade_7
START /Wait c:\Winsteps\Winsteps.exe BATCH=Y exam1.txt exam1g8.out.txt pselect=$8*
Title=READING_DIF_grade_8
START /Wait c:\Winsteps\Winsteps.exe BATCH=Y exam1.txt exam1g10.out.txt pselect=$10*
Title=READING_DIF_grade_10
START /Wait c:\Winsteps\Winsteps.exe BATCH=Y exam2.txt exam2g2.out.txt pselect=$2*
Title=MATH_DIF_grade_2
START /Wait c:\Winsteps\Winsteps.exe BATCH=Y exam2.txt exam2g3.out.txt pselect=$3*
Title=MATH_DIF_grade_3
START /Wait c:\Winsteps\Winsteps.exe BATCH=Y exam2.txt exam2g4.out.txt pselect=$4*
Title=MATH_DIF_grade_4
START /Wait c:\Winsteps\Winsteps.exe BATCH=Y exam2.txt exam2g5.out.txt pselect=$5*
Title=MATH_DIF_grade_5
START /Wait c:\Winsteps\Winsteps.exe BATCH=Y exam2.txt exam2g6.out.txt pselect=$6*
Title=MATH_DIF_grade_6
START /Wait c:\Winsteps\Winsteps.exe BATCH=Y exam2.txt exam2g7.out.txt pselect=$7*
Title=MATH_DIF_grade_7
START /Wait c:\Winsteps\Winsteps.exe BATCH=Y exam2.txt exam2g8.out.txt pselect=$8*
Title=MATH_DIF_grade_8
START /Wait c:\Winsteps\Winsteps.exe BATCH=Y exam2.txt exam2g10.out.txt pselect=$10*
Title=MATH_DIF_grade_10

```

If you want to combine all the output files into one file:

```
copy /Y *.out.txt all.txt
```

Example 2. To estimate the expected values of the eigenvalues in Table 23.0 by simulation

Your files are in folder: D:\rasch_simulation

Winsteps.exe is in folder C:\Winsteps

The output is Table 23.0

The batch file is:

```

D:
MKDIR \rasch_simulation
CD \rasch_simulation
REM
START /WAIT C:\Winsteps\winsteps.exe BATCH=YES marlon0.txt marlon0.out.txt PFILE=pf.txt IFILE=if.txt SFILE=sf.txt
REM
set /a test=1
:loop
REM
START /WAIT C:\Winsteps\winsteps.exe BATCH=YES marlon0.txt marlon0%loop%.out.txt PAFILE=pf.txt IAFILE=if.txt SAFILE=sf.txt
SIFILE=SIFILE%test%.txt SISEED=0
REM
START /WAIT C:\Winsteps\winsteps.exe BATCH=YES marlon0.txt data=SIFILE%test%.txt SIFILE%test%.out.txt TFILE=* 23.0 *
REM
set /a test=%test%+1
if not "%test%"=="11" goto loop
PAUSE

```

Difficulty running Batch or Command files?

Microsoft Windows is designed to run interactively, not in batch mode. Microsoft are not consistent with the way they implement batch files in different versions of Windows. So our challenge is to discover a method of running batch files that works for the version of Windows we happen to have. Since Windows is very bad at running batch or command files. You need to validate your instructions one step at a time:

First make sure that your batch file runs *without* "BATCH=YES" so that you can see Winsteps in operation.

Paths with blanks? Put in quotes:

```
START /WAIT "e:\my folder\Winsteps folder\Winsteps.exe" BATCH=YES . . . .
```

i) Run Winsteps in standard mode from the [DOS command prompt](#).

ii) Have the full paths to everything in your batch or command file, e.g., called mybatch.cmd,

```
START /WAIT c:\Winsteps\Winsteps BATCH=YES c:\Winsteps\examples\example0.txt c:\Winsteps\examples\example0.out.txt
```

also have full paths to everything in your Winsteps control file, e.g.,

```
DATA = c:\Winsteps\examples\mydata.txt
```

Note: In this Batch command:

```
START /WAIT c:\Winsteps\Winsteps BATCH=YES c:\Winsteps\examples\controlfile.txt outputfile.txt  
file "outputfile.txt" will be placed in directory "c:\Winsteps\examples\"
```

iii) Windows "Start" menu. "Run". Copy and paste the following line into the Windows Run box on the Windows Start menu.

Click OK:

```
c:\Winsteps\Winsteps c:\Winsteps\examples\example0.txt c:\Winsteps\examples\example0.out.txt table=1
```

Does Winsteps start in the ordinary way? This tests the Windows command line interface.

iv) Windows "Start" menu. "Run". Copy and paste the following line into the Run box. Click OK:

```
c:\Winsteps\Winsteps BATCH=YES c:\Winsteps\examples\exam 15.txt c:\Winsteps\examples\exam 15.out.txt table=1
```

Does the Winsteps icon appear on the Task bar and then disappear? This tests Winsteps background processing.

v) On your desktop, right-click, "New", "Text document". Double-click on icon. Paste in:

```
START /WAIT c:\Winsteps\Winsteps c:\Winsteps\examples\example0.txt c:\Winsteps\examples\example0.out.txt table=1
```

"Save as" Test.cmd. Double-click on Test.cmd

Does Winsteps run in the ordinary way? This test the Windows START function. If this fails, "Save as" Test.bat instead of Test.cmd.

vi) On your desktop, right-click, "New", "Text document". Double-click on icon. Paste in:

```
START /WAIT c:\Winsteps\Winsteps BATCH=YES c:\Winsteps\examples\exam 15.txt c:\Winsteps\examples\exam 15.out.txt  
table=1
```

"Save as" Test2.cmd. Double-click on Test2.cmd (or "Save as" Test2.bat if that works better on your computer.)

Does the Winsteps icon flash on the task bar line, and then disappear? Winsteps has run in background.

vii) Now build your own .cmd batch file, using lines like:

```
START /WAIT c:\Winsteps\Winsteps BATCH=YES c:\Winsteps\examples\example0.txt c:\Winsteps\examples\example0.out.txt
```

viii) If your command line contains sub-lists, indicate those with commas, e.g.,

```
IWEIGHT=* 23,2.5 47,1.3 *
```

Running Winsteps within other Software

Automating the standard version of Winsteps is straightforward using the control instruction BATCH=YES. Winsteps will run under Windows in background.

Let's assume your software is written in Visual Basic (or any other programming, database or statistical language)

- (a) write out a Winsteps control file as a .txt file
- (b) write out a Winsteps data file as a .txt file
- (c) "shell" out to
"START /WAIT Winsteps BATCH=YES controlfile.txt outputfile.txt data=datafile.txt ifile=ifile.txt pfile=pfile.txt"
- (d) read in the ifile.txt, pfile.txt or whatever Winsteps output you need to process.

This is being done routinely by users of [SAS](#).

Running Winsteps within R

Use the R "system" command. To test that Winsteps runs correctly, specify:

```
try(system("c:/Winsteps/Winsteps.exe BATCH=NO yourcontrolfile youroutputfile", intern = TRUE, ignore.stderr = TRUE))
```

or, if the file names contain spaces, then use single quotes and double quotes:

```
try(system("c:/Winsteps/Winsteps.exe BATCH=NO "your control file" "your output file" ", intern = TRUE, ignore.stderr = TRUE))
```

For regular use:

```
try(system("START /WAIT c:/Winsteps/Winsteps.exe BATCH=YES yourcontrolfile youroutputfile", intern = TRUE, ignore.stderr = TRUE))
```

or

```
try(system("START /WAIT c:/Winsteps/Winsteps.exe BATCH=YES "your control file" "your output file" ", intern = TRUE, ignore.stderr = TRUE))
```

or

```
shell("c:/Winsteps/winsteps.exe control.txt output.txt")
this automatically waits for Winsteps to complete before returning to an R prompt
```

See also [r2Winsteps](#)

11.8 BOXSHOW= draw boxes around Output Tables

Output Tables can be output with or without box borders. Here is an example with [Table 13](#). Excel "Data", "[Text to columns](#)" is easier with BOXSHOW=No.

BOXSHOW= Yes

ENTRY	TOTAL	TOTAL		MODEL	INFIT	OUTFIT	PT-MEASURE	EXACT	MATCH				
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	TAP
18	0	35	6.13	1.84		MAXIMUM MEASURE	.00	.00	100.0	100.0	4-1-3-4-2-1-4		
15	1	35	4.80	1.07	.74	-.1	.11	-.6	.32	.24	97.1	97.0	1-3-2-4-1-3
16	1	35	4.80	1.07	.74	-.1	.11	-.6	.32	.24	97.1	97.0	1-4-2-3-1-4
17	1	35	4.80	1.07	.74	-.1	.11	-.6	.32	.24	97.1	97.0	1-4-3-1-2-4
14	3	35	3.37	.70	1.56	1.2	1.49	.8	.22	.38	85.3	92.0	1-4-2-3-4-1

BOXSHOW= No

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PT-MEASURE CORR.	EXP.	EXACT OBS%	MATCH EXP%	TAP
18	0	35	6.13	1.84			MAXIMUM MEASURE		.00	.00	100.0	100.0	4-1-3-4-2-1-4
15	1	35	4.80	1.07	.74	-.1	.11	-.6	.32	.24	97.1	97.0	1-3-2-4-1-3
16	1	35	4.80	1.07	.74	-.1	.11	-.6	.32	.24	97.1	97.0	1-4-2-3-1-4
17	1	35	4.80	1.07	.74	-.1	.11	-.6	.32	.24	97.1	97.0	1-4-3-1-2-4
14	3	35	3.37	.70	1.56	1.2	1.49	.8	.22	.38	85.3	92.0	1-4-2-3-4-1

BOXSHOW= No and [HEADER=No](#)

18	0	35	6.13	1.84			MAXIMUM MEASURE		.00	.00	100.0	100.0	4-1-3-4-2-1-4
15	1	35	4.80	1.07	.74	-.1	.11	-.6	.32	.24	97.1	97.0	1-3-2-4-1-3
16	1	35	4.80	1.07	.74	-.1	.11	-.6	.32	.24	97.1	97.0	1-4-2-3-1-4
17	1	35	4.80	1.07	.74	-.1	.11	-.6	.32	.24	97.1	97.0	1-4-3-1-2-4
14	3	35	3.37	.70	1.56	1.2	1.49	.8	.22	.38	85.3	92.0	1-4-2-3-4-1

Example with **Excel**:

1. In your Winsteps control file, or in the "Specification" menu dialog box:
BOXSHOW=NO
2. Output the Winsteps Table
3. Copy and paste the relevant lines from your Winsteps Output Table into an Excel worksheet
4. In Excel, "Data", "Text to columns"
5. Excel: Delimited with spaces, or Fixed Width
6. Excel: Click to add or remove field separators
7. Using the Table above, in Excel:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	TAP
2														
3	18	0	35	6.13	1.84	MAXI	MUM	M	EASURE	0	0	100	100	4-1-3-4-2-1-4
4	15	1	35	4.8	1.07	.74	-0.1	0.11	-0.6	0.32	0.24	97.1	97.1	1-3-2-4-1-3
5	16	1	35	4.8	1.07	.74	-0.1	0.11	-0.6	0.32	0.24	97.1	97.1	1-4-2-3-1-4
6	17	1	35	4.8	1.07	.74	-0.1	0.11	-0.6	0.32	0.24	97.1	97.1	1-4-3-1-2-4
7	14	3	35	3.37	.70	1.56	1.2	1.49	0.8	0.22	0.38	85.3	92.0	1-4-2-3-4-1

11.9 BYITEM= display graphs for items

In the bit-mapped [graphs](#) produced by the Graphs pull-down menu, the empirical item characteristic curves can be produced at the grouping level or the item level. When [ISGROUPS=0](#), the item-level and grouping-level curves are the same.

BYITEM = Yes show empirical curves at the item level.

BYITEM = No show empirical curves at the grouping level.

11.10 CATREF= reference category for Table 2

If a particular category corresponds to a criterion level of performance, choose that category for CATREF=.

Table 2, "most probable responses/scores", maps the items vertically and the most probable responses, expected scores, and Rasch-Thurstone thresholds (50% cumulative probabilities) horizontally. Generally, the vertical ordering is item difficulty measure. If, instead, a particular category is to be used as the reference for sorting, give its value as scored and recoded.

Special uses of CATREF= are:

CATREF=-3 for item entry order

CATREF=-2 for item measure order

CATREF=-1 for items measure order with ISGROUPS=

CATREF=0 for item measure order

CATREF=1 ... 32767 for item measure order based on this category.

Example 1: You have 4-point partial-credit items, entered in your data as A,B,C,D, and then scored as 1,2,3,4. You wish to list them based on the challenge of category C, rescored as 3,

CODES=ABCD ; original responses
 NEWScore=1234 ; rescored values
 REScore=2 ; rescore all

CATREF=3 ; Table 2 reference category
 ISGROUPS=0 ; partial credit: one item per grouping

If, for an item, the category value "3" is eliminated from the analysis or is the bottom category, the nearest higher category is used for that item.

Example 2: You have 6 3-category items in Grouping 1, and 8 4-category items in Grouping 2. You wish to list them in Table 2.2 by measure within grouping, and then by measure overall.

CODES=1234
 NI= 14
 ISGROUPS= 11111122222222
 TFILE=*
 2.2 0 0 0 -1 -1 ; means CATREF=-1
 2.2 0 0 0 0 ; last 0 means CATREF=0
 *

11.11 CFILE= scored category label file

Rating (or partial credit) scale output is easier to understand when the categories are shown with their substantive meanings. Use CFILE= to label categories using their scored values, i.e., after rescoring. Use CLFILE= to label categories using their original codes, i.e., before any rescoring.

Labels for categories, after they have been scored, can be specified using CFILE= and a file name, or CFILE=* and placing the labels in the control file. Each category number is listed (one per line), followed by its descriptive label. If the observations have been rescored (NEWScore=) or keyed (KEYn=), then use the recoded category value in the CFILE= specification. When there are different category labels for different ISGROUPS= of items, specify an example item from the grouping, followed immediately by "+" and the category number. Blanks or commas can be used as separators between category numbers and labels.

CFILE= file name	file containing details
CFILE = *	in-line list
CFILE = ?	opens a Browser window to find the file

Example 1: Identify the three LFS categories, 0=Dislike, 1=Don't know, 2=Like.

CODES=012
 CFILE=*
 0 Dislike
 1 Don't know
 2 Like
 *

The labels are shown in Table 3.2 as:

```

-----
CATEGORY OBSERVED AVGE  INFIT  OUTFIT  STRUCTURE
 LABEL      COUNT  MEASURE MNSQ   MNSQ   MEASURE
-----
    0         378    -.87   1.08   1.19   NONE    Dislike
    1         620     .13    .85    .69    -.85    Don't know
    2         852     2.23   1.00   1.46    .85     Like
-----

```

Example 2: Items 1-10 (Grouping 1) are "Strong Disagree, Disagree, Agree, Strongly Agree". Items 11-20 (Grouping 2) are "Never, Sometimes, Often, Always".

NI=20
 CODES=1234
 ISGROUPS=11111111112222222222

```

CFILE=*
7+1 Strongly Disagree ; We could use any item number in Grouping 1, i.e., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
7+2 Disagree ; item 7 has been chosen
7+3 Agree
7+4 Strong Agree
13+1 Never ; We could use any item number in Grouping 2, i.e., 11, 12, 13, 14, 15, 16, 17, 18, 19, 20
13+2 Sometimes ; item 13 has been chosen
13+3 Often
13+4 Always
*

```

Example 3: To enter CFILE= information on the DOS Prompt or Extra Specifications lines, using commas instead of blanks as separators:

```
C:>Winsteps SF.TXT SFO.TXT CFILE=* 1,Dislike 2,Don't-know 3,Like *
```

Example 4: Some items have one rating scale definition, but most items have another rating scale definition. But each item is calibrated with its own structure: ISGROUPS=0

```

NI=20
CODES=1234
ISGROUPS=0
CFILE=*
1 Strongly Disagree This scale is used by most items
2 Disagree
3 Agree
4 Strong Agree
16+1 Never 16 is one item using the other scale
16+2 Sometimes
16+3 Often
16+4 Always
17+1 Never 17 is another item using the other scale
17+2 Sometimes
17+3 Often
17+4 Always
.... for all the other items using the other scale
*

```

Example 5: Several categories are collapsed into one category. The original codes are A-H. After rescaling there is only a dichotomy: 0, 1.

```

NI=30
CODES =ABCDEFGH
NEWSCORE=00011110
CFILE=*
0 Fail Specify the categories as recoded
1 Pass
*

```

11.12 CHART= graphical plots in Tables 10, 13-15

The measures and fit statistics in the Tables can be displayed as graphical plots in subtables [6.2](#), 10.2, 13.2, 14.2, 15.2, 17.2, 18.2, 19.2, 25.2, 26.2.

CHART=N Omit the graphical plots.

CHART=Y Include graphical plots

```

          PUPIL FIT GRAPH:  OUTFIT ORDER
+-----+-----+-----+-----+-----+-----+-----+-----+
|ENTRY| MEASURE |  INFIT MEAN-SQUARE | OUTFIT MEAN-SQUARE |          |
|NUMBR| -      + |0      0.7 1 1.3    2|0      0.7 1 1.3    2| PUPIL    |
+-----+-----+-----+-----+-----+-----+-----+

```

```

| 72|*      |      : . :      * |A      : . :      *| JACKSON, SOLOMON |
| 47| *     |      : . *:     |J      : . :      *| VAN DAM, ANDY   |
| 53|*      |      : *. :     |K      : . :      *| SABOL, ANDREW   |
| 32| *     |      *: . :     |w      * . :     | ROSSNER, JACK   |
| 21| *     | *      : . :     |a *    : . :     | EISEN, NORM L.  |
+-----+

```

The fit information is shown in graphical format to aid the eye in identifying patterns and outliers. The fit bars are positioned by [FITLOW=](#) and [FITHIGH=](#). They may also be repositioned using [TFILE=](#).

11.13 CHISQUARE= in IFILE= and PFILE=

	In IFILE= and PFILE= , report Infit and Outfit fit statistics as
CHISQUARE=No	Mean-square fit statistics = Chi-square statistics divided by their degrees of freedom.
CHISQUARE=Yes	Chi-square fit statistics
	The chi-square degrees of freedom are reported in IFILE=, PFILE= columns INDF and OUTDF. The two-sided chi-square probabilities are reported adjacent to the chi-squares when LOCAL=Prob

Computation: Winsteps and Facets usually report mean-square statistics. These are chi-squares divided by their degrees of freedom. Now the chi-squares and their d.f. are also reported. For the computation of the mean-squares and d.f., see www.rasch.org/rmt/rmt34e.htm together with www.rasch.org/rmt/rmt162g.htm

So that $q_i^2 = 2 / \text{d.f.}$

And chi-square = mean-square * d.f.

Example:

LOCAL=Prob

CHISQUARE=Yes

IFILE=if.txt

Then, in file: if.txt:

```

;ENTRY MEASURE ST  COUNT  SCORE MODLSE IN.CHI IN.PROB OUT.MS OUT.PR  INDF  OUTDF G M R NAME
  4  -4.40  1  35.0   32.0   .81  6.77 .9809   .35 .9103  7.52  1.00 1 R . 4= 1-3-4

```

Infit Chi-square is 6.77 with 7.52 d.f., 2-sided probability = .9809

11.14 CLFILE= codes label file

Rating (or partial credit) scale output is easier to understand when the categories are shown with their substantive meanings. Use [CFILE=](#) to label categories using their scored values, i.e., after rescaling. Use [CLFILE=](#) to label categories using their original codes, i.e., before any rescaling. Labels for the original categories in the data can be specified using [CLFILE=](#) and a file name, or [CLFILE=*](#) and placing the labels in the control file. Each category number is listed (one per line), followed by its descriptive label. Original category values are used. There are several options:

CLFILE= file name	file containing details
CLFILE = *	in-line list
CLFILE = ?	opens a Browser window to find the file

Its format is:

CLFILE=*

item entry number + category code category label

%item label + category code category label

*

XWIDE=2 ;observations are two columns wide

CODES = " 0 1 299" ; codes are 0, 1 , 2, 99


```
CLFILE=*
99 Strongly Agree ; original code of 99 has the label "Strongly Agree"
2 Agree ; original code of blank+2 (or 2+blank) has the label "Agree"
2+99 Heartily Agree ; for item 2, code 99 has the label "Heartily Agree"
3+0 Disagree ; for item 3, code 0 means "Disagree"
*
```

Example 1: Identify the three LFS categories, D=Dislike, N=Don't know, L=Like.

```
CODES =DNL
NEWSCORE=012
CLFILE=*
D Dislike
N Neutral
L Like
*
```

The labels are shown in Table 3.2 as:

CATEGORY LABEL	OBSERVED COUNT	AVGE MEASURE	INFIT MNSQ	OUTFIT MNSQ	STRUCTURE MEASURE	
0	378	-.87	1.08	1.19	NONE	Dislike
1	620	.13	.85	.69	-.85	Don't know
2	852	2.23	1.00	1.46	.85	Like

Example 2: Items 1-10 (Grouping 1) are "Strong Disagree, Disagree, Agree, Strongly Agree". Items 11-20 (Grouping 2) are "Never, Sometimes, Often, Always".

```
NI=20
CODES =ABCD
NEWSCORE=1234
ISGROUPS=11111111112222222222
CLFILE=*
7+A Strongly Disagree ; 7 is any item in Grouping 1
7+B Disagree
7+C Agree
7+D Strong Agree
13+A Never ; 13 is any item in Grouping 2
13+B Sometimes
13+C Often
13+D Always
*
```

Example 3: To enter CLFILE= information on the DOS Prompt or Extra Specifications lines, using commas instead of blanks as separators:

```
C:>Winsteps SF.TXT SFO.TXT CLFILE=* D,Dislike N,Don't-know L,Like *
```

Example 4: One grouping of items has a unique response format, but the other groupings all have the same format. Here, each grouping has only one item, i.e., ISGROUPS=0

```
NI=20
CODES=1234
ISGROUPS=0
CLFILE=*
1 Strongly Disagree ; This rating scale is used by most items
2 Disagree
3 Agree
4 Strong Agree
16+1 Never ; 16 is the one item using this rating scale
16+2 Sometimes
16+3 Often
```

16+4 Always

*

Example 5: Several categories are collapsed into one category. The original codes are A-H. After rescaling there is only a dichotomy: 0, 1.

```

NI=30
CODES = ABCDEFGH
NEWSCORE=00011110
CFILE=*
0 Fail Specify the categories as recoded
1 Pass
*
; or
CLFILE=*
A Fail
B Fail
C Fail
D Pass
E Pass
F Pass
G Pass
H Pass
*

```

Example 6: Identifying the distractors for a multiple-choice MCQ item.

Here is item 5:

5. The shape of the relationship between raw score and measures on the latent variable is

CODES = abcd

```

CLFILE=*
5+a diatonic
5+b harmonic
5+c monotonic
5+d synchronic
*

```

Example 7: Using the item labels

```

CLFILE=*
%APMED+1 Attentive
%APMED+2 Interested
%APMED+3 Residual
%APMED+9 NOT ASKED THAT YEAR
*

```

11.15 CMATRIX= category matrix = Yes

A category matrix (confusion matrix, matching matrix) compares the observed categorization (classification) of observations with the predicted categorization (classification). One matrix is produced in [Table 3.2](#), etc., for each rating-scale group defined in [ISGROUPS=](#).

CMATRIX = N Do not output the category matrix.

CMATRIX = Y Output the category matrix. For interpretation, see [Table 3.2](#)

Category Matrix : Confusion Matrix : Matching Matrix					
Predicted Scored-Category Frequency					
Obs Cat Freq	0	1	2	TOTAL	
0	.50	2.06	.44	3.00	
1	2.03	20.83	12.15	35.00	
2	.47	12.12	24.41	37.00	

11.16 CODERANGE = numerical range of data codes

Says what characters to recognize as valid codes in your data file, together with [CODES=](#). Characters in your data not included in CODES= or CODERANGE= are given the [MISSCORE=](#) value. The highest category number 32767. CODERANGE= can be up to 4,000 characters long. If CODES= and CODERANGE= are both specified, then the data are matched to CODES= first. CODES= can be rescored. CODERANGE= cannot be rescored or keyed. Do **not** use CODERANGE= with [KEY=](#), [IVALUE=](#)

CODERANGE= lowest category number - highest category number	lowest and highest category numbers are in the range 0-32767. Highest number must be higher or equal to lowest number
---	---

Examples from [CODES=](#) rewritten for CODERANGE=.

Example 1: A test has four response choices. These are "1", "2", "3", and "4". All other codes in the data file are to be treated as "item not administered". Each response uses 1 column in your data file. Data look like: 134212342132 . 3212343221

```
XWIDE=1 one character wide (the standard)
CODERANGE=1-4
```

Example 2: There are four response choices. Each response takes up 2 columns in your data file and has leading 0's, so the codes are "01", "02", "03" and "04". Data look like: 0302040103020104040301

```
XWIDE=2 two characters wide
CODERANGE=1-4 ; 01 is 1, 02 is 2, 03 is 3, 04 is 4.
```

Example 3: There are four response choices entered on the file with leading blanks, so that codes are " 1", " 2", " 3", and " 4". Data look like: 3 2 4 2 1 3 2

```
XWIDE=2 two characters wide
CODERANGE=1-4 ; 1 is 1, 2 is 2, 3 is 3, 4 is 4.
```

Example 4: Your data is a mixture of both leading blanks and leading 0's in the code field, e.g. "01", " 1", " 2", "02" etc. The numerical value of a response is calculated, where possible, so that both "01" and " 1" are analyzed as 1.

Data look like: 02 1 20102 1 2 01

```
XWIDE=2 two characters wide
CODERANGE=1-4
```

Example 5: Your valid data are 1,2,3,4,5 and your missing data codes are 7,8,9 which you want reported separately on the [distractor tables](#).

```
CODES = 789
NEWSCORE = XXX ; missing values scored with non-numeric values
CODERANGE= 1-5 ; data matched to CODERANGE= cannot be rescored
```

Example 6-8 must use CODES= (non-numeric codes)

Example 9: The valid responses are [percentages](#) in the range 00 to 99.

```
XWIDE = 2 two columns each percent
CODERANGE= 0-99
also recommended:
ISRANGE = *
1 0 99 ; defines the full range of the rating scale, regardless of the data
*
SFUNCTION=4 ; defines a smooth function for the Rasch-Andrich thresholds
```

Example 10: Codes are in the range 0-254.

```
XWIDE=3 ; 3 characters per response:
CODERANGE= 0-254
also recommended:
ISRANGE = *
1 0 254 ; defines the full range of the rating scale, regardless of the data
```

```
*
SFUNCTION=4 ; defines a smooth function for the Rasch-Andrich thresholds
```

Example 11 must use CODES= (numeric codes are rescored)

Example 12: Codes in the range 0 - 1000.

```
XWIDE=4 ; 4 characters per response
CODERANGE= 0-1000
also recommended:
ISRANGE = *
1 0 1000 ; defines the full range of the rating scale, regardless of the data
*
SFUNCTION=4 ; defines a smooth function for the Rasch-Andrich thresholds
```

11.17 CODES= valid data codes

Says what characters to recognize as valid codes in your data file, together with [CODERANGE=](#). If [XWIDE=1](#) (the standard), use one column/character per legitimate code. If [XWIDE=2](#), use two columns/characters per valid code. Characters in your data not included in [CODES=](#) or [CODERANGE=](#) are given the [MISSCORE=](#) value. The highest category number 32767. [CODES=](#) can be up to 4,000 characters long. If [CODES=](#) and [CODERANGE=](#) are both specified, then the data are matched to [CODES=](#) first. [CODES=](#) can be rescored. [CODERANGE=](#) cannot be rescored.

Initially, Winsteps assumes that all items share the same response structure.

If there are 2 response codes in [CODES=](#), then this is the Rasch dichotomous model

If there are 3 or more response codes in [CODES=](#), then this the Andrich rating-scale model.

To override those,

If [ISGROUPS=0](#), then this is the Masters partial-credit model. Each item is modeled to have its own rating-scale structure.

If [ISGROUPS = AABACDDCADEEA](#), then the items are grouped to share rating scales.

The "A" group are items 1,2,5, 10,11,15 - they share the same rating scale.

The "B" group are items 3,4 - they share the same rating scale.

The "C" group

See also these Examples written with [CODERANGE=](#)

Example 1: A test has four response choices. These are "1", "2", "3", and "4". All other codes in the data file are to be treated as "item not administered". Each response uses 1 column in your data file. Data look like: 134212342132 . 3212343221

```
XWIDE=1 one character wide (the standard)
CODES=1234 four valid 1-character response codes
```

Example 2: There are four response choices. Each response takes up 2 columns in your data file and has leading 0's, so the codes are "01", "02", "03" and "04". Data look like: 0302040103020104040301

```
XWIDE=2 two characters wide
CODES=01020304 four valid 2-character response codes
```

Example 3: There are four response choices entered on the file with leading blanks, so that codes are " 1", " 2", " 3", and " 4". Data look like: 3 2 4 2 1 3 2

```
XWIDE=2 two characters wide
CODES=" 1 2 3 4" " required: blanks in 2-character responses
```

Note: when [XWIDE=2](#) or more, both [CODES=](#) and the data value are left-aligned before matching, so both " 1" and "1 " in [CODES=](#) match both " 1" and "1 " in your data file.

Example 4: Your data is a mixture of both leading blanks and leading 0's in the code field, e.g. "01", " 1", " 2", "02" etc. The numerical value of a response is calculated, where possible, so that both "01" and " 1" are analyzed as 1.

Data look like: 02 1 20102 1 2 01

```
XWIDE=2 two characters wide
CODES=" 1 2 3 401020304" two characters per response
```

Example 5: Your valid data are 1,2,3,4,5 and your missing data codes are 7,8,9 which you want reported separately on the [distractor tables](#).

```
CODES = 12345789
```

NEWSCORE = 12345XXX ; missing values scored with non-numeric values

Example 6: The valid responses to an attitude survey are "a", "b", "c" and "d". These responses are to be recoded "1", "2", "3" and "4". Data look like: abdbabcd

```
CODES =abcd four valid response codes
NEWSCORE=1234 new values for codes
RESCORE=2 rescore all items
```

Typically, "abcd" data implies a multiple choice test. Then KEY1= is used to specify the correct response. But, in this example, "abcd" always mean "1234", so that the RESCORE= and NEWSCORE= options are easier to use.

Example 7: Five items of 1-character width, "abcd", then ten items of 2-character width "AA", "BB", "CC", "DD". These are preceded by person-id of 30 characters. Data look like:

```
George Washington Carver III dabcdBBAACCAADDBBCCDDBBAA
FORMAT=(30A1,5A1,10A2) Name 30 characters, 5 1-chars, 10 2-chars
XWIDE =2 all converted to 2 columns
CODES ="a b c d AABCCDD" "a" becomes "a "
NEWSCORE="1 2 3 4 1 2 3 4 " response values
RES EDFILE= edit data file CORE=2 rescore all items
NAME1=1 name starts column 1 of reformatted record
ITEM1=31 items start in column 31
NI=15 15 items, all XWIDE=2
```

Example 8: Items are to rescored according to Type A and Type B. Other items to keep original scoring.

```
IREFER = AAAAAAABBBBBBBBCCCCCCC ; 3 item types
CODES = 1234 Original codes in the data file
IVALUEA = 1223 Recode Type A items
IVALUEB = 1123 Recode Type B items
IVALUEC = 1234 Recode Type * item. Can be omitted
```

Example 9: The valid responses are [percentages](#) in the range 00 to 99.

```
XWIDE = 2 two columns each percent
; for data with leading zeroes,
CODES = 0001020304050607080910111213141516171819+
+2021222324252627282930313233343536373839+
+4041424344454647484950515253545556575859+
+6061626364656667686970717273747576777879+
+8081828384858687888990919293949596979899
or, for data with blanks,
CODES =" 0 1 2 3 4 5 6 7 8 910111213141516171819+
+2021222324252627282930313233343536373839+
+4041424344454647484950515253545556575859+
+6061626364656667686970717273747576777879+
+8081828384858687888990919293949596979899"
```

also recommended:

```
ISRANGE = *
1 0 99 ; defines the full range of the rating scale, regardless of the data
*
SFUNCTION=4 ; defines a smooth function for the Rasch-Andrich thresholds
```

Example 10: Codes are in the range 0-254.

```
XWIDE=3 ; 3 characters per response:
; for data with leading zeroes,
CODES="000001002003004005006007008009010011012013014015016017018019020021022023+
+024025026027028029030031032033034035036037038039040041042043044045046047+
+048049050051052053054055056057058059060061062063064065066067068069070071+
+072073074075076077078079080081082083084085086087088089090091092093094095+
+096097098099100101102103104105106107108109110111112113114115116117118119+
+120121122123124125126127128129130131132133134135136137138139140141142143+
+144145146147148149150151152153154155156157158159160161162163164165166167+
+168169170171172173174175176177178179180181182183184185186187188189190191+
```

```

+192193194195196197198199200201202203204205206207208209210211212213214215+
+216217218219220221222223224225226227228229230231232233234235236237238239+
+240241242243244245246247248249250251252253254"
;or for data with blanks
CODES=" 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23+
+ 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47+
+ 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71+
+ 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95+
+ 96 97 98 99100101102103104105106107108109110111112113114115116117118119+
+120121122123124125126127128129130131132133134135136137138139140141142143+
+144145146147148149150151152153154155156157158159160161162163164165166167+
+168169170171172173174175176177178179180181182183184185186187188189190191+
+192193194195196197198199200201202203204205206207208209210211212213214215+
+216217218219220221222223224225226227228229230231232233234235236237238239+
+240241242243244245246247248249250251252253254"

```

also recommended:

```

ISRANGE = *
1 0 254 ; defines the full range of the rating scale, regardless of the data
*
SFUNCTION=4 ; defines a smooth function for the Rasch-Andrich thresholds

```

Example 11: The data include negative values, -10, -9, through to positive codes 9, 10. Winsteps can only analyze data that are positive integers, so the data would need to be rescored:

```

XWIDE=3 ; each observation is 3 characters wide in the data file
; the next line is so that I can be sure that each code is 3 characters wide
;      123123123123123123123123123123123123123123123123123123123123123
CODES  ="-10-9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10 "
NEWSCORE="0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 "

```

also recommended:

```

ISRANGE = *
1 0 20 ; defines the full range of the rating scale, regardless of the data
*
SFUNCTION=4 ; defines a smooth function for the Rasch-Andrich thresholds

```

Example 12: Codes in the range 0 - 1000.

; CODES= is impractical. Use [CODERANGE=](#) 0-1000

11.18 CONVERGE= select convergence criteria

This selects which of [LCONV=](#) and [RCONV=](#) set the convergence criterion. See [convergence considerations](#).

CONVERGE=L	LCONV= for "Logit change size" controls convergence. Iteration stops when the biggest logit change is less or equal to LCONV=, or when the biggest logit change size increases (divergence).
CONVERGE=R	RCONV= for "Residual size" controls convergence. Iteration stops when the biggest residual score is less or equal to RCONV=, or when the biggest residual size increases (divergence).
CONVERGE=E	Either LCONV= for "Logit change size" or RCONV= for "Residual size" controls convergence. Iteration stops when the biggest logit change is less or equal to LCONV=, or when the biggest residual score is less or equal to RCONV=, or when both the biggest logit change size increases and the biggest residual size increases (divergence).
CONVERGE=B	Both LCONV= for "Logit change size" and RCONV= for "Residual size" controls convergence. Iteration stops when both the biggest logit change is less or equal to LCONV= and the biggest residual score is less or equal to RCONV=, or when both the biggest logit change size increases and the biggest residual size increases (divergence).
CONVERGE=F	Force both LCONV= for "Logit change size" and RCONV= for "Residual size" to control convergence. Iteration stops when both the biggest logit change is less or equal to LCONV= and the biggest residual score is less or equal to RCONV=.

Example 1: We want to be take a conservative position about convergence, requiring both small logit changes and small residual sizes when iteration ceases.

CONVERGE=Both

Example 2: We need very high precision, then specify:

CONVERGE=BOTH ; both score-residual and logit-change criteria

RCONV=.001 ; at most, one tenth of the smallest score residual needed.

LCONV=.00001 ; at most, one tenth of the highest precision to be reported.

These values are much, much smaller than the natural precision of the data, which is .5 raw score points, or the logit precision (S.E.) of the ability measures.

Example 3: We want to set the convergence criteria to match BIGSTEPS version 2.59

CONVERGE=B ; the criteria were LCONV= and RCONV=

RCONV= 0.5 ; the BIGSTEPS standards or whatever value you used

LCONV= .01

Example 4: We want to set the convergence criteria to match Winsteps version 3.20

CONVERGE=E ; the criterion was LCONV or RCONV

RCONV= 0.5 ; the 3.20 standards or whatever value you used

LCONV= .01

Example 5: We want the convergence criteria to match Winsteps version 2.85

CONVERGE= F ; force both LCONV and RCONV to be met

RCONV= 0.5 ; the 2.85 standards or whatever value you used

LCONV= .01

You may also want:

WHEXACT=NO ; centralized Wilson-Hilferty was the default

Example 6: *Question:* With anchored analyses, iterations never stop!

ITERATION	JMLE	MAX SCORE RESIDUAL*	MAX LOGIT CHANGE	LEAST CONVERGED EXID	BYCASE	CAT	CATEGORY RESIDUAL	STEP CHANGE
1		-239.04	.5562	1993	392*	6	85.70	-.5960
2		-105.65	-.1513	1993	392*	4	-28.92	.2745
.....								
18		-5.35	-.0027	2228	352*	3	2.35	.0146
19		-5.16	.0029	2228	352*	3	2.31	.0106
20		-5.05	.0025	2228	352*	3	2.28	.0055
21		-5.00	.0010	2228	352*	3	2.26	.0075
22		-4.99	-.0008	2228	352*	3	2.25	.0025
.....								
170		-5.00	-.0011	1377	352*	3	1.96	.0109
171		-5.00	.0018	187	352*	3	1.96	.0019
.....								

The standard convergence criteria in Winsteps are preset for "free" analyses. With anchored analyses, convergence is effectively reached when the logit estimates stop changing in a substantively meaningful way. This has effectively happened by iteration 20. Note that the logit changes are less than .01 logits - i.e., even the biggest change would make no difference to the printed output (which is usually reported to 2 decimal places)

To have the current Winsteps do this automatically, set

CONVERGE=L

LCONV=.005 ; set to stop at iteration 22 - to be on the safe side.

11.19 CSV= comma-separated values in output files

To facilitate importing the [IFILE=](#), [ISFILE=](#), [PFILE=](#), [SFILE=](#) and [XFILE=](#) files into spreadsheet and database programs, the fields can be separated by commas, and the character values placed inside " " marks.

CSV=No or Fixed	Use fixed field length format (the standard) (.txt)
CSV=Yes or CSV=,	Separate values by commas (or their international replacements). If QUOTES=Yes, character fields in " " marks
CSV=Tab or CSV=	Separate values by tab characters. If QUOTES=Yes, character fields in " " marks
CSV=SPSS	SPSS format (.sav)
CSV=Excel or XL	Excel format (.xls, .xlsx)
CSV=Rstat	R-statistics format (.rda)
CSV=Webpage	Webpage format (.htm, .html)

Examples:

Fixed space:

```
; MATCH Chess Matches at the Venice Tournament, 1971 Feb 11 0:47 2004
;ENTRY MEASURE STTS COUNT SCORE ERROR IN.MSQ IN.ZSTD OUT.MS OUT.ZSTD DISPL PTME WEIGHT
DISCR G M NAME
1 .87 1 2.0 2.0 .69 1.17 .47 1.17 .47 .01 1.00 1.00
1.97 1 R I0001
```

Tab-delimited:

```
"MATCH Chess Matches at the Venice Tournament, 1971 Feb 11 0:47 2004"
";" "ENTRY" "MEASURE" "STATUS" "COUNT" "SCORE" "ERROR" "IN.MSQ" "IN.ZSTD" .....
" " 1 .87 1 2.0 2.0 .69 1.17 .47 1.17 .47 .01 1.00 1.00 1.97 "1" "R" "I0001"
```

Comma-separated:

```
"MATCH Chess Matches at the Venice Tournament, 1971 Feb 11 0:47 2004"
";" "ENTRY" "MEASURE" "STATUS" "COUNT" "SCORE" "ERROR" "IN.MSQ" "IN.ZSTD" ,...
" " ,1 ,.87 ,1 ,2.0 ,2.0 ,.69 ,1.17 ,.47 ,1.17 ,.47 ,.01 ,1.00 ,1.00 ,1.97 ,"1" ,"R" ,"I0001"
```

SPSS format: This is the SPSS .sav file format.

11.20 CURVES= probability curves for Table 21

CURVES= specifies which curves in Table 21 are to display. "1" to display the curves. "0" to omit the curve. Formerly also controlled Table 2, now done with [T2SELECT=](#).

CURVES=000 indicates no curves are to be drawn - Table 21 will be skipped, unless [STEPT3=N](#), in which case only the structure summaries are output.

CURVES=101 displays subtables 21.1, 21.3 (and 21.4, 21.6, ...) when Table 21 is selected.

CURVES=	Table 21 displays ...	
100	21.1 , 21.4, ...	Category probability curves (modes, structure calibrations)
010	21.2 , 21.5, ...	Expected score ogive (means. model Item Characteristic Curve).
001	21.3 , 21.6, ...	Cumulative category probability curves (medians, shows Rasch-Thurstone thresholds = 50% cumulative probabilities)

11.21 CUTHI= cut off responses with high expectations

Use this if careless responses are evident. CUTHI= cuts off the top left-hand corner of the Scalogram in Table 22.

Eliminates (cuts off) observations where examinee ability measure is CUTHI= logits or more higher than item difficulty measure, so the examinee has a high probability of success. Removing off-target responses takes place after [PROX](#) has converged. After elimination, PROX is restarted, followed by [JMLE](#) estimation and fit calculation using only the reduced set of responses. This may mean that the original score-based ordering is changed.

Usually with [CUTLO=](#) and CUTHI=, misfitting items aren't deleted - but miskeys etc. must be corrected first. Setting CUTLO= and CUTHI= is a compromise between fit and missing data. If you loose too much data, then increase the values. If there is still considerable misfit or skewing of [equating](#), then decrease the values.

Here are the usual effects of [CUTLO=](#) and CUTHI=

1. Fit to the Rasch model improves.
2. The count of observations for each person and item decreases.
3. The [variance](#) in the data explained by the measures decreases.

Polytomous items: [CUTLO=](#) and CUTHI= trim the data relative to the item difficulty, so they tend to remove data in high and low categories. You can adjust the item difficulty relative to the response structure using [SAFILE=](#).

Example 1: Eliminate responses where examinee measure is 3 or more logits higher than item measure, to eliminate their worst careless wrong responses:

CUTHI= 3

This produces a scalogram with eliminated responses blanked out:

RESPONSES SORTED BY MEASURE:

```

KID      TAP
          111111111
        123745698013245678
-----
15  111      11100000 observations for extreme scores remain
14  111      1110000000
28  111      111010000000
30  1111 11110000000000
27  1111111000000000000

```

Example 2. At www.rasch.org/rmt/rmt62a.htm the originator of this approach suggests CUTLO= -1 and CUTHI=2

11.22 CUTLO= cut off responses with low expectations

Use this if guessing or [response sets](#) are evident. CUTLO= cuts off the bottom right-hand corner of the Scalogram in Table 22.

Eliminates (cuts off) observations where examinee ability measure is CUTLO= logits or more lower than item difficulty measure, so that the examinee has a low probability of success. The elimination of off-target responses takes place after [PROX](#) has converged. After elimination, PROX is restarted, followed by [JMLE](#) estimation and point-measure and fit calculation using only the reduced set of responses. This may mean that the original score-based ordering is changed.

CUTLO= is equivalent to Waller's procedure in Waller, M.I. (1976) "Estimating Parameters in the Rasch Model: Removing the Effects of Random Guessing", Report No. ETS-RB-76-0, Educational Testing Service, Princeton, N.J.

<http://files.eric.ed.gov/fulltext/ED120261.pdf>

Usually with CUTLO= and [CUTHI=](#), misfitting items aren't deleted - but miskeys etc. must be corrected first. Setting CUTLO= and CUTHI= is a compromise between fit and missing data. If you loose too much data, then increase the values. If there is still considerable misfit or skewing of [equating](#), then decrease the values.

Here are the usual effects of CUTLO= and [CUTHI=](#)

1. Fit to the Rasch model improves.
2. The count of observations for each person and item decreases.
3. The [variance](#) in the data explained by the measures decreases.

CUTLO= is equivalent to the procedure outlined in Bruce Choppin. (1983). *A two-parameter latent trait model*. (CSE Report No. 197). Los Angeles, CA: University of California, Center for the Study of Evaluation, and the procedure in David Andrich, Ida Marais, and Stephen Humphry (2012) *Using a Theorem by Andersen and the Dichotomous Rasch Model to Assess the Presence of Random Guessing in Multiple Choice Items*. Journal of Educational and Behavioral Statistics, 37, 417-442.

Polytomous items: CUTLO= and [CUTHI=](#) trim the data relative to the item difficulty, so they tend to remove data in high and low categories. You can adjust the item difficulty relative to the response structure using [SAFILE=](#).

Example 1: Disregard responses where examinees are faced with too great a challenge, and so might guess wildly, i.e., where examinee measure is 2 or more logits lower than item measure:

```
CUTLO= -2 ; 12% success
```

This is equivalent to a "Optimum Appropriateness Measurement" (OAM) model in which it is assumed that persons might guess on all the items, so all responses in guessing situations are eliminated.

GUTTMAN SCALOGRAM OF RESPONSES:

```
PERSON | ITEM
| 12 22 1 3231311 1322112 2 113322
|62257012473946508491143795368350281
|-----|
147 +0001111110011010100111111000001000 154
130 +1100101111001110110101000000010100 135
93 +011110010100111010000001101000010 098
129 +1111101110110110110000000000001010 134
134 +000110101110010001110010001000010 139
133 +100100111110000000101010011000000 138
137 +100000011110100101101000010000000 143
141 +11010001001110000110000000001010 147
138 +1010010101110100000000010010 144
113 +101100100010111100000000000010 118
144 +100000101000100010001000010001 151
114 +1000010010000100 119
|-----|
| 12 22 1 3231311 1322112 2 113322
|62257012473946508491143795368350281
```

Example 2: Richard Gershon applied this technique in [Guessing and Measurement](#) with CUTLO=-1 ; 27% success

Example 3: We have some misbehaving children in our sample, but don't want their behavior to distort our final report.

An effective approach is in two stages:

Stage 1. calibrate the items using the good responses

Stage 2. anchor the items and measure the students using all the responses.

In Stage 1, we trim the test. We want to remove the responses by children that are so off-target that successes are probably due to chance or other off-dimensional behavior. These responses will contain most of the misfit. For this we analyze the data using

CUTLO= -2 (choose a suitable value by experimenting)

CUTLO= -1.39 ; 20% success

CUTLO= -1.10 ; 25% success

write an item file from this analysis:

[IFILE=if.txt](#)

In Stage 2. Anchor all the items at their good calibrations:

[IAFILE=if.txt](#)

Include all the responses (omit CUTLO=)

We can now report all the children without obvious child mis-behavior distorting the item measures.

11.23 DATA= name of data file

Your data can be the last thing in the control file (which is convenient if you only have a small amount of data), but if you have a large amount of data, you can place it in a separate file, and then use DATA= to say where it is. [FORMAT=](#) reformats these records. [MFORMS=](#) enables multiple reformatting.

DATA= file name	file containing details
DATA = file name + file name+ ...	multiple data files
DATA = ?	opens a Browser window to find the file

Example 1: Read the observations from file "A:\PROJECT\RESPONSE.TXT".

DATA=A:\PROJECT\RESPONSE.TXT

Example 2: Read scanned MCQ data from file DATAFILE.txt in the current directory.

DATA=DATAFILE.txt

You may specify that several data files be analyzed together in one run, by listing their file names, separated by "+" signs. The list, e.g., FILE1.TXT+MORE.TXT+YOURS.D, can be up to 200 characters long. The layout of all data files must be identical.

Example 3: A math test has been scanned in three batches into files "BATCH.1", "BATCH.2" and "BATCH.3". They are to be analyzed together.

DATA=BATCH.1+BATCH.2+BATCH.3

11.24 DATESHOW= show date in Table headings

The date of the analysis is usually reported in Winsteps table headings. The date can be removed with DATESHOW=No

DATESHOW=Yes

```
TABLE 23.16 KNOX CUBE TEST                EXAM1-PLUS.OUT  Jul 30  2:06 2013
INPUT: 35 KID  18 TAP  REPORTED: 35 KID  18 TAP  2 CATS          WINSTEPS 3.80.0
-----
```

DATESHOW=No

```
TABLE 23.16 KNOX CUBE TEST                EXAM1-PLUS.OUT
INPUT: 35 KID  18 TAP  REPORTED: 35 KID  18 TAP  2 CATS          WINSTEPS 3.80.0
-----
```

11.25 DELIMITER= or SEPARATOR= data field delimiters

It is often convenient to organize your data with delimiters, such as commas, semi-colons or spaces, rather than in fixed column positions. However, often the delimiter (a Tab, space or comma) only takes one column position. In which case, it may be easier to include it in the [CODES=](#) or use [MFORMS=](#) or [FORMAT=](#).

To check that your data file has decoded properly, look at [RFILE=](#)

To do this, specify the following command DELIMITER= value (or SEPARATOR= value). This value is the separator.

Examples: DELIMITER= " " fixed-field values

DELIMITER= "," comma-separated values CSV. The , must be ","

DELIMITER=BLANK blank-separated values

or DELIMITER=SPACE space-separated values

DELIMITER=TAB tab-separated values

DELIMITER=";" semi-colon separated values. The ; must be ";", otherwise it is treated as a comment.

When decoding delimited values, leading and trailing blanks, and leading and trailing quotation marks, "" and '' in each value field are ignored. Responses are left-aligned, and sized according to XWIDE=.

For [NAME1=](#) and [ITEM1=](#), specify the value number in the data line, starting with 1 as the leftmost value. FORMAT= does not apply to this data design.

Combine your person name and demographic information into one field that is to be referenced by [NAME1=](#).

Example 1 of a data line:

; the following is ONE data line:

```
"01"; 02; "01"; "01"; "01"; 00; 02; 00; "01"; 02; 02; 02; 02; 00; 02; "01"; "01"; 02; 02; 00; 02; "01"; 00; 02; 00; ROSSNER, MARC DANIEL
```

;which decodes as:

```
01020101010002000102020202000201010202000201000200ROSSNER, MARC DANIEL
```

```
ITEM1=1      ; item responses start in first field
NI=25       ; there are 25 responses, i.e., 25 response fields
NAME1=26    ; the person name is in the 26th field
DELIMITER = ";" ; the field delimiters are semi-colons
XWIDE=2     ; values are right-aligned, 2 characters wide.
CODES=000102 ; the valid codes.
NAMLEN=20   ; override standard person name length of 30 characters.
```

Example 2 of a data line:

; the following is ONE data line:

```
ROSSNER - MARC DANIEL, "01", 02 , "01", "01", "01", 00, 02, 00, "01", 02, 02, 02, 02, 00, 02, "01", "01", 02, 02, 00, 02, "01", 00, 02, 00
```

;which decodes as:

```
01020101010002000102020202000201010202000201000200ROSSNER - MARC DANIEL
```

```
ITEM1=2      ; item responses start in second field
NI=25       ; there are 25 responses, i.e., 25 response fields
NAME1=1     ; the person name is in the 1st field
DELIMITER = "," ; the field delimiters are commas (so no commas in names)
XWIDE=2     ; values are right-aligned, 2 characters wide.
CODES=000102 ; the valid codes
NAMLEN=20   ; override standard person name length of 30 characters.
```

Example: Here is the data file, "Book1.txt"

```
fred,1,0,1,0
george,0,1,0,1
```

Here is the control file:

```
name1=1 ; first field
item1=2 ; second field
ni=4   ; 4 fields
data=book1.txt
codes=01
delimiter = ","
&END
looking
viewing
peeking
seeking
END LABELS
```

Here is the reformatted file from the [Edit Pull-Down menu](#): *View Delimiter File*:

```
1010fred
0101george
```

Suggestion:

If you data file can be conveniently organized in columns:

```
fred ,1,0,1,0
```

```
george,0,1,0,1
```

then process the data file as fixed width fields

```
NAME1 = 1 ; start of person label "fred"  
NAMELENGTH = 6 ; up to first comma  
ITEM1 = 8 ; first column of numbers  
XWIDE = 2 ; number + comma  
CODES = "0,1,0 1 " ; the data codes "0 1 " are for the last codes on the line.
```

this will then analyze the data as a standard fixed-column rectangular data matrix.

11.26 DIF= columns within person label for Table 30

DIF= specifies the part of the person label which is to be used for classifying persons in order to identify Differential Item Function (DIF) - uniform or [non-uniform](#) - using the column [selection rules](#). See also [DIF Table](#) and [DIF and DPF considerations](#).

DIF= file name	file containing details
DIF= *	in-line list
DIF= \$S1W1	field in person label. DIF fields must be inside the person label defined by NAME1= and NAMELENGTH=
DIF= MA2	person ability levels for non-uniform DIF
DIF= \$S1W1 + MA2	person classes or groups

DIF= location is usually column number **within person label** field. DIF=1 means "DIF selection character is first character of person label."

Example 1: Columns 18-20 of the person label (in columns 118-120 of the data record) contain a district code:

```
NAME1=101 ; person label starts in column 101  
NAMELENGTH = 20 ; person label must include all the DIF codes  
DIF = $S18W3 ; district starts in column 18 of person label with a width of 3  
or  
@district = 18W3 ; district starts in column 18 of person label with a width of 3  
DIF = @district ; DIF classifier  
tfile=*  
30 Table 30 for the DIF report (or use Output Tables menu)  
*
```

Example 2: DIF by Gender+Grade: Column 1 of the person labels contains gender (M or F) and Columns 7-8 contain grade level (K-12).

```
DIF = 1W1 + 7W2
```

Example 3: I have tab-separated data and my DIF indicator is in a separate field from the Person label.

```
Solution: for the DIF analysis, do a separate run of Winsteps. At the "Extra Specifications" prompt:  
NAME1=(location of DIF indicator)  
DIF=$S1W1
```

Example 4: Columns 18-20 of the person label (in columns 118-120 of the data record) contain a district code. Column 21 of the person label (in column 121 of the data record) has a gender code. Three independent DIF analyses are needed: district, gender, district+gender

```
NAME1=101 ; person label starts in column 101  
NAMELENGTH = 21 ; person label must include all the DIF codes  
DIF = *  
$S18W3 ; start in person label column 18 with a width of 3 - district  
$S21W1 ; start in person label column 21 with a width of 1 - gender  
$S18W3+ $S21W1 ; DIF by district+gender
```

*
 tfile=*
 30 Table 30 for the DIF report (or use Output Tables menu)
 *

Example 5: An investigation of [non-uniform DIF](#) with high-low ability classification for the KCT data.
 ; action the following with the Specification pull-down menu
 @SEX = \$S9W1 ; the sex of participants is in column 9 of the person label
 DIF = @SEX + MA2 ; look for non-uniform DIF (gender + two ability strata): [MA2](#)
 PSUBTOT = @SEX + MA2 ; summary statistics by gender and ability strata
 Tfile=* ; This is more easily actioned through the [Output Tables Menu](#)
 30 ; Table 30 - DIF report
 28 ; Table 28 - Person subtotals for DIF classifications
 *

Table 30: DIF specification is: DIF=@SEX+MA2

KID	DIF	DIF	KID	DIF	DIF	DIF	JOINT	TAP			
CLASS	MEASURE	S.E.	CLASS	MEASURE	S.E.	CONTRAST	S.E.	t	d.f.	Number	Name
F1	-1.86	1.09	M1	-4.54	1.15	2.68	1.59	1.69	8	6	6= 3-4-1

Table 28: Subtotal specification is: PSUBTOTAL=@SEX+MA2

KID	MEAN	S.E.	OBSERVED	MEDIAN	REAL		
COUNT	MEASURE	MEAN	P.SD		SEPARATION	CODE	
4	-2.08	.90	.89		.00	F1	<- Non-extreme
6	-2.82	.41	.91	-2.86	.32	M1	

Example 6: With Example0.txt (the Liking for Science rating scale data) you want to see if any items were biased against names starting with any letter of the alphabet, then:

run example0.txt
 request the DIF Table (Table 30) from the Output Tables menu
 specify: \$S1W1
 a DIF table is produced.

The equivalent DIF specification is: DIF=\$S1W1

Positive DIF size is higher ACT difficulty measure

KID	DIF	DIF	KID	DIF	DIF	DIF	JOINT	ACT			
CLASS	MEASURE	S.E.	CLASS	MEASURE	S.E.	CONTRAST	S.E.	t	d.f.	Number	Name
R	-.06	.54	W	.89	2.05	-.95	2.12	-.45	8	1	WATCH BIRDS
R	-.06	.54	L	-.65	.75	.59	.92	.64	12	1	WATCH BIRDS
R	-.06	.54	S	-.42	.57	.36	.78	.46	18	1	WATCH BIRDS
R	-.06	.54	H	-1.63	1.13	1.57	1.25	1.26	11	1	WATCH BIRDS
R	-.06	.54	D	.12	.86	-.18	1.01	-.18	11	1	WATCH BIRDS

.....

Example 7: The DIF demographic column is not part of the person label. Use the FORMAT= instruction to rearrange the record.

Was:
 0 1 2 3
 123456789012345678901234567890
 AxxxxxxxxxxxxxxxxxxxxxBCyyyyyyyyy ; original record layout
 ITEM1=2
 NI=19 ; the x's
 NAME1=21 ; starts with B

To move column 1 after column 21 with XWIDE=1:

```
xxxxxxxxxxxxxxxxxxxxxxxxBACyyyyyyy ; after internal reformatting
FORMAT = (T2,20A,T1,1A,T22,100A)
ITEM1 = 1 ; references position in the reformatted record.
NI = 19 ; the x's
NAME1 = 20 ; starts with B - references position in the reformatted record.
DIF = S2W1 ; code A
```

The first reformatted record is shown on your screen so you can check that it is how you want it.

Example 8: We only want DIF between person classification groups A and B, not the other person groups. Classification group code is in column 3 of the person label.

In the [Specification](#) pull down menu

```
PSELECT=??{AB}
```

Output Table menu: Table 30.

11.27 DISCRIMINATION= item discrimination

Rasch models assert that items exhibit the model-specified item discrimination. Empirically, however, item discriminations vary. During the estimation phase of Winsteps, all item discriminations are asserted to be equal, of value 1.0, and to fit the Rasch model. But empirical item discriminations never are exactly equal, so Winsteps can also report an estimate of those discriminations post-hoc (as a type of fit statistic). The amount of the departure of a discrimination from 1.0 is an indication of the degree to which that item misfits the Rasch model.

DISCRIM=NO Do not report an estimate of the empirical item discrimination.

DISCRIM=YES Report an estimate of the empirical item discrimination in the IFILE= and [Tables 6.1](#), 10.1, etc.

An [estimated discrimination](#) of 1.0 accords with Rasch model expectations for an item of this difficulty. A value greater than 1 means that the item discriminates between high and low performers more than expected for an item of this difficulty. A value less than 1 means that the item discriminates between high and low performers less than expected for an item of this difficulty. In general, the geometric mean of the estimated discriminations approximates 1.0, the Rasch item discrimination.

Rasch analysis requires items which provide indication of relative performance along the latent variable. It is this information which is used to construct measures. From a Rasch perspective, over-discriminating items are tending to act like switches, not measuring devices. Under-discriminating items are tending neither to stratify nor to measure.

Over-discrimination is thought to be beneficial in many raw-score and IRT item analyses. High discrimination usually corresponds to low MNSQ values, and low discrimination with high MNSQ values. In Classical Test Theory, Guttman Analysis and much of Item Response Theory, the ideal item acts like a switch. High performers pass, low performers fail. This is perfect discrimination, and is ideal for sample stratification, but such an item provides no information about the relative performance of low performers, or the relative performers of high performers.

Winsteps reports an approximation to what the discrimination parameter value would have been in a 2-PL IRT program, e.g., BILOG for MCQ, or PARSCALE for partial credit items. IRT programs artificially constrain discrimination values in order to make them estimable, so Winsteps discrimination estimates tend to be wider than 2-PL estimates. For the lower asymptote, see [ASYMPTOTE=](#).

The algebraic representation of the discrimination and lower asymptote estimate by Winsteps are similar to 2-PL/3-PL IRT, but the estimation method is different, because Winsteps does not change the difficulties and abilities from their 1-PL values. Consequently, in Winsteps, discrimination and asymptotes are indexes, not parameters as they are in 2-PL/3-PL.

A Rasch-Andrich threshold discrimination is also reported, see [Table 3.2](#).

With DISCRIM=YES,

ENTRY	RAW		INFIT	OUTFIT	SCORE	ESTIM					
NUMBER	SCORE	COUNT	MEASURE	ERROR	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	DISCR	ACTS

23	40	74	2.19	.21	2.42	6.3	4.13	8.9	.00	.09	WATCH A RAT
17	93	74	.16	.19	.65	-2.7	.59	-2.5	.70	1.20	WATCH WHAT ANIMALS EAT

11.28 DISFILE= category/distractor/option count file

DISFILE=filename produces an output file containing the counts for each distractor or option or category of each item, similar to [Table 10.3](#). See also [Distractor Analysis](#)

DISFILE=? opens a Browse window

[DISOPTION=](#) Code or [DISOPTION=](#) Score summary. These control the contents of the output fields.

With [DISOPTION=](#) Code

ITEM CODE	VALUE	SCORE	UNWTD	UNWTD %	WTD	WTD %	AVGE	MEAS	P.SD	MEAS	S.E.	MEAS	INFT	MNSQ	OUTF	MNSQ	PTMA
LABEL																	
13 *	-1	-1	17	56.67	17.0	56.67	1.14	1.99	.50	.00	.00	.46					
nh01 supermarket																	
13 a	0	0	2	15.38	2.0	15.38	-1.07	.12	.12	.68	.61	-.31					
nh01 supermarket																	
13 b	0	0	1	7.69	1.0	7.69	-.13	.00	.00	1.73	1.55	.14					
nh01 supermarket																	
13 c	1	1	8	61.54	8.0	61.54	-.58	.82	.31	1.35	1.40	-.11					
nh01 supermarket																	
13 d	0	0	2	15.38	2.0	15.38	.14	.60	.60	2.26	2.43	.36					
nh01 supermarket																	

With [DISOPTION=](#) Score

ITEM CODE	VALUE	SCORE	UNWTD	UNWTD %	WTD	WTD %	AVGE	MEAS	P.SD	MEAS	S.E.	MEAS	INFT	MNSQ	OUTF	MNSQ	PTMA
LABEL																	
13	-1	-1	17	56.67	17.0	56.67	1.14	1.99	.50	.00	.00	.46					
nh01 supermarket																	
13	0	0	5	38.46	5.0	38.46	-.40	.68	.34	1.41	1.53	.11					
nh01 supermarket																	
13	1	1	8	61.54	8.0	61.54	-.58	.82	.31	1.35	1.40	-.11					
nh01 supermarket																	

This file contains 1 heading lines (unless [HLINES=N](#)), followed by one line for each [CODES=](#) of each item (whether the code is observed for the item or not) containing:

Start:	End:	Heading	Description
1	10	ITEM	Item entry number
11	20	CODE (blank for DISOPTION= S)	Response code in CODES= or *** = missing (not subtotal)
21	25	VALUE (blank for DISOPTION= S)	Numerical value of the response code (directly or with KEYn= , NEWSCORE= , IVALUEx= , MISSCORE= etc.). -1 means "response is ignored"
27	30	SCORE	Score for category (DISOPTION=C) or summed across categories (DISOPTION=S) used for estimating the Rasch item measure (after recounting, weighting, etc.). -1 means "response is ignored"
31	40	UNWTD	Count of all valid response codes in data set (excludes missing except for "MISSING" lines)
41	50	UNWTD %	Percent of UNWTD responses to this item.
51	60	WTD	Count of all valid response codes (weighted with PWEIGHT=) in data set (excludes missing except for "MISSING" lines)
61	70	WTD %	% of WTD responses to this item

71	80	AVGE MEAS	Average person measure for WTD responses
81	90	P.SD MEAS	P.SD is the population standard deviation of the person measures for WTD responses
91	100	S.E. MEAS	Standard Error of average person measure for WTD responses = $\sqrt{P.SD (AVGE MEAS) / (WTD - 1)}$
101	101	INFT MNSQ	Infit Mean-square of weighted responses, omitting responses in extreme scores
111	120	OUTF MNSQ	Outfit Mean-square of weighted responses, omitting responses in extreme scores
121	130	PTMA	PTMA is the point-correlation between this code or summary score, scored 1, or its non-occurrence, scored 0, and the person raw scores or measures chosen by PTBISERIAL= . The computation is described in Correlations .
132	-	LABEL	Item label

DISOPTION=Code - Since the DISFILE= has the same number of CODES= and MISSING "****" entries for every item, the repeated fields are filled out with "0" for any unobserved response codes.

DISOPTION=Score - Only observed scores are reported.

Percentages for non-missing response-scores are computed based on total count of non-missing response scores. Percentages for missing response-scores are computed based on total count of all response scores.

When [CSV=Y](#), commas separate the values with quotation marks around the "Item label", response in CODES=, and ***. When [CSV=T](#), the commas are replaced by tab characters.

Example: You wish to write a file on disk called "DISCOUNT.TXT" containing the item distractor counts from Table 14.3, for use in constructing your own tables:

`DISFILE=DISCOUNT.TXT`

11.29 DISOPTION= distractor file option = C

DISOPTION= Code	The distractor file DISFILE= reports by item response code in CODES=
DISOPTION= Score	The distractor file DISFILE= reports by item response score values of codes

11.30 DISTRACTOR= output option counts in Tables 10, 13-15

This variable controls the reporting of counts of option, distractor or category usage in [Table 10.3](#) etc. The standard is [DISTRACTOR=YES](#), if more than two values are specified in [CODES=](#).

DISTRACTOR=No	Omit the option or distractor Table 10.3 etc.
DISTRACTOR=Yes	Include Table 10.3 etc, i.e., counts, for each item, for each of the values in CODES= , and for the number of responses counted as MISSSCORE=

Distractor Analysis - Multiple-Choice (MCQ) Distractors

The distractor table reports what happened to the original observations after they were scored dichotomously.

Item writers often intend the MCQ options to be something like:

- A. Correct option (scored 1) - high-ability respondents - highest positive point-biserial
- B. Almost correct distractor (scored 0) - almost-competent respondents - somewhat positive point-biserial
- C. Mostly wrong distractor (scored 0) - slightly competent respondents - zero to negative point-biserial
- D. Completely wrong distractor (scored 0) - ignorant respondents - highly negative point-biserial

The distractor table reports whether this happened.

Example of a good set of distractors:

ITEM	CODE	SCORE	AVGE MEAS	PTMEA	
1	D	0	-1.59	-0.76	Most incorrect choice
1	C	0	-0.44	-0.15	
1	B	0	0.22	0.10	
1	A	1	2.37	0.45	Most correct choice

You will obtain exactly the same Rasch measures if you score your MCQ items in advance and submit a 0-1 dataset to Winsteps.

11.31 DPF= columns within item label for Table 31

DPF= specifies the part of the item label which is to be used for classifying items in order to identify Differential Person Function (DPF) - uniform or non-uniform - using the column [selection rules](#). See also [DPF Table](#) and [DIF and DPF considerations](#).

DPF= file name	file containing details
DPF= *	in-line list
DPF= \$S1W1	field in item label
DPF= MA2	item difficulty levels for non-uniform DPF
DPF= \$S1W1 + MA2	item groups

See [ISUBTOTAL](#) for format.

See [DPF Table 31](#).

Example 1: Columns 3 and 4 of the item label (between &END and END LABELS) contains content-area code:

DPF = \$S3E4 ; start in column 3 and end in column 4 of item label

tfile=*

31 ; Table 31 is DPF Table (or use Output Tables menu)

*

Example 2: Columns 3 of the item label contains a content code. Column 5-6 have a complexity code. Two independent DIF analyses are needed:

DPF = *

\$S3W1 ; content analysis

\$S5W2 ; complexity

*

tfile=*

31 ; Table 31 is DPF Table (or use Output Tables menu)

*

11.32 EDFILE= edit data file

This permits the replacement of data values in your data file with other values, without altering the data file. Data values are in the original data file format, specified in [CODES=](#). If specified as [decimals](#), they are rounded to the nearest integers. Additional persons can be entered beyond those in [DATA=](#) or after [END LABELS](#). Values can be in any order, but are more speedily process in person-entry order. Person labels can be inserted or changed using [PLFILE=](#).

EDFILE= file name	file containing details
EDFILE= file name + file name + ...	multiple files
EDFILE = *	in-line list
EDFILE = ?	opens a Browser window to find the file

Suggestion: sort the data lines as "person-entry-number ascending" for faster processing.

Its format is:

EDFILE=* person entry number or "person selection" *	item entry number or "item selection"	insertion or replacement data value
--	---	-------------------------------------

Ranges are permitted for entry numbers: first-last.

Person and item selections must be in quotation marks "", and follow the selection rules:

Selection rules:	
Control characters match label or name. They start at the first column of the label or name.	
?	matches any character
*	matches any string of characters - must be last selection character. If * is in the first column, then every available person or item is selected.
A	matches A in the person label, and similarly all other characters except { }
{..}	braces characters which can match a single character: {ABC} matches A or B or C.
{..-..}	matches single characters in a range. {0-9} matches digits in the range 0 to 9.
{..-..}	matches a single "-" {AB-} matches A or B or "-".
{~ABX}	omits persons or items which match A or B or X
@fieldname=	positions the next selection character at the start of the specified field

Example 1: In your MCQ test, you wish to correct a data-entry error. Person 23 responded to item 17 with a D, not whatever is in the data file.

```
EDFILE=*
23 17 D      ; person 23, item 17, data value of D
*
```

Example 2: Person 43 failed to read the attitude survey instructions correctly for items 32-56. Mark these missing.

```
EDFILE=*
43 32-56 " " ; person 43, items 32 to 56, blanks are missing data.
*
```

Example 3: Persons 47-84 are to be given a rating of 4 on item 16.

```
EDFILE=*
47-84 16 4 ; persons 47 to 84, item 16, data value of 4
*
```

Example 4: Items 1-10 are all to be assigned a datum of 1 for the control sub-sample, persons 345-682.

```
EDFILE=*
345-682 1-10 1 ; persons 345-682, items 1 to 10, data value 1.
*
```

Example 5: Missing data values are to be imputed with the values nearest to their expectations.

- Produce [PFILE=](#), [IFILE=](#) and [SFILE=](#) from the original data (with missing).
- Use those as [PAFILE=](#), [IAFILE=](#), [SAFILE=](#) anchor files with a data set in which all the original non-missing data are made missing, and vice-versa - it doesn't matter what non-missing value is used.
- Produce [XFILE=](#) to obtain a list of the expected values of the originally missing data.
- Use the [EDFILE=](#) command to impute those values back into the data file. It will round expected values to the nearest integer, for us as a category value.

```
EDFILE=*
17 6 2.6 ; persons 17, item 6, expected value 2.6, imputed as category "3".
*
```

Example 6: All responses to item 6 for males "M" in column 6 of person label are to be coded as "missing", character ".":

```
EDFILE=*
"?????M" 6 .
*
```

Example 7: We want to do Examples 1, 2, 3, 4, 5, 6 all at once to our dataset:

```
EDFILE=*
23 17 D ; person 23, item 17, data value of D
43 32-56 " " ; person 43, items 32 to 56, blanks are missing data.
47-84 16 4 ; persons 47 to 84, item 16, data value of 4
345-682 1-10 1 ; persons 345-682, items 1 to 10, data value 1.
17 6 2.6 ; persons 17, item 6, expected value 2.6, imputed as category "3".
"?????M" 6 .
*
```

Example 8: We want person group X (in column 4 of the person label) except for subgroups 11, 24 (in columns 6, 7 of the person label):

```
PSELECT = "???X?{12}{14}" ; this selects X and 11, 14, 21, 24
EDFILE=*
"?????14" "?" . ; convert 14 to missing data
"?????21" "?" . ; convert 21 to missing data
*
```

Example 9: Exceedingly unexpected responses are to be coded "missing". (It is easier, but not as exact, to use [CUTLO=](#) and [CUTHI=](#) to trim the observations).

Either

[Extract into Excel](#) the list of unexpected observations from Table 6.6 or Table 10.6.

Or

Output the [XFILE=](#) to Excel
Sort by unexpectedness (standardized residual)
Delete all rows except for the responses you want to code missing

Then

Rearrange the columns: Person Item
In the third column put the desired missing data code.
Copy-and-paste the three columns into a text file.
In your Winsteps control file:
EDFILE = (text file name)
Rerun the analysis

Example 9: All data in a separate EDFILE= data file.

NAME1 = 1 ITEM1 = 31 NI = (number of items) CODES = ABCD EDFILE= * 1 1 A 1 5 B 2 3 A 2 10 C *	NAME1 = 1 ITEM1 = 31 NI = (number of items) CODES = ABCD EDFILE= eddata.txt DATA = personlabels.txt ; list of person labels &END END LABELS
---	--

&END	and in another file, eddata.txt,
....	1 1 A
END LABELS	1 5 B
(list of person labels or nothing)	2 3 A
	2 10 C

Example 10: Item bank recalibration for computer-adaptive tests (CAT) or similar.

Here is a method is that maintains the accuracy of previously-reported person measures as much as possible:

1. collect up all the relevant data and format the data into a rectangular dataset or equivalent. EDFILE= is useful for this.
2. [anchor all the persons](#) at their report measures
3. [anchor all items](#) at their item-bank difficulties, and rating-scale structures (if polytomies) [SAFILE=](#)
4. analyze the dataset
5. the item displacements tell us which items have drifted by how much.
6. items with displacements of more than 0.5 logits, that are also bigger than the item S.E.s, are candidates for recalibration.
7. unanchor all the displaced items. Keep everything else anchored
8. reanalyze the dataset. The displaced items will now have revised difficulties in the context of the anchored persons.

11.33 EFILE= expected scores on items

EFILE=filename produces an output file containing the Rasch-model expected score on each item for the measures in the operational range of the set of items.

Subsets of items can be reported by using [IDELETE=](#) or [ISELECT=](#) from the [Specification menu](#).

Options using ETYPE=

ETYPE=E	Expected score on the item for the ability measure according to the Rasch model
ETYPE=P	Most probable category on the item for the ability measure according to the Rasch model. If more than one category, then middle or immediately higher category is shown.
ETYPE=T	Thurstone category on the item for the ability measure according to the Rasch model. If more than one category, then middle or immediately higher category is shown.

Example 1: ETYPE=E with dichotomous 0-1 items.

```
; EXPECTED SCORES FOR Knox Cube Test (Best Test Design p.31)
MEASURE SCORE      4      5      14      15      16      17
  -7.00   .11   .07   .04   .00   .00   .00   .00
  -6.85   .12   .08   .04   .00   .00   .00   .00
  -6.70   .14   .09   .05   .00   .00   .00   .00
  -6.55   .16   .10   .06   .00   .00   .00   .00
.....
   7.70   5.74   1.00   1.00   .98   .92   .92   .92
   7.85   5.78   1.00   1.00   .98   .93   .93   .93
   8.00   5.81   1.00   1.00   .99   .94   .94   .94
```

MEASURE -7.00	ability measure in logits or user-scaled units (UJMEAN= , USCALE=)
SCORE .11	expected raw score = sum of expected scores on the items with item weighting if IWEIGHT=
4, 5, 14, 15, ...	item entry numbers
.07, .04, .00, ...	Rasch-model expected score or category on the item for the MEASURE

Example 2: Reckase Chart of Expected Scores on All Items for All Measures.

Reckase, M. D. (2001). Innovative methods for helping standard-setting participants to perform their task: The role of feedback regarding consistency, accuracy, and impact. In: Gregory J. Cizek (Ed.), *Setting performance standards: Concepts methods and perspectives* (pp. 159-173). Mahwah, NJ: Erlbaum.
 MacCann & Stanley, "Rasch Standard Setting", *Practical Assessment Research & Evaluation*, 11, 2, 2, Table 1.

Example 3: Angoff standard-setting ratings to measures to cut-scores

In <https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1188&context=par> "The Use of Rasch Modeling To Improve Standard Setting", their Table 1 is the EFILE= output
 Their Table 2 is deduced from the EFILE= output

11.34 END LABELS or END NAMES

The first section of a control file contains control variables, one per line, and ends with [&END](#). This is followed by the second section of item labels, one per line, matching the items in the analysis. This sections ends with END LABELS or END NAMES, which mean the same thing. The data can follow as a third section, or the data can be in a separate file specified by the control variable [DATA=](#).

```
TITLE = "5 item test"
ITEM1 = 1
NI = 5
.....
&END
Addition ; label for item 1
Subtraction ; label for item 2
Multiplication ; label for item 3
Division ; label for item 4
Geometry ; label for item 5
END LABELS
..... ; data here
```

11.35 EQFILE= code equivalences

This specifies that different demographic or item-type codes are to be reported as one code. This is useful for Tables [27](#), [28](#), [30](#), [31](#), [33](#) and Use EQFILE=filename or EQFILE=*, followed by a list, followed by a *. These values can be overwritten from the equivalence boxes when invoking the Tables from the Output Tables menu.

EQFILE= file name	file containing details
EQFILE = *	in-line list
EQFILE = ?	opens a Browser window to find the file

The format is

[@Field name](#) = \$S1W1 ; user defined field name and location, see [selection rules](#).

EQFILE=* ; start of list

[@Field name](#) ; field to be referred to

Base Code Code Code Code ; code list

Base Code Code Code Code

Base Code Code Code

[@Field name](#) ; field to be referred to

Base Code Code Code Code Code ; code list

Base Code Code Code Code

Base Code Code Code Code

* ; end of list

where [@Field name](#) is the name of field in the person or item label, such as

@GENDER = \$S1W1 ; M or F

@STRAND = \$S10W2 ; 01 to 99

Base is the demographic or item-type code to be reported. It need not be present in a label

Code is a demographic or item-type code to be included with the Base code.

Example: gender in the data file is coded 1=Male, 2=Female, but to be reported as M or F

```
@GENDER = $S1W1 ; or wherever it is in the person label
EQFILE=*
@GENDER
M 1
F 2
*
```

11.36 ETYPE= option for EFILE=

See [EFILE=](#)

11.37 EXCAT= extreme category adjustment

When using Rasch measures for prediction, slightly better results have been obtained when the measure estimates are based on adjusted data. A productive adjustment makes observations in extreme categories slightly less extreme.

EXCAT= 0	observations in the top and bottom categories of a rating scale or dichotomy are not adjusted
EXCAT= 0.25 (or any value between 0 and 1)	observations in the top and bottom categories are made more central by EXCAT= score points

Example: EXCAT=0.25

Dichotomies: 0,1 data are analyzed as 0.25,0.75 data. The observed raw scores are the sums of these values. The category frequencies are: 0 becomes $0.75*0$ and $0.25*1$, 1 becomes, $0.25*0$, $0.75*1$.

Rating scale: 1,2,3,4 data are analyzed as 1.25,2,3,3.75 data. The observed raw scores are the sums of these values. The category frequencies are: 1 becomes $0.75*1$ and $0.25*2$, 4 becomes, $0.25*3$, $0.75*4$.

11.38 EXCELNUMBER= Excel converts numeric labels into numbers

When person and items labels are output to Excel using [IFILE=](#) and [PFILE=](#), those numeric labels can be stored as characters or numerical values.

EXCELNUMBER = No (default)	All person and items labels become Excel text. A blank may be appended to the label to force Excel to do this.
EXCELNUMBER = Yes	<p>Strictly numerical labels become Excel numerical values.</p> <p>Strictly numerical labels have</p> <ul style="list-style-type: none"> i) at least one numeral ii) the first character is one of .,+,-0123456789 iii) the remaining characters are .,0123456789 <p>Examples are:</p> <p>1234 +1234.45 -123,784</p> <p>Excel stores and displays numerical values using its own format rules.</p> <p>All character labels become Excel text.</p> <p>Example are:</p> <p>Arthur The First of June</p> <p>All other labels become Excel text. A blank may be appended to the label to force text in Excel.</p>

	<p>Examples are: 2+2= (which Excel might otherwise display as 4) 23E2 (which Excel might otherwise display as 2300)</p>
--	---

11.39 EXTRSC= extreme score correction for extreme measures

EXTRSCORE= is the fractional score point value to subtract from perfect (maximum possible) scores, and to add to zero (minimum possible) scores, in order to estimate finite values for extreme scores (formerly MMADJ=). Look at the location of the E's in the tails of the test ogive in Table 20. If they look too far away, increase EXTRSC= by 0.1. If they look too bunched up, reduce EXTRSCORE= by 0.1.

The measure corresponding to an extreme (maximum or minimum) score is not estimable, but the measure corresponding to a score a little less than maximum, or a little more than minimum, is estimable, and is often a useful measure to report.

This is how Winsteps handles extreme scores. In Rasch theory, an extreme score (maximum possible score or minimum possible score) on a set of items corresponds to an infinite ability measure (theta). This is impractical and also misleading in most situations. So Winsteps takes the following action:

If the score is maximum possible, then Winsteps estimates the ability measure for the score (maximum possible - EXTRSCORE) where EXTRSCORE is a small adjustment to the score. Usually EXTRSCORE = 0.3 score-points. The reasonable range of EXTRSCORE is 0.1 to 0.5.

If the score is minimum possible, then Winsteps estimates the ability measure for the score (minimum possible + EXTRSC)

So in one line, this becomes:

Score for estimation = Maximum (Minimum(observed score, maximum possible score - EXTRSCORE) , minimum possible score + EXTRSCORE)

The actual ability measure corresponding to an adjusted extreme score or any other score depends on the spread of the items, but it is at least $\ln((\text{score for estimation} - \text{minimum possible score}) / (\text{maximum possible score} - \text{score for estimation}))$ away from the mean item difficulty.

Rasch programs differ in the way they estimate measures for extreme scores. Adjustment to the value of EXTRSC= can enable a close match to be made to the results produced by other programs.

There is no "correct" answer to the question: "How large should extreme score adjustments be?" The most conservative value, and that recommended by Joseph Berkson, is 0.5. Some work by John Tukey indicates that 0.167 is a reasonable value. The smaller you set EXTRSC=, the further away measures corresponding to extreme scores will be located from the other measures. The technique used here is Strategy 1 in www.rasch.org/rmt/rmt122h.htm.

See also: G.H. Fischer , On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. Psychometrika 46 (1981), pp. 59–77

Treatment of Extreme Scores	Tables	Output files
Placed at extremes of map	1, 12, 16	
Reported by estimated measure	2, 3, 13, 14, 15, 17, 18, 19, 20, 22, 25, 28, 29, 30, 31, 33, 34, 35, 36	IFILE= , ISFILE= , PFILE= , RFILE= , XFILE=
Omitted	4, 5, 6, 7, 8, 9, 10, 11, 21, 23, 24, 26	SFILE=

Example 1: You wish to estimate conservative finite measures for extreme scores by subtracting 0.4 score points from each perfect score and adding 0.4 score points to each zero person score.

EXTRSCORE=0.4

Example 2: With the standard value of EXTRSCORE=, this Table is produced:

```

+-----+
|ENTRY  RAW                | INFIT | OUTFIT |PTBSE|      |

```


NUMBER	SCORE	COUNT	MEASURE	ERROR	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	PERSON
46	60	20	7.23	1.88	MAXIMUM ESTIMATED MEASURE					XPQ003
94	62	21	5.83	1.12	.44	-.9	.08	-.6	.62	XPQ011
86	18	6	5.11	1.90	MAXIMUM ESTIMATED MEASURE					XPQ009
64	50	17	4.94	1.09	.53	-.7	.13	-.6	.60	XPQ006

Here, the 5.11 corresponds to a perfect score on 6 easier items. The 5.83 was obtained on 21 harder items (perhaps including the 6 easier items.) To adjust the "MAXIMUM ESTIMATED" to higher measures, lower the value of EXTRSCORE=, e.g., to EXTRSCORE=0.2

Example 3. EXTRSCORE= is applied to the response string as observed. So, imagine that Mary is administered 5 dichotomous items out of a pool of 100 items and scores R on the 5 items. Her estimate will be based on:
Score for estimation = Maximum (Minimum(R, 5 - EXTRSCORE) , 0 + EXTRSCORE)

11.40 FITHIGH= higher bar in charts

Use FITHIGH= to position the higher acceptance bars in [Tables like 6.2](#). Use FITLOW= to position the lower bar.

FITHIGH=0 cancels the instruction.

Example: We want the lower mean-square acceptance bar to be shown at 1.4
FITHIGH=1.4 ; show higher fit bar
CHART=YES ; produce Tables like 6.2

ACTS FIT GRAPH: MISFIT ORDER

ENTRY	MEASURE	INFIT MEAN-SQUARE	OUTFIT MEAN-SQUARE	
NUMBER	- + 0.0	1 1.4	2 0.0	1 1.4 2 ACTS
23	* :	. :	* :	* WATCH A RAT 0
5	* :	. :	* :	* FIND BOTTLES AND CANS 0
20	* :	. *:	:	* WATCH BUGS 0
18	* :	. *:	:	* GO ON PICNIC 0
8	* :	. *:	:	* LOOK IN SIDEWALK CRACKS 0

11.41 FITI= item misfit criterion

Specifies the minimum *t* standardized fit value at which items are selected for reporting as misfits. For [Table 10](#), the table of item calibrations in fit order, an item is omitted only if the absolute values of both *t* standardized fit statistics are less than FITI=, and both mean-square statistics are closer to 1 than (FITI=)/10, and the item point-biserial correlation is positive. The 26 most underfitting items (high mean-squares) and 26 most overfitting items (low mean-squares) are always reported.

For [Table 11](#), the diagnosis of misfitting items, all items with a *t* standardized fit greater than FITI= are reported. Selection is based on the OUTFIT statistic, unless you set OUTFIT=N in which case the INFIT statistic is used. If MNSQ=YES, then selection is based on the mean-square value: 1 + FITI=/10.

Example 1: You wish to focus on grossly "noisy" items in Tables 10 and 11.
FITI=4 an extreme positive value

Example 2: You wish to include all items in Tables 10 and 11.
FITI= 0

11.42 FITLOW= lower bar in charts

Use FITLOW= to position the lower acceptance bars in [Tables like 6.2](#). Use FITHIGH= to position the higher bar.

FITLOW=0 cancels the instruction.

Example: We want the lower mean-square acceptance bar to be shown at 0.6
FITLOW=0.6 ; show lower fit bar
CHART=YES ; produce Tables like 6.2

PUPIL FIT GRAPH: MISFIT ORDER

ENTRY NUMBER	MEASURE		INFIT MEAN-SQUARE			OUTFIT MEAN-SQUARE			PUPIL	
	-	+	.0	0.6	1	:2	.0	0.6		1
46	*		:	.	*	T	:	.	*	MULLER, JEFF
26	*		:	*	.	U	:	.	*	NEIMAN, RAYMOND
30	*		:	.	*	V	:	.	*	NORDGREN, JAN SWEDE
23		*	:	*	.	c	*	:	.	VROOM, JEFF
22	*		*	:	.	b	*	:	.	HOGAN, Kathleen
21	*		*	:	.	a	*	:	.	RISEN, NORM L.

11.43 FITP= person misfit criterion

Specifies the minimum *t* standardized fit value at which persons are selected for reporting as misfits. For [Table 6](#), person measures in fit order, a person is omitted only if the absolute values of both *t* standardized fit statistics are less than FITP=, and both mean-square statistics are closer to 1 than (FITP=)/10, and the person point-biserial correlation is positive. The 26 most underfitting persons (high mean-squares) and 26 most overfitting persons (low mean-squares) are always reported.

For [Table 7](#), the diagnosis of misfitting persons, with a *t* standardized fit greater than FITP= are reported. Selection is based on the OUTFIT statistic, unless you set OUTFIT=N in which case the INFIT statistic is used. If MNSQ=YES, then selection is based on the mean-square value: 1 + FITP=/10.

Example 1: You wish to examine wildly guessing persons in Tables 6 and 7.
FITP= 3 an extreme positive value

Example 2: You wish to include all persons in Tables 6 and 7.
FITP= 0

11.44 FORMAT= reformat data

Enables you to process awkwardly formatted data! But [MFORMS=](#) is easier

FORMAT= is rarely needed when there is one data line per person.

Place the data in a separate [DATA=](#) file, then the Winsteps screen file will show the first record before and after FORMAT=. The formatted data records are also shown from the [Edit menu](#), Formatted Data=.

Control instructions to pick out every other character for 25 two-character responses, then a blank, and then the person label:

```
xWIDE=1
data=datafile.txt
format=(T2,25(1A,1X),T90,1A,T11,30A)
```

This displays on the Winsteps screen:

```
Opening: datafile.txt
Input Data Record before FORMAT=:
 1         2         3         4         5         6         7
123456789012345678901234567890123456789012345678901234567890
-----
01xx 1x1 100020001020202000201010202000201000200ROSSNER, MARC DANIEL
Input Data Record after FORMAT=:
1x11102012222021122021020 L
^I                ^N^P
```

^I is Item 1= column
^N is the last item according to NI=
^P is Name 1= column

FORMAT= enables you to reformat one or more data record lines into one new line in which all the component parts of the person information are in one person-id field, and all the responses are put together into one continuous item-response string. A FORMAT= statement is required if

- 1) each person's responses take up several lines in your data file.
- 2) if the length of a single line in your data file is more than 10000 characters.
- 3) the person-id field or the item responses are not in one continuous string of characters.
- 4) you want to rearrange the order of your items in your data record, to pick out sub-tests, or to move a set of connected forms into one complete matrix.
- 5) you only want to analyze the responses of every second, or nth, person.

FORMAT= contains up to 512 characters of reformatting instructions, contained within (..), which follow special rules. Instructions are:

nA	read in n characters starting with the current column, and then advance to the next column after them. Processing starts from column 1 of the first line, so that 5A reads in 5 characters and advances to the sixth column.
nX	means skip over n columns. E.g. 5X means bypass this column and the next 4 columns.
Tc	go to column c. T20 means get the next character from column 20. T55 means "tab" to column 55, not "tab" passed 55 columns (which is TR55).
TLc	go c columns to the left. TL20 means get the next character the column which is 20 columns to the left of the current position.
TRc	go c columns to the right. TR20 means get the next character the column which is 20 columns to the right of the current position.
/	means go to column 1 of the next line in your data file.
n(..)	repeat the string of instructions within the () exactly n times.
,	a comma is used to separate the instructions.
	Set <u>XWIDE=2</u> and you can reformat your data from original 1 or 2 column entries. Your data will all be analyzed as XWIDE=2. Then:
nA2	read in n pairs of characters starting with the current column into n 2-character fields of the formatted record. (For responses with a width of 2 columns.)
A1	read in n 1-character columns, starting with the current column, into n 2-character fields of the formatted record.

Always use nA1 for person-id information. Use nA1 for responses entered with a width of 1-character when there are also 2-character responses to be analyzed. When responses in 1-character format are converted into 2-character field format (compatible with XWIDE=2), the 1-character response is placed in the first, left, character position of the 2-character field, and the second, right, character position of the field is left blank. For example, the 1-character code of "A" becomes the 2-character field "A ". Valid 1-character responses of "A", "B", "C", "D" must be indicated by CODES="A B C D " with a blank following each letter.

ITEM1= must be the column number of the first item response in the formatted record created by the FORMAT= statement. NAME1= must be the column number of the first character of the person-id in the formatted record.

Example 1: Each person's data record file is 80 characters long and takes up one line in your data file. The person-id is in columns 61-80. The 56 item responses are in columns 5-60. Codes are "A", "B", "C", "D". No FORMAT= is needed. Data look like:

xxxxDCBDABCADCBACDADABDADCDADDCCDADDCAABCADCCBBADACDBBADCZarathrustra-Xerxes

Without FORMAT=

```
XWIDE=1 response width (the standard)
ITEM1=5 start of item responses
NI=56 number of items
NAME1=61 start of name
NAMLEN=20 length of name
CODES=ABCD valid response codes
```

With FORMAT=

Example 5: There are five records or lines in your data file per person. There are 100 items. Items 1-20 are in columns 25-44 of first record; items 21-40 are in columns 25-44 of second record, etc. The 10 character person-id is in columns 51-60 of the last (fifth) record. Codes are "A", "B", "C", "D". Data look like:

```

XXXXXXXXXXXXXXXXXXXXXXXXXXXXACDBACDBACDCABACDACD
XXXXXXXXXXXXXXXXXXXXXXXXXXXXDABCBACDBACDCABACDA
XXXXXXXXXXXXXXXXXXXXXXXXXXXXACDBACDBACDCABACDACD
XXXXXXXXXXXXXXXXXXXXXXXXXXXXDABCBACDBACDCABACDA
XXXXXXXXXXXXXXXXXXXXXXXXXXXXABCDBACDBACDCABACDADxxxxxxMary-Jones

```

becomes:

```
ACDBACDBACDCABACDADDDABCBACDBACDCABACDAACDBACDBACDCABACDACDDABCBACDBACDCABACDAABCBACDBACDCABACDADMary-Jones
```

```

FORMAT=(4(T25,20A,/,),T25,20A,T51,10A)
ITEM1=1 start of item responses
NI=100 number of item responses
NAME1=101 start of person name in formatted record
NAMLEN=10 length of person name
CODES=ABCD valid response codes

```

Example 6: There are three lines per person. In the first line from columns 31 to 50 are 10 item responses, each 2 columns wide. Person-id is in the second line in columns 5 to 17. The third line is to be skipped. Codes are "A", "B", "C", "D". Data look like:

```

XXXXXXXXXXXXXXXXXXXXXXXXXXXX A C B D A D C B A DXXXXXXXXXX
xxxxJoseph-CarlosXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

```

becomes:

Columns:

```

1-2-3-4-5-6-7-8-9-0-1234567890123
A C B D A D C B A DJoseph-Carlos

```

```

FORMAT=(T31,10A2,/,T5,13A1,/)
ITEM1=1 start of item responses
NI=10 number of items
XWIDE=2 2 columns per response
NAME1=11 starting "A" of person name
NAMLEN=13 length of person name
CODES='A B C D ' valid response codes

```

If the third line isn't skipped, format a redundant extra column in the skipped last line. Replace the first control variable in this with:

```
FORMAT=(T31,10A2,/,T5,13A1,/,A1) last A1 unused
```

Example 7: Pseudo-random data selection

To skip every other record, use (for most situations):

```
FORMAT=(500A, /) ; skips every second record of two
```

or

```
FORMAT=(/, 500A) ; skips every first record of two
```

You have a file with 1,000 person records. This time you want to analyze every 10th record, beginning with the 3rd person in the file, i.e., skip two records, analyze one record, skip seven records, and so on. The data records are 500 characters long.

```

XWIDE = 1
FORMAT = (/,/,500A,/,/,/,/,/,/,/)

```

or

```

XWIDE = 2
FORMAT = (/,/,100A2,300A1,/,/,/,/,/,/,/) ; 100 2-character responses, 300 other columns

```

Example 8: Test A, in file EXAM10A.TXT, and TEST B, in EXAM10B.TXT, are both 20 item tests. They have 5 items in common, but the distractors are not necessarily in the same order. The responses must be scored on an individual test basis. Also the validity of each test is to be examined separately. Then one combined analysis is wanted to equate the

tests and obtain bankable item difficulties. For each file of original test responses, the person information is in columns 1-25, the item responses in 41-60.

The combined data file specified in EXAM10C.TXT, is to be in RFILE= format. It contains

Person information 30 characters (always)
Item responses Columns 31-64

The identification of the common items is:

Test Item Number (=Location in item string)							
Bank	1	2	3	4	5	6-20	21-35
A:	3	1	7	8	9	2, 4-6, 10-20	
B:	4	5	6	2	11		1, 3, 7-10, 12-20

I. From Test A, make a response (RFILE=) file rearranging the items with FORMAT=.

```

; This file is EXAM10A.TXT
&INST
TITLE="Analysis of Test A"
RFILE=EXAM10AR.TXT ; The constructed response file for Test A
NI=20
FORMAT=(25A,T43,A,T41,A,T47,3A,T42,A,T44,3A,T50,11A)
ITEM1=26 ; Items start in column 26 of reformatted record
CODES=ABCD# ; Beware of blanks meaning wrong!
; Use your editor to convert all "wrong" blanks into another code,
; e.g., #, so that they will be scored wrong and not ignored as missing.
KEYFRM=1 ; Key in data record format
&END
Key 1 Record CCBDACABDADCBCABBCCA
BANK 1 TEST A 3 ; first item name
.
BANK 20 TEST A 20
END NAMES
Person 01 A BDABCDDBDACBCACBDBA
.
Person 12 A BADCACADCDABDDDCBACA

```

The RFILE= file, EXAM10AR.TXT, is:

```

Person 01 A 00001000010010001001
Person 02 A 00000100001110100111
.
Person 12 A 00100001100001001011

```

II. From Test B, make a response (RFILE=) file rearranging the items with FORMAT=. Responses unique to Test A are filled with 15 blank responses to dummy items.

```

; This file is EXAM10B.TXT
&INST
TITLE="Analysis of Test B"
RFILE=EXAM10BR.TXT ; The constructed response file for Test B
NI=35
FORMAT=(25A,T44,3A,T42,A,T51,A,T100,15A,T41,A,T43,A,T47,4A,T52,9A)
; Blanks are imported from an unused part of the data record to the right!
; T100 means "go beyond the end of the data record"
; 15A means "get 15 blank spaces"
ITEM1=26 ; Items start in column 26 of reformatted record
CODES=ABCD# ; Beware of blanks meaning wrong!
KEYFRM=1 ; Key in data record format
&END

```

```

Key 1 Record                                CDABCDDBABCADCBCBDCAD
BANK 1   TEST B 4
.
BANK 5   TEST B 11
BANK 6   TEST A 2
.
BANK 20  TEST A 20
BANK 21  TEST B 1
.
BANK 35  TEST B 20
END NAMES
Person 01 B                                BDABDDCDBBCCCCDAACBC
.
Person 12 B                                BADABBADCBA DBBBBBBB

```

The RFILE= file, EXAM10BR.TXT, is:

```

Person 01 B                                10111                                010101001000100
Person 02 B                                00000                                010000000001000
.
Person 11 B                                00010                                0010000000000100
Person 12 B                                00000                                000101000101000

```

III. Analyze Test A's and Test B's RFILE='s together:

```

; This file is EXAM10C.TXT
&INST
TITLE="Analysis of Tests A & B (already scored)"
NI=35
ITEM1=31 ; Items start in column 31 of RFILE=
CODES=01 ; Blanks mean "not in this test"
DATA=EXAM10AR.TXT+EXAM10BR.TXT ; Combine data files

; or, first, at the DOS prompt,
; C:> COPY EXAM10AR.TXT+EXAM10BR.TXT EXAM10AB.TXT(Enter)
; then, in EXAM10C.TXT,
; DATA=EXAM10AB.TXT

PFILE=EXAM10CP.TXT ; Person measures for combined tests
IFILE=EXAM10CI.TXT ; Item calibrations for combined tests
tfile=* ; List of desired tables
3 ; Table 3.1 for summary statistics, 3.2, ...
10 ; Table 10 for item structure
*
PRCOMP=S ; Principal components/contrast analysis with standardized residuals
&END
BANK 1   TEST A 3 B 4
.
BANK 35  TEST B 20
END NAMES

```

Shortening FORMAT= statements

If the required FORMAT= statement exceeds 512 characters, consider using this technique:

Relocate an entire item response string, but use an IDFILE= to delete the duplicate items, i.e., replace them by blanks. E.g., for Test B, instead of

```

FORMAT=(25A, T44,3A,T42,A,T51,A, T100,15A, 41,A,T43,A,T47,4A,T52,9A)
NI=35

```

Put Test 2 as items 21-40 in columns 51 through 70:

```

FORMAT=(25A, T44,3A,T42,A,T51,A, T100,15A, T41,20A)
NI=40

```

Blank out (delete) the 5 duplicated items with an IDFILE= containing:

24-26
22
31

11.45 FORMFD= the form feed character

Do not change FORMFD= unless you have problems printing the tables or importing them into some other program.

The form feed character indicates the start of a new page of print-out. The DOS standard is Ctrl+L (ASCII 12) which is what represented by ^ (Shift+6). The DOS standard is understood by most word-processing software and PC printers as the instruction to skip to the top of a new page, i.e., form feed. The ASA (FORTRAN) form feed character is 1.

Notepad does not have a "form feed" or page advance feature. You must put extra blank lines in the output files.

Example 1: You want your EPSON LQ-500 printer to form feed automatically at each new page of output. (You have already set the printer to use compressed print, at 15 cpi, because output lines contain up to 132 characters):

FORMFD=^ (the standard)

Example 2: Your main-frame software understands a "1" in the first position of a line of print-out to indicate the top of a new page:

FORMFD=1

11.46 FRANGE= half-range of fit statistics on plots

Specifies the t standardized fit Y-axis half-range, (i.e. range away from the origin), for the t standardized fit plots. FRANGE= is in units of t standardized fit (i.e., expected mean = 0, standard deviation = 1).

Example: You want the fit plots to display from -3 to +3 units of t standardized fit:

FRANGE=3

11.47 FREQUENT= heading in Tables 1, 12, 16

Please see [MORE=](#)

11.48 FSHOW= show files created from control file

When files are created from a control variable in the control file are not immediately displayed in a window unless FSHOW=Yes.

Files can be displayed from the [Edit Menu](#).

Example: The [PFILE=](#) is specified in the control file, and is to be displayed as soon as it is created:

PFILE= examineelist.txt

FSHOW=Yes ; display immediately

11.49 GRFILE= probability curve coordinate output file

If GRFILE=filename is specified, a file is output which contains a list of measures relative to item difficulty (x-axis coordinates) and corresponding expected scores and category probabilities (y-axis coordinates) to enable you to use your own plotting program to produce item-category plots like those in the Graphs menu.

GRFILE=? opens a Browse window

Column Heading	Column Contents	Fixed Columns
----------------	-----------------	---------------

ITEM (or ITEM= value)	An example item number from the response-structure grouping - see ISGROUPS=	6
MEAS (Measure)	The measure relative to item difficulty (user-rescaled by USCALE=) with UDECIMALS= decimal places. Add the item difficulty to MEAS for absolute measure relative to the latent variable. The plotted range is at least MRANGE= away from its center.	7
SCOR (Score)	Expected score for the Modeled Item Characteristic Curve (ICC) or Item Response Function (IRF).	7
INFO (Information)	Statistical information for Item Information Function. Sum these for all items for the Test Information Function (TIF).	7
0 (Category number)	Probability of observing lowest category	7
1, 2, etc. (Category numbers)	Probability of observing higher categories	7
If CSV=Y , values are separated by commas. When CSV=T , values are separated by tab characters.		

For the numbers plotted for empirical ICCs, see the "[Empirical ICC](#)" plot, "Copy data to Clipboard"

Example 1: You wish to write a file on disk called "MYDATAGR.txt" containing x- and y-coordinates for plotting your item's category response curves and its item information function.

GRFILE=MYDATAGR.txt

With CSV=Y

```
"PROBABILITY CURVES FOR LIKING FOR SCIENCE (Wright & Masters p.18) Jul 4 16:03 2000"
;ITEM,MEAS,SCOR,INFO,0,1,2
1,-3.00,.11,.10,.89,.10,.00
1,-2.94,.12,.11,.89,.11,.00
1,-2.88,.12,.11,.88,.12,.00
1,-2.82,.13,.12,.87,.12,.00
1,-2.76,.14,.12,.87,.13,.00
1,-2.70,.14,.13,.86,.14,.00
```

With CSV=N (fixed spacing)

```
; PROBABILITY CURVES FOR LIKING FOR SCIENCE (Wright & Masters p.18) Nov 23 23:31 2008
ACT MEAS SCOR INFO 0 1 2
1 -3.00 .11 .10 .89 .10 .00
1 -2.94 .12 .11 .89 .11 .00
1 -2.88 .12 .11 .88 .12 .00
1 -2.82 .13 .12 .87 .12 .00
1 -2.76 .14 .12 .87 .13 .00
```

Example 2: I want the information function for every item in my test relative to the latent variable (absolute x-axis). All my items share the same rating scale.

1. output the IFILE= to Excel
2. delete all columns except the item difficulties
3. transpose the item difficulties into a row

4. output the GRFILE= to Excel
5. delete unwanted columns and unwanted top row
6. move MEAS into the first column, INFO in the second column.
7. insert two blank rows at the top
8. copy the item difficulty row from the IFILE worksheet into the top row of the GRFILE= worksheet starting in the third column
9. in each cell of the (information x item difficulty) rectangle, add the item difficulty from row 1 to the MEAS value in column 1
10. for each item, plot the INFO (column 2, same for all items) against the item's difficulty (MEAS) column

11.50 GROUPS= or ISGROUPS= assigns items to rating scale groupings

See [ISGROUPS=](#)

11.51 GRPFROM= location of ISGROUPS

Only use this if you have too many items to put conveniently on one line of the ISGROUPS= control variable.

Instructs where to find the [ISGROUPS=](#) information.

GRPFROM=N

ISGROUPS= is a control variable before &END (the standard).

GRPFROM=Y

ISGROUPS= information follows just after &END, but before the item names. It is formatted exactly like a data record. It is helpful to enter "ISGROUPS=", for reference, where the person name would go.

Example: An attitude survey of 10 items with 3 rating scale definitions. Items 1 through 3 on Rating Scale 1, items 4 through 6 on Rating Scale 2 and items 7 through 10 on Rating Scale 3. The ISGROUPS= information is formatted like a data record and entered after &END and before the item names. The responses are in columns 1-10, and the person-id in column 11 onwards.

```

NAME1=11 start of person-id
ITEM1=1 start of responses
NI=10 number of items
CODES=12345 valid responses
GRPFRM=Y ISGROUPS= formatted like data
&END
1112223333 ISGROUPS= information
Item name 1 item names
|
Item name 10
END NAMES
2213243223 John Smith first data record
|

```

11.52 GUFIL= (G0ZONE=, G1ZONE=) Guttmanized response file

This writes out the response file edited to more closely match an ideal Guttman scalogram. It is in a format close to the original data file, with items and person in entry order.

GUFIL=? opens a Browse window

Outlying 1's are converted to 0's according to G0ZONE=

Outlying 0's are converted to 1's according to G1ZONE=

This removes unlikely 1's in the G0ZONE (e.g., lucky guesses) and unlikely 0's in the G1ZONE (e.g. careless mistakes)

It is also useful for imputing theories about item hierarchies.

G0ZONE= sets the % of observed 0's, starting from the "difficult" side of the Guttman scalogram, among which all 1's are turned to 0's. (The item hierarchy is constructed with the current data, but can be enforced through anchoring.) Standard value is 50.

G1ZONE= sets the % of observed 1's, starting from the "easy" side of the Guttman scalogram, among which all 0's are turned to 1's. Standard value is 50.

Example:

```
GUFILE= guttman.txt
G0ZONE = 20%
G1ZONE = 40%
```

Original data (Guttman ordered)

```
11100110011001001010
becomes
11111110011001001000
```

The file format matches the input data file if both are in fixed-field format.

When GUFILE= is written with CSV=Y, comma-separated or CSV=T, tab-separated, the item responses precede the person label.

Example: KCT.txt Guttmanized with fixed field format:

```
Richard M 111111100000000000
Tracie F 111111111100000000
Walter M 111111111001000000
```

KCT.txt Guttmanized with comma-separated, CSV=Y, format:

```
1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,Richard M
1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,Tracie F
1,1,1,1,1,1,1,1,1,0,0,1,0,0,0,0,0,0,0,0,0,0,Walter M
```

11.53 HEADER= display or suppress subtable headings

Subtables are usually displayed with two heading lines, showing Table number, Title, date, etc.

To display these (the standard), specify HEADER=YES.

To suppress these, specify HEADER=NO. The heading lines are replaced by 2 blank lines.

Example: In Table 7.2, the person misfits. Heading lines are not wanted between persons.

HEADER=NO

11.54 HIADJ= correction for top rating scale categories

The Rasch model models the measure corresponding to a top rating (or partial credit) scale category as infinite. This is difficult to think about and impossible to plot. Consequently, graphically in Table 2.2 and numerically in Table 3.1 a measure is reported corresponding to a top category. This is the measure corresponding to an imaginary rating HIADJ= rating points below the top category. The corresponding instruction for the bottom category is [LOWADJ=](#).

Example: The standard spread in Table 2.2 is based on HIADJ=0.25. You wish the top category number to be printed more to the right, further away from the other categories.

HIADJ=0.1

11.55 HLINE= heading lines in output files

To facilitate importing the [IFILE=](#), [PFILE=](#), [SFILE=](#) and [XFILE=](#) files into spreadsheet and database programs, the heading lines can be omitted from the output files.

HLINE=Y Include heading lines in the output files (the standard)

In IFILE= and PFILE=, specifying HLINE=Y also puts ";" at the start of missing, deleted and extreme lines.

For simulated data files, [SIFILE=](#), HLINE=Y puts the person generating value in each data line.

HLINE=N Omit heading lines.

Example: I want a tab-separated score-to-measure file, without the column headings:

SCOREFILE=mysc.txt

HLINE=NO

CSV=TAB

```
0 -6.46 1.83 .28 217 85 1 2.9 1 2.9 1
1 -5.14 1.08 .81 278 50 0 .0 1 2.9 3
2 -4.22 .86 1.29 321 40 1 2.9 2 5.7 4
```

with column headings, HLINE=YES, the standard:

```
;" "KID" "SCORE FILE FOR"
"; TABLE OF SAMPLE NORMS (500/100) AND FREQUENCIES CORRESPONDING TO COMPLETE TEST"
";SCORE" "MEASURE" "S.E." "INFO" "NORMED" "S.E." "FREQUENCY" "%" "CUM.FREQ." "%" "PERCENTILE"
0 -6.46 1.83 .28 217 85 1 2.9 1 2.9 1
```

11.56 IAFILE= item anchor file

The [IFILE=](#) from one analysis can be used unedited as the item anchor file, [IAFILE=](#), of another.

IAFILE= file name	file containing details
IAFILE = *	in-line list
IAFILE = \$\$S1W1	field in item label

The item parameter values (deltas) can be anchored (fixed) using [IAFILE=](#). Anchoring facilitates [equating](#) test forms and building item banks. The items common to two test forms, or in the item bank and also in the current form, can be anchored at their other form or bank calibrations. Then the measures constructed from the current data will be [equated](#) to the measures of the other form or bank. Other measures are [estimated](#) in the frame of reference defined by the anchor values. The anchored values are imputed (inserted) in place of the estimated-from-the-data values. Mathematically, the anchor values are treated as though they, like the estimated-from-the-data values, are the best available estimates of the true values.

[Displacements](#) are reported, indicating the differences between the anchored values and the freely estimated values. If these are large, please try changing the setting of [ANCESTIM=](#).

For polytomies (rating scales, partial credit), [IAFILE=](#) must have [SAFILE=](#). The [IFILE=](#) and the [SFILE=](#) are really one file. For dichotomies, the [SFILE=](#) is uninformative, so it can be ignored. For polytomies, the [IFILE=](#) and the [SFILE=](#) form a pair, and so do the [IAFILE=](#) and the [SAFILE=](#). For polytomies, anchoring with the [IAFILE=](#) without the [SAFILE=](#) is usually meaningless. The items are not completely anchored. Use [IAFILE=](#) and [SAFILE=](#) if you need the polytomous item in one analysis to be identical in thresholds and overall difficulty to the same item in another analysis. Use only [SAFILE=](#) if you need the polytomous item in one analysis to be identical in thresholds to the same item in another analysis, but the overall item difficulties can differ.

[IAFILE=?](#) opens a Browse window

In order to anchor items, a data file must be created of the following form:

1. Use one line per item (or item range) to be anchored.
2. Type the sequence number of the item in the current analysis, a blank, and the measure-value at which to anchor the item (in logits if [UASCAL=1](#), or in your user-rescaled [USCALE=](#) units otherwise). [Arithmetical expressions](#) are allowed.

Further values in each line are ignored. An IFILE= works well as an IAFILE=.

3. If the same item appears more than once, the first anchor value is used. When an IFILE= will be used as an IAFILE=, be sure to output the measures with many decimal places: [UDECIMALS=4](#)

[UIMEAN=](#) and [UPMEAN=](#) are ignored when there are anchor values, [IAFILE=](#) or [PAFILE=](#)

Stopping estimation: usually Winsteps estimation converges successfully by itself. If it does not, the [ctrl+F](#) stops estimation. If this happens repeatedly for an analysis, then you can explicitly tell Winsteps to do whatever you see when you decide to end estimation. For instance, if you decide to stop estimation when the biggest change to logit estimates is less than .01 logits,

[LCONV=.01](#)

[CONVERGE=L](#)

With anchor values, which usually mean that sum score residuals will never be zero, this choice makes sense.

Examples:

2 3.47 ; anchors item 2 at 3.47 logits (or USCALE= values)

10-13 1.3 ; items 10, 11, 12, 13 are each anchored at 1.3 logits

2 5.2 ; item 2 is already anchored. This item anchoring is ignored

1-50 0 ; all the unanchored items in the range 1-50 are anchored at 0.

Anything after ";" is treated as a comment.

IAFILE = filename

Item anchor information is in a file containing lines of format

item entry number anchor value

item entry number anchor value

IAFILE=*

Item anchor information is in the control file in the format

IAFILE=*

item entry number anchor value

item entry number anchor value

*

IAFILE=\$SnnEnn or IAFILE=\$SnnWnn or [@Field](#)

Item anchor information is in the item labels using the column [selection rules](#). Blanks or non-numeric values indicate no anchor value.

Example 0: only one item is to be anchored:

Slow method - include in your control file:

CONVERGE= L ; Convergence decided by logit change

LCONVERGE=.00001 ; Set logit convergence tight because of anchoring

IAFILE = * ; Item anchor file to preset the difficulty of an item

6 0.25 ; Item 6 exactly at 0.25 logit point.

*

Faster method:

1) do a standard unanchored analysis

2) output Table 14 items

3) see the measure for item 6 (for me it is 1.30)

4) edit your control file so that UIMEAN = wanted value - current value = 1.30 - 0.25 = 1.05

5) do the standard unanchored analysis again: item 6 is now 0.25

Example 1: The third item is to be anchored at 1.5 logits, and the fourth at 2.3 logits.

1. Create a file named, say, "ANC.FIL"

2. Enter the line "3 1.5" into this file, which means "item 3 in this test is to be fixed at 1.5 logits".

3. Enter a second line "4 2.3" into this file, which means "item 4 in this test is to be fixed at 2.3 logits".

3. Specify, in the control file,

IAFILE=ANC.FIL

CONVERGE=L ; only logit change is used for convergence

LCONV=0.005 ; logit change too small to appear on any report.

or place directly in the control file:

```
IAFILE=*  
3 1.5  
4 2.3  
*
```

```
CONVERGE=L ; only logit change is used for convergence  
LCONV=0.005 ; logit change too small to appear on any report.
```

or in with the item labels:

```
IAFILE=$S10W4 ; location of anchor value in item label  
CONVERGE=L ; only logit change is used for convergence  
LCONV=0.005 ; logit change too small to appear on any report.  
&END  
Zoo  
House 1.5 ; item label and anchor value  
Garden 2.3  
Park  
END LABELS
```

To check: "A" after the measure means "anchored"

```
+-----+  
|ENTRY  RAW          | INFIT | OUTFIT | PTMEA | |  
|NUMBER SCORE COUNT MEASURE ERROR|MNSQ  ZSTD|MNSQ  ZSTD|CORR. |DISPLACE| ITEMS |  
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+  
| 3      32      35      1.5A  .05| .80  -.3| .32   .6| .53| .40| House |
```

Example 2: The calibrations from one run are to be used to anchor subsequent runs. The items have the same numbers in both runs. This is convenient for generating tables not previously requested.

1. Perform the calibration run, say,
C:> Winsteps SF.TXT SOME0.TXT IFILE=ANCHORS.SF TABLES=111

2. Perform the anchored runs, say,
C:> Winsteps SF.TXT MORE0.TXT IFILE=ANCHORS.SF TABLES=0001111
C:> Winsteps SF.TXT CURVES0.TXT IFILE=ANCHORS.SF CURVES=111

Example 3: Score-to-measure Table 20 is to be produced from known item and rating scale structure difficulties.

Specify:
IAFILE=iafile.txt ; the item anchor file
SAFILE=safile.txt ; the structure/step anchor file (only for polytomies)
TFILE=*
20 ; the score table
*

```
CONVERGE=L ; only logit change is used for convergence  
LCONV=0.005 ; logit change too small to appear on any report.  
STBIAS=NO ; anchor values do not need estimation bias correction.  
The data file comprises two dummy data records, so that every item has a non extreme score, e.g.,
```

For dichotomies:
CODES = 01
Record 1: 10101010101
Record 2: 01010101010

For a rating scale from 1 to 5:
CODES = 12345
Record 1: 15151515151
Record 2: 51515151515

Example 4. Anchoring polytomous items for the Rating Scale Model

```

CODES = 012 ; 3 category Rating Scale Model
IAFILE=*
1 2.37 ; anchor item 1 at 2.37 logits
2 -1.23
*

SAFILE=*
0 0 ; the bottom category is always anchored at 0
1 -2.34 ; Andrich threshold (step difficulty) from category 0 to 1
2 2.34 ; Andrich threshold (step difficulty) from category 2 to 3
*
```

Example 5. Anchoring polytomous items for the Partial Credit and Grouped-Items models

```

CODES = 012 ; 3 category Rating Scale Model
ISGROUPS=0
IAFILE=*
1 2.37 ; anchor item 1 at 2.37 logits
2 -1.23
*

SAFILE=*
; for item 1, relative to the difficulty of item 1
1 0 0 ; the bottom category is always anchored at 0
1 1 -2.34 ; Andrich threshold (step difficulty) from category 0 to 1
1 2 2.34 ; Andrich threshold (step difficulty) from category 2 to 3
; for item 2, relative to the difficulty of item 2
2 0 0 ; the bottom category is always anchored at 0
2 1 -1.54 ; Andrich threshold (step difficulty) from category 0 to 1
2 2 1.54 ; Andrich threshold (step difficulty) from category 2 to 3
*
```

11.57 IANCHQU= anchor items interactively

Items to be anchored can be entered interactively by setting IANCHQ=Y. If you specify this, you are asked if you want to anchor any items. If you respond "yes", it will ask if you want to read these anchored items from a file; if you answer "yes" it will ask for the file name and process that file in the same manner as if [IAFILE=](#) had been specified. If you answer "no", you will be asked to enter the sequence number of each item to be anchored, one at a time, along with its logit (or user-rescaled by [USCALE=](#) or [UASCALE=](#)) value. When you are finished, enter a zero.

Example: You are doing a number of analyses, anchoring a few, but different, items each analysis. You don't want to create a lot of small anchor files, but rather just enter the numbers at the terminal, so specify:

```

IANCHQ=Y
CONVERGE=L ; only logit change is used for convergence
LCONV=0.005 ; logit change too small to appear on any report.
```

You want to anchor items 4 and 8.

Winsteps asks you:

```

DO YOU WANT TO ANCHOR ANY ITEMS?
respond YES(Enter)
DO YOU WISH TO READ THE ANCHORED ITEMS FROM A FILE?
respond NO(Enter)
INPUT ITEM TO ANCHOR (0 TO END):
respond 4(Enter) (the first item to be anchored)
INPUT VALUE AT WHICH TO ANCHOR ITEM:
respond 1.45(Enter) (the first anchor value)
INPUT ITEM TO ANCHOR (0 TO END): 8(Enter)
INPUT VALUE AT WHICH TO ANCHOR ITEM:-0.23(Enter)
INPUT ITEM TO ANCHOR (0 TO END): 0(Enter) (to end anchoring)
```

11.58 ICORFILE= item residual correlation file

This writes out the Table of inter-item correlations which is the basis of the principal components analysis of residuals.

The basis for the correlations is controlled by [PRCOMP=](#).

PRCOMP=R correlates the raw residuals

PRCOMP=S correlates the standardized residuals

PRCOMP=O correlates the observations

ICORFILE=? opens a Browse window

Extreme scores: minimum possible (0) and maximum possible (perfect) person scores are omitted from the computation of the correlations. Items with extreme scores are not reported. Their correlation is 0.

Missing data: for these Winsteps substitutes their expectations when possible. For residuals and standardized residuals, these are 0. Persons with extreme scores (minimum possible or maximum possible): Winsteps drops these from the correlation computation. The reason for these choices is to make the principal components analysis of residuals as meaningful as possible.

Expected correlations: the expected inter-item correlation between any two items is $-1/(L-1)$. Anscombe & Tukey, 1963, p. 144, *courtesy of Larry Ludlow*.

ICORFILE= file name

Example 1: Write out the Table of inter-item residual correlations. [PRCOMP=R](#). [ICORFILE=file.txt](#) - Then file.txt contains, for SF.txt, fixed-field columns 12 characters wide.

ITEM	ITEM	CORRELATION
1	2	-.171665
1	3	-.144027
1	4	-.009290
1	5	-.103947

Example 2: When [ICORFILE=](#) is selected on the [Output Files](#) menu or [MATRIX=YES](#), the Data Format: Matrix option can be selected:



This produces 10-character wide columns in fixed-field format

1	2	3	4	...
1.000000	-.171665	-.144027	-.009290	...
-.171665	1.000000	.175921	.039622	...
-.144027	.175921	1.000000	-.101622	...
-.009290	.039622	-.101622	1.000000	...
-.103947	.099335	-.021021	.145339	...
.107936	-.009881	-.039799	-.109941	...
.108457	-.061624	.015850	-.001168	...

Example 3: Is there a way to include extreme scores?

1. All the correlations relating to extreme scores are zero.
2. If you select [ICORFILE=](#) from the [Output Files](#) menu, and click on "matrix". You will get all pairs of non-extreme correlations.
3. You can obtain a matrix of raw or standardized response-residuals including extreme scores, by using [IPMATRIX=](#) from the [Output Files](#) menu (check the extreme score boxes). Then use this to compute the correlations with a standard statistical package etc. You will notice that the residuals for extreme scores may not be zero. This is because they are computed based on the measures shown for extreme scores in the measure Tables.

11.59 ICOLORFILE= colors for item labels in PKMAPs

Item labels (identifiers) in [Table 36](#) PKMAPs are colored using default colors. These can be replaced with other colors using commands in [PKMAP=](#) file.

Instead, item identifiers can be colored by matching with the identifiers in ICOLORFILE=. The matching color is the first in the list that matches the start of the 1R value on the PKMAP.

```
ICOLORFILE=* or filename
Item identifier = color
Item identifier = color
....
*
```

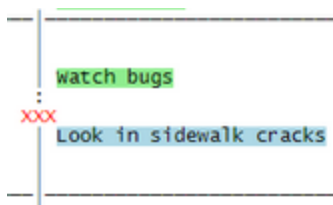
Example: in the Liking for Science analysis: [Example0.txt](#)

```
ASCII=Webpage or ASCII=Doc
; use item label on PKMAP.
PKMAP=*
1R = #NAME#
*
```

```
ICOLORFILE=* or filename
Wat = lightgreen
Loo = lightblue
....
*
```

Output Tables menu: Table 36 PKMAPS

Excerpt from webpage output: "Watch bugs" matches "Wat" in ICOLORFILE= which has color "lightgreen"



watch bugs
xxx Look in sidewalk cracks

Color Names



11.60 IDELETE= item one-line item deletion

A one-line list of items to be deleted or reinstated can be conveniently specified with IDELETE=. This has the same functionality as [IDFILE=](#). [ISELECT=](#) is another way of deleting items.

The formats are:

IDELETE= 3	3 is an item entry number: delete item 3	<table border="1"> <thead> <tr> <th>ENTRY NUMBER</th> <th>TOTAL SCORE</th> <th>COUNT</th> <th>MEAS</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>7333</td> <td>10000</td> <td>-.88</td> </tr> <tr> <td>2</td> <td>8274</td> <td>10000</td> <td>-1.48</td> </tr> <tr> <td>3</td> <td>7754</td> <td>10000</td> <td>-1.11</td> </tr> <tr> <td>4</td> <td>5783</td> <td>10000</td> <td>-.04</td> </tr> <tr> <td>5</td> <td>7873</td> <td>10000</td> <td>-1.15</td> </tr> <tr> <td>.....</td> <td>660</td> <td>6564</td> <td>-1.43</td> </tr> <tr> <td>MEAN</td> <td>5612.2</td> <td>10000</td> <td>.00</td> </tr> <tr> <td>S.D.</td> <td>1803.1</td> <td>.0</td> <td>1.00</td> </tr> </tbody> </table>	ENTRY NUMBER	TOTAL SCORE	COUNT	MEAS	1	7333	10000	-.88	2	8274	10000	-1.48	3	7754	10000	-1.11	4	5783	10000	-.04	5	7873	10000	-1.15	660	6564	-1.43	MEAN	5612.2	10000	.00	S.D.	1803.1	.0	1.00
ENTRY NUMBER	TOTAL SCORE		COUNT	MEAS																																		
1	7333		10000	-.88																																		
2	8274		10000	-1.48																																		
3	7754		10000	-1.11																																		
4	5783		10000	-.04																																		
5	7873		10000	-1.15																																		
.....	660	6564	-1.43																																			
MEAN	5612.2	10000	.00																																			
S.D.	1803.1	.0	1.00																																			
IDELETE= 6 1	delete items 6 and 1																																					
IDELETE= 2-5	delete items 2, 3, 4, 5																																					
IDELETE= +3-10	delete all items, then reinstate items 3 to 10																																					
IDELETE= 4-20 +8	delete items 4-20 then reinstate item 8																																					
IDELETE= 3,7,4,10	delete items 3, 7, 4, 10. Commas, blanks and tabs are separators. At the "Extra information" prompt, use commas as separators																																					
IDELETE= (blank)	In the Specification Menu dialog box, resets temporary item deletions																																					

Example 1: After an analysis is completed, delete all items except for one subtest in order to produce a score-to-measure Table for the subtest.

In the Specifications pull-down box:

IDELETE = +11-26 ; the subtest is items 11-26

Screen displays: CURRENTLY REPORTABLE ITEMS = 16

In the [Output Tables](#) menu (or [SCOREFILE=](#))

[Table 20](#). Measures for all possible scores on items 11-26.

Example 2: 9 common items. 3 items on Form A. 4 items on Form B. Score-to-measure tables for the Forms.

For Form A: in the Specifications pull-down box:

IDELETE = 13-16 ; deletes Form B items

In the [Output Tables](#) menu:

[Table 20](#). Measures for all possible scores on items in Form A.

For Form B: in the Specifications pull-down box:

IDELETE= ; to reset all deletions

then

IDELETE = 10-12 ; deletes Form A items

In the [Output Tables](#) menu:

[Table 20](#). Measures for all possible scores on items in Form B.

11.61 IDELQU= delete items interactively

[IDELETE=](#) and/or [IDFILE=](#) at Extra Specifications? are easier.

If your system is interactive, items to be deleted or selected can be entered interactively by setting IDELQU=Y. If you specify this, you will be asked if you want to delete any items. If you respond "yes", it will ask if you want to read these deleted items from a file; if you answer "yes" it will ask for the file name and process that file in the same manner as if IDFILE= had been specified. If you answer "no", you will be asked to enter the sequence number or numbers of items to be deleted or selected one line at a time, following the rules specified for IDFILE=. When you are finished, enter a zero.

Example: You are doing a number of analyses, deleting different items each analysis.

NI=60

ITEM1=30

IDELQU=Y

&END

Winsteps asks you:

DO YOU WANT TO READ SOME TAP DELETIONS FROM A FILE? (Y/N)

Y

WHAT IS THE NAME OF THE TAP DELETE FILE?

MYIDFILE.TXT

Processing TAP IDELQU=MYIDFILE.TXT

DO YOU WANT TO READ MORE TAP DELETIONS FROM A FILE? (Y/N)

N

DO YOU WANT TO TYPE IN MORE TAP DELETIONS? (Y/N)

Y

TYPE IN THE DELETIONS IN IDELETE= FORMAT:

3 4-10 +6

DO YOU WANT TO TYPE IN MORE TAP DELETIONS? (Y/N)

N

11.62 IDFILE= item deletion file

Deletion or selection of items from a test for an analysis, but without removing the responses from your data file, is easily accomplished by creating a file in which each line contains the sequence number or numbers of items to be deleted or selected. Specify this file by means of the control variable, IDFILE=, or enter the deletion list in the control file using IDFILE=*. Your control file should include item labels for all items, including the ones you are deleting. It has the same functionality as [IDDELETE=](#). [ISELECT=](#) is another way of deleting items. IDFILE=filename can be action from the [Specification menu](#) dialog box.

IDFILE= file name	file containing list of item entry numbers to be deleted or reinstated. Each line in the file is in the IDDELETE= format
IDFILE = ?	opens a Browser window to find the file
IDFILE = *	in-line list in the control file
IDFILE=* 4 18-23 *	At the Extra Specifications prompt
IDFILE=	In the Specifications Menu dialog box, reinstates all temporary deletions

a) Delete an item: enter the item number. E.g., to delete item 5, enter
5

b) Delete a range of items: enter the starting and ending item number on the same line separated by a blank or dash. E.g., to delete items 13 through 24
13-24
or
13 24

c) Select an item for analysis: enter a plus sign then the number.
E.g., to select item 19 from a previously deleted range
+19

d) Select a range of items for analysis: enter a plus sign, the starting number, a blank or dash, then the ending number.
E.g., to select items 17 through 22
+17-22
or
+17 22

e) If a + selection is the first entry in the deletion file, then all items are deleted before the first selection is undertaken, so that the items analyzed will be limited to those selected, e.g,
if +10-20 is the only line in the item deletion file for a 250 item test, it means
1-250 ; delete all 250 items
+10-20 ; reinstate items 10 through 20.

f) You may specify an item deletion list in your control with
IDFILE=*(
List)
*

e.g.,

IDFILE=*

17 ; delete item 17

2 ; delete item 2
*

g) each line can have multiple instructions:

IDFILE=*
17 2 +10
3
*

Item Deletion and Reinstatement Sequence	
Control File or at Extra Specifications: Order of Processing	Here, 123 is an example item entry number. It can be any item entry number.
<u>IDDELETE</u> =list	+123 deletes all items only if it is the first deletion instruction in the first of these deletion commands that is active. +123 always reinstates item 123.
IDFILE=* list * or filename	
<u>IDELQU</u> =Yes	
<u>ISELECT</u> =selection	selects the specified items and permanently deletes all other items
Specification Menu Dialog Box	
IDDELETE=(nothing), IDFILE=(nothing)	reinstates all temporary deletions
ISELECT=(nothing) or ISELECT=*	reinstates all temporary deselections
IDDELETE=list, IDFILE=filename	temporary deletions, in addition to previous temporary deletions, if any. If the first deletion instruct starts +123 then temporarily deletes all items and reinstates item 123 if not permanently deleted
ISELECT=selection	temporary deselections, in addition to previous temporary deselections, if any

Example 1: You wish to delete the fifth and tenth through seventeenth items from an analysis, but then keep item fourteen.

1. Create a file named, say, ITEM.DEL

2. Enter into the file, the lines:

5
10-17
+14

3. Specify, in the control file,

NI=50
ITEM1=63
IDFILE=ITEM.DEL
TABLES=1110111
&END

or, specify in the control file,

NI=50
ITEM1=63
IDFILE=*
5
10-17
+14
*
TABLES=1110111
&END

Example 2: The analyst wants to delete the most misfitting items reported in Table 10.

1. Set up a standard control file.

2. Specify

IDFILE=*

*

3. Copy the target portion of Table 10.

4. Paste it between the ""

5. Delete characters before the entry numbers.

6. Type ; after the entry numbers to make further numbers into comments.

TITLE = 'Example of item deletion list from Table 10'

IDFILE = *

Delete the border character before the entry number

```
; ENTRY      RAW      INFIT      OUTFIT
;  NUM  SCORE  COUNT  MEASURE  ERROR  MNSQ  ZSTD  MNSQ  ZSTD  PTBSE  ACTS      G
;  5 ; 2    4    .00    1.03 1.48  1.8 1.50  1.8 A-.83  FIND BOTTLES AND CANS  0
;  8 ; 2    4    .00    1.03 1.40  1.6 1.43  1.6 B-.71  LOOK IN SIDEWALK CRACKS 0
;  4 ; 3    4    .00    .62 1.33  .7 1.49  .9 C-.21  WATCH GRASS CHANGE     0
;  9 ; 4    4    .00    .74 1.51  .8 1.57  .9 D-.59  LEARN WEED NAMES       0
; 20 ; 1    4    .00    1.03 1.12  .5 1.14  .6 E-.05  WATCH BUGS             0
; 24 ; 6    4    .30    1.03 1.15  .6 1.13  .5 F-.15  FIND OUT WHAT FLOWERS LIVE ON 0
```

Enter the ; to make details to right of entry numbers into comments

*

Example 3: The analyst want to delete item 4 and items 18 to 23 on the DOS control (or Extra Specifications) line:

Extra specifications? IDFILE=* 4 18-23 * (Enter)

or

C:>Winsteps CONTROL.FIL OUTPUT.FIL IDFILE=* 4 18-23 *

Example 3. In Winsteps analyses, the data file is not changed.

1. To delete items or persons:

[IDFILE=](#) (list of item entry numbers)

[PDFILE=](#) (list of person entry numbers)

2. To edit the data file:

[EDFILE=](#)

(person entry number) (item entry number) (new data value or missing)

For instance in an analysis of [Exam1.txt](#), we can place these instructions in the Winsteps control file:

IDFILE=*

4 ; delete item 4

8 ; delete item 8

*

PDFILE=-*

6 ; delete person 6

*

EDFILE=*

1 2 0 ; person 1 on item 2 scored a 0

3 4 -1 ; person 3 on item 4 scored -1 = missing, not administered

*

11.63 IDROPEXTREME= drop items with extreme scores

Unanchored items with extreme (zero, minimum possible or perfect, maximum possible) scores provide no information for estimating person measures, but they are reported and included in summary statistics. To remove them:

IDROPEXTREME = No ; do not drop extreme items (standard)

IDROPEXTREME = Yes or All ; drop items with zero (minimum possible) and perfect (maximum possible) scores

IDROPEXTREME = Zero or Low or Bottom or Minimum ; drop items with zero or minimum-possible scores

IDROPEXTREME = Perfect or High or Top or Maximum ; drop items with perfect or maximum-possible scores

Example: The instrument contains items asking about very rare conditions (scored "0" - not observed). These are skewing the survey summary statistics:

IDROPEXTREME = Minimum ; items about conditions never observed in the sample are dropped.

11.64 IFILE= item output file

IFILE=filename produces an output file containing the information for each item. This file contains 4 heading lines (unless [HLINES=N](#) or [ROW1HEADING=N](#)), followed by one line for each item containing the following fields and the standard field selection. To change the output-field selection, go to the [Output File dialog](#). IFILE=, Field selection, Make default., or [IOFSFIELDS=](#).

IFILE=? opens a Browse window

"Status=-2 to -6" means that there are no measurable responses by those items in this analysis. The items may be listed in the IFILE= and in Table 14, but all the numbers shown are default values. They have no meaning. Please do not include those items in summary statistics.

Columns:		with "Select All Fields" using Output File Field Selection		
Start	End	Label	Format	Description
1	1		A1	Blank or ";" if HLINES=Y and there are no responses or deleted or extreme (status: 0,-1, -2, -3)
2	6	ENTRY	I5	1. The item sequence entry number
7	14	MEASURE	F8.2	2. Item's MLE estimated difficulty calibration user-rescaled by UMEAN= , USCALE= , UDECIM= . Measures for deleted or inestimable items are shown as 9999.
15	17	STATUS	I3	3. The item's status: 2 = Anchored (fixed) measure 1 = Estimated measure 0 = Extreme maximum measure (estimated using EXTRSC=) for extreme minimum raw score -1 = Extreme minimum measure (estimated using EXTRSC=) for extreme maximum raw score (usually 0) -2 = No responses available for measure -3 = Deleted by user. PDELETE= , PDFILE= , IDDELETE= , IDFILE= , PSELECT= , ISELECT= -4 = Inestimable: high (all responses in the same category with ISGROUPS=0 or CUTHI=) -5 = Inestimable: low (all responses in the same category with ISGROUPS=0 or CUTLO=) -6 = Anchored (fixed) measure with extreme (minimum or maximum) observed raw score -7 to -16 = Temporarily deselected by Specification box with iSELECT= (usual STATUS - 10) -17 to -26 = Temporarily deleted by Specification box with IDDELETE= (usual STATUS - 20) -27 to -36 = Temporarily deselected and deleted by Specification box with iSELECT= and IDDELETE= (usual STATUS - 30)
18	25	COUNT	F8.1	4. The number of responses used in calibrating (TOTAL=N) or the observed count (TOTAL=Y)
26	34	SCORE	F9.1	5. The raw score used in calibrating (TOTAL=N) or the observed score (TOTAL=Y)

35	41	MODLSE REALSE	F7.2	6. Standard error of the MLE or WLE item difficulty estimate adjusted by REALSE= and user-rescaled by USCALE= , UDECIM=
42	48	IN.MSQ IN.CHI	F7.2	7. Item infit: mean square infit. Chi-square = IN.MSQ* INDF If CHISQUARE=Yes, IN.CHI = Infit Chi-square
49	55	IN.ZSTD, ZEMP , LOG, PROB	F7.2	8. Item infit: t standardized, locally t standardized, log-scaled or probability of mean-square/chi-square (LOCAL=)
56	62	OUT.MS OUT.CHI	F7.2	9. Item outfit: mean square outfit. Chi-square = OUT.MS * OUTDF If CHISQUARE=Yes, OUT.CHI = Outfit Chi-square
63	69	OUT.ZSTD, ZEMP , LOG, PROB	F7.2	10. Item outfit: t standardized, locally t standardized, log-scaled or probability of mean-square/chi-square (LOCAL=)
70	76	DISPLACE	F7.2	11. Item displacement (user-rescaled by USCALE= , UDECIM=)
77	83	PBSA PBSX PTMA PTMX	F7.2	12. Item by test-score correlation: point-biserial or point-measure. PTBIS= PBSA = Point-Biserial correlation including all responses in the raw score PBSX = Point-Biserial correlation excluding the current item's response from the raw score PTMA = Point-Measure correlation including all responses for the measure PTMX = Point-Biserial excluding the current item's response from the measure
84	90	WEIGHT	F7.2	13. Item weight IWEIGHT=
91	96	OBSMA	F6.1	14. Observed percent of observations within 0.5 score-points of their expected values
97	102	EXPMA	F6.1	15. Expected percent of observations within 0.5 score-points of their expected values
103	109	DISCRIM	F7.2	16. Item discrimination (this is not a parameter estimate, merely a descriptive statistic) DISCRIM=
110	115	LOWER	F6.2	17. Item lower asymptote: ASYMPTOTE= Yes
116	121	UPPER	F6.2	18. Item upper asymptote: ASYMPTOTE= Yes
122	127	PVALU	F6.2	19. Item proportion-correct-values or average ratings: PVALUE= Yes:
128	133	PBA-E PBX-E PMA-E PMX-E	F6.2	20. Expected value of Item by test-score correlation. PTBIS= PBA-E = Expected value of Point-Biserial including all responses in the raw score PBX-E = Expected value of Point-Biserial excluding the current response from the raw score PMA-E = Expected value of Point-Measure including all responses in the measure PMX-E = Expected value of Point-Biserial excluding the current response from the measure
134	139	RMSR	F6.2	21. Root-mean-square residual RMSR=
140	147	WMLE	F8.2	22. Warm's (Weighted) Mean Likelihood Estimate (WLE) of Item Difficulty user-rescaled by UMEAN= , USCALE= , UDECIM=
148	153	INDF	F6.2	23. degrees of freedom of Infit mean-square
154	159	OUTDF	F6.2	24. degrees of freedom of Outfit mean-square
160	167	QCMLE	F8.2	25. Quasi-CMLE estimates for dichotomous data. 0 otherwise.
168	168		1X	Blank
169	169	G	A1	26. Grouping to which item belongs (G) ISGROUPS=
170	170		1X	Blank
171	171	M	A1	27. Model used for analysis (R=Rating, S=Success, F=Failure) MODELS=
172	172		1X	Blank
173	173	R	A1	28. Recoding/Rescoring indicator: "." = only CODES= "A" = ALPHANUM= "K" = KEY1=

				"N" = RESCORE=2 and NEWSCORE= "1" = RESCORE=1 and NEWSCORE= Others = IREFER=
174	174		1X	Blank
175	205	NAME	A30+	29. Item name or label: use LFILE= for different item names
The format descriptors are:				
	In = Integer field width n columns			
	Fn.m = Numeric field, n columns wide including n-m-1 integral places, a decimal point and m decimal places			
	An = Alphabetic field, n columns wide			
	nX = n blank columns.			

When CSV=Y, commas separate the values, which are squeezed together without spaces between. Quotation marks surround the "Item name", e.g., 1,2,3,4,"Name". When CSV=T, the commas are replaced by tab characters.

Example: You wish to write a file on disk called "ITEM.CAL" containing the item statistics for use in updating your item bank, with values separated by commas:

```
IFILE=ITEM.CAL
CSV=Y
```

When W300=Yes, then this is produced in Winsteps 3.00, 1/1/2000, format:

Columns:				
Start	End	Label	Format	Description
1	1		A1	Blank or ";" if HLLINES=Y and there are no responses or deleted (status = -2, -3)
2	6	ENTRY	I5	1. The item sequence number
7	14	MEASURE	F8.2	2. Item's MLE estimated calibration (user-rescaled by UMEAN=, USCALE=, UDECIM)
15	17	STATUS	I3	3. The item's status: 3 = Anchored (fixed) measure with extreme (minimum or maximum) observed raw score 2 = Anchored (fixed) measure 1 = Estimated measure 0 = Extreme minimum (estimated using EXTRSC=) -1 = Extreme maximum (estimated using EXTRSC=) -2 = No responses available for measure -3 = Deleted by user -4 = Inestimable: high (all responses in the same category with ISGROUPS=0 or CUTHI=) -5 = Inestimable: low (all responses in the same category with ISGROUPS=0 or CUTLO=) -6 = Deselected
18	23	COUNT	I6	4. The number of responses used in calibrating (TOTAL=N) or the observed count (TOTAL=Y)
24	30	SCORE	I6	5. The raw score used in calibrating (TOTAL=N) or the observed score (TOTAL=Y)
31	37	MODLSE REALSE	F7.2	6. Item calibration's standard error with REALSE= and user-rescaled by USCALE=, UDECIM=
38	44	IN.MSQ	F7.2	7. Item mean square infit
45	51	ZSTD, ZEMP , LOG	F7.2	8. Item infit: t standardized (ZSTD), locally t standardized (ZEMP) or log-scaled (LOG)
52	58	OUT.MS	F7.2	9. Item mean square outfit

59	65	ZSTD, ZEMP , LOG	F7.2	10. Item outfit: standardized (ZSTD), locally standardized (ZEMP) or log-scaled (LOG)
66	72	DISPLACE	F7.2	11. Item displacement (user-rescaled by USCALE=, UDECIM=)
73	79	PBSA PBSX PTMA PTMX	F7.2	12. Item by test-score correlation: point-biserial or point-measure. PTBIS= PBSA = Point-Biserial correlation including all responses in the raw score PBSX = Point-Biserial correlation excluding the current item's response from the raw score PTMA = Point-Measure correlation including all responses for the measure PTMX = Point-Biserial excluding the current item's response from the measure
80	80		1X	15. Blank
81	81	G	A1	16. Grouping to which item belongs, ISGROUPS=
82	82		1X	17. Blank
83	83	M	A1	18. Model used for analysis (R=Rating, S=Success, F=Failure) MODELS=
84	84		1X	19. Blank
85	132+	NAME	A30+	18. Item name

Example of IFILE= (to see other fields: [Output File dialog](#))

```
; ACT LIKING FOR SCIENCE (Wright & Masters p.18) Aug 8 22:17 2013
;ENTRY MEASURE ST COUNT SCORE MODLSE IN.MSQ IN.ZST OUT.MS OUT.ZS DISPL PBSA WEIGHT OBSMA EXPMA DISCRM LOWER UPPER PVALU PBA-E
RMSR WMLE G M R NAME
1 -.40 1 75.0 109.0 .21 .55 -3.48 .49 -2.53 .00 .69 1.00 77.0 61.7 1.52 .00 2.00 1.45 .53
.42 -.39 1 R . Watch birds
2 -.71 1 75.0 116.0 .22 .93 -.39 .72 -1.02 .00 .66 1.00 74.3 64.4 1.26 .00 2.00 1.55 .50
.52 -.70 1 R . Read books on animals
```

- Example: If you have your anchor values connected to unique item IDs (item labels), then
- (1) put your item IDs and anchor values in an Excel table with the item IDs in the first column
 - (2) run an unanchored analysis in Winsteps. Output the IFILE= to Excel.
 - (3) VLOOKUP the item IDs in (2) using the Table in (1) to place the anchor values in the Measure column
 - (4) Save the entry number and measures from (2) to use as an anchor file.

11.65 ILFILE= item label file

Useful item identification greatly enhances the output tables.

You usually specify items labels between &END and END LABELS in the control file. You may also use ILFILE= for the initial or additional sets of item labels.

ILFILE= file name	file containing details
ILFILE = *	in-line list
ILFILE = ?	opens a Browser window to find the file
(between &END and END LABELS or END NAMES)	in-line list

Example: You wish to use abbreviated item labels for the initial analysis, but then switch to longer item labels to produce the final reports.

In your control file specify the shorter labels, one per line,

- (a) between &END and END LABELS
- or (b) between ILFILE=* and * in the control file
- or (c) in a separate file identified by ILFILE=*

You can switch to the longer labels, in a file called "Longer.txt" by using the "Specification" menu item, and entering ILFILE=Longer.txt

If you have ILFILE= in your control file and your data is also in your control file , be sure that there is an "END LABELS" before your data.

Example 1: 4 item arithmetic test with items in ILFILE=*

```
NI=4
ILFILE=*
Addition ; labels for the 4 items
Subtraction
Multiplication
Division
*
&End
END LABELS
```

Example 2: One set of item labels, then another

```
NI=4
&End
Addition ; labels for the 4 items
Subtraction
Multiplication
Division
END LABELS
```

Contents of "labels.txt"

```
Plus
Minus
Product
Quotient
```

After standard reporting:

[Specification](#) pull-down menu box

```
ILFILE=labels.txt
```

or

```
ILFILE=(full path)\labels.txt
```

OK

further reporting

11.66 IMAP= item label on item maps Tables 1, 12

This specifies what part of the item label is to be used on the item map. The length of IMAP= overrides [NAMLMP=](#).

It's format is IMAP = \$S..W.. or \$S..E. etc. using the column [selection rules](#).

Example: Item type is in columns 3 and 4 of the item label. Item content area is in columns 7 and 8.

```
IMAP= $S3W2+"/".$S7W2
```

```
tfile=*
```

```
12 ; Item maps in Table 12 (or use Output Tables menu)
```

```
*
```

If the item label is "KH323MXTR", the item label on the map will be "32/XT"

11.67 INUMB= label items by sequence numbers

Are item names provided, or are they the entry sequence numbers?

INUMB=Y

a name is given to each item based on its sequence number in your data records. The names are "I0001", "I0002", ..., and so on for the NI= items. This is a poor choice as it produces non-informative output.

INUMB=N, the standard

Your item names are entered (by you) after the "&END" at the end of the control variables. Entering detailed item names makes your output much more meaningful to you.

The rules for supplying your own item names are:

1. Item names are entered, one per line, generally directly after &END.
2. Item names begin in column 1.
3. Up to 300 characters (or [ITLEN=](#)) in each item name line will be used, but you may enter longer names in the control file for your own record keeping.
4. The item names must be listed in exactly the same order as their responses appear in your data records.
5. There should be the same number of item names as there are items specified by [NI=](#). If there are too many or too few names, a message will warn you and sequence numbers will be used for the names of any unnamed items. You can still proceed with the analysis.
6. Type END NAMES starting in column 1 of the line after the last item name.

Example: An analysis of 4 items for which you supply identifying labels.

```
; these lines can start at any column
  NI=4 four items
ITEM1=10 responses start in column 10
INUMB=N item names supplied (the standard)
  &END
My first item name ; must start at column 1.
My second item label
My third item identifier
My fourth and last item name
END NAMES ; must start at column 1, in capital letters
Person A 1100 data records
  |
Person Z 1001
```

11.68 IOFSFIELDS= field selection for IFILE=

IOSFIELDS= is set using the [Output Field selection](#) dialog box.

Field Number

- 1 Flag extremes with ;
- 2 Entry number
- 3 Measures
- 4 Status
- 5 Count of observations
- 6 Raw score
- 7 Standard error
- 8 Infit mean-square
- 9 Infit t standardized
- 10 Outfit mean-square
- 11 Outfit t standardized
- 12 Displacement
- 13 Correlation
- 14 Weight
- 15 Observed matches
- 16 Expected match
- 17 Discrimination
- 18 Lower asymptote
- 19 Upper asymptote
- 20 P-value: average rating
- 21 Expected correlation
- 22 Isgroups= group
- 23 Models= model
- 24 Name or Label
- 25 Include deleted

26 RMS Residual
 27 Recode/score
 28 WMLE measure

Example: Output into the [FILE=](#) file only the Entry number and the WMLE measures.
 IOSFIELDS = NYNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNY

11.69 IPEXTREME= placement of extreme scores on maps

IPEXTREME= Yes	place persons and items with extremes scores at the extreme top and bottom in Tables 1, 12, 16.
IPEXTREME= No	place persons and items with extremes scores at their estimated measures in Tables 1, 12, 16.

Example 1: Exam1.txt. Table 1.0

IPEXTREME=Yes. Persons and items with extreme scores (here in red) placed at the extreme ends of the map.

```

MEASURE                                KID - MAP - TAP
5                                     <more>|<rare>
+ 1-3- 1-4- 1-4- 4-1-
|
4                                     T+
Ric Sus |S 1-4-
3                                     Fra +
|
2                                     Ber Bet Els Tho S+ 1-3- 1-4-
|
1                                     Bri Car Dav Joe Zul + 1-3-
|
0                                     M+M
Ann Aud Bar Dor Jam Kim Lin Mik Ron Tra Wal Wil |
-1                                     +
-2                                     Jan Lis Rod | 2-4-
Bre Pet S+
| 1-4-
-3                                     Ada Ric +
Bla Mar |S 1-3- 3-4-
-4                                     T+ 1-4- 2-1-
Don | 1-3-
-5                                     Hel + 1-2- 1-4 2-3
<less>|<frequent>

```

IPEXTREME=No. Persons and items with extreme scores (here in red) placed at their estimated measures on the map.

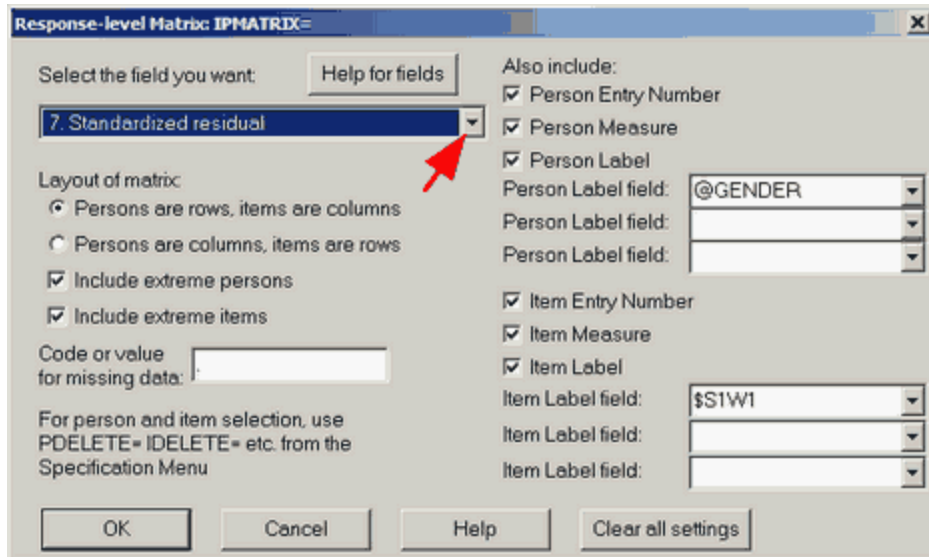
```

MEASURE                                KID - MAP - TAP
6                                     <more>|<rare>
+ 4-1-3-
|
5                                     + 1-3-2- 1-4-2- 1-4-3-
|
4                                     T+
Ric Sus |S 1-4-2-
3                                     Fra +
|
2                                     Ber Bet Els Tho S+ 1-3-2- 1-4-3-
|
1                                     Bri Car Dav Joe Zul + 1-3-1-
|
0                                     +
Ann Aud Bar Dor Jam Kim Lin Mik Ron Tra Wal Wil M|
-1                                     +M
Jan Lis Rod | 2-4-3-
-2                                     Bre Pet +
S| 1-4-2-
-3                                     Ada Ric +
Bla Mar | 1-3-2- 3-4-1
-4                                     + 1-4-3- 2-1-4
Don | 1-3-4
-5                                     T+S
|
-6                                     +
Hel | 1-2-4 1-4 2-3

```

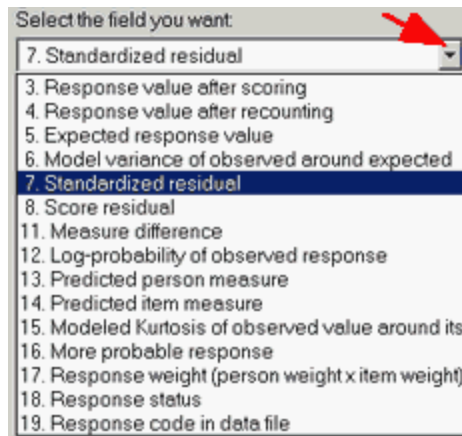
11.70 IPMATRIX= response-level matrix

IPMATRIX= is only available from the Output Files pull-down menu. It constructs a rectangular matrix in which there is one row for each person and one column for each item, or vice-versa. The entries in the matrix are selected from the following screen:



The first rows and columns can be the entry numbers, measures, labels and/or fields from the labels.

The matrix contents are selected from the "Select the field you want" pull-down menu (red arrow above). The identifying numbers in the drop-down list are those for [XFILE=](#).



The "Person Label field" and "Item Label field" boxes enable you to [select columns](#) from the person and item labels to use as demographics. These can be typed in or chosen from the pull-down menu.



Depending on [CSV=](#), data values are separated by "tab" or comma characters. In fixed field format, all fields are 7 characters long, separated by a blank. Missing data codes are "." as standard, but can be any character, or nothing.

Example:

```

"Entry"      1      2      3      4  <- Item entry numbers
  1      .162    .162      .    .482  <- standardized residuals. "." is missing or
inestimable data
  2      .042    .042    .044    .126
  3      .042    .042    .044    .126
  4      .225    .225    .236    .671
  5      .042    .042    .044    .126
  6      .042    .042    .044    .126
  7      .006    .006    .006    .017
  8      .042    .042    .044    .126
  9      .042    .042    .044    .126
^
Person entry numbers

```

11.71 IPRFILE= change codes for blocks of persons and items

IPRFILE= enables rectangles of item and person observations to be changed based on their data codes. This is done after reading in the data files and performing [EDFILE=](#) (if specified).

IPRFILE= file name	file containing details
IPRFILE= file name + file name + ...	multiple files
IPRFILE = *	in-line list
IPRFILE = ?	opens a Browser window to find the file

IPRFILE= cannot add extra items. Please increase [N=](#) to add extra items. Please specify the highest person beyond those in the data file by means of [EDFILE=](#).

Several selections and recodings can be specified. An observation is recoded according to the first selection that it matches.

The format is:

IPRFILE=* (or IPRFILE= filename)	item-person-recode sublist or file
#ITEMS	start of item list (if omitted, select all items)
Item number	select this item
Item number - Item number	select this range of items
- Item number	exclude this item from this selection
- Item number - Item number	exclude this range of items item from this selection
"Item selection value"	select items using ISELECT= values, in quotation marks
- "Item selection value"	exclude these items
#PERSONS	start of person list (if omitted, select all persons)
Person number	select this person
Person number - Person number	select this range of persons
- Person number	exclude this person from this selection
- Person number - Person number	exclude this range of persons item from this selection
"Person selection value"	select persons using PSELECT= values, in quotation marks
- "Person selection value"	exclude these persons
#RECODE="codes"	a string of codes which align with the CODES= codes, and indicate the new code values. If a code in #RECODE= is not in CODES=, then it is interpreted as "change that valid code to missing data".
#MISSING="code"	this is the code in CODES= to replace "missing-data" in the selection. If this code is not in CODES=, then missing-data in the selection continues to be missing-data.
#	next selection begins (if any)

#ITEMS	start of list for items and persons not selected yet
.....	
#RECODE="codes"	list of codes for the second selection
#	next selection begins (if any)
#ITEMS	start of list for items and persons not selected yet
.....	
*	end of IPRECORD=

Example 1: Convert all responses to items 3 and 7-10 by persons 4 and 13-22 into missing (not administered) data. Other missing responses are to be treated as incorrect.

```

NI = 16 ; 16 item test
CODES = ABCDM ; the data are multiple-choice responses
IPRFILE = * ; start of rectangular recode
#ITEMS ; list of items to select
3
7-10
#PERSONS ; list of persons to select
4
13-22
; CODES = ABCDM ; this comment is a reminder of the CODES=
#RECODE = MMMM ; A,B,C, D will all be converted to invalid code
*
IREFER = CBADBACDDCCABDBA ; same as scoring key (for convenience)
; CODES = ABCDM ; comment: another reminder
IVALUEA = 1000M ; A scored 1. M scored "M", non-numeric, "missing, not administered"
IVALUEB = 0100M ; B scored 1
IVALUEC = 0010M ; C scored 1
IVALUED = 0001M ; D scored 1
MISSING-SCORED = 0 ; all responses not in CODES= scored 0.

```

Example 2: Convert all "C" responses to items 3 and 7-10 by persons 4 and 13-22 into "A". Convert all "C" responses to the other items by persons 4 and 13-22 into "B". Convert all other "C" responses into "D".

```

CODES = ABCD ; valid data codes
IPRFILE = * ; start of rectangular recode
#ITEMS ; list of items to select
3
7-10
#PERSONS ; list of persons to select
4
13-22
; CODES = ABCD ; this comment is a reminder of the CODES=
#RECODE = ABAD ; C is converted into A
# ; end of this recode: start of next recode
; #ITEMS ; commented out: applies to all items

```



```

#PERSONS
4
13-22
#RECODE = ABBD ; C is converted into B
; #ITEMS ; commented out: applies to all items
; #PERSONS ; commented out: applies to all persons
#RECODE = ABDD ; C is converted into D
*

```

Example 3: Select everyone with a person label beginning "A". Make their responses to items with labels starting "Z" into missing data.

```

CODES = ABCD ; valid data codes
IPRFILE = * ; start of rectangular recode
; #ITEMS ; commented out - all items
"Z" ; item labels starting Z
#PERSONS ; list of persons to select
"A" ; person labels starting A
; CODES = ABCD ; this comment is a reminder of the CODES=
#RECODE = . . . . ; All responses made missing.
*

```

11.72 IREFER= identifying items for recoding

Responses are revalued according to the matching codes in [VALUE=](#). If this implies that the items may have different rating (or partial credit) scale structures, so [ISGROUPS=](#) may also be required.

IREFER= *file name	file containing details
IREFER = *	in-line list
IREFER = codes	codes for item groups

IREFER= has three forms: IREFER=AABBCDAAD and IREFER=* list * and IREFER=*filename

Valid one-character IREFER= codes include: !#\$%&-'./123456789<>@ABCDEFGHIJKLMNPOQRSTUVWXYZ^_`~

A-Z are the same as a-z.

Characters with ASCII codes from 129-255 can also be used, but display peculiarly: ãĎÑÒÓÔÕÖ×ØÙÚÛÜÝÞßàáâã etc.

When [XWIDE=2](#) or more, then either (a) Use one character per XWIDE and blanks,

```

NI=8
XWIDE=2
IREFER=' A B C D D C B A '

```

or (b) Use one character per item with no blanks

```

NI=8
XWIDE=2
NEWSCORE= ' ABCDDCBA '

```

Item identifying codes can be letters or numbers. "A" is the same as "a", etc.

Example 1. There are 3 item types. Items are to rescored according to Type A and Type B. Other items to keep original scoring.

```
CODES = 1234
IREFER = AAAAAAABBBBBBBB***** 3 item types: ("a" is the same as "A" in these codes)
IVALUEA = 1223 Recode Type A items
IVALUEB = 1123 Recode Type B items
IVALUE* = 1234 Recode Type * item. Can be omitted
```

Or

```
IREFER=*
1-8 A
9-16 B
17-23 *
*
```

Or

```
IREFER=*filename.txt
```

in filename.txt:

```
1-8 A
9-16 B
17-23 *
```

Example 2. There are 3 item types. Responses are 2 characters wide. Items are to rescored according to Type A and Type B. Other items to keep original scoring.

```
XWIDE=2
CODES = '1 2 3 4 '
IREFER = AAAAAAABBBBBBBB***** 3 item types
IVALUEA = '1 2 2 3 ' Recode Type A items
IVALUEB = '1 1 2 3 ' Recode Type B items
IVALUE* = 1234 Recode Type * item. Can be omitted
```

Example 3: All items are to be rescored the same way

```
NI = 100 100 ITEMS
IREFER=*
1-100 X FOR ALL 100 ITEMS, reference is X
*
Codes = 12345 rescore 12345
IVALUEX = 12223 into 12223
```

Example 4: Items are to be rescored in 3 different ways, forward scoring, reversed scoring and dichotomized scoring. Then the items are to be divided into 4 rating scale structures.

```
ISGROUPS=11112223312444 ; 4 RATING SCALE GROUPINGS
IREFER =AAAABBBCCABBBB ; 3 RECODINGS
CODES =01234 ; ORIGINAL CODES IN DATA
IVALUEA =01234 ; ORIGINAL CODING MAINTAINED - THIS LINE CAN BE OMITTED
IVALUEB =43210 ; CODING IS REVERSED
IVALUEC =*112* ; DICHOTOMIZED WITH EXTREME CATEGORIES MARKED MISSING
```

Example 5: Multiple-choice test with 4 options, ABCD

```
IREFER=ABCDABADCD ; SCORING KEY
CODES =ABCD ; VALID DATA CODES
IVALUEA=1000 ; A SCORED 1
IVALUEB=0100 ; B SCORED 1
IVALUEC=0010 ; C SCORED 1
IVALUED=0001 ; D SCORED 1
MISSCORE=0 ; EVERYTHING ELSE IN THE DATA SCORED 0
```

Example 6: A 10 item multiple choice test with 4 partial credit performance items scores 0,1,2,3.

```
NI = 14 ; 14 items
; 12345678901234
ISGROUPS=DDDDDDDDDD0000 ; 10 Dichotomies in Group "D". 4 Partial Credit items each in its own group "0"
IREFER =ABCDABADC1111 ; ITEM REFERENCE FOR THE 14 ITEMS
CODES =ABCD0123 ; VALID DATA CODES
```

```

IVALUEA=10000000 ; A SCORED 1
IVALUEB=01000000 ; B SCORED 1
IVALUEC=00100000 ; C SCORED 1
IVALUED=00010000 ; D SCORED 1
IVALUE1=00000123 ; PERFORMANCE ITEMS: 0,1,2,3
MISSING-SCORED=0 ; EVERYTHING ELSE IN THE DATA SCORED 0

```

Example 7: Partial credit items are to be rescored:

```

IREFER = ABCAAB ; there are six items items 1, 4 and 5 have one rescoring. Items 2 and 6
another rescoring. Item 3 a third rescoring
CODES = 1234 ; the original scoring
IVALUEA = 1123 ; the new scoring for A-type items
IVALUEB = 1334 ; for B-type items
IVALUEC = 1234 ; original scoring maintained

```

Now there is a decision:

does 134 mean the same as 123, so that 2 is a structural zero, if so `STKEEP=NO`

does 134 mean the same as 1(2)34, so that 2 is an incidental or sampling zero, if so `STKEEP=YES`

Example 8: An MCQ item where two boxes can be checked for one item, e.g., choose 2 options out of 5.

Load the data into Excel. Each one-box item has a separate column. The two-box item is converted into 5 columns, one for each option. For example, 2 out 5 becomes 5 dichotomies, one for each option: A-E, Each dichotomy is scored 1 if the response is right, 0 if wrong, so the two correct answers are scored 1 if selected, 0 if not the three incorrect answers are scored 0 if selected, 1 if not then the two correct dichotomies are weighted 1 and the three incorrect dichotomies are weighted 0 Use Winsteps Excel/RSSST menu to create the Winsteps data file.

11.73 ISELECT= item selection criterion

Items to be selected may be specified by using the ISELECT= instruction to match characters within the item name. Items deleted by [DELETE=](#) or [IDFILE=](#) are never selected by ISELECT=.

This can be done before analysis in the control file or with "Extra specifications". It can also be done after the analysis using the "Specification" pull-down menu.

Selection rules:	
Control characters match label or name. They start at the first column of the label or name.	
?	matches any character
*	matches any string of characters - must be last selection character. If * is in the first column, then every available person or item is selected.
A	matches A in the person label, and similarly all other characters except { }
{..}	braces characters which can match a single character: {ABC} matches A or B or C.
{..-..}	matches single characters in a range. {0-9} matches digits in the range 0 to 9.
{..-..}	matches a single "-" {AB--} matches A or B or "-".
{~ABX}	omits persons or items which match A or B or X
@fieldname=	positions the next selection character at the start of the specified field

Each ISELECT= performed using the "Specification" pull-down menu selects from all those analyzed. For incremental selections from the [Specification menu](#), i.e., selecting from those already selected, specify +ISELECT=

Example 0: In [Example 10](#), select bank 6:

[@BANK=5](#) ; Bank number is in column 5 of the item label

ISELECT=@BANK=6

Example 1: Select for analysis only items with M in the 5th column of item name.

ISELECT=????M* M in column means Math items

0001M2x4 ; *selected*

0002R the cat ; *omitted*

END NAMES

Example 2: Select for analysis all items with code "A4" in columns 2-4 of their names.

ISELECT="?A4*" quotes because a blank is included. A is in col. 2 etc.

ZA4PQRS selected

Example 3: Select all Math (M in column 2) items which are Addition or Division (A or D in column 4):

ISELECT="?M?{AD}*"

1M3A456 2+2 selected

1M5D689 23/7 selected

1H2A123 George omitted (History, American)

Example 3: Select codes A, 1,2,3,4,5,6 in column 3:

ISELECT=?{A1-6}*

Example 4: Select "-" in columns 2 and 3:

ISELECT="?- "

Example 5: Select "-" or "x" in column 2 with " " in column 3:

ISELECT="?{-x} "

Example 6: Analyze only math (column 4 or person-id). Then report only items in Topic C (column 1). Then only report items in Strands 4 and 5 (column 2) in Topic C.

ISELECT=????M* in the Control file or at the Extra Specifications prompt.

ISELECT=C* using the Specification pull-down menu, after the analysis

+ISELECT=?{45}* using the Specification pull-down menu.

Example 7: 4 forms in one analysis with common items. Produce a score table for each form.

In the item labels, columns 1-4 of item label are form indicator. Items can be in more than one form

1xx for form 1 items

12xx for form 1 and 2 items

xx3x for form 3

xx34 for form 3 and 4

etc.

Then, use the [Specification](#) menu box:

ISELECT = 1??? will select all items on form 1

Output Tables: [Table 20](#) score table

ISELECT = * select all items

ISELECT = ???? will select all items on form 2

etc.

Example 8: Omit unwanted items, marked by "x" in the first character of the item label:

ISELECT={~x}

Example 9: If the field to be selected is already identified, then this simplifies counting the column:

@Dx = 16E17 ; \$C44W2

16E17 means that the field is in columns 16 and 17 of the item label. Look at a data record to see what is actually there.

ISELECT=????????????23 ; selects "23" in columns 16 and 17 of the item label

Example 10: We want item type group X (in column 4 of the item label) except for sub-types 11, 24 (in columns 6, 7 of the item label):

```

ISELECT = "??X?{12}{14}" ; this selects X and 11, 14, 21, 24
EDFILE=*
"?" "????14" . ; convert 14 to missing data
"?" "????21" . ; convert 21 to missing data
*
```

Example 11: multiple levels in the Specification menu box:

Between each step, in the Specification menu box, please specify ISELECT= (nothing) to reinstate all the items.

Example using Example0.txt and the Specification menu box:

```

iselect=L
"ISELECT=L*" selected 5 ACT
CURRENTLY REPORTABLE ACT = 5

iselect=W
"ISELECT=W*" selected 0 ACT
CURRENTLY REPORTABLE ACT = 5
... not done

iselect=
"ISELECT=*" selected 25 ACT
CURRENTLY REPORTABLE ACT = 25

iselect=W
"ISELECT=W*" selected 7 ACT
CURRENTLY REPORTABLE ACT = 7
```

11.74 ISFILE= item structure output file

Do not use this file for anchor values. Instead use [IFILE=](#) (becomes [IAFILE=](#)) and [SFILE=](#) (becomes [SAFILE=](#)).

Communicating the functioning of a rating scale is challenging, especially if your audience think of its categories as separate and equally-spaced points on the latent variable.

If you want to communicate the categories as points, then the best points for the intermediate categories are the locations on the latent variable at which the probability of observing each category is the highest. These are also the points where the expected score on the item is the category value. In Winsteps these are the "AT CAT" measures in the ISFILE= output file. These points are at infinity for the extreme categories, so Winsteps reports the measures for expected scores of "lowest category + 0.25" (= CAT + 0.25) and "highest category - 0.25" (= CAT - 0.25).

The Rasch-Thurstone thresholds (50%PRB in ISFILE=) dichotomize the rating scale at each category boundary into 50% probability of being observed below the category and 50% probability of being observed in or above the category.

The points on the latent variable where the expected scores are 2.5, etc., are called the CAT-0.5 points in ISFILE=.

The Rasch item difficulty (in [IFILE=](#)) is the point on the latent variable at which the highest and lowest categories are equally probable.

ISFILE=? opens a Browse window

ISFILE=filename produces an output file containing the category structure measure (Andrich threshold) information for each item. All measures are added to the corresponding item's calibration and rescaled by [USCALE=](#) and [UDECIMALS=](#). This file contains 4 heading lines (unless [HLINES=N](#) or [ROW1HEADING=N](#)), followed by one line for each item containing:

[DISFILE=](#), [DISOPTION=](#)Score shows the contents of each ISFILE= category.

Columns:

Start	End	Heading	In Table	Description
1	1			Blank or ";" if no responses or deleted (status = -2, -3)
2	6	ENTRY		The item sequence number
7	11	STATUS STAT		2. The item's status 1 = Estimated calibration 2 = Anchored (fixed) calibration 3 = Anchored (fixed) calibration with extreme (minimum or maximum) observed raw score 0 = Extreme minimum (estimated using EXTRSC=) -1 = Extreme maximum (estimated using EXTRSC=) -2 = No responses available for calibration -3 = Deleted by user
12	16	MAXIMUM MAX		Number of active categories
17	21	CAT BOT		Lowest active category number, bottom category
22	29	BOT+.25	2.2	Measure for an expected score of bottom category + LOWADJ= Useful as a measure for a performance in the bottom category, for which the performance range extends to -infinity.
The following fields are repeated for the remaining active categories:				
30	34	CAT		Active category number
35	39	ORDINAL ORD		Ordered category number in structure = "Step counting from bottom category"
40	47	THRESHOL D THRESH		Rasch-Andrich threshold (step difficulty) relative to item difficulty = Rasch parameter Fj. Use this or SFILE= for anchoring in SAFILE= .
48	55	I+THRESH I+THRSH	2.4	item measure + Rasch-Andrich threshold = Structure measure = Step calibration = Delta = Rasch parameter Dij. Do not use for anchoring. Use SFILE= for SAFILE= and IFILE= for IAFILE= .
56	63	S.E.		Rasch-Andrich threshold's standard error, with the item difficulty S.E. assumed to be 0.0.
64	71	CAT-0.5	2.2	Measure for an expected score of category - 0.5 score points. This is the Rasch-half-point threshold, the boundary between categories when conceptualized as average performances. It is not a model parameter. TOP-0.5 for the highest (top) category,
72	79	AT CAT (TOP-0.25)	2.2	Measure for an expected score of category score points (AT CAT). This is the measure corresponding to a category when predicting for an individual or sample about which nothing else is known. For items with fewer rating-scale categories, the AT CAT value shown for the highest category is the TOP-0.25 value.
80	87	PR50%	2.3	Probability 50%. Measure at Rasch-Thurstone threshold = 50% cumulative probability.
Only for the top (highest) category				
		TOP-0.25	2.2	For the highest (top) category this value corresponds to the top category value less HIADJ= , the measure for an expected score of HIADJ= score points less than the top category value. This is useful as a measure for a performance in the top category, for which the performance range extends to infinity. For items with fewer rating-scale categories, the AT CAT value shown for the highest category is the TOP-0.25 value.
		ITEM		Item label

The "AT CAT" values in the ISFILE= are based on the Rasch-model. They are the points on the "expected score" ogive for the rating scale (also called "the model ICC") at which the expected score = the category number. This is also the point at which the probability of observing the category is highest.

For extreme categories (top and bottom of the rating scale), the model values are infinite, so an adjustment is made. The "AT CAT" values correspond to expected scores bottom+0.25 score points and top-0.25 score points. These provide reasonable estimates for performance in the extreme categories of the rating scale. The adjustment of 0.25 can be changed with [LOWADJ=](#) and [HIADJ=](#). The "AT CAT" values are plotted on [Table 2.2](#).

Since the ISFILE= has the same number of category entries for every item, the repeated fields are filled out with "0" for any further categories up to the maximum categories for any item.

When [CSV=Y](#), commas separate the values with quotation marks around the "Item name". When CSV=T, the commas are replaced by tab characters.

When [STKEEP=YES](#) and there are intermediate [null](#) categories, i.e., with no observations, then the Rasch-Andrich threshold into the category is set about 40 logits above the previous threshold. The threshold out of the category, and into the next category, is set about 40 logits above. The exact values depend on the category frequencies of the observed categories. Thus:

Rasch-Andrich Thresholds for Unobserved Categories		
Category	in Table 3.2	in SFILE=
0 (observed)	NULL	0.00
1 (observed)	-1.00	-1.00
2 (unobserved)	NULL	39.00
3 (observed)	1.00	-38.00
Total:	0.00	0.00

Meanings of the columns

There are several ways of conceptualizing the [category boundaries or thresholds](#) of a rating (or partial credit) scale item. Imagine a rating (or partial credit) scale with categories, 1, 2, 3:

From the "expected score ogive", also called the "model item characteristic curve"

Average rating:	Measure (must be ordered)
1.25	Measure for an expected score of 0.25 (BOT+.25) when LOWADJ=0.25
1.5	Measure for an expected score of category - 0.5 score points (CAT-0.5)
2.0	Measure for an expected score of category score points (AT CAT)
2.5	Measure for an expected score of category - 0.5 score points (CAT-0.5)
2.75	Measure for an expected score of category score points (AT CAT) Since this is the top extreme category the reported values is for TOP-0.25 when HIADJ=0.25

From the "category probability curves" relative to the origin of the measurement framework (need not be ordered)

1-2 equal probability	Structure measure = Rasch-Andrich threshold + item measure = Dij (MEASURE)
standard error	Rasch-Andrich threshold's standard error (ERROR)

2 maximum probability	Measure for an expected score of category score points (AT CAT) - (yes, same as for the ogive)
2-3 equal probability	Structure measure = Rasch-Andrich threshold + item measure = Dij (MEASURE)
standard error	Rasch-Andrich threshold's standard error (ERROR)

From the "cumulative probability curves" (preferred by Thurstone) (must be ordered)

Category 1 at .5 probability	Measure at the Rasch-Thurstone threshold = 50% cumulative probability (50%PRB)
Category 1+2 at .5 probability	Measure at the Rasch-Thurstone threshold = 50% cumulative probability (50%PRB)

Example 1: You wish to write a file on disk called "ITEMST.FIL" containing the item statistics reported in Table 2.2, for use in constructing your own tables:

ISFILE = ITEMST.FIL

ISGROUPS = 0 ; each item has its own "partial credit" scale

LOWADJ = 0.25 ; the standard for the low end of the rating scale

HIADJ = 0.25 ; the standard for the high end of the rating scale

For column definitions, see above.

```
; ACT ITEM-STRUCTURE FILE (not for anchoring: use SFILE=) FOR LIKING FOR SCIENCE (Wright & Masters p.18) Mar 24 20:55 2015
;ENTRY STAT MAX CAT BOT+.25 CAT ORD THRESH I+THRSH S.E. CAT-0.5 AT CAT PR50% CAT ORD THRESH I+THRSH S.E. CAT-0.5 TOP-
.25 PR50% ACT
1 1 2 0 -2.68 1 1 -1.03 -1.50 .14 -1.76 -.47 -1.61 2 2 1.03 .55 .13 .81
1.74 .66 Watch birds
2 1 2 0 -3.01 1 1 -1.03 -1.83 .14 -2.09 -.80 -1.94 2 2 1.03 .23 .13 .49
1.41 .33 Read books on animals
```

Example 2: To produce a Table of expected measures per item-category similar to Pesudovs, K., E. Garamendi, et al. (2004). "The quality of life impact of refractive correction (QIRC) questionnaire: Development and validation." Optometry and Vision Science 81(10): 769-777, write the ISFILE= to Excel. Then delete or hide unwanted columns.

Item number	Response category				
	1	2	3	4	5
1	60.51	45.06	29.61	29.61	29.61
2	65.11	49.66	34.21	34.21	34.21
3	56.71	41.26	25.81	25.81	25.81

Example 3: To plot the operating range of each item using Excel.

"Output Files", "ISFILE=", output to Excel.

"Output Files", "IFILE=", output to Excel.

Paste the "MEASURE" column from the Excel IFILE= into the Excel ISFILE=

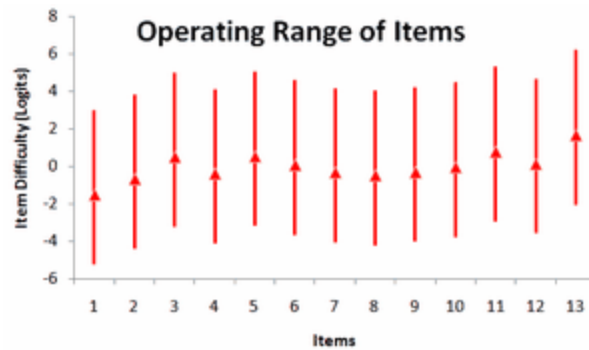
Arrange the columns: TOP-.25, BOT+.25, MEASURE

If some items have fewer categories, then paste their highest values into the TOP-.25 column

Draw a "hi-lo-close" plot with TOP-.25, BOT+.25, MEASURE for each item

If some items have fewer categories, then use their highest values in ISFILE=

This is one I have drawn from Exam12.txt with Exam12lo.txt+Exam12hi.txt



Example 4: To produce ISFILE= without the addition of item difficulties (similar to Table 3.2 for a rating scale)

We can do this by writing out the SFILE=sf.txt from the original analysis. Then do the analysis again with SAFILE=sf.txt and IAFILE=*

1-number of items 0
*

11.75 ISGROUPS= or GROUPS= assigns items to rating scale groupings

Winsteps chooses the model-family based on [MODELS=](#)

Models=R (or MODELS= is omitted) is the rating-scale family of models which includes the Andrich Rating-Scale Model, the Masters Partial-Credit Model, the Grouped Rating-Scale Model and the Rasch Dichotomous Model.

Within Models=R, the choice of model is decided by ISGROUPS= or GROUPS= or IGROUPS=

If ISGROUPS= is omitted, then

- 1) If there are only two categories in the data the dichotomous model
- 2) If there are more than two categories the Andrich Rating-Scale Model

ISGROUPS= *file name	file containing details
ISGROUPS = *	in-line list
ISGROUPS = codes	codes for item groups

If ISGROUPS=0 then
the Masters Partial Credit Model

If ISGROUPS=(something else)
the Grouped Rating-Scale Model

Please look at the output in [Table 3.2](#), 3.3, to see exactly how the data have been modeled.

Items in the same "grouping" share the same dichotomous, rating scale or partial credit response structure. For tests comprising only dichotomous items, or for tests in which all items share the same rating (or partial credit) scale definition, all items belong to one grouping, i.e., they accord with the simple dichotomous Rasch model or the Andrich "Rating Scale" model. For tests using the "Masters' Partial Credit" model, each item comprises its own grouping (dichotomous or polytomous). For tests in which some items share one polytomous response-structure definition, and other items another response-structure definition, there can be two or more item groupings. Groups are called "blocks" in the PARSCALE software. All the items in the same group of items must have the same number of categories. If an item has a different number of categories, place it in a different item group, or in an item group by itself. In ISGROUPS=, an item in a item group by itself can be given its own item group code or group "0". Item difficulty measures are the locations on the latent variable (Rasch dimension) where the highest and lowest categories of the item are equally probable, regardless of the number of categories the item has.

$$\log \left(\frac{P_{nij}}{P_{ni(j-1)}} \right) = B_n - D_{gi} - F_{gj}$$

where P is a probability, and the Rasch parameters are B_n , the ability of person, D_{gi} , the difficulty of item i of grouping g, and F_{gj} , the Rasch-Andrich threshold between categories j-1 and j of grouping g. If there is only one grouping, this is the Andrich "rating scale" model. If each item forms a grouping of its own, i.e., $g=i$, this is the Masters' "partial credit" model. When several items share the rating scale, then this could be called an item-grouping-level Andrich rating-scale model, or an item-grouping-level Masters' partial-credit model. They are the same thing.

ISGROUPS= also acts as [IREFER=](#), when [IVALUE=](#) is specified, but IREFER= is omitted.

ISGROUPS= has three forms: **ISGROUPS=1101110** and **ISGROUPS=* list *** and **ISGROUPS=*filename**

ISGROUPS=""	(standard if only one model specified with MODELS=) All items belong to one grouping. This is sometimes called "Andrich's Rating Scale Model"
ISGROUPS=0	(standard if MODELS= specifies multiple models) Each item has a grouping of its own, so that a different response structure is defined for each item, as in the "Masters' Partial Credit model". This is also used for rank order data.
ISGROUPS= some combination of numbers and letters	Use only one letter or number per grouping, regardless of the XWIDE= setting. Valid are: 0's, 1's, 2's, A's, B's, etc. ("a" is the same as "A"), also #, @, !, & etc. Items are assigned to the grouping label whose location in the ISGROUPS= string matches the item sequence number. Items in groupings labeled "1", "2", "A", etc. share the response-structure definition with other items in the same labeled grouping, which is the "Grouped Rating Scale Model". Each item assigned to grouping 0 is allocated to a "partial credit" grouping by itself, which is the same as having only one item in a group.
ISGROUPS= with extreme items	If there are items with minimum-possible or maximum-possible total raw scores, then please group those items with other items which do not have extreme scores, so that Winsteps knows which categories should have been observed for the extreme item. This does not change the response structure for the non-extreme items.

Valid one-character ISGROUPS= codes include: !#\$%&-. /123456789<>@ABCDEFGHIJKLMN O PQRSTU VWXYZ^_ |~

A-Z are the same as a-z

For the ISGROUPS= specification, "0" has the special meaning: "this is a one item grouping" - and can be used for every 1 item grouping.

NI=9

ISGROUPS=0

is the same as

ISGROUPS=000000000

is the same as

ISGROUPS=123456789

Characters with ASCII codes from 129-255 can also be used, but display peculiarly: ÏÐÑÒÓÔÕÖ×ØÙÚÛÜÝÞßàáâã etc.

When XWIDE=2 or more, then

either (a) Use one character per XWIDE= and blanks,

NI=8

XWIDE=2

ISGROUPS=' 1 0 1 0 1 0 1 1'

or (b) Use one character per item with no blanks

NI=8

XWIDE=2

ISGROUPS='10101011'

ISGROUPS=
item number grouping code
item number-item number grouping code
*

Each line has an item number, e.g., 23, or an item number range, e.g., 24-35, followed by a space and then a grouping code, e.g., 3. The items can be out of order. If an item is listed twice, the last entry dominates.

ISGROUPS=*file name
This has a file with the format of the ISGROUPS=* list.

Particularly with ISGROUPS=0, some extreme categories may only be observed for persons extreme scores. To reinstate them into the measurement analysis, see [Extreme Categories: Rescuing the Dropped](#).

Groupings vs. Separate Analyses

ISGROUPS= is very flexible for analyzing together different item response structures in one analysis. Suppose that an attitude survey has 20 Yes-No items, followed by 20 5-category Likert (Strongly disagree - disagree - neutral - agree - strongly agree) items, followed by 20 3-category frequency (never - sometimes - often) items. When possible, we analyze these together using ISGROUPS=. But sometimes the measurement characteristics are too different. When this happens, the fit statistics stratify by item type: so that, say, all the Yes-No items overfit, and all the Frequency items underfit. Then analyze the test in 3 pieces, and [equate](#) them together - usually into the measurement framework of the response structure that is easiest to explain. In this case, the Yes-No framework, because probabilistic interpretation of polytomous logits is always difficult to explain or perform.

The "equation" would be done by cross-plotting the person measures for different item types, and getting the slope and intercept of the conversion from that. Drop out of the "equation" any noticeably off-trend-line measures. These are person exhibiting differential performance on the different item types.

Example 1: Responses to all items are on the same 4-point rating scale, an Andrich "Rating Scale" model,
ISGROUPS=" "

Example 2: An arithmetic test in which partial credit is given for an intermediate level of success on some items. There is no reason to assert that the intermediate level is equivalent for all items. 0=No success, 1=Intermediate success (or complete success on items with no intermediate level), 2=Complete success on intermediate level items.

CODES=012 valid codes
ISGROUPS=0 each item has own response structure, i.e., Masters' Partial Credit model
or
ISGROUPS=
1 0 ; item 1 is in Grouping 0, no other items mentioned, so all assumed to be in Grouping 0
*

Example 3: An attitude survey consists of four questions on a 0,1,2 rating scale (grouping 1), an Andrich "Rating Scale" model, followed by three 0,1 items (grouping 2), an other Andrich "Rating Scale" model, and ends with one 0,1,2,3,4,5 question (grouped by itself, 0), a Masters' "Partial Credit" model.

NI=8 number of items
CODES=012345 valid codes for all items
ISGROUPS=11112220 the item groupings
or
ISGROUPS=
1-4 1
5-7 2
8 0 ; this line is optional, 0 is the standard.
*

When XWIDE=2, use two columns for each ISGROUPS= code. Each ISGROUPS= code must be one character, a letter or number, specified once in the two columns, e.g. " 1" or "1 " mean "1", and " 0" or "0 " mean "0".

Example 4: You wish most items on the "Liking for Science" Survey to share the same rating scale, an Andrich "Rating Scale" model, (in Grouping A). Items about birds (1, 10, 21) are to share a separate response structure, another Andrich "Rating Scale" model, (in Grouping B). Items 5 (cans) and 18 (picnic) each has its own response structure, i.e., the "Masters' Partial Credit" model, (Grouping 0).

```
NI=25 number of items
XWIDE=2
CODES=000102 valid codes for all items
ISGROUPS=' B AAA0 AAAAB AAAAAA0 AAB AAAA'
item groupings - use only one letter/number codes.
```

or

```
ISGROUPS=* ; XWIDE=2 is not a worry here, but use one letter/number codes.
```

```
1-25 A ; most items in grouping A
```

```
1 B ; item 1 transferred to grouping B
```

```
10 B
```

```
21 B
```

```
5 0 ; grouping 0 means item is by itself
```

```
18 0
```

*

Example 5: Four separate tests of patient performance (some on a 4-point rating scale, some on a 5-point rating scale) are to be Rasch-analyzed. All 500 persons were given all 4 tests. I analyzed each separately, to get an idea of good-fitting and bad-fitting items, etc. Now, I'd like to run all 4 tests together using a partial credit model.

There is no problem running all four tests together. Put them all in one file, or use [MFORMS=](#). If you intend every item of every test to have its own rating scale (i.e., a strict partial-credit model), use [ISGROUPS=0](#). But if you intend items on test 1 to share the same rating scale, similarly test 2 etc. (i.e., a test-level partial-credit model), then specify [ISGROUPS=11111112222233334444....](#) matching the grouping number indicators to the count of items in each test.

Example 6: Items are to be rescored in 3 different ways, and then the items are to be divided into 4 rating scale structures.

```
ISGROUPS=11112223312444 ; 4 RATING SCALE GROUPINGS
IREFER =AAAABBBCCABBBB ; 3 RECODINGS
CODES =01234 ; ORIGINAL CODES IN DATA
IVALUEA =01234 ; ORIGINAL CODING MAINTAINED - THIS LINE CAN BE OMITTED
IVALUEB =43210 ; CODING IS REVERSED
IVALUEC =*112* ; DICHOTOMIZED WITH EXTREME CATEGORIES MARKED MISSING
```

Example 7: A five-item test.

Item 1: Dichotomy already scored 0-1 ; let's call this a "D" (for dichotomy) group item

Item 2: Dichotomy already scored 0-1 ; this is another "D" (for dichotomy) group item. Under the Rasch model, all dichotomies have the same response structure.

Item 3: Partial credit polytomy already scored 0-1-2 ; this is an "0" type item. "0" means "this item has its own response structure".

Item 4: Rated polytomy already scored 1-2-3-4 ; let's call this an "R" group items

Item 5: Rated polytomy already scored 1-2-3-4 with the same rating scale as item 4, so this is another "R" group item,

```
CODES = 01234 ; this is all possible valid codes in the data
```

```
ISGROUPS = DD0RR ; Winsteps detects from the data which are the responses for each item-group and what they mean.
```

Example 8: An attitude survey in which the rating-scale items are worded alternately positively and negatively.

```
CODES = 12345 ; the rating scale of 5 categories
```

```
; rescore the items
```

```
IREFER = PNPNPNPNP ; 10 item, alternately positively-worded and negatively-worded
```

```
IVALUEP = 12345 ; the positive orientation
```

```
IVALUEN = 54321 ; the negative orientation
```

```
; model the rating scales
```

```
GROUPS = PNPNPNPNP ; model the two item orientations with separate rating scale structures
```

Example 9: A test with 5 dichotomous items and 5 partial credit items:

```
CODES = 01234 ; all possible item scores
```

```
GROUPS = 1111100000 ; 5 dichotomies share the same response structure
```


sort column
END NAMES

Example 2: The item name contains several important classifiers. Table 15 is required for each one:

```
TFILE=*
15 --- 1 ; sort starts with column 1 of item name
15 --- 6 ; sort starts with column 6
15 --- 13 ; sort starts with column 13 of the item name and goes to the end of the item name
- entered as place-holders, see TFILE=
*
```

```
&END
MCQU Geogrp 1995-0234
sort column
    sort column
        sort column
|
END NAMES
```

Example 3: A version of Table 15, sorted on item name column 13, is to be specified on the DOS command line or on the Extra Specifications line. Commas are used as separators, and "-" as place-holders:

```
TFILE=* 15,-,-,13 *
```

11.77 ISRANGE= category range of a rating scale

By default, Winsteps deduces the range of categories of an item group from the data. ISRANGE= or ICRANGE= states the range of category numbers explicitly. This is useful with [SFUNCTION=](#)

ISRANGE= file name	file containing details
ISRANGE = *	in-line list
ISRANGE = ?	opens a Browser window to find the file

```
ISRANGE=* (or filename)
(item number in item group, ISGROUP=) (lowest category number) (highest category number)
```

```
...
*
```

Example: For the Olympic Ice-Skating in [Exam15.txt](#), the rating scale range is 0-60 though the observed range is 29-60. All items (raters) share the same rating scale.

```
ISRANGE=*
2 00 60 ; 2 is an example rater number in the group
*
```

11.78 ISUBTOTAL= columns within item label for Table 27

This specifies what part of the data record is to be used to classify items for subtotal in Table 27.

ISUBTOTAL= file name	file containing details
ISUBTOTAL = *	in-line list
ISUBTOTAL = \$S1W1	field in item label

Format 1: ISUBTOTAL = \$S.W.. or \$S.E.. using the column [selection rules](#).

\$S.W.. e.g., \$S2W13 means that the label to be shown on the map starts in column 2 of the item label and is 13 columns wide.

\$S.E.. e.g., \$S3E6 means that the label to be shown on the map starts in column 3 of the item label and ends in column 6.

These can be combined, and constants introduced, e.g,
 ISUBTOTAL=\$S3W2+/"'+\$S7W2

If the item label is "KH323MXTR", the sub-grouping will be shown as "32/XT"

Format 2: ISUBTOTAL=*

This is followed by a list of sub-groupings, each on a new line using the column [selection rules](#):

```
ISUBTOTAL=*
$$S1W1+$S7W2 ; Subtotals reported for item classifications according to these columns
$$S3E5 ; Subtotals reported for item classifications according to these columns
*
```

Example: Subtotal by first letter of item name:

```
ISUBTOTAL=$S1W1
or
@SGROUP=$S1W1
ISUBTOTAL=@SGROUP
```

```
TFILE=*
27 ; produce the subtotal report
*
```

Here is a subtotal report (Tables 27) for items beginning with "R"

```
"R" SUBTOTAL FOR 8 NON-EXTREME ITEMS
+-----+
|          RAW          MODEL          INFIT          OUTFIT          |
|          SCORE        COUNT    MEASURE    ERROR          MNSQ    ZSTD    MNSQ    ZSTD    |
+-----+-----+-----+-----+-----+-----+-----+-----+
| MEAN          28.1          25.0          4.04    3.48          .91    -.5    1.04    .0    |
| P.SD           5.5           .0           6.63    .14          .31    1.1    .54    1.4    |
| MAX.           38.0          25.0          16.30   3.82          1.61    2.0    2.37   3.4    |
| MIN.           19.0          25.0          -6.69   3.38          .64   -1.6    .60   -1.2    |
+-----+-----+-----+-----+-----+-----+-----+-----+
| REAL RMSE    3.63 TRUE SD    5.54 SEPARATION 1.52 PUPIL RELIABILITY .70 |
| MODEL RMSE   3.48 TRUE SD    5.64 SEPARATION 1.62 PUPIL RELIABILITY .72 |
| S.E. OF PUPIL MEAN = 2.50 |
| WITH 2 EXTREME = TOTAL 10 PUPILS MEAN = 3.05, P.SD = 28.19 |
| REAL RMSE    8.88 TRUE SD    26.75 SEPARATION 3.01 PUPIL RELIABILITY .90 |
| MODEL RMSE   8.83 TRUE SD    26.77 SEPARATION 3.03 PUPIL RELIABILITY .90 |
| S.E. OF PUPIL MEAN = 9.40 |
+-----+-----+-----+-----+-----+-----+-----+-----+
| MAXIMUM EXTREME SCORE: 1 PUPILS |
| MINIMUM EXTREME SCORE: 1 PUPILS |
| LACKING RESPONSES: 1 PUPILS |
| DELETED: 1 PUPILS |
```

11.79 ITEM= title for item labels

Up to 12 characters to use in table headings to describe the kind of items, e.g.

ITEM=MCQ.

11.80 ITEM1= column number of first response

Specifies the column position where the response-string begins in your data file record, or the column where the response-string begins in the new record formatted by [FORMAT=](#).

If you have the choice, put the person-identifiers first in each record, and then the item responses with each response taking one column.

Error messages regarding **ITEM1=** may be because your control file is not in "Text with line breaks" format.

It is easy to miscount the **ITEM1=** column. Scroll to the top of the **Winsteps** screen and check column positions:

```
Input in process..
Input Data Record:
      1      2
1234567890123456789012345678
Richard M 111111100000000000
^P      ^I      ^N
      35 KID      Records Input.
```

^P marks the Name1=1 column position with ^.

^I marks the Item1=11 column position with ^.

^N marks the NI=18 column position with ^.

Example 1: The string of 56 items is contained in columns 20 to 75 of your data file.

```
ITEM1=20 response to item 1 in column 20
NI=56 for 56 items
XWIDE=1 one column wide (the standard)
```

Example 2: The string of 100 items is contained in columns 30 to 69 of the first record, 11 to 70 of the second record, followed by 10 character person i.d.

```
XWIDE=1 one column wide (the standard)
FORMAT=(T30,40A,/,T11,60A,10A) two records per person
ITEM1=1 item 1 in column 1 of reformatted record
NI=100 for 100 items
NAME1=101 person id starts in column 101
NAMLEN=10 person id starts in 10 columns wide
```

11.81 ITLEN= maximum length of item label

ITLEN= specifies the maximum number of columns in the control file that are to be used as item names. The maximum possible is 300 characters. ITLEN= can be specified in the [Specification](#) box. Its length cannot exceed the original maximum label length. If your item labels are longer than 300 characters, please use [ITEM1=](#) and ITLEN= to pick out the important part of the label, and use [FORMAT=](#) if the important parts are too far apart.

Example 1: You only wish to use the first five columns of item information to identify the items on the output:

```
NI=4
ITLEN=5
&END
AX123 This part is not shown on the output
BY246 Trial item
AZ476 This item may be biased
ZZ234 Hard item at end of test
END NAMES
```

Example 2: Your item names may be up to 50 characters long:

```
NI=4
ITLEN=50
&END
This item demonstrates ability for constructive reasoning
This item flags rigid thinking
This item detects moral bankruptcy
This item is a throw-away
END NAMES
```


11.82 IVALUEx= recoding of data

Responses are revalued according to the matching codes in [IREFER=](#) (or [ISGROUPS=](#) if IREFER= is omitted). Items in IREFER= not referenced by an IVALUEx= are not recoded.

IVALUEa= is the same as IVALUEA=

The recoded values in IVALUEx= line up vertically with the response codes in [CODES=](#), if a data value does not match any value in CODES= it will be treated as missing.

Valid one-character IVALUE= codes include: 123456789ABCDEFGHIJKLMNQRSTUWXYZ^_~!#\$%&.-/;<>@

A-Z are the same as a-z.

Characters with ASCII codes from 129-255 can also be used, but their appearance depends on the font chosen:

À Á Â Ã Ä Å Æ Ç È É Ê Ë Ì Í Î Ï Ñ Ò Ó Ô Õ Ö × Ø Ù Ú Û Ü Ý Þ ß à á â ã ä å æ ç è é ê ë ì í î ï ð ñ ò ó ô õ ö ÷ ø ù ú û ü ý þ ÿ
, f , ... † ‡ % † ¢ £ ¤ ¥ ¦ § ¨ © ª « ¬ ® ¯ ± ² ³ ´ µ ¶ · ¸ ¹ º » ¼ ½ ¾ ¿

When XWIDE=2 or more, then either (a) Use one character per XWIDE and blanks,

```
NI=8
XWIDE=2
IREFER=' A B C D D C B A'
```

or (b) Use one character per item with no blanks

```
NI=8
XWIDE=2
RESCORE='ABCDDCBA'
```

Layout is:

```
NI = 17
IREFER = AACADBCDEEEDAABC ; the recoding type designators for the 17 items
; see the vertical line up here
CODES = 0123456 ; valid codes across all items
IVALUEA = 012**** ; recodes for Grouping A
IVALUEB = *1224** ; recodes for Grouping B: "2" and "3" recoded to "2"
IVALUEC = *122226 ; 1-2-6 acts as 1-2-3 because STKEEP=NO
IVALUED = 012333*
IVALUEE = 00123**
STKEEP=NO ; missing intermediate codes are squeezed out
```

Example 1: Items identified by Y and Z in IREFER= are to be recoded.

Y-type items are 1-3, 7-8. Z-type items are 4-6, 9-10. All items have their own rating (or partial credit) scales,

```
NI = 10
IREFER = YYYZZZYYZZ ; items identified by type: item 1 is Y, item 4 is Z etc.
CODES = ABCD ; original codes in the data
IVALUEY= 1234 ; for Y-type items, this converts A to 1, B to 2, C to 3, D to 4
IVALUEZ= 4321 ; for Z-type items, this converts A to 4, B to 3, C to 2, D to 1
ISGROUPS=0 ; allow each item to have its own rating (or partial credit) scale structure
```

Example 2: Items identified by 1, 2, 3 in ISGROUPS= are to be recoded and given their own rating (or partial credit) scales

Y-type items are 1-3, 7-8. Z-type items are 4-6, 9-10.

```
NI = 10
ISGROUPS = YYYZZZYYZZ
CODES= ABCD ; original codes in the data
IVALUEY= 1234
IVALUEZ= 4321
```

or

```
NI = 10
```

```
ISGROUPS = YYYZZZZYYYZZ
IREFER = YYYZZZZYYYZZ
CODES= ABCD ; original codes in the data
IVALUEY= 1234
IVALUEZ= 4321
```

Example 3: All items are to be recoded the same way.

```
NI = 100 ; 100 ITEMS
IREFER=*
1-100 X ; FOR ALL 100 ITEMS, reference is X
*
```

```
Codes = 12345 ; rescore 12345
IVALUEX = 12223 ; into 12223
```

11.83 IWEIGHT= item (variable) weighting

IWEIGHT= allows for differential weighting of items. The standard weights are 1 for all items. To change the weighting of persons, specify [PWEIGHT=](#).

IWEIGHT of 2 has the same effect on person estimation as putting the item and its responses into the analysis twice. It does not change an item scored 0-1 into an item scored 0-2. When IWEIGHT is 0 for an item, the measure and fit statistics are reported for the item, but the item does not influence the measures or fit statistics of the other items or persons. IWEIGHT= applies to everything except the dimensionality computations (Tables [23](#), [24](#)) where PWEIGHT= is set to 1 and IWEIGHT= is set to 1 for all persons and items.

IWEIGHT= file name	file containing details
IWEIGHT = *	in-line list
IWEIGHT = \$\$S1W1	field in item label

Raw score, count, and standard error of measurement reflect the absolute size of weights as well as their relative sizes. Measure, infit and outfit and correlations are sensitive only to relative weights.

Weighting is treated for estimation as that many independent observations. So, if you weight all items by two, you will divide the S.E. by the square-root of 2, but will not change the measures or fit statistics.

If you want to do different weighting at different stages of an analysis, one approach is to use weighting to estimate all the measures. Then anchor them all ([IFILE=](#) and [IAFILE=](#) etc.) and adjust the weighting to meet your "independent observation" S.E. and reporting requirements.

If you want the standard error of the final weight-based measure to approximate the S.E. of the unweighted measure, then ratio-adjust case weights so that the total of the weights is equal to the total number of independent observations.

Formats are:

```
IWEIGHT=file name the weights are in a file of format:
item number weight
```

```
IWEIGHT=*
item number weight
```

```
....
*
```

```
IWEIGHT=$$S...$W... or $$S...$E... weights are in the item labels using the column selection rules, e.g., starting in column S... with a width of W... or starting in column S and ending in column E. This can be expanded, e.g.,
IWEIGHT = $$S23W1+"."+$S25W2
places the columns next to each other (not added to each other)
```

Example 0: What happens when an item is weighted

```
IWEIGHT=*
1 0.5
```

*

- 1) [Table 14](#): the item scores and counts are not changed
- 2) [Table 18](#): the person scores and counts are changed. Item 1 score and count is weighted by 0.5
- 3) [Table 3.2](#): the response counts for the rating scale are weighted

Example 1: In a 20-item test, item 1 is to be given a weight of 2.5, all other items have a weight of 1.

```
IWEIGHT=*  
1 2.5  
2-20 1  
*
```

A better weighting, which would make the reported person standard errors more realistic by maintaining the original total sum of weights at 20, is:

```
IWEIGHT=*  
1 2.33 ; 2.5 * 0.93  
2-20 0.93 ; the sum of all weights is 20.0  
*
```

or adjust the weights to keep the sample-based "test" [separation](#) and reliability about the same - so that the reported statistics are still reasonable:

e.g., original sample "test" reliability (person separation index) = .9, separation coefficient = 3, but separation coefficient with weighting = 4

Multiply all weights by $(3/4)^2$ to return separation coefficient to about 3.

Example 2: The item labels contain the weights in columns 16-18.

```
IWEIGHT= $$S16W3 ; or $$S16E18  
&END  
Item 1 Hello 0.5  
Item 2 Goodbye 0.7  
.....  
END NAMES
```

Example 3: Item 4 is a pilot or variant item, to be given weight 0, so that item statistics are computed, but this item does not affect person measurement.

```
IWEIGHT=*  
4 0 ; Item 4 has weight 0, other items have standard weight of 1.  
*
```

Example 4: We have 24 0/1 items and 5 0/1/2/3 items. We want them to weight equally. There are several concerns here. These may require different weights for different purposes, i.e., several runs.

(a) Raw score reporting. For 24 items of 0/1 and 5 items of 0/0.33/0.67/1. Then

```
IWEIGHT=*  
1-24 1  
25-29 0.333  
*
```

This will give the reportable raw scores you want, 0-29, but incorrect reliabilities (too small).

(b) Statistical information. The total amount of overall statistical information can be maintained approximately by maintaining the total raw score. So original ordinal unweighted raw score range = 0 - $(24 \times 1 + 5 \times 3) = 39$. New raw score in (a) = 29. So we need to up the weights by $39/29 = 1.345$. This will give peculiar-looking raw scores, but a better estimate of fit.

```
IWEIGHT=*  
1-24 1.345  
25-29 0.448  
*
```

The Rasch measures for (a) and (b) will be the same, but the standard errors, reliabilities and fit statistics will differ.

(c) Reliability maintenance. If you want to maintain the same person "test" reliability (i.e., measure reproducibility), then

approximate weighting factor = (1 - weighted person model reliability) / (1 - unweighted person model reliability)
 IWEIGHT=*
 1-24 3 * weighting factor
 25-29 1 * weighting factor
 *

(d) recoding the 0/1 data into 0/3 to give equal weighting with 0/1/2/3 is not recommended because of the two unobserved categories, 1, 2, which change the slope of the model ICC, so distorting the measures, altering the fit and creating artificially high reliabilities.

Example 5: The items are to be weighted in accordance with Jian Tao, Ning-Zhong Shi and Hua-Hua Chang (2012). Item-Weighted Likelihood Method for Ability Estimation in Tests Composed of Both Dichotomous and Polytomous Items. Journal of Educational and Behavioral Statistics, 2012, 37, 2, 298-315.

In their empirical example, items 1-50 are dichotomies. Item 51 is a 0-4 polytomy. Their weighting is:
 IWEIGHT=*
 1-50 0.9924 ; this is the weighting for the 50 dichotomous items
 51 1.3795 ; this is the weighting for the 1 polytomous (0-4) item
 *

Example 6: A demonstration:

1. Analyze exam1.txt in the Winsteps Examples folder. Save the person measures. Table 17.

```
-----
|ENTRY  TOTAL  TOTAL      MODEL|  INFIT |  OUTFIT |PT-MEASURE|EXACT MATCH|  |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP. |OBS%  EXP%| KID  |
|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|      7      14      18      3.73      .94|.95  .1|.39  -.1|.69  .69|85.7  88.8| Susan  F|
```

2. Add these lines to the exam1.txt control file:

IWEIGHT=*
 13 2 ; weight item 13 double
 *

3. Analyze exam1.txt in the Winsteps Examples folder. Compare the person measures. Table 17, with Table 17 from 1.

```
-----
|ENTRY  TOTAL  TOTAL      MODEL|  INFIT |  OUTFIT |PT-MEASURE|EXACT MATCH|  |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP. |OBS%  EXP%| KID  |
|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|      7      15      19      3.85      .91|.92  .0|.37  -.1|.67  .66|86.7  89.2| Susan  F|
```

There are also differences in the item Table 14:

```
-----
|ENTRY  TOTAL  TOTAL      MODEL|  INFIT |  OUTFIT |PT-MEASURE|EXACT MATCH|  |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP. |OBS%  EXP%|WEIGH| TAP  |
|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|     13      7      35      1.87      .54|.49 -1.9|.24  -.6|.67  .42|94.1  86.4| 2.00| 1-4-3-2-4  |
```

11.84 KEYFROM= location of KEYn

KEYFROM= instructs where to find the [KEYn=](#) information. *Only use this if you have the scoring Key conveniently in data-record format.*

KEYFROM= 0	KEY1= through KEY99=, if used, are control variables before &END.
KEYFROM= 1	KEY1 information follows after &END, but before the item names. The key is formatted exactly like a data record. It is helpful to place the name of the key, e.g. "KEY1=", where the person name would usually go, for reference.
KEYFROM= n (a positive number up to 99)	KEY1, then KEY2, and so on up to KEYn (where n is a number up to 99) follow &END, but placed before the item names. Each key is formatted exactly like a data record. It is

	helpful to place the name of the key, e.g. "KEY2", where the person name would usually go.
SPFILE= filename	KEY1= , KEY2= , etc., in a separate file in control-file format

Example: KEY1 and KEY2 information are to follow directly after &END

```

NAME1=1  start of person-id (the standard)
ITEM1=10 start of response string
NI=7    number of items
CODES=abcd  valid codes
KEYFROM=2  two keys in data record format
&END
KEY1=****bacddba keys formatted as data
KEY2=****cdbbaac
Item 1 name  item names
|
Item 7 name
END NAMES
Mantovanibbacdba first data record
|  subsequent data records

```

11.85 KEYn= scoring key

Usually only [KEY1=](#) is needed for an MCQ scoring key.

Up to 99 keys can be provided for scoring the response choices, with control variables [KEY1=](#) through [KEY99=](#). Usually [KEY1=](#) is a character string of "correct" response choices. The standard is one column per correct response, or two columns if [XWIDE=2](#).

As standard, responses matching the characters in [KEY1=](#) are scored 1. Other valid responses are scored 0. [KEY2=](#) through [KEY99=](#) are character strings of successively "more correct" response choices to be used when more than one level of correct response choice is possible for one or more items. The standard score value for [KEY2=](#) is 2, and so on up to the standard score value for [KEY99=](#) which is 99. The values assigned to these keys can be changed by means of [KEYSCR=](#). If [XWIDE=1](#), only the values assigned to [KEY1=](#) through [KEY9=](#) can be changed, [KEY10=](#) through [KEY99=](#) retain their standard values of 10 through 99. If [XWIDE=2](#), the all [KEYn=](#) values can be changed.

More complicated scoring can be done with [IREFER=](#)

Example 1: A key for a 20-item multiple choice exam, in which the choices are coded "1", "2", "3" and "4", with one correct choice per item, scored 1. Wrong, invalid and missing responses are scored 0.

```

CODES = 1234    ; valid codes
KEY1  = 31432432143142314324 ; correct answers scored 1
; incorrect responses in CODES= automatically scored 0
MISSING-SCORED = 0 ; scoring of responses not in CODES=

```

Example 2: A 20-item MCQ test with responses entered as "a", "b", "c", "d". Wrong responses scored 0. Invalid and missing responses are scored "not administered", and so excluded from the analysis

```

CODES=abcd    ; valid responses
KEY1 =cadcbdcbadcadbcadcbd ; ; correct answers scored 1
; incorrect responses in CODES= automatically scored 0
MISSING-SCORED = -1 ; scoring of responses not in CODES= (default)

```

Example 3: A 20 item multiple choice exam with two somewhat correct response choices per item. One of the correct choices is "more" correct than the other choice for each item, so the "less correct" choice will get a score of "1" (using [KEY1=](#)) and the "more correct" choice will get a score of "2" (using [KEY2=](#)). All other response choices will be scored "0":

```

CODES=1234    valid responses
KEY1=23313141324134242113  assigns 1 to these responses
KEY2=31432432143142314324  assigns 2 to these responses
;      0 is assigned to other valid responses
ISGROUPS = 0 ; for the Partial Credit model
MISSING-SCORED = 0 ; scoring of responses not in CODES=

```

Example 4: A 100 item multiple choice test key.

```
CODES= ABCD
KEY1 = BCDADDCBBADCDACBCDADDCBBADCDACBCDADDCB+
+DBBADCDACBCDADDCBBADCDACBCDADDCBBADCCD+
+ACBCDADDCBBADCDACBC ; continuation lines
```

Example 5: Multiple score key for items 1 to 10. Items 11 to 15 are on a rating scale of 1 to 5

```
CODES = abcd12345
KEY1 = bacdbaddcd*****
RESCORE= 111111111100000 ; RESCORE= signals when to apply KEY1=
ISGROUPS= 111111111122222 ; 1 indicates dichotomy, 2 indicates rating scale
```

Example 6: A 12 item test. Items 1-5 are MCQ with responses entered as "ABCD", with one of those being correct: Item 1, correct response is B. Item 2 is C. 3 is D. 4 is A. 5 is C. Then items 6-10 are partial-credit performance items rated 0-5. Items 11-12 are two dichotomous items, scored 0,1.

```
CODES =ABCD012345 ; screen for all valid codes in the data file
KEY1 = BCDAC***** ; Key1= automatically has the value "1", etc.
RESCORE= 111110000000 ; "1" selects the KEY1=
ISGROUPS=111110000022 ; 1 indicates MCQ, 0 indicates partial credit, 2 indicates dichotomy
```

Example 7: Item 1 has correct options B,D,E. Other items have only one correct option.

```
ITEM1=1
NI=10
NAME1=11
NAMELENGTH=8
CODES = "ABCDE"
; B,D,E ARE ALL CORRECT FOR ITEM 1
KEY1 = BBADCBDDC
KEY2 = D*****
KEY3 = E*****
KEYSCR = 111 ; ALL KEYS ARE WORTH 1 POINT
&END
Item 1
Item 2
Item 3
Item 4
Item 5
Item 6
Item 7
Item 8
Item 9
END NAMES
123456789
ADABADCCD Person 1 ; this scores as 001000000
BBDDCAAB Person 2 ; 110100000
CDCCBABA Person 3 ; 000001000
DCBCCACDC Person 4 ; 100010011
EDCDCDCBD Person 5 ; 100110000
```

Example 8: the answer to each one of the options, but the options differ across items. Item 1 has options ABC with correct option B. Item 2 has options CDEF with correct option E.

```
CODES = ABCDEF ; all options
KEY1 = BE ; correct options
```

11.86 KEYSR= reassign scoring keys

This is only needed for complicated rescoring.

Specifies the score values assigned to response choices which match **KEY1=** etc. To assign responses matching key to the "missing" value of -1, make the corresponding **KEYSCR=** entry blank or some other non-numeric character.

When **XWIDE=1**, each value can only take one position, so that only **KEY1=** through **KEY9=** can be reassigned. **KEY10=** through **KEY99=** can also be used but keep their standard values of 10 through 99.

When **XWIDE=2**, each value takes two positions, and the values corresponding to all keys, **KEY1=** through **KEY99=**, can be reassigned.

Example 1: Three keys are used, and **XWIDE=1**.
Response categories in **KEY1=** will be coded "1"
Response categories in **KEY2=** will be coded "2"
Response categories in **KEY3=** will be coded "3"
KEYSCR=123 (standard)

Example 2: Three keys are used, and **XWIDE=1**.
Response categories in **KEY1=** will be coded "2"
Response categories in **KEY2=** will be coded "1"
Response categories in **KEY3=** will be coded "1"
KEYSCR=211

Example 3: Three keys are used, and **XWIDE=2**
Response categories in **KEY1=** will be coded "3"
Response categories in **KEY2=** will be coded "2"
Response categories in **KEY3=** will be coded "1"
KEYSCR=030201
or
KEYSCR=" 3 2 1"

Example 4: Three keys are used, and **XWIDE=1**
Response categories in **KEY3=** will be coded "1"
Response categories in **KEY6=** will be coded "missing"
Response categories in **KEY9=** will be coded "3"
KEY3=BACDCACDBA response keys
KEY6=ABDADCDCAB
KEY9=CCBCBBBBCC
KEYSCR=xx1xxXxx3 scores for keys

The "x"s correspond to unused keys, and so will be ignored.
The "X" corresponds to specified **KEY6=**, but is non-numeric and so will cause responses matching **KEY6=** to be ignored, i.e. treated as missing.

Example 5: Some items in a test have two correct answers, so two keys are used. Since both answers are equally good, **KEY1=** and **KEY2=** have the same value, specified by **KEYSCR=**. But some items have only one correct answer so in one key "**", a character not in **CODES=**, is used to prevent a match.
CODES=1234
KEY1 =23313141324134242113
KEY2 =31*324321*3142314* ; * is not in CODES=**
KEYSCR=11 both KEYS scored 1

Example 6: More than 9 **KEYn=** lines, together with **KEYSCR=**, are required for a complex scoring model for 20 items, but the original data are only one character wide.

Original data: Person name: columns 1-10
20 Item responses: columns 21-40

Looks like: M. Stewart.....1321233212321232134

Solution: reformat from **XWIDE=1** to **XWIDE=2**

```
TITLE="FORMAT= from XWIDE=1 to =2"  
FORMAT=(10A1,10X,20A1) ; 10 of Name, skip 10, 20 of responses
```

```

NI=20
NAME1=1
ITEM1=11   Responses in column 11 of reformatted record
XWIDE=2
CODES="1 2 3 4 "   Original response now "response blank"
KEY1 ="1 2 1 3 2 1 2 3 1 4 3 2 1 1 1 1 1 2 1 1 " Keying 20 items
KEY2 ="2 1 2 1 1 2 1 1 2 1 1 1 2 3 3 2 2 * 2 1 "
. . . .
KEY10="3 3 3 2 3 3 4 2 3 * * * 4 4 4 4 4 4 4 "
KEYSCR="1 2 3 2 2 2 3 4 1 4 " ; Renumbering 10 KEYn=
&END

```

11.87 LCONV= logit change at convergence

Measures are only reported to 2 decimal places, so a change of less than .005 logits will probably make no visible difference.

Specifies what value the largest change in any logit estimate for a person measure or item calibration or rating (or partial credit) scale structure calibration must be less than, in the iteration just completed, for iteration to cease. The current largest value is listed in Table 0 and displayed on your screen. See [convergence considerations](#).

The standard setting is [CONVERGE="E"](#), so that iteration stops when either [LCONV=](#) or [RCONV=](#) is satisfied. (Note: this depends on Winsteps version - and may explain differences in converged values.)

Example 1: To set the maximum change at convergence to be less or equal to .001 logits:

```

LCONV=.001
RCONV=0 ; set to zero, so does not affect convergence decision
CONVERGE=Logit

```

Example 2: To set the maximum change at convergence to less or equal to .1 user-scale [USCALE=](#) units:

```

USCALE = 50 ; 50 user-scaled units per logit
LCONV = .002 ; logit value = .002*50 = 0.1 user-scaled units

```

11.88 LESS= heading in Tables 1, 12, 16

Please see [MORE=](#)

11.89 LINELENGTH= length of printed lines

The misfitting responses, name maps, scalogram, and option frequency tables can be output in any convenient width up to 3003 characters. Specify [LINELENGTH=0](#) for a large page width (503 characters).

Example: You want to print the map of item names with up to 100 characters per line.

```

LINELENGTH=100 set line length to 100 characters

```

11.90 LOCAL= locally restandardize fit statistics

LOCAL=N accords with large-sample statistical theory.

Standardized fit statistics test report on the hypothesis test: "Do these data fit the model (perfectly)?" With large sample sizes and consequently high statistical power, the hypothesis can never be accepted, because all empirical data exhibit some degree of misfit to the model. This can make *t* standardized statistics meaninglessly large. *t* standardized statistics are reported as unit normal deviates. Thus $ZSTD=2.0$ is as unlikely to be observed as a value of 2.0 or greater is for a random selection from a normal distribution of mean 0.0, standard deviation, 1.0. *ZSTD* (standardized as a z-score) is used of a *t*-test result when either the *t*-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's [t-statistic](#) distribution value has been adjusted to a unit normal value.

Control	Column Heading in Table	Explanation
---------	-------------------------	-------------

LOCAL=No	ZSTD standardized like a Z-score	t-standardized fit statistics are computed in their standard form. Even the slightest item misfit in tests taken by many persons will be reported as very significant misfit of the data to the model. Columns reported with this option are headed "ZSTD" for model-exact standardization. This is a "significance test" report on "How unexpected are these data if the data fit the model perfectly?" Adjusted by WHEXACT=
LOCAL=Log	LOG $\log_e(\text{mean-square})$	Instead of t-standardized statistics, the natural logarithm of the mean-square fit statistic is reported. This is a linearized form of the ratio-scale mean-square. Columns reporting this option are headed "LOG", for mean-square logarithm.
LOCAL=Yes	ZEMP locally standardized ZSTD	t-standardized fit statistics are transformed to reflect their level of unexpectedness in the context of the amount of disturbance in the data being analyzed. The model-exact t standardized fit statistics are divided by their local standard deviation . Thus their transformed standard deviation becomes 1.0. Columns reported with this option are headed "ZEMP" for "empirically re-standardized to match a unit-normal (Z) distribution". The effect of the local-rescaling is to make the fit statistics more useful for interpretation. The meaning of ZEMP statistics is an "acceptance test" report on "How unlikely is this amount of misfit in the context of the overall pattern of misfit in these data?" Adjusted by WHEXACT=
LOCAL=Prob	PROB probability of mean-square	2-sided probability of the mean-square (CHISQUARE=No) or chi-square (CHISQUARE=Yes). The degrees of freedom are in the IFILE= or PFILE= . The chi-square is mean-square * d.f.

The ZSTD t-standardized-as-a-Z-score fit statistics test a null hypothesis. The usual null hypothesis is "These data fit the Rasch model exactly after allowing for the randomness predicted by the model." Empirical data never do fit the Rasch model exactly, so the more data we have, the more certain we are that the null hypothesis must be rejected. This is what your fit statistics are telling you. But often we don't want to know "Do these data fit the model?" Instead, we want to know, "Is this item behaving much like the others, or is it very different?"

Ronald A. Fisher ("Statistical Methods and Scientific Inference" New York: Hafner Press, 1973 p.81) differentiates between "tests of significance" and "tests of acceptance". "Tests of significance" answer hypothetical questions: "how unexpected are the data in the light of a theoretical model for its construction?" "Tests of acceptance" are concerned with whether what is observed meets empirical requirements. Instead of a theoretical distribution, local experience provides the empirical distribution. The "test" question is not "how unlikely are these data in the light of a theory?", but "how acceptable are they in the light of their location in the empirical distribution?"

This also parallels the work of Shewhart and W.E. Deming in quality-control statistics. They construct the control lines on their quality-control plots based on the empirical "common-cause" variance of the data, not on a theoretical distribution or specified tolerance limits.

So, in Winsteps, you can specify LOCAL=Yes to test a different null hypothesis for "acceptance" instead of "significance". This is not "cheating" as long as you inform the reader what hypothesis you are testing. The revised null hypothesis is: "These data fit the Rasch model exactly after allowing for a random normal distribution of standardized fit statistics equivalent to that observed for these data."

The ZEMP transformed standardized fit statistics report how unlikely each original standardized fit statistic ZSTD is to be observed, if those original standardized fit statistics ZSTD were to conform to a random normal distribution with the same variance as that observed for the original standardized fit statistics.

To avoid the ZEMP values contradicting the mean-square values, Winsteps does separate adjustments to the two halves of the ZSTD distribution. ZEMP takes ZSTD=0 as the baseline, and then linearly adjusts the positive and negative halves of the ZSTD distribution independently, giving each half an average sum-of-squares of 1.0 away from 0. When the two halves are put together, the model distribution of ZEMP is $N[0,1]$, and the empirical distribution of ZEMP approximates a mean of 0 and a standard deviation of 1. Usually there is no ZSTD with value exactly 0.000. Algebraically:

MATRIX = NO is the list format

```
Item Item Correlation
1 2 -.04
1 3 .05
```

MATRIX = YES is the matrix format

```
1.0000 -.0451 .0447 .0095 .....
-.0451 1.0000 -.0448 -.2024 .....
.0447 -.0448 1.0000 -.0437 .....
.....
```

11.95 MAXPAGE= the maximum number of lines per page

For no page breaks inside Tables, leave MAXPAGE=0

If you prefer a different number of lines per page of your output file, to better match the requirements of your printer or word processor, give that value (see [Using a Word Processor or Text Editor](#)). If you prefer to have no page breaks, and all plots at their maximum size, leave MAXPAGE=0.

On [Table 1](#) and similar Tables, MAXPAGE= controls the length of the Table.

Example: You plan to print your output file on standard paper with 60 lines per page (pages are 11 inches long, less 1 inch for top and bottom margins, at 6 lpi):

```
MAXPAGE=60 (set 60 lines per page)
FORMFD=^ (standard: Word Processor form feed)
```

11.96 MFORMS= reformat input data

MFORMS= supports the reformatting of input data records, and also [equating](#) multiple input files in different formats, such as alternate forms of the same test. Data after [END NAMES](#) or END LABELS is processed first, as is data specified by [DATA=](#) in the core control file.

MFORMS= file name	file containing details
MFORMS= file name + file name + ...	multiple file names
MFORMS = *	in-line list
MFORMS = ?	opens a Browser window to find the file

Question: I've been trying to analyze two test forms, sharing 1-50 as common and 51-70 as unique. The common items also have [IWEIGHT=](#). I am trying to estimate the difficulties for each form. I want to retain the [KEY1=](#) so I can have option information (so using [CODES=ABCD](#)), but also enable test form comparisons. I've tried building the combined control and data file using MFORMS=, but no success yet. Any advice would be great.

Reply: The items become

- 1-50 common
- 51-70 form 1
- 71-90 form2

and the Item labels, KEY1= and IWEIGHT= match these numbers

Combine the data files by

- a) copying your form 1 data into the combined file followed by
- b) items 1-50 of your form 2 data
- c) and then items 51-70 of your form2 data into the position for items 71-90 of the combined data.

To do this copying, please use "rectangular copying" (alt+Mouse). This is available in Microsoft Word and also the freeware NotePad++ <http://notepad-plus-plus.org/>

Data reformatted by MFORMS= can be accessed, viewed, edited and "saved as" permanently using the "Edit" pull-down menu. It has a file name of the form: ZMF.....txt

Procedure:

1. Make a list of all your different items. Usually we list the common items first.
2. For each item, list its entry number (sequence number) on each form.
3. Use MFORMS=. Transform the list of item numbers into MFORMS= instructions.

Here is the layout:

MFORMS=*	
DATA=forma.txt	; the name of an input data file
L=2	; there are 2 lines in input data file for each data record
I1 = 20	; response to item 1 of the combined test is in column 20 of the input data file
I3-5 = 21	; items 3, 4, 5 are in columns 21, 22, 23 of the input data file
I16-20=11	; items 16, 17,18, 19, 20 are in columns 11, 12, 13, 14, 15
P1=9	; the first character of person label is in column 9 of the input data file
P3-8=1	; person label characters 3 through 8 start in column 1
C20-24="FORMA"	; put in columns 20-24 the letters FORMA
C40-90 = 2:1	; put in columns 40-90 the characters in the second line of the data record
#	; end of definition - start of next file reformat
DATA=formb.txt	; name of input data file
P3-7=1	; information for columns 3-7 of person label starts in column 1 of data record
.....	
*	; end of mforms= command

Details:	
MFORMS=* or filename	* instructions follow in control file, and end with another * or instructions are in a file.
data=filename	name of input file name to be reformatted. The reformatted records are placed in a temporary work file. This may be accessed from the Edit pull-down menu, and saved into a permanent file. This temporary file is processed after any Data Files specified with the master data= instruction and in the same way, e.g., any FORMAT= command will be applied also to the temporary work file.
L=nnn	nnn is the count of lines in each input data record. If L=1 this can be omitted. L=4 means that 4 input data lines are processed for each data record output.
Cnnn= column in original data record	nnn is the column number in the formatted data record. XWIDE= does not apply. C10-12 means columns 10, 11, 12 of the formatted record. C1= refers to column 1 of the formatted data record. This can also be used to move item and person information.
Innn= column in original data record	nnn is the starting item number in the formatted data record. nnn-mmm are the starting to ending item numbers in the formatted data record. XWIDE= is applied, so that I3-5= with XWIDE=2 means 6 characters. I1= points to column Item1= in the formatted data record.
Pnnn= column in original data record	nnn is the starting column number in the person label in the formatted person label.

	XWIDE= is not applied. P6-8= always means 3 columns starting in column 6. P1= points to column Name1= in the formatted data record.
....=nnn	nnn is the starting column of the only, or the first, line in the input data record.
...=m:nnn	m is the line number in each data record nnn is the starting column number of that line
...="xxx"	xxx is a character constant to be placed in the formatted data record. Note: for I18-20="abc" with XWIDE=2, then response to Item 18 is "ab", 19 is "c ", 20 is " ".
#	end of processing of one file, start of the next
*	end of Mforms= processing

Check that everything is correct, by looking at the [Edit menu](#), MFORMS= file.

Example 1: See [Exam10c.txt](#)

Example 2: Three data files with common items and one MCQ scoring key. See below for constructing a combined scoring key.

Item bank	Datafile1.txt	Datafile2.txt	Datafile3.txt
1-3	1-3	1-3	1-3
4-6	4-6	-	-
7-9	-	4-6	-
10-12	-	-	4-6

Datafile1.txt: (Items 1-6)

TOMY ABCDAB
BILL BCDADD

Datafile2.txt (Items 1-3 and 7-9)

TOTO BBADAB
MOULA BADADD

Datafile3.txt (Items 1-3 and 10-12)

IHSANI ACCDAB
MALIK CBDDCD

Control file:

```
TITLE="Multiple MCQ forms with one scoring key"
NI=12 ; 12 ITEMS IN TOTAL
ITEM1=11
NAME1=1
CODES="ABCD"
KEY1=BACCADACADDA
mforms=*
data=datafile1.txt ; name of data file
L=1 ; one line per person
P1-10=1 ; person label in columns 1-10
I1-3=11 ; items 1-3 in columns 11-13
I4-6=14 ; items 4-6 in columns 14-16
#
data=datafile2.txt
L=1
P1-10=1
I1-3=11
I7-9=14 ; items 8-9 in columns 14-16
#
data=datafile3.txt
```

```

L=1
P1-10=1
I1-3=11
I10-12=14 ; items 10-12 in columns 14-16
*
&END
; item identification here
END NAMES

```

Here is how the data appear to Winsteps for analysis:

```

TOMY      ABCDAB
BILL      BCDADD
TOTO      BBA   DAB
MOULA     BAD   ADD
IHSANI    ACC   DAB
MALIK     CBD   DCD

```

Example 3: Test 1 is a 4-item survey. Test 2 is a 4-item survey its first two items are the last two items on Test 1. Those items share the same rating scale, called "C". The other 4 items also share another rating scale, called "U".

Item bank	Datafile1.txt	Datafile2.txt
1-2	1-2	-
3-4	3-4	1-2
5-6	-	3-4

Datafile1.txt: (Items 1-4)

```

TOMY      3241
BILL      3352

```

Datafile2.txt (Items 3-4 and 5-6)

```

TOTO      3325
MOULA     2143

```

TITLE="Combing two surveys"

NI=6 ; 6 ITEMS IN TOTAL

ITEM1=11

NAME1=1

CODES="12345" ; all valid codes on all the survey items

ISGROUPS = UUUUUU ; assigns items to rating scales

mforms=*

data=datafile1.txt ; name of data file

L=1 ; one line per person

P1-10=1 ; person label in columns 1-10

I1-4=11 ; items 1-4 in columns 11-14

#

data=datafile2.txt

L=1

P1-10=1

I1-2=13 ; items 1-2 on the second survey are items 3 and 4 on the first survey

I3-4=15 ; items 3-4 in columns 15-16

#

&END

; item identification here

END NAMES

Here is how the data appear to Winsteps for analysis:

```

TOMY      3241 ; items 1-2 unique, items 3-4 common
BILL      3352
TOTO      3325 ; items 3-4 common, items 5-6 unique
MOULA     2143

```

Example 4: Test 1 is a 4-item survey. Test 2 is a 4-item survey with two items in common with Test 1 which are to be anchored to their Test 1 values.

Item bank	data1.txt	data2.txt
1	1	-
2	2	3
3	3	-
4	4	2
5	-	1
6	-	4

Test 1 has 4 rating scale items. Each item has its own partial-credit structure:

```

title = "Test 1"
item1 = 1 ; items start in column 1
ni = 4 ; 4 items
name1 = 5 ; person label starts in column 5
namlen = 14 ; length of person name
codes = 01234 ; rating scale
ISGROUPS = 0 ; each item has its own rating scale structure
stkeep = YES ; this is probably what you want for these type of data
data = data1.txt
ifile = itemslif.txt ; item calibrations from Test 1 for Test 2 (output)
sfile = itemslsf.txt ; structure calibrations from Test 1 for Test 2 (output)
&END
Test 1 item 1
Test 1 item 2
Test 1 item 3
Test 1 item 4
END NAMES

```

data1.txt is:

```

1234Person 1-1
3212Person 1-2
.....

```

Test 2 has 4 items. 1 and 4 are new - we will call these items 5 and 6 of the combined Test 1 and 2. Item 2 is Test 1 item 4, and item 3 is Test 1 item 2.

```

title = "Test 2 (formatted to match Test 1)"
item1 = 1 ; items start in column 1
ni = 6 ; 4 items in Test 1 + 2 more in Test 2
name1 = 7 ; person label starts in column 7
namlen = 14 ; length of person name
codes = 01234 ; rating scale
stkeep = YES ; this is probably what you want for these type of data
ISGROUPS = 0 ; each item has its own rating scale structure
iafile = itemslif.txt ; item calibrations from Test 1 (input - unchanged)
safile = itemslsf.txt ; structure calibrations from Test 1 (input - unchanged)
MFORMS = * ; reformat the Test 2 data to align with Test 1
data = data2.txt ; the name of an input data file
L = 1 ; there is 1 line in input data file for each data record
I2 = 3 ; response to item 2 of Test 1 is in 3 of the data2.txt file
I4 = 2 ; response to item 4 of Test 1 is in 2 of the data2.txt file
I5 = 1 ; item 5 is in column 1 of data2.txt
I6 = 4 ; item 6 is in column 4 of data2.txt
P1-14 = 5 ; the first character of person label is in column 5 of data2.txt for 14 columns.
* ; end of mforms= command
&END
Test 1 item 1 (blank in Test 2)
Test 1 item 2 (Test 2 item 3)
Test 1 item 3 (blank in Test 2)

```

```

Test 1 item 4 (Test 2 item 2)
Item 5 (not in Test 1, Test 2 item 1)
Item 6 (not in Test 1, Test 2 item 4)
END NAMES

```

data2.txt is:

```

5426Person 2-1
1234Person 2-2
....

```

The formatted file (see Edit pull-down menu MFORMS=) is

```

2 456Person 2-1
3 214Person 2-2
....

```

Example 5. For my computer-adaptive test (CAT), I want a new MFORMS= specification for each respondent.

In your Winsteps control file, put
MFORMS=myrespondents.txt (or any other file name)

Then myrespondents.txt can contain an unlimited number of MFORMS instructions

```

DATA=respondent1.txt
....
#
DATA=respondent2.txt
....
#
DATA=respondent3.txt
....
#
.....
.....
#

```

Example 6: Constructing a combined scoring key.

1. Construct MFORMS= as above.
2. Put each data files scoring key in the format of that data file in a separate data file.
3. In your control file, include:

CODES= list of valid data codes, e.g., CODES=ABCDE

IREFER=*

1-number of items A ; all items will be scored by IVALUEA=

*

IVALUEA=10000

IVALUEB=01000

IVALUEC=00100

IVALUED=00010

IVALUEE=00001

Run your control file with MFORMS=, using the data files containing only the keys.

Look at Edit menu, MFORMS=. It will have a data record for each files key, in the combined file format.

Now, Output Files menu, XFILE=,

clear all settings

check: item entry number

response code in data file

OK

Output to text file space separated.

You will see something like:

```

; RESIDUAL FILE FOR YOUR TEST
ITEM C

```


- 11 A
- 12 B
- 13 E
- 14 A

Delete the two heading lines.
The other lines are the scoring key.

In IREFER=, replace
1-number of items A
with all these lines, or you can also put them in a separate file,, and specify IREFER=theirfilename

IREFER is acting like a key, by referring each observation to the IVALUE which matches its scoring key

Now run your Winsteps control file with the correct data files (delete the scoring keys in the data files).

Winsteps should produce the correct output.

11.97 MHSlice= Mantel-Haenszel slice width

Differential item functioning (DIF) is investigated using log-odds estimators in [Table 30.1Mantel-Haenszel](#) (1959) for dichotomies or [Mantel](#) (1963) for polytomies. The sample is divided into difference classification groups (also called reference groups and focal groups) which are shown in [Table 30](#) and specified with [DIF=](#).

MHSlice= specifies the width of the slice (in logits) of the latent variable be included in each cross-tab. The lower end of the lowest slice is always the lowest observed person measure.

MHSlice = 0 bypasses Mantel-Haenszel or Mantel computation.

MHSlice = .1 logits and smaller. The latent variable is stratified into thin slices. This corresponds to the slicing by raw scores with complete data

MHSlice = 1 logit and larger. The latent variable is stratified into thick slices.

For each slice, a cross-tabulation is constructed for each pair of person classifications against each scored response level. An odds-ratio is computed from the cross-tab. Zero and infinite ratios are ignored. A homogeneity chi-square is also computed when possible.

Example: The sample size is small, but an approximate Mantel-Haenszel estimate is wanted:
MHSlice=3 ; this probably divides the sample into 3 strata.

11.98 MHZERO= adjustment to zero cells in Mantel cross-tabulation

The Mantel-Haenszel and Mantel DIF statistics are based on 2x2 cross-tabulations of response frequencies for the Reference and Focal groups.

	Low Category	High Category
Focal Group	FL	FH
Reference Group	RL	RH

If all 4 cells of the cross-tab are non-zero, then the computation proceeds.
If FL+RL = 0 and/or FH+RH=0 then the computation is skipped.
Otherwise FL = maximum of FL + MHZERO and similarly for the other 4 cells, and the computation proceeds.

The default is MHZERO = 0, no adjustment.
SAS uses MHZERO = 0.5

Example: with [Exam1.txt](#)

MHZERO=0

KID	KID	DIF	JOINT	Rasch-Welch		Mantel-Haenszel		Size	Active	TAP		
CLASS	CLASS	CONTRAST	S.E.	t	d.f.	Prob.	Chi-squ	Prob.	CUMLOR	Slices	Number	Name
F	M	1.23	2.11	.58	21	.5677	.0000	1.000		5	4	1-3-4
F	M	1.92	2.06	.93	20	.3626				5	5	2-1-4
F	M	-1.22	1.28	-.95	31	.3484	.6901	.4061		5	6	3-4-1
F	M	-2.20	1.46	-1.51	31	.1414	1.0057	.3159		5	7	1-4-3-2
F	M	-1.86	1.09	-1.71	31	.0982	.6901	.4061		5	8	1-4-2-3

MHZERO=0.5

KID	KID	DIF	JOINT	Rasch-Welch		Mantel-Haenszel		Size	Active	TAP		
CLASS	CLASS	CONTRAST	S.E.	t	d.f.	Prob.	Chi-squ	Prob.	CUMLOR	Slices	Number	Name
F	M	1.23	2.11	.58	21	.5677	.2222	.6374	1.39	10	4	1-3-4
F	M	1.92	2.06	.93	20	.3626	.6000	.4386	-.69	10	5	2-1-4
F	M	-1.22	1.28	-.95	31	.3484	.0360	.8496	-1.30	10	6	3-4-1
F	M	-2.20	1.46	-1.51	31	.1414	.4792	.4888	-2.23	10	7	1-4-3-2
F	M	-1.86	1.09	-1.71	31	.0982	.0360	.8496	-1.30	10	8	1-4-2-3

Notice that CUMLOR, the cumulative log-odds ratio, the size of the DIF according to Mantel-Haenszel/Mantel, cannot be computed for these data with MHZERO=0, and generally approximates the Rasch DIF CONTRAST when MHZERO=0.5

If Rasch and Mantel-Haenszel/Mantel disagree, then there is probably not enough data to draw strong conclusions.

11.99 MISSCORE= scoring of missing data codes

This is NOT the missing-value code in your data. All codes NOT in [CODES=](#) are missing value codes. Use this control specification when you want missing data to be treated as valid responses. Winsteps and Missing Data: No Problem!

Winsteps processes one observation at a time. For each observation, X_{ni} by person n on item i , it computes an expectation Eni , based on the current person measure estimate B_n and the current item measure Di and, if relevant, the current rating (or partial credit) scale structure (calibrations) $\{F_k\}$. P_{nik} is the probability of observing category k for person n on item i .

In this computation it skips over, omits, ignores "missing" data.

It then compares $\sum(X_{ni})$ with $\sum(Eni)$ for each person n , and adjusts B_n .

It then compares $\sum(X_{ni})$ with $\sum(Eni)$ for each item i , and adjusts Di

It then compares the count of $(X_{ni}=k)$ with the sum (P_{nik}) for each k , and adjusts F_k

These sums and counts are only over the observed data. There is no need to impute missing data.

There are no pairwise, listwise or casewise deletions associated with missing data.

MISSCORE= says what to do with characters that are not valid response codes, e.g. blanks and data entry errors. Usually any characters not in [CODES=](#) are treated as missing data, and assigned a value of -1 which mean "ignore this response." This is usually what you want when such responses mean "not administered". If they mean "I don't know the answer", you may wish to assign missing data a value of 0 meaning "wrong", or, on a typical attitude survey, 3, meaning "neutral" or "don't know".

MISSING=0 is the same as MISSCORE=0 meaning that all codes in the data not listed in [CODES=](#) are to be scored 0.

Non-numeric codes included in [CODES=](#) (without rescaling/recoding) or in [NEWSCORE=](#) or [IVALUE=](#) are always assigned a value of "not administered", -1.

Example 0a: In my data file, missing data are entered as 9. I want to score them 0, wrong answers. Valid codes are 0 and 1.

```
CODES = 01 ; do not specify a 9 as valid
MISSCORE = 0 ; specifies that all codes not listed in CODES=, e.g., 9's. are to be scored 0.
```

Example 0b: In my data file, missing data are entered as 9. I want to ignore them in my analysis. Valid codes are 0 and 1.

```
CODES = 01 do not specify a 9 as valid
; the following line is the standard, it can be omitted.
MISSCORE = -1 specifies that all codes not listed in CODES=, e.g., 9's.
are to be treated as "not administered"
```

Example 1: Assign a code of "0" to any responses not in CODES=

```
MISSCORE=0 missing responses are scored 0.
```

Example 2: In an attitude rating scale with three categories (0, 1, 2), you want to assign a middle code of "1" to missing values

```
MISSCORE=1 missing responses scored 1
```

Example 3: You want blanks to be treated as "wrong" answers, but other unwanted codes to be ignored items, on a questionnaire with responses "Y" and "N".

```
CODES="YN " blank included as valid response
NEWSCORE=100 new response values
RESCORE=2 rescore all items
MISSCORE=-1 ignore missing responses (standard)
```

Example 4: Your optical scanner outputs an "@" if two bubbles are marked for the same response. You want to ignore these for the analysis, but you also want to treat blanks as wrong answers:

```
CODES ="1234 " blank is the fifth valid code
KEY1 =31432432143142314324 correct answers
MISSCORE=-1 applies to @ (standard)
```

Example 5: Unexpected codes are scored "wrong", but 2's to mean "not administered".

```
CODES = 012
NEWSCORE= 01X ; X is non-numeric, matching 2's ignored
MISSCORE= 0 ; all non-CODES= responses scored 0
```

Example 6: You have a long 4-option MCQ test with data codes ABCD. Most students do not have the time to complete all the items. This requires a two-stage item analysis:

1. Estimate the item difficulties without allowing missing data to bias them

Missing data = -1 ; not administered. Perform the analysis estimate the item difficulties. Save them with IFILE=if.txt

2. Estimate the person measures lowering the estimates if there is missing data

Missing data = 0 ; incorrect. Anchor the item difficulties with IAFILE=if.txt to estimate the person measures and report them.

So, in this Example:

Stage 1. *Item calibration:*

Deliberately skipped responses are coded "S" and scored incorrect. The student could not answer the question. Not-reached items are coded "R" and scored "not administered". This prevents easy items at the end of the test being calibrated as "very difficult".

```
CODES="ABCD S"
KEY1="CDBAD....."
MISSCORE=-1
IFILE=ITEMCAL.TXT ; write out the item calibrations
```

Stage 2. *Person measurement:*

The convention with MCQ tests is that all missing responses are scored incorrect when measuring the persons.

```
IAFILE=ITEMCAL.TXT ; anchor on the Stage 1 item calibrations
CODES="ABCD S"
KEY1="CDBAD....."
MISSCORE=0 ; all missing data are scored incorrect
```

11.100 MJMLE= or MUCON= maximum number of JMLE iterations

[JMLE](#) iterations may take a long time for big data sets, so initially set this to -1 for no JMLE iterations. Then set MJMLE= to 10 or 15 until you know that more precise measures will be useful. The number of [PROX](#) iterations, [MPROX=](#), affects the number of JMLE iterations but does not affect the final estimates.

MJMLE= (or MUCON=) specifies the maximum number of JMLE iterations to be performed. Iteration will always cease when both [LCONV=](#) and [RCONV=](#) criteria have been met, see [CONVERGE=](#). To specify no maximum number limitation, set MJMLE=0. Iteration always be stopped by Ctrl with F, see ["Stopping Winsteps"](#).

Example 1: To allow up to 4 iterations in order to obtain rough estimates of a complex rating (or partial credit) scale:
MJMLE=4 ; 4 JMLE iterations maximum

Example 2: To allow up to as many iterations as needed to meet the other convergence criteria:
MJMLE=0 ; Unlimited JMLE iterations

Example 3: Perform no JMLE iterations, since the PROX estimates are good enough.
MJMLE=-1 ; No JMLE iteration

Example 4: Run as quick estimation as possible to check out control options.
MPROX=-1 ; minimal PROX estimation iterations
MJMLE=-1 ; no JMLE iterations

Example 5: To perform an exact number of JMLE iterations.
a) set the convergence criteria very tight, so that they are never satisfied:
CONVERGE=B
RCONV=.00001
LCONV=.00001
b) set the maximum number of JMLE iterations equal to the desired number:
MJMLE = 99 (or whatever)

11.101 MNSQ= show mean-square or standardized fit statistics

The mean-square or t standardized fit statistics are shown in [Tables 7, 11](#) to quantify the unexpectedness in the response strings, and in [Tables 4, 5, 8, 9](#) for the fit plots.

MNSQ=N Show standardized (ZSTD) fit statistics. ZSTD (standardized as a z-score) is used of a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's [t-statistic](#) distribution value has been adjusted to a unit normal value.

MNSQ=Y Show mean-square fit statistics. Use [LOCAL=L](#) for log scaling.

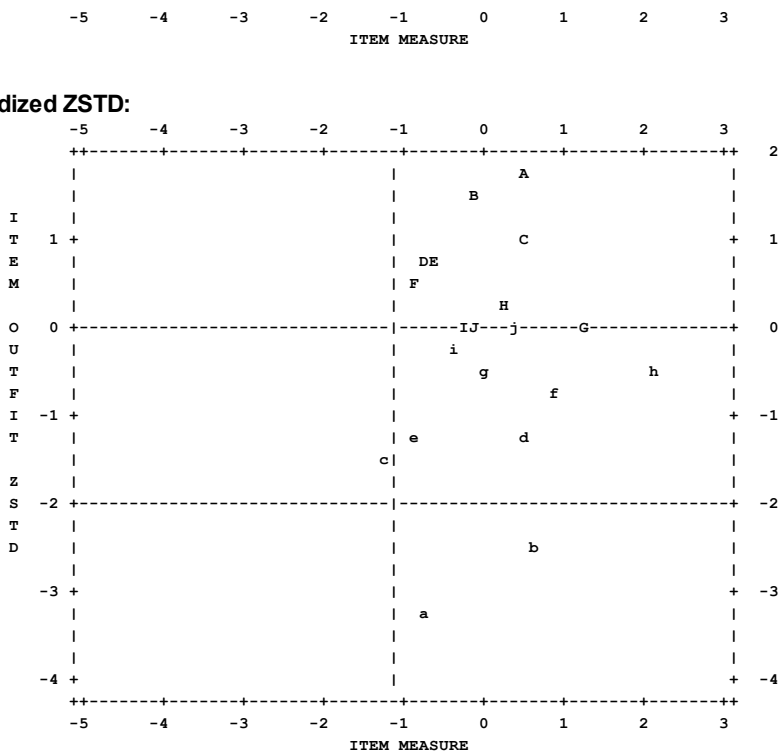
TABLE 7.1 TABLE OF POORLY FITTING PERSONS (ITEMS IN ENTRY ORDER)

NUMBER	NAME	POSITION	MEASURE	INFIT (MNSQ)	OUTFIT
17	Rod	M	-1.41	2.4	A 2.2
RESPONSE:	1:	0 0 2 4 1	4 3 1 3 3	1 4 3 2 3	3 1 4 2 1
Z-RESIDUAL:		-2 2		-2	-2 2 -2

Mean-square:

TABLE 9.1

	-5	-4	-3	-2	-1	0	1	2	3
2 +	+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+								
I									
T									
E						B	A		
M						FDE	C		
1 +	+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+								
O						e	d f		
U						c	b		
T						a		h	
F									
I 0 +	+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+								
T									



11.102 MODELS= assigns model types to items

This command is for experimental use only. Please do not use in production situations.

Winsteps chooses the model-family based on MODELS=

Models=R (or MODELS= is omitted) is the rating-scale family of models which includes the Andrich Rating-Scale Model, the Masters Partial-Credit Model, the Grouped Rating-Scale Model and the Rasch Dichotomous Model.

MODELS= *file name	file containing details
MODELS = *	in-line list
MODELS = codes	codes for item groups

Within Models=R, the choice of model is decided by [ISGROUPS=](#)

If ISGROUPS= is omitted, then

- 1) If there are only two categories in the data the dichotomous model
- 2) If there are more than two categories the Andrich Rating-Scale Model

If ISGROUPS=0 then

the Masters Partial Credit Model

If ISGROUPS=(something else)

the Grouped Rating-Scale Model

Please look at the output in [Table 3.2](#), 3.3, to see exactly how the data have been modeled.

Winsteps estimates calibrations for four different ordered response category structures. Dichotomies are always analyzed using the Rasch dichotomous model, regardless of what model is specified for polytomies. Items are assigned to the model for which the serial location in the MODELS= string matches the item sequence number.

The item grouping default becomes the "Partial Credit Model" each item with its own rating scale: MODEL=R and [ISGROUPS=0](#).

Do not specify MODELS= unless you intend to use the "S" or "F" models.

MODELS=R (standard) is the default option, specifying standard Rasch analyses using the Rasch dichotomous model, Andrich "Rating Scale" model and Masters' "Partial Credit" model, see [ISGROUPS=](#).

Andrich D. (1978) A rating scale formulation for ordered response categories. *Psychometrika*, 43, 561-573.

Masters G.N. (1982) A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.

MODELS=S uses the Rasch dichotomous model and the Glas-Verhelst "Success" (growth) model, also called the "Steps" Model (Verhelst, Glas, de Vries, 1997). If and only if the person succeeds on the first category, another category is offered until the person fails, or the categories are exhausted, e.g. an arithmetic item, on which a person is first rated on success on addition, then, if successful, on multiplication, then, if successful, on division etc. "Scaffolded" items can function this way. This is a continuation ratio model parameterized as a Rasch model with missing data on unreached categories.

Verhelst N.D., Glas C.A.W. & De Vries H.H. (1997) A Steps model to analyze partial credit. In W.J. van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 123 - 138) New York: Springer.

Recommendation: Instead of an S-type polytomy, model each step to be a dichotomy: 1=succeeded, 0=failed, missing=not reached. This facilitates much more powerful diagnosis of the functioning of the success process.

MODELS=F uses the Rasch dichotomous model and the Linacre "Failure" (mastery) model. If a person succeeds on the first category, top rating is given and no further categories are offered. On failure, the next lower category is administered until success is achieved, or categories are exhausted. This is a continuation ratio model parameterized as a Rasch model with missing data on unreached categories. The Success and Failure model computations were revised at Winsteps version 3.36, August 2002.

Recommendation: Instead of an F-type polytomy, model each step to be a dichotomy: 1=succeeded, 0=failed, missing=not reached. This facilitates much more powerful diagnosis of the functioning of the failure process.

MODELS= has three forms: MODELS=RRSSFR and MODELS=* list * and MODELS=*filename. When only one letter is specified with MODELS=, e.g., MODELS=R, all items are analyzed using that model. Otherwise MODELS=some combination of R's, F's, S's, and G's, e.g., MODELS=RRSF

When XWIDE=2 or more, then

either (a) Use one character per XWIDE and blanks,

```
NI=8
```

```
XWIDE=2
```

```
MODELS=' R S R F R S R R ' ; this also forces ISGROUPS=0 to be the default
```

or (b) Use one character per item with no blanks

```
NI=8
```

```
XWIDE=2
```

```
RESCORE='RSRFRSRR' ; this also forces ISGROUPS=0 to be the default
```

Example 1: All items are to be modeled with the "Success" model.

```
MODELS=S ; the Success model
```

Example 2: A competency test consists of 3 success items followed by 2 failure items and then 10 dichotomies. The dichotomies are to be reported as one grouping.

```
NI=15 fifteen items
```

```
MODELS=SSSFRRRRRRRRRR ; matching models: ; forces ISGROUPS=0 to be the default
```

```
ISGROUPS=0000011111111111 ; dichotomies grouped: overriding the default ISGROUPS=0
```

or

```
MODELS=*
```

```
1-3 S
```

```
4 F
```

```
5 F
```

```
6-15 R
```

```
*
```

11.103 MODFROM= location of MODELS

This command has not proved productive. It is maintained for backwards compatibility.

Instructs where to find the MODELS= information.

Only use this if you have too many items to put conveniently on one line of the [MODELS=](#) control variable. It is easier to use "+" [continuation lines](#)

MODFROM=N MODELS= is a control variable before &END (the standard).

MODFROM=Y

MODELS= information follows just after &END but before the item names. It is formatted exactly like a data record. It is helpful to enter "MODELS=" where the person name would go.

Example: A test consists of 10 three-category items. The highest level answer is scored with KEY2=. The next level with KEY1=. Some items have the "Success" structure, where the higher level is administered only after success has been achieved on the lower level. Some items have the "Failure" structure, where the lower level is administered only after failure at the higher level. The MODELS=, KEY1=, KEY2= are formatted exactly like data records. The data records are in a separate file.

```

NAME1 = 5  start of person-id
ITEM1 = 20 start of responses
NI = 10  ten items
CODES = ABCDE valid codes
MODFRM = Y  MODELS= in data format
KEYFRM = 2  two keys in data format
DATA = DATAFILE location of data
;          1          2 columns
;2345678901234567890
&END
MODELS=          SSSFFFSSSS  data format
KEY1=            BCDABCDABC  starts in column ITEM1 = 20
KEY2=            ABCDDBCBAA
Item name 1 first item name
|
Item name 10
END NAMES

```

11.104 MORE= heading in Tables 1, 12, 16

[MORE=](#), [LESS=](#), [RARE=](#), [FREQ=](#) control the column headings in [Tables 1, 12](#) and [16](#). The first 8 characters of each variable are displayed, and sometimes more. Please change these headings to match the meaning of more and less of your latent variable.

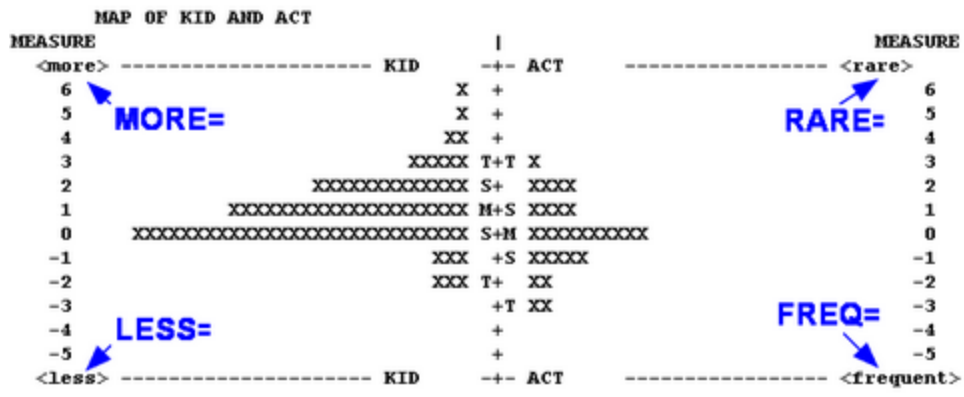
<i>The default values are:</i>	<i>Meaning relative to the latent variable</i>
MORE = <more>	More ability, higher ability
LESS = <less>	Less ability, lower ability
RARE= <rare>	Rarer success, higher item difficulty, harder item
FREQ= <frequent>	More frequent success, lower item difficulty, easier item

Example 1: Table 1.1. Default values:

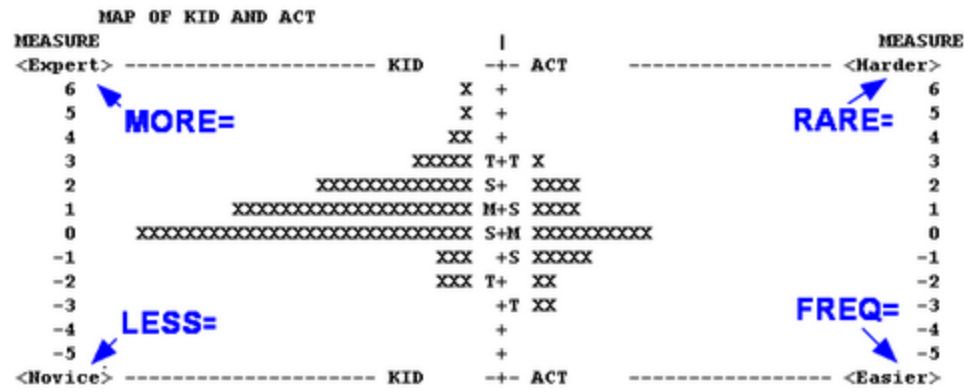
[PERSON=](#) KID

[ITEM=](#) ACT

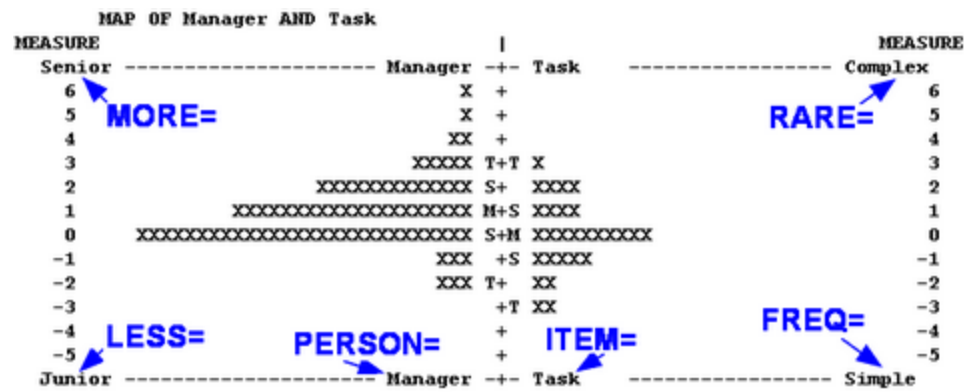
[MAXPAGE=](#) 40



Example 2:
 MORE= <Expert>
 LESS= <Novice>
 RARE= <Harder>
 FREQ= <Easier>
 PERSON= KID
 ITEM= ACT
 MAXPAGE= 40



Example 3:
 MORE= Senior
 LESS= Junior
 RARE= Complex
 FREQ= Simple
 PERSON= Manager
 ITEM= Task
 MAXPAGE= 40



11.105 MPROX= maximum number of PROX iterations

Specifies the maximum number of [PROX](#) iterations to be performed. The PROX, "normal approximation", algorithm provides starting values for the JMLE algorithm, in order to speed up estimation. PROX has little influence on the final JMLE estimates.

MPROX= 1 or more	PROX iterations are performed so long as inestimable parameters have been detected in the previous iteration, because inestimable parameters are always dropped before the next iteration. At least 2 PROX iterations will be performed. PROX iteration ceases when the specified number of iterations have been done or the spread of the persons and items no longer increases noticeably (0.5 logits). The spread is the logit distance between the top 5 and the bottom 5 persons or items. Then JMLE iterations begin.
MPROX= 0	Continue PROX iterations indefinitely until you intervene with Ctrl+f
MPROX= -1 (or there are anchored items or persons)	No PROX estimation is done

Example 1: To set the maximum number of PROX iterations to 20, in order to speed up the final JMLE estimation of a symmetrically-distributed set of parameters,
MPROX=20

Example 2: To minimize the influence of PROX on the final JMLE estimates,
MPROX=-1

11.106 NAME1= first column of person label

NAME1= gives the column position where the person label information starts in your data file or in the new record formatted by [FORMAT=](#).

It is easy to miscount the NAME1= column. Scroll to the top of the Winsteps screen and check column positions:

```

Input in process..
Input Data Record:
      1      2
1234567890123456789012345678
Richard M 111111100000000000
^P      ^I      ^N
      35 KID      Records Input.

```

^P marks the Name1=1 column position with ^.

^I marks the Item1=11 column position with ^.

^N marks the NI=18 column position with ^.

Example 1: The person-id starts in column 10, data responses are 1 column wide, in columns 1-8:

```

NAME1=10 starting column of person-id
XWIDE=1 width of response
NI=8 number of responses

```

Example 2: The person-id in column 10, there are 4 data responses are 2 columns wide, in columns 1-8:

```

NAME1=10 starting column of person-id
XWIDE=2 width of response
NI=4 number of responses

```

Example 3: The person id starts in column 23 of the second record.

`FORMAT=(80A,/,80A)` concatenate two 80 character records
`NAME1=103` starts in column 103 of combined record

Example 4: The person id starts in column 27 of a record with `XWIDE=2` and `FORMAT=`.

This becomes complicated, see [FORMAT=](#)

Example 5: How do I add convenient IDs (e.g., 1, 2, 3...) to the control file?

Give `NAME1=` a value beyond the end of the data record.

`NAME1=100` ; a number that is bigger than the longest data record
`NAMELENGTH = 5` ; allows numbers up to 99999

11.107 MRANGE= half-range of measures on plots

Specifies the measure half-range, (i.e., range away from the origin or [UMEAN=](#)), of the maps, plots and graphs. This is in logits, unless [USCALE=](#) is specified, in which case it must be specified in the new units defined by `USCALE=`. To customize particular tables, use the [Specification pull-down menu](#), or see [TFILE=](#).

Example 1: You want the vertical range on the [Table 1](#) map to be -5 to +5

`MRANGE=5`

Example 1: You want to see the category probability curves in the range -3 to +3 logits:

`MRANGE=3`

Example 2: With `UMEAN=500` and `USCALE=100`, you want the category probability curves to range from 250 to 750:

`UMEAN=500` ; item mean calibration

`USCALE=100` ; value of 1 logit

`MRANGE=250` ; to be plotted each way from `UMEAN=`

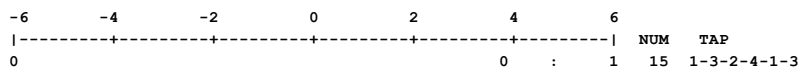
11.108 MTICK= measure distance between tick marks on plots = 0, auto-size

Specifies the measurement distance between tick marks in Tables [1](#), [2](#), [3.2](#), [4](#), [5](#), [8](#), [9](#), [12](#), [16](#), and more. This is in logits, unless [USCALE=](#) is specified, in which case it must be specified in the new units defined by `USCALE=`. To customize particular tables, use the [Specification pull-down menu](#).

Example 1: You want the vertical range on the [Table 2.2](#) map to be -6 to +6 logits with tick marks every 2 logits:

`MRANGE=6`

`MTICK=2`



11.109 MUCON= or MJMLE= maximum number of JMLE iterations

See [MJMLE=](#)

11.110 NAMLEN= length of person label

Use this if too little or too much person-id information is printed in your output tables.

`NAMLEN=` allows you define the length of the person-id name with a value in the range of 1 to 30 characters. This value overrides the value obtained according to the rules which are used to calculate the length of the person-id. These rules are:

- 1) Maximum person-id length is 300 characters
- 2) Person-id starts at column [NAME1=](#)
- 3) Person-id ends at [ITEM1=](#) or end of data record.
- 4) If `NAME1=` equals `ITEM1=` then length is 30 characters.

Example 1: The 9 characters including and following NAME1= are the person's Social Security number, and are to be used as the person-id.

NAMLEN=9

Example 2: We want to show the responses in Example0.txt as the person label to help diagnose the fit statistics:

ITEM1 = 1
 NI = 25
 NAME1 = 1
 NAMLEN = 25

```

+-----+
|ENTRY  RAW      MODEL|  INFIT  |  OUTFIT  |PTMEA|
|NUMBER SCORE  COUNT MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR. | KID
+-----+
|   12   17    25   -.85    .36|1.90   2.7|2.86   3.9|A .24| 0100220101210011021000101|
|    6   24    25   -.02    .34|1.48   1.7|1.83   2.3|B .30| 1011211011121010122101210|
|   15   27    25    .31    .33| .86   -.5|1.72   1.9|C .10| 1111111111111111112211111|
|    7   44    25   2.62    .47|1.71   1.7|1.00   .3|D .41| 22200222022222222222222|
|   14   23    25   -.14    .34|1.62   2.1|1.53   1.6|E .54| 2110020212022100022000120|

```

11.111 NAMLMP= name length on map for Tables 1, 12, 16

Person and item labels are often too long to display nicely on the maps in Tables 1, 12, 16. NAMLMP= (name length on maps) limits the number of characters of item and person labels displayed.

NAMLMP=0 to ignore this instruction.

For item labels, NAMLMP= is ignored when IMAP= is specified.

For person labels, NAMLMP= is ignored when PMAP= is specified.

Example: The 9 characters of the person label are the person's Social Security number, and are to be displayed on the person maps.

NAME1 = 23 ; start of person label

NAMLMP=9 ; number of characters of person label to show on the person maps. The item labels will also be truncated to display the first 9 characters.

11.112 NEWSCORE= recoding values

NEWSCORE= says which values must replace the original codes when RESCORE= is used. If XWIDE=1 (the standard), use one column per code. If XWIDE=2, use two columns per code. The length of the NEWSCORE= string must match the length of the CODES= string. For examples, see RESCORE=. NEWSCORE= is ignored when KEYn= is specified.

The responses in your data file may not be coded as you desire. The responses to some or all of the items can be rescored or keyed using RESCORE=. RESCORE= and NEWSCORE= are ignored when KEYn= is specified, except as below.

RESCORE=" " or 2 or is omitted

All items are recoded using NEWSCORE=. RESCORE=2 is the standard when NEWSCORE= is specified.

RESCORE= some combination of 1's and 0's

Only items corresponding to 1's are recoded with NEWSCORE= or scored with KEYn=. When KEYn= is specified, NEWSCORE= is ignored.

If some, but not all, items are to be recoded or keyed, assign a character string to RESCORE= in which "1" means "recode (key) the item", and "0" (or blank) means "do not recode (key) the item". The position of the "0" or "1" in the RESCORE= string must match the position of the item-response in the item-string.

Example 0: I want to collapse data, rating scale 1,2,3,4,5 6 to 0 through 3 is 0, 4 through 6 is 1

CODES = 123456
 NEWSCORE = 000111 ; the codes align vertically

Example 1: The original codes are "0" and "1". You want to reverse these codes, i.e., 1 0 and 0 1, for all items.

```

XWIDE=1 one character wide responses (the standard)
CODES =01 valid response codes are 0 and 1 (the standard)
NEWSCORE=10 desired response scoring
RESCORE=2 rescore all items - this line can be omitted
or
NI = 100 100 ITEMS
IREFER=*
1-100 X FOR ALL 100 ITEMS, reference is X
*
Codes = 01 recode 01
IVALUEX = 10 into 10

```

Example 2: Your data is coded "0" and "1". This is correct for all 10 items except for items 1 and 7 which have the reverse meaning, i.e. 1 0 and 0 1.

```

NI=10 ten items
CODES =01 the standard, shown here for clarity

```

(a) old method - which still works:

```

NEWSCORE=10 revised scoring
RESCORE=1000001000 only for items 1 and 7

```

(b) new method - recommended:

```

IVALUE1 =10 revised scoring
IVALUE0 =01 scoring unchanged, so this line can be omitted.
IREFER =1000001000 only for items 1 and 7

```

If XWIDE=2, use one or two columns per RESCORE= code, e.g., " 1" or "1 " mean recode (key). " 0" or "0 " mean do not recode (key).

Example 3: The original codes are "0" and "1". You want to reverse these codes, i.e., 1 0 and 0 1, for items 1 and 7 of a ten item test.

```

NI =10 ten items
XWIDE =2 two characters wide
CODES = " 0 1" original codes
NEWSCORE=" 1 0" new values
RESCORE =" 1 0 0 0 0 0 1 0 0 0" rescore items 1 & 7

```

Example 4: The original codes are "0", "1", and "2". You want 0 0, 1 1, and 2 1 for all items

```

XWIDE=1 one character wide (standard)
CODES =012 valid codes
NEWSCORE=011 desired scoring

```

Example 5: The original codes are "0", "1", and "2". You want to make 0 2, 1 1, and 2 0, for even-numbered items in a twenty item test.

```

NI=20 twenty items
CODES =012 three valid codes
NEWSCORE=210 desired scoring
RESCORE=01010101010101010101 rescore "even" items

```

Example 6: The original codes are "0", "1", "2", "3" and some others. You want to make all non-specified codes into "0", but to treat codes of "2" as missing.

```

CODES = 0123 four valid codes
NEWSCORE= 01X3 response code 2 will be ignored
MISSCORE=0 treat all invalid codes as 0

```

Example 7: The original codes are "0", "1", "2", "3". You want to rescore some items selectively using KEY1= and KEY2= and to leave the others unchanged - their data codes will be their rating values. For items 5 and 6, 0 0, 1 0, 2 1, 3 2; for item 7, 0 0, 1 0, 2 0, 3 1. Responses to other items are already entered correctly as 0, 1, 2, or 3.

```

CODES =0123 valid codes
RESCORE=0000111000 rescore items 5,6,7
KEY1 =****223*** keyed for selected items
KEY2 =****33X*** the X will be ignored
^ read these columns vertically

```

11.113 NI= number of items

NI= is the total number of items to be read in (including those to be deleted by [IDFILE=](#) etc.). The maximum value of NI= is about 65,000 for one column responses or about 32,500 for two column responses in the standard program. NI= is usually the length of your test (or the total number of items in all test forms to be combined into one analysis).

Example: If there are 230 items in your test, enter

```
NI=230 ; 230 items
```

It is easy to miscount the NI= column. Scroll to the top of the Winsteps screen and check column positions:

```
Input in process..
Input Data Record:
      1      2
1234567890123456789012345678
Richard M 111111100000000000
^P      ^I      ^N
      35 KID      Records Input.
```

^P marks the Name1=1 column position with ^.

^I marks the Item1=11 column position with ^.

^N marks the NI=18 column position with ^.

11.114 NORMAL= normal distribution for standardizing fit

The standard generally matches the statistics used in BTM and RSA.

Specifies whether distribution of squared residuals is hypothesized to accord with the chi-square or the normal distribution. Values of t standardized fit statistics are obtained from squared residuals by means of one of these distributions.

NORMAL=N	t standardized fit statistics are obtained from the squared residuals by means of the chi-square distribution and the Wilson-Hilferty transformation (the standard).
NORMAL=Y	t standardized fit statistics are obtained from the squared residuals by means of an asymptotic normal distribution (F.A.G. Windmeijer, The asymptotic distribution of the sum of weighted squared residuals in binary choice models, Statistica Neerlandica, 1990, 44:2, 69-78).

11.115 OFILE= report output file

A. Output Tables requested from the [Output Tables menu](#) are displayed in temporary files. These can be "Saved As" permanent files.

B. Output Tables can be requested in the control file or at the [Extra Specifications](#) prompt using [TABLES=](#) or [TFILE=](#). These are output to a report file.

1) if no report output file is specified, then the tables are output to a temporary file which is displayed after the analysis phase is completed.

2) if a report output file is specified when Winsteps is [launched](#) interactively, then the tables are output to the specified file which is displayed after the analysis phase is completed.

3) if a report output file is specified when Winsteps is launched in [Batch](#) mode, then the tables are output to the specified file.

4) report output files can be specified in the control file or at the [Extra Specifications](#) prompt using OFILE=. These can be displayed from the [Edit menu](#). When Winsteps is launched interactively, the last output report file containing a Table is displayed.

Example 1: Winsteps is launched interactively, but we want the Output Tables to be written to file "mytables.txt". Press enter when Winsteps asks for a Report Output file name.

```
TITLE="My analysis"
TABLES=1110110101
OFILE=mytables.txt
.....
```

Example 2: We want Winsteps to analyze the data once, but produce the same Output Tables twice in different scaling written to the same report output files:

```
TITLE="My analysis"
TFILE=*
OFILE="mytable1s.txt"
UMEAN=500 ; item mean calibration
USCALE=100 ; value of 1 logit
MRANGE=250 ; to be plotted each way from UMEAN=
1 ; Table 1
UMEAN=0 ; item mean calibration
USCALE=1 ; value of 1 logit
MRANGE=5 ; to be plotted each way from UMEAN=
1 ; Table 1 again
*
```

11.116 OSORT= option/distractor sort

Specifies the order in which categories, options and distractors are listed within items in the Item Category/Option/Distractor Tables, such as [Table 13.3](#)

OSORT = D Options within items are listed in their Data order in CODES= (the standard).
 OSORT = S or V or " " Score or Value: Options are listed in their Score Value order.
 OSORT = A or M Average or Measure: Options are listed in their [Average Measure](#) order.

Example: List distractors in original [CODES=](#) order:

```
OSORT=D
CODES=000102
XWIDE=2
```

ACTS CATEGORY/OPTION/Distractor FREQUENCIES: MISFIT ORDER										
ENTRY	DATA	SCORE	DATA	AVERAGE	S.E.	OUTF				
NUMBER	CODE	VALUE	COUNT	%	MEASURE	MEAN	MNSQ	ACT		
23	A 00	0	44	59	9.57	1.37	1.7	WATCH A RAT	00	dislike
	01	1	20	27	10.35	3.79	2.4		01	neutral
	02	2	11	15	8.92	6.48	7.6		02	like
	MISSING	***	1	1						
5	B 00	0	47	63	9.57	1.48	1.8	FIND BOTTLES AND CANS	00	dislike
	01	1	19	25	8.34	3.45	2.9		01	neutral
	02	2	9	12	13.14	8.10	6.1		02	like

11.117 OUTFIT= sort misfits on infit or outfit

Other Rasch programs may use *infit*, *outfit* or some other fit statistic. There is no one "correct" statistic. Use the one you find most useful.

Specifies whether mean-square infit or mean-square outfit is used as your output sorting and selection criterion for the diagnosis of misfits in Tables [6](#), [7](#), [10](#), [40](#), [41](#).

OUTFIT=Yes or Both or 2	For each person, the greater of the outfit mean-square and infit mean-square is used as the fit statistic for sorting and selection (the standard).
OUTFIT=No or Infit or 1	Infit mean-square only is used as the fit statistic for sorting and selection.
OUTFIT=Outfit or 0	Outfit mean-square only is used as the fit statistic for sorting and selection.

Example: We want Table 6 sorted by Outfit Mean-square descending:

OUTFIT = Outfit

Output Tables: 6

11.118 PAFILE= person anchor file

The [PFILE=](#) from one run can be used unedited as the person anchor file, PAFILE=, of another.

PAFILE= file name	file containing details
PAFILE = *	in-line list
PAFILE = \$S1W1	field in person label

The person parameter values (thetas) can be anchored (fixed) using PAFILE=. Person anchoring can also facilitate test form [equating](#). The persons common to two test forms can be anchored at the values for one form. Then the measures constructed from the second form will be [equated](#) to the measures of the first form. Other measures are [estimated](#) in the frame of reference defined by the anchor values.

[Displacements](#) are reported, indicating the differences between the anchored values and the freely estimated values. If these are large, please try changing the setting of [ANCESTIM=](#).

PAFILE=? opens a Browse window

In order to anchor persons, a data file must be created of the following form:

1. Use one line per person (or person range) to be anchored.
2. Type the sequence number of the person in the current analysis, a blank, and the measure-value at which to anchor the person (in logits if [UASCALE=1](#), or in your user-rescaled [USCALE=](#) units otherwise). [Arithmetical expressions](#) are allowed. Further values in each line are ignored. An IFILE= works well as an IAFILE=.
3. If the same person appears more than once, the first anchor value is used.

[UIMEAN=](#) and [UPMEAN=](#) are ignored when there are anchor values, [IAFILE=](#) or [PAFILE=](#)

Examples:

```
PAFILE=*
2 3.47 ; anchors person 2 at 3.47 logits (or USCALE= values)
10-13 1.3 ; persons 10, 11, 12, 13 are each anchored at 1.3 logits
2 5.2 ; person 2 is already anchored. This person anchoring is ignored
1-50 0 ; all the unanchored persons in the range 1-50 are anchored at 0.
*
```

Anything after ";" is treated as a comment.

PAFILE = filename

Person anchor information is in a file containing lines of format

```
person entry number   anchor value
person entry number   anchor value
```

PAFILE=*

Person anchor information is in the control file in the format

```
PAFILE=*
person entry number   anchor value
person entry number   anchor value
*
```

PAFILE=\$SnnEnn or PAFILE=\$SnnWnn

Person anchor information is in the person data records using the column [selection rules](#), e.g., starting in column Snn and ending in column Enn or of width Wnn. Blanks or non-numeric values indicate no anchor value. PAFILE=\$S10E12 or PAFILE=\$S10W2 means anchor information starts in column 10 and ends in column 12 of the person's data record (not person label). This can be expanded, e.g, PAFILE = \$S23W1+"."+\$S25W2 places the columns next to each other (not added to each other)

Example 1: The first 3 persons are to be anchored at pre-set values

```
TITLE = "PAFILE TEST"
ITEM1 = 1
NI = 4
NAME1 = 6
CODES = 01
PAFILE=*
1 0.6 ; PERSON 1 ANCHORED AT 0.6 LOGITS
2 2.4 ; PERSON 2 ANCHORED AT 2.4 LOGITS
3 -1.8 ; PERSON 3 ANCHORED AT -1.8 LOGITS
*
&END
Item 1
Item 2
Item 3
Item 4
END LABELS
0101 Person 1
1110 Person 2
1000 Person 3
0110 Person 4
```

Example 2: The third person in the test is to be anchored at 1.5 logits, and the eighth at -2.7.

1. Create a file named, say, "PERSON.ANC"
2. Enter the line "3 1.5" into this file, meaning "person 3 is fixed at 1.5 logits".
3. Enter the line "8 -2.7", meaning "person 8 is fixed at -2.7 logits".
4. Specify, in the control file,
 - PAFILE=PERSON.ANC
 - CONVERGE=L ; only logit change is used for convergence
 - LCONV=0.005 ; logit change too small to appear on any report.

or, enter directly into the control file

```
PAFILE=*
3 1.5
8 -2.7
*
CONVERGE=L ; only logit change is used for convergence
LCONV=0.005 ; logit change too small to appear on any report.
```

or, include the anchor information in the data record

```
PAFILE=$S1E4
NAME1=5
ITEM1=11
NI=12
CONVERGE=L ; only logit change is used for convergence
LCONV=0.005 ; logit change too small to appear on any report.
&END
.....
END LABELS
Fred 111010111001 ; this is the first data line
Mary 011010111100
1.5 Jose 101001111010 ; this data line has an PAFILE= anchor value
Jo 111110011101
```

etc.

To check: "A" after the measure means "anchored"

```
+-----+
|ENTRY  RAW          | INFIT |  OUTFIT |PTMEA|          |
|NUMBER SCORE COUNT  MEASURE  ERROR|MNSQ  ZSTD|MNSQ  ZSTD|CORR.|DISPLACE| PERSONS |
+-----+-----+-----+-----+-----+-----+-----+
|   3    32    35    1.5A    .05|.80  -.3|.32  .6|.53|.40| Jose  |
```

Example 2: Matching person labels.

Winsteps expects the person entry numbers for the common persons to be the same in both tests. If they are not, then here is a procedure using Excel.

Run the two analyses independently.
Output the two PFILE= to Excel files.
Sort the two Excel Pfiles by person label.
Copy and paste the two files into one excel worksheet:

The columns are:
Columns A,B,C: Test 1: person id, entry number, measure
Columns D,E :Test 2: person id, entry number

The person labels will not align, because some people are in Test 1 and not Test 2, and some people are in Test 2 but not Test 1.

We need to apply anchor values from Test 1 in the Test 2 analysis.

Now we want Excel to do a look-up for us:
Paste into cell F1:
=VLOOKUP(D1,A\$1:C\$999,3,FALSE)
This looks up the person label in cell D1 in column A, and puts the measure in column C into column F1.
Change 999 to match the number of rows in column A

Copy and paste cell F1 into F2:F999 (to the bottom row of Test 2)
The anchor measures are now in column F, with #N/A where there is no measure.

Copy and paste columns E and F into a new text file: test2pafile.txt.
Add to your Test 2 control file:
PAFILE=test2pafile.txt
The lines containing #N/A will be ignored by Winsteps.
Perform the Test 2 analysis.
Check that all is correct by looking at the entry-order measure Table for Test 2.

11.119 PAIRED= correction for paired comparison data

Paired comparison data is entered as only two observations in each row (or each column). The raw score of every row (or column) is identical. In the simplest case, the "winner" receives a '1', the "loser" a '0', and all other column (or rows) are left blank, indicating missing data.

Example 1: Data for a chess tournament is entered. Each row is a player. Each column a match. The winner is scored '2', the loser '0' for each match. For draws, each player receives a '1'. See [Bradley-Terry model](#).
PAIRED=YES ; paired comparisons
CODES=012 ; valid outcomes
NI=56 ; number of matches

Example 2: From a research paper - mcg.lboro.ac.uk/mji/files/BERJpre.pdf
"for each question used in the main study (38 in total)"
"Judges ... 250 pairwise comparisons of question responses via the comparative judgement website. 5000 pairwise judgement decisions were collected in total." = 20 judges.

So we follow www.rasch.org/rmt/rmt113o.htm with Winsteps:

1. The 38 questions are 38 Winsteps items
2. The 5000 judgements are 5000 "persons". Each person has two responses: "1" (yes/more/better) and "0" (no/less/worse)
3. Each of the 20 judges has a code in the person label, so we can summarize the judges with Winsteps [Table 28](#)
4. Since the data are pairwise, in the Winsteps control file,
PAIRED=YES

11.120 PANCHQU= anchor persons interactively

If your system is interactive, persons to be anchored can be entered interactively by setting PANCHQ=Y before the &END line. If you specify this, you will be asked if you want to anchor any persons. If you respond "yes", it will ask if you want to read these anchored persons from a file; if you answer "yes" it will ask for the file name and process that file in the same manner as if [PAFILE=](#) had been specified. If you answer "no", you will be asked to enter the sequence number of each person to be anchored, one at a time, along with the logit (or user-rescaled by [UASCALE=](#), [USCALE=](#)) calibration. When you are finished, enter a zero.

Example: You are doing a number of analyses, anchoring a few, but different, persons each analysis. This time, you want to anchor person 4.

Enter on the DOS control line, or in the control file:

```
PANCHQ=Y
CONVERGE=L ; only logit change is used for convergence
LCONV=0.005 ; logit change too small to appear on any report.
```

You want to anchor person 4:

```
Winsteps asks you:
DO YOU WANT TO ANCHOR ANY PERSONS?
respond YES(Enter)
DO YOU WISH TO READ THE ANCHORED PERSONS FROM A FILE?
respond NO(Enter)
INPUT PERSON TO ANCHOR (0 TO END):
respond 4(Enter) (the first person to be anchored)
INPUT VALUE AT WHICH TO ANCHOR PERSON:
respond 1.45(Enter) (the first anchor value)
INPUT PERSON TO ANCHOR (0 TO END): 0(Enter) (to end anchoring)
```

11.121 PCORFIL= person residual correlation file

This writes out the Table of inter-person correlations which is the basis of the principal components analysis of residuals. The basis for the correlations is controlled by [PRCOMP=](#).

PRCOMP=R correlates the raw residuals

PRCOMP=S correlates the standardized residuals

PRCOMP=O correlates the observations

Extreme scores: minimum possible (0) and maximum possible (perfect) item scores are omitted from the computation of the correlations. Persons with extreme scores are not reported. Their correlation is 0.

Missing data: for these Winsteps substitutes their expectations when possible. For residuals and standardized residuals, these are 0. Items with extreme scores (minimum possible or maximum possible): Winsteps drops these from the correlation computation. The reason for these choices is to make the principal components analysis of residuals as meaningful as possible.

PCORFILE=? opens a Browse window

Example 1: Write out the Table of inter-person residual correlations. PCORFIL=file.txt - Then file.txt contains, for SF.txt, in 12-character, fixed-format, columns:

PERSON	PERSON	CORRELATION
1	2	.028295
1	3	.376179
1	4	.457773
1	5	.212332
1	6	-.452587
1	7	.338937

.....

Example 2: When PCORFILE= is selected on the [Output Files](#) menu or [MATRIX=](#)YES, the Data Format: Matrix option can be selected:



This produces in 10-character wide columns:

```

      1      2      3      4
1. 000000  .028295  .376179  .457773
   .028295  1.000000  .460451  -.314252
   .376179  .460451  1.000000  -.083707
   .457773  -.314252  -.083707  1.000000
   .212332  -.258320  -.009482  .270427
  -.452587  .271853  -.443830  -.118054
   .338937  -.556438  .204127  .348083

```

11.122 PDELETE= person one-line item deletion

A one-line list of persons to be deleted or reinstated can be conveniently specified with PDELETE=. It has the same functionality as [PDFILE=](#). [PSELECT=](#) is another way of deleting persons.

The formats are:																																										
PDELETE= 3	3 is an person entry number: delete person 3	<table border="1"> <thead> <tr> <th>ENTRY NUMBER</th> <th>TOTAL SCORE</th> <th>COUNT</th> <th>MEAN</th> </tr> </thead> <tbody> <tr><td>1</td><td>7333</td><td>10000</td><td>-.85</td></tr> <tr><td>2</td><td>8274</td><td>10000</td><td>-1.48</td></tr> <tr><td>3</td><td>7754</td><td>10000</td><td>-1.11</td></tr> <tr><td>4</td><td>5783</td><td>10000</td><td>-.04</td></tr> <tr><td>5</td><td>7873</td><td>10000</td><td>-1.15</td></tr> <tr><td>.....</td><td></td><td></td><td></td></tr> <tr><td>660</td><td>6564</td><td>10000</td><td>-.42</td></tr> <tr><td>MEAN</td><td>5612.2</td><td>10000</td><td>.00</td></tr> <tr><td>S.D.</td><td>1803.1</td><td>.0</td><td>1.00</td></tr> </tbody> </table>	ENTRY NUMBER	TOTAL SCORE	COUNT	MEAN	1	7333	10000	-.85	2	8274	10000	-1.48	3	7754	10000	-1.11	4	5783	10000	-.04	5	7873	10000	-1.15				660	6564	10000	-.42	MEAN	5612.2	10000	.00	S.D.	1803.1	.0	1.00
ENTRY NUMBER	TOTAL SCORE		COUNT	MEAN																																						
1	7333		10000	-.85																																						
2	8274		10000	-1.48																																						
3	7754		10000	-1.11																																						
4	5783		10000	-.04																																						
5	7873		10000	-1.15																																						
.....																																										
660	6564	10000	-.42																																							
MEAN	5612.2	10000	.00																																							
S.D.	1803.1	.0	1.00																																							
PDELETE= 6 1	delete persons 6 and 1																																									
PDELETE= 2-5	delete persons 2, 3, 4, 5																																									
PDELETE= +3-10	delete all persons, then reinstate persons 3 to 10																																									
PDELETE= 4-20 +8	delete persons 4-20 then reinstate person 8																																									
PDELETE= 3,7,4,10	delete persons 3, 7, 4, 10. Commas, blanks and tabs are separators. At the "Extra information" prompt, use commas as separators																																									
PDELETE= (blank)	In the Specification Menu dialog box, resets temporary person deletions																																									

Example 1: After an analysis is completed, delete cases 16, 25 and 87 from the reporting.

In the Specification pull-down box:

PDELETE=16 25 87

Example 2: Delete all except persons 5-10 and report

Specification menu box: PDELETE=+5-10

Output Tables Menu

Now reinstate person 11 and report persons 5-11.

Specification menu box: PDELETE= 11 +11 ; person 11 is already deleted, but prevents deletion of all except +11.

Output Tables Menu

11.123 PDELQU= delete persons interactively

[PDELETE=](#) and/or [PDFILE=](#) at Extra Specifications? are easier.

Persons to be deleted or selected can be entered interactively by setting PDELQU=Y. If you specify this, you will be asked if you want to delete any persons. If you respond "yes", it will ask if you want to read these deleted persons from a file; if you answer "yes" it will ask for the file name and process that file in the same manner as if [PDFILE=](#) had been specified. If you answer "no", you will be asked to enter the sequence number or numbers of persons to be deleted or selected one line at a time, following the rules specified for [IDFILE=](#). When you are finished, enter a zero.

Example: You are doing a number of analyses, deleting a few, but different, persons each analysis. You don't want to create a lot of small delete files, but rather just enter the numbers at the terminal, so specify:

PDELQU=Y

Winsteps asks you:

DO YOU WANT TO READ SOME KID DELETIONS FROM A FILE? (Y/N)

Y

WHAT IS THE NAME OF THE KID DELETE FILE?

MYPDFILE.TXT

Processing KID PDELQU=MYPDFILE.TXT

DO YOU WANT TO READ MORE KID DELETIONS FROM A FILE? (Y/N)

N

DO YOU WANT TO TYPE IN MORE KID DELETIONS? (Y/N)

Y

TYPE IN THE DELETIONS IN PDELETE= FORMAT:

1 5 9

DO YOU WANT TO TYPE IN MORE KID DELETIONS? (Y/N)

N

11.124 PDFILE= person deletion file

Deletion or selection of persons from a test to be analyzed, but without removing their responses from your data file, is easily accomplished by creating a file in which each line contains the sequence number of a person or persons to be deleted or selected (according to the same rules given under [IDFILE=](#)), and then specifying this file by means of the control variable, PDFILE=, or enter the deletion list in the control file using PDFILE=*. PDFILE=filename can be action from the [Specification menu](#) dialog box. See also [PDELETE=](#).

PDFILE= file name	file containing list of person entry numbers to be deleted or reinstated. Each line in the file is in the PDELETE= format.
PDFILE = ?	opens a Browser window to find the file
PDFILE = *	in-line list in the control file
PDFILE=* 4 18-23 *	At the Extra Specifications prompt
PDFILE=	In the Specifications Menu dialog box, reinstates all temporary deletions

Example 1: You wish to delete the fifth and tenth persons from this analysis.

1. Create a file named, say, "PERSON.DEL"

2. Enter into the file, the lines:

5

10

3. Specify, in the control file,

PDFILE=PERSON.DEL

or, enter directly into the control file,

PDFILE=*

5

10

*

Example 2: The analyst wants to delete the most misfitting persons reported in Table 6.

1. Set up a standard control file.

2. Specify

PDFILE=*

*

3. Copy the target portion of Table 6.

4. Paste it between the ""

5. Delete characters before the entry numbers.

6. Type ; after the entry numbers to make further numbers into comments.

Example 3:

TITLE = 'Example of person deletion list from Table 6'

```

PDFILE = *
Delete the border character before the entry number
; ENTRY   RAW                INFIT      OUTFIT
;  NUM  SCORE COUNT MEASURE  ERROR MNSQ  ZSTD MNSQ  ZSTD PTBSE  PUP
   73   ; 21    22     .14    .37 .95   -.3 1.03   .2 B-.19  SAN
   75   ; 16    22    -.56    .39 .95   -.3 1.03   .2 C-.19  PAU
      Enter the ; to make other numbers into comments
*      ; closure of PDFILE=*

```

Example 4: My sample is large enough (n>2000). I want to exclude in the analysis the persons and items that have infit or outfit statistics of less than -2 or greater than +2.

1. Perform your analysis.
2. Output files menu: PFILE to Excel.
3. In Excel sort the persons in Outfit ZSTD order.
4. Delete all person rows you want to keep in your analysis.
5. Copy the column of entry numbers of persons you want to delete.
6. Create a new text file "delete.txt"
7. Paste the list of entry numbers into the text file
8. In your Winsteps control file, add
PDFILE = delete.txt
9. Perform your revised analysis

11.125 PDROPEXTREME= drop persons with extreme scores

Unanchored persons with extreme (zero, minimum possible or perfect, maximum possible) scores provide no information for estimating item measures, but they are reported and included in summary statistics. To remove them:

PDROPEXTREME = No ; do not drop extreme persons (standard)

PDROPEXTREME = Yes or All ; drop zero (minimum possible) and perfect (maximum possible) scores

PDROPEXTREME = Zero or Low or Bottom or Minimum ; drop zero or minimum-possible scores

PDROPEXTREME = Perfect or High or Top or Maximum ; drop perfect or maximum-possible scores

Example 1: The data file contains many data records of persons who did not attempt this test and so were scored 0. They are skewing the test statistics:

PDROPEXTREME = Zero

Example 2: On a satisfaction survey, hurried respondents mindlessly code every item "highly satisfied".

PDROPEXTREME = Perfect

11.126 PERSON= title for person labels

Up to 12 characters to use in table headings to describe the kind of persons, e.g.

PERSON=KID

11.127 PFILE= person output file

PFILE=filename produces an output file containing the information for each person. This file contains 4 heading lines (unless [HLINES=N](#) or [ROW1HEADING=N](#)), followed by one line for each person containing the following fields and the standard field selection. To change the output-field selection, go to the [Output File dialog](#), PFILE=, Field selection, Make default, or [POFSFIELDS=](#).

PFILE=? opens a Browse window

"Status=-2 to -6" means that there are no measurable responses by those persons in this analysis. The persons may be listed in the PFILE= and in Table 18, but all the numbers shown are default values. They have no meaning. Please do not include those persons in summary statistics.

Columns:		with 'Select All" fields using Output File Field Selection			
Start	End	Label	Format	Description	
1	1		A1	Blank or ";" if HLINES=Y and there are no responses or deleted or extreme (status =0,-1, -2, -3)	
2	6	ENTRY	I5	1. The person sequence entry number	
7	14	MEASURE	F8.2	2. MLE person ability estimate user-rescaled by UMEAN= , USCALE= , UDECIM= . Measures for deleted or inestimable persons are shown as 9999.	
15	17	STATUS	I3	3. The person's status 2 = Anchored (fixed) measure 1 = Estimated measure 0 = Extreme maximum measure (estimated using EXTRSC=) for extreme maximum raw score -1 = Extreme minimum measure (estimated using EXTRSC=) for extreme minimum raw score (usually 0) -2 = No responses available for measure -3 = Deleted by user -4 = Inestimable: high (all responses in the same category with ISGROUPS=0 or CUTHI=) -5 = Inestimable: low (all responses in the same category with ISGROUPS=0 or CUTLO=) -6 = Anchored (fixed) measure with extreme (minimum or maximum) observed raw score -7 to -16 = Temporarily deselected by Specification box with PSELECT= (usual STATUS - 10) -17 to -26 = Temporarily deleted by Specification box with PDELETE= (usual STATUS - 20) -27 to -36 = Temporarily deselected and deleted by Specification box with PSELECT= and PDELETE= (usual STATUS - 30)	
18	25	COUNT	F8.1	4. The number of responses used in calibrating (TOTAL=N), or the observed count (TOTAL=Y)	
26	34	SCORE	F9.1	5. The raw score used in calibrating (TOTAL=N) or the observed score (TOTAL=Y)	
35	41	MODLSE REALSE	F7.2	6. Standard error of the MLE or WLE person ability estimate adjusted by REALSE= and user-rescaled by USCALE= , UDECIM=	
42	48	IN.MSQ IN.CHI	F7.2	7. Person infit: mean square infit. Chi-square = IN.MSQ* INDF If CHISQUARE=Yes, IN.CHI = Infit Chi-square	
49	55	IN.ZSTD, ZEMP , LOG, PROB	F7.2	8. Person infit: t standardized, locally t standardized,, log-scaled or probability (LOCAL=)	
56	62	OUT.MS OUT.CHI	F7.2	9. Person outfit: mean square outfit. Chi-square = OUT.MSQ* OUTDF If CHISQUARE=Yes OUT.CHI = Outfit Chi-square	
63	69	OUT.ZSTD, ZEMP , LOG, PROB	F7.2	10. Person outfit: t standardized, locally t standardized, log-scaled or probability (LOCAL=)	
70	76	DISPLACE	F7.2	11. Person displacement (user-rescaled by USCALE= , UDECIM=)	
77	83	PTBS, PTMEAS	F7.2	12. Person by test-score correlation: point-biserial, or point-measure (PTBIS=). This is 0.00 if inestimable.	
84	90	WEIGHT	F7.2	13. Person weight (PWEIGHT=)	
91	96	OBSMA	F6.1	14. Observed percent of observations matching prediction	
97	102	EXPMA	F6.1	15. Expected percent of observations matching prediction	
103	108	PVALUE	F6.2	16. P-value: proportion correct or average rating (PVALUE=)	
109	114	PME-E	F6.2	17. Expected value of Person by test-score correlation. This is 0.00 if inestimable.	

115	120	RMSR	F6.2	18. RMSR: root-mean-square residual (RMSR=)
121	128	WMLE	F8.2	19. Warm's (Weighted) Mean Likelihood Estimate (WLE) of person Ability user-rescaled by UMEAN= , USCALE= , UDECIM=
129	134	INDF	F6.2	20. degrees of freedom of Infit mean-square
135	140	OUTDF	F6.2	21. degrees of freedom of Outfit mean-square
141	148	QCMLE	F8.2	22. Quasi-CMLE estimates for dichotomous data. 0 otherwise.
149	149		1X	Blank
150	150+	NAME	A30+	23. Person name: change NAME1= and NAMELENGTH= to alter this.

The format descriptors are:

In	= Integer field width n columns
Fn.m	= Numeric field, n columns wide including n-m-1 integral places, a decimal point and m decimal places
An	= Alphabetic field, n columns wide
nX	= n blank columns.

When CSV=Y, commas separate the values, which are squeezed together without spaces between. Quotation marks surround the "Person name", e.g., 1,2,3,4,"Name". When CSV=T, the commas are replaced by tab characters.

When W300=Yes, then this is produced in Winsteps 3.00, 1/1/2000, format:

Columns:				
Start	End	Label	Format	Description
1	1		A1	Blank or "," if HLINES=Y and there are no responses or deleted (status = -2, -3)
2	6	ENTRY	I5	1. The person sequence number
7	14	MEASURE	F8.2	2. Person's ability estimate (user-rescaled by UMEAN=, USCALE=, UDECIM)
15	17	STATUS	I3	3. The person's status: 3 = Anchored (fixed) measure with extreme (minimum or maximum) observed raw score 2 = Anchored (fixed) measure 1 = Estimated measure 0 = Extreme minimum (estimated using EXTRSC=) -1 = Extreme maximum (estimated using EXTRSC=) -2 = No responses available for measure -3 = Deleted by user -4 = Inestimable: high (all responses in the same category with ISGROUPS=0 or CUTHI=) -5 = Inestimable: low (all responses in the same category with ISGROUPS=0 or CUTLO=) -6 = Deselected
18	23	COUNT	I6	4. The number of responses used in measuring (TOTAL=N) or the observed count (TOTAL=Y)
24	30	SCORE	I6	5. The raw score used in calibrating (TOTAL=N) or the observed score (TOTAL=Y)
31	37	MODLSE REALSE	F7.2	6. Person ability standard error adjusted by REALSE= and user-rescaled by USCALE= , UDECIM=
38	44	IN.MSQ	F7.2	7. Person mean square infit
45	51	ZSTD, ZEMP , LOG	F7.2	8. Person infit: t standardized, locally t standardized, or log-scaled (LOCAL=)
52	58		F7.2	9. Person mean square outfit (OUT.MS)

59	65	ZSTD, ZEMP , LOG	F7.2	10. Person outfit: t standardized, locally t standardized, or log-scaled (LOCAL=)
66	72	DISPLACE	F7.2	11. Person displacement (user-rescaled by USCALE=, UDECIM=)
73	79	PTBS, PTME	F7.2	12. Person by test-score correlation: point-biserial, or point-measure
80	80		1X	13. Blank
81	112+	NAME	A30+	14. Person name

Example of standard PFILE= (to see other fields: [Output File dialog](#))

```
; PERSON Survey Analysis Dec 7 19:21 2009
;ENTRY MEASURE ST COUNT SCORE ERROR IN.MSQ IN.2ST OUT.MS OUT.2S DISPL FIMEAS WEIGHT OBSMA EKPGA FVALU FME-E RMSR NAME
1 .95 1 4.0 19.0 .88 1.24 .65 2.67 1.63 .00 -.87 1.00 50.0 66.0 4.75 .42 .66 M Richard
2 .26 1 4.0 19.0 .79 1.31 .63 1.90 1.22 .00 .90 1.00 50.0 60.3 4.50 .35 .72 F Tracie
3 -.92 1 4.0 16.0 .78 .49 -.78 .80 -.25 .00 .87 1.00 50.0 53.9 4.00 .18 .45 M Walter
4 -1.59 1 4.0 15.0 .86 1.00 .22 1.12 .40 .00 .62 1.00 25.0 50.2 3.75 .10 .58 M Blaise
5 -.92 1 4.0 16.0 .78 .21 -1.71 .32 -1.64 .00 -.87 1.00 100.0 53.9 4.00 .18 .29 M Ron
6 1.85 1 4.0 20.0 1.02 .20 -1.10 .19 -.77 .00 .88 1.00 100.0 74.5 5.00 .47 .22 M William
```

Example: You wish to write a file on disk called "STUDENT-PF.txt" containing the person statistics for import later into a student information database:

PFILE=STUDENT-PF.txt

11.128 PKMAP= customize Table 36 diagnostic maps

PKMAP= can be used to customize [Table 36](#) diagnostic maps. These maps are based on the KIDMAP - www.rasch.org/rmt/rmt82k.htm

PKMAP= customization, see also [ICOLORFILE=](#) for more color options

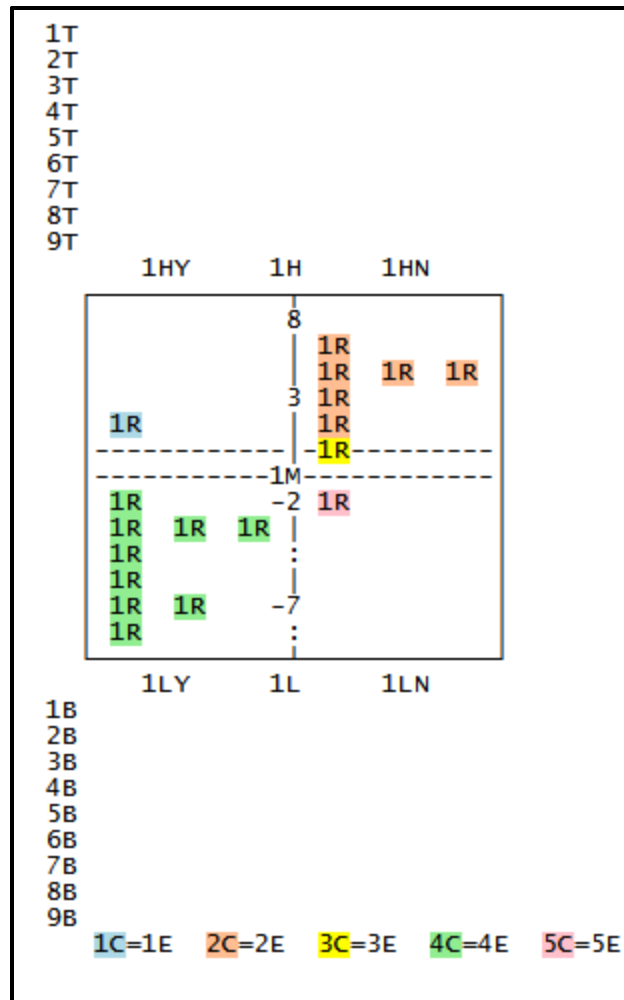


Figure: PKMAP field numbers and positions

Format in Winsteps control file:

ASCII=Webpage ; for the colored output
 PKMAP=* ; inline instructions for the PKMAP output in [Table 36](#)
 (Field number)= value

 *

PKMAP Size

The internal height of PKMAP grid box is 61 or [MAXPAG=](#) value - 8.
 The internal width of the PKMAP grid box is [LINELENGTH=](#) value - 6.

Format in text file:

use Edit menu
 Edit/Create file with NOTEPAD to modify these settings.
 In text file, for example:
 1D = Yes
 Save text file as MyPKMAP.txt in the same folder as the Control file
 Specification menu: PKMAP = MyPKMAP.txt

Field code	Values	Meaning
1T-9T = Top headings 1HY, 1H, 1HN = High side of grid	(blank)	do not display the field

1M = Measure = Middle (ability) 1LY, 1L , 1LN = Low side of grid 1B-9B = Bottom footers 1E-5E = Color explanations	\	blank line
	text	display the text in the field location
	#nnn# (can be included in text)	display the statistic in the field location
1R = Response = Rating	(blank) or text or #nnn#, e.g., #LABEL# #CODE# #CAT#	(as above) Response code in the data file Scored category number For coloring based on 1R value, see ICOLORFILE=
1G = Grid for PKMAP layout	Top left cell of grid is: HY High measure, Yes reached HN High measure, Not reached LY Low measure, Yes reached LN Low measure, Not reached	Controls layout of the PKMAP by identifying the top left quadrant
1C-5C = Color codes	HTML color codes, e.g., light-green #E7C4A3	The color codes for text in the grid cells: 1C = High Yes (Current level) 2C = High No (Next level) 3C = 50-50 4C = Low Yes (Current level) 5C = Low No (Next level)
1P = Placement of response	T = Rasch-Thurstone 50% threshold H or F = Half-point and Full-point thresholds	Vertical measure at which to place a response.
1D = Display all the field numbers	"Y" = Display field numbers and values at start of output	The field codes and their values
1F = Minimum mean-square to report	0 = all mean-squares 1.5 = only persons with infit or outfit mean- squares greater than 1.5 are reported	Selects only Kidmaps for misfitting response strings
1S = Sort instructions	n = Entry number descending N = Entry number ascending m = Measure descending M = Measure ascending f = Mean-square fit descending F = Mean-square fit ascending fl, Fl = use Infit mean-square fo, FO = use Outfit mean-square fB, FB = use bigger of both mean- squares a = Alphabetical label descending A = Alphabetical label ascending a\$S3W10, A\$S3W10 = Alphabetical sort on part of person label	Order of the PKMAP Kidmaps in the output table.
1V-9V = Criterion lines	numbers followed by L or U	logit L or uscale U values for horizontal = lines on grid
1A = Ability lines	00000 = no lines 10000 = Line at ability level 01000 = Line 1 S.E. above ability level 00100 = Line 1 S.E. below ability level 01100 = Both 1 S.E. lines (default) 00010 = Line 2 S.E. above ability level 00001 = Line 2 S.E. below ability level These can be combined	show lines around 1M ability level
Field number not in PKMAP= list	none	(ignored)
Control variable= PSELECT=???A; selection PDELETE=(list of entry numbers)	value	control variable for the PKMAP display

PDFILE=(file of entry numbers) ASCII=Webpage		
T36	-	output the Table 36 PKMAPs

#nnn#	Value: 1= to 8= Person. 9= Item
#NAME# or #LABEL#	Name or label
#ENTRY#	Entry number
#MEASURE#	Measure (user-rescaled by UMEAN= , USCALE= , UDECIM=)
#COUNT#	Total number of scored responses
#SCORE#	Total raw score
#ERROR#	Standard error of measure (user-rescaled by USCALE= , UDECIM=)
#INMNSQ#	Infit mean-square
#INZSTD#	Standardized Infit statistic: t standardized, locally t standardized, or log-scaled (LOCAL=)
#OUTMNSQ#	Outfit mean-square
#OUTZSTD#	Standardized Outfit statistic: t standardized, locally t standardized, or log-scaled (LOCAL=)
#TITLE#	TITLE= of this analysis
#Y#	Measure-units for each row (line) advance on vertical y-axis.
\$\$S1W1# #@fieldname#	Display the label substring
##	#
#CENTER#	Center this field (1T-9T and 1B-9B only)

PKMAP= default values for dichotomies (True/False, Multiple-Choice)

```
PKMAP = *
1D = "No"
1F = "0"
1S = "N"
1T = " Name: #NAME#"
2T = " Ref. Number: #ENTRY#           Measure: #MEASURE# S.E. #ERROR# Score: #SCORE#"
3T = " Test: #TITLE#"
4T = ""
1G = "HY"
1P = "Half-point threshold"
1R = "#ENTRY#.#CAT#"
1HY = "Hard items answered correctly"
1H = "\<b>\<i>-Harder-\</b>\</i>"
1HN = "Hard items answered incorrectly"
1M = "\<FONT COLOR="RED">XXX\</FONT>"
1LY = "Easy items answered correctly"
1L = "\<b>\<i>-Easier-\</b>\</i>"
1LN = "Easy items answered incorrectly"
1B = ""
2B = "#CENTER#Each row is #Y# logits" "units" if USCALE<>0
3B = ""
1E = "\<SPAN CLASS=1C>>blue\</SPAN>=Unexpected Yes."
2E = " \<SPAN CLASS=2C>>orange\</SPAN>=No."
3E = " \<SPAN CLASS=3C>>yellow\</SPAN>=50/50."
4E = " \<SPAN CLASS=4C>>green\</SPAN>=Yes."
5E = " \<SPAN CLASS=5C>>pink\</SPAN>=Unexpected No."
1C = "lightblue"
2C = "#FFBD91"
3C = "yellow"
4C = "lightgreen"
5C = "pink"
; = = = "default values above. User-set values below."
*
```

PKMAP= default values for polytomies (Rating Scale, Partial Credit)

```
PKMAP = *
```

```

1D = "No"
1F = "0"
1S = "N"
1T = " Name: #NAME#"
2T = " Ref. Number: #ENTRY#           Measure: #MEASURE# S.E. #ERROR# Score: #SCORE#"
3T = " Test: #TITLE#"
4T = "\ "
1G = "HY"
1P = "Full-point threshold"
1R = "#ENTRY#. #CODE#"
1HY = "Hard levels reached"
1H = "\<b>\<i>-Higher-\</b>\</i>"
1HN = "Hard levels not reached"
1M = "\<FONT COLOR="RED">XXX\</FONT\>"
1LY = "Easy levels reached"
1L = "\<b>\<i>-Lower-\</b>\</i>"
1LN = "Easy levels not reached"
1B = "\ "
2B = "#CENTER#Each row is #Y# logits" "units" if USCALE<>0
3B = "\ "
1E = "\<SPAN CLASS=1C\>blue\</SPAN\>=Unexpected Yes."
2E = " \<SPAN CLASS=2C\>orange\</SPAN\>=No."
3E = " \<SPAN CLASS=3C\>yellow\</SPAN\>=50/50."
4E = " \<SPAN CLASS=4C\>green\</SPAN\>=Yes."
5E = " \<SPAN CLASS=5C\>pink\</SPAN\>=Unexpected No."
1C = "lightblue"
2C = "#FFBD91"
3C = "yellow"
4C = "lightgreen"
5C = "pink"
; = = = "default values above. User-set values below."
*

```

HTML codes can be used with `<` and `>` replacing `<` and `>`. HTML codes are ignored for text output.

Color: `...` for HTML colors

or `...` for the colors defined in 1C, 2C, 3C, 4C, 5C

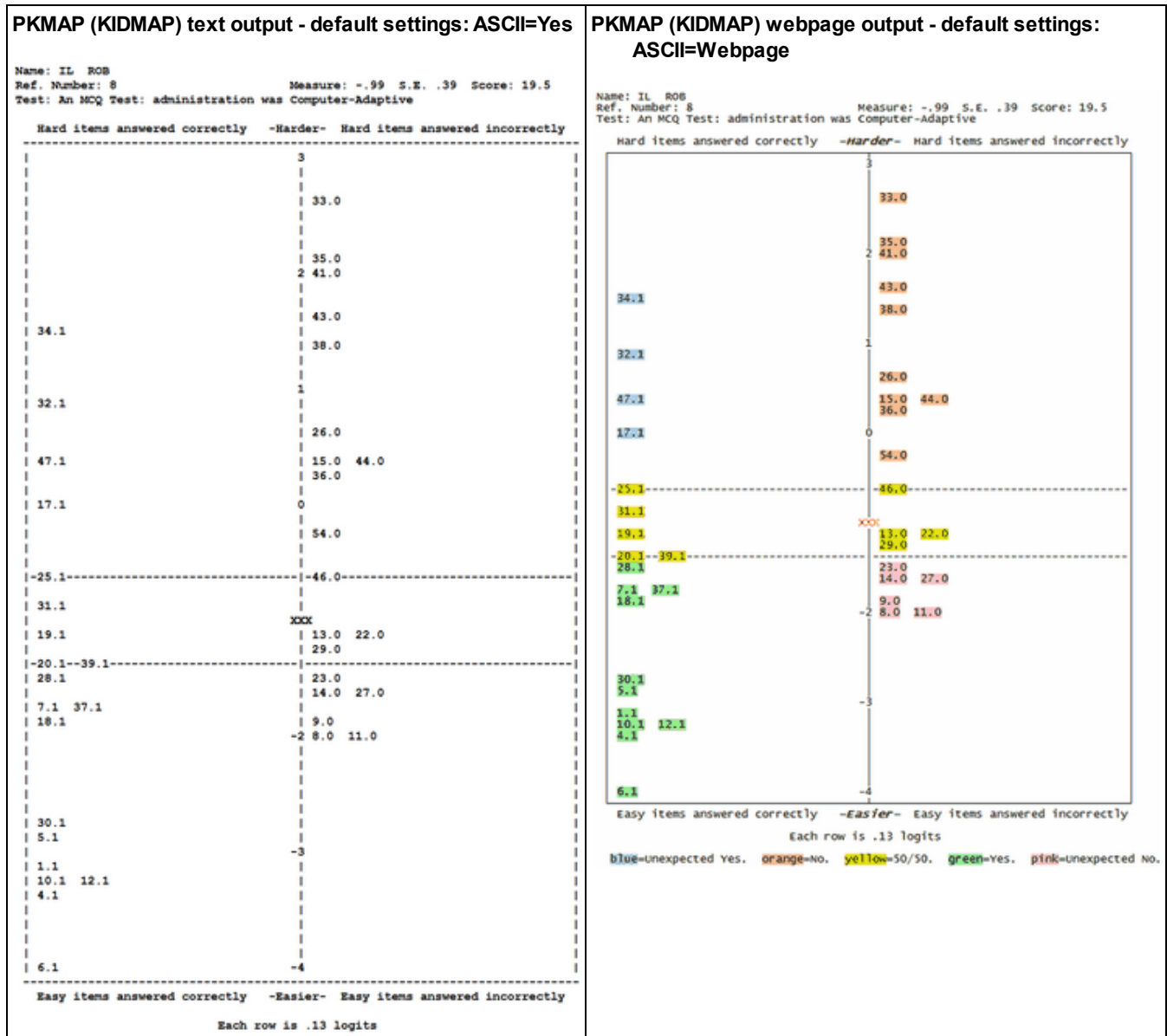
A list of color names is shown at http://www.w3schools.com/colors/colors_names.asp - Hex values can also be used.

These color names and hex values are used for 1C-5C.

Bold: `...` but bold for 1M and 1R may cause misalignment. Use a color instead.

Italic: `<i>...</i>` but bold for 1M and 1R may cause misalignment. Use a color instead.

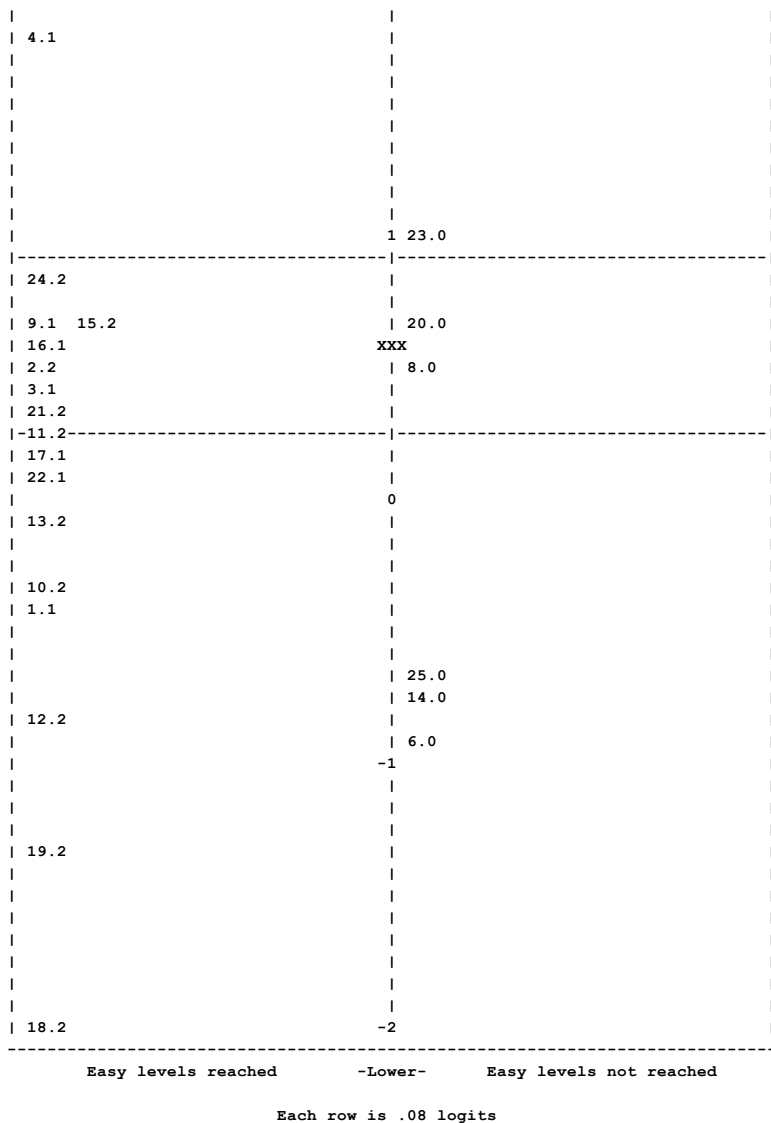
Example 1: Exam5.txt - dichotomy (Yes/No, True/False, Multiple-Choice Question) -



Example 2: Example0.txt - "Liking for Science" rating-scale

Name: M Rossner, Marc Daniel
 Ref. Number: 1
 Measure: .61 S.E. .34 Score: 30
 Test: LIKING FOR SCIENCE (Wright & Masters p.18)

Hard levels reached	-Higher-	Hard levels not reached
	3	
5.1		
7.2		
	2	



Left column: success = higher rating on the rating scale

Item 18 (bottom left of PKMAP) is the easiest item. Person 1 rated in the top category 2. - a success - an "easy level reached"

"18.2" on the PKMAP is positioned at the ability level corresponding to that rating on that item.

"18.0" would have been much lower in the right-hand column.

Item 5 (top left of PKMAP) is the hardest item. Person 1 rated in the middle category 1. - a partial success - a "hard level reached"

"5.1" on the PKMAP is positioned at the ability level corresponding to that rating on that item.

"5.2" would have been much higher in the left-hand column.

"5.0" would have been much lower in the right-hand column.

Right column: failure = bottom rating on the rating scale

Item 23 (upper right) is the second hardest item. Person 1 rated in the bottom category 0. - a failure

"23.0" on the PKMAP is positioned at the ability level corresponding to that rating on that item.

"23.2" would have been much higher in the left-hand column.

Arbitrarily, 5.1 is in the left-column as we usually think of someone climbing the rating scale, rather than falling down the rating scale.

11.129 PLFILE= person label file

PLFILE= enables person labels to be inserted or changed after a data file has been processed. It complements [EDFILE=](#) which performs the same function for response-level data.

PLFILE= file name	file containing person entry numbers and labels
PLFILE= file name + file name + ...	multiple files
PLFILE = *	in-line list
PLFILE = ?	opens a Browser window to find the file

Suggestion: sort the PLFILE= data lines as "person-entry-number ascending" for faster processing.

Its format is:

PLFILE=*	
person entry number	insertion or replacement person label
*	

Ranges are permitted for entry numbers: first-last.

Example: Computer-adaptive test (CAT) data are analyzed. The responses are in EDFILE= in any order of person-entry-number item-entry-number response-code.

The person labels are in any order of person-entry-number person-label.

```

PLFILE=*
103 George
23-37 Anonymous
*

Title = CAT data
Item1 =1
NI=500 ; highest item entry number
Name1=502
NAMELENGTH=30 ; length of longest person label or more
CODES=ABCD
KEY1=BBACDBACD.... ; mcq scoring key from item bank
EDFILE=catdata.txt
&END
; item labels here or use ILFILE=
END NAMES
(no data here)

```

11.130 PMAP= person label on person map: Tables 1, 16

This specifies what part of the data record is to be used on the person map: [Table 1](#) and [Table 16](#).

Its format is PMAP = \$S..W.. or \$S..E.. using the column [selection rules](#).

\$S..W.. e.g., \$S2W13 means that the person label to be shown on the map starts in column 2 of the person label and is 13 columns wide.

\$S..E.. e.g., \$S3E6 means that the person label to be shown on the map starts in column 3 of the person label and ends in column 6.

These can be combined, and constants introduced, e.g.,

```
PMAP= $S3W2+"/"+$S7W2
```

If the person label is "KH323MXTR", the person label on the map will be "32/XT"

The length of PMAP= overrides NAMLMP=

11.131 POSFIELDS= field selection for PFILE=

POSFIELDS= is set using the [Output Field selection](#) dialog box.

Field Number

- 1 Flag extremes with ;
- 2 Entry number
- 3 Measures
- 4 Status
- 5 Count of observations
- 6 Raw score
- 7 Standard error
- 8 Infit mean-square
- 9 Infit t standardized
- 10 Outfit mean-square
- 11 Outfit t standardized
- 12 Displacement
- 13 Correlation
- 14 Weight
- 15 Observed matches
- 16 Expected match
- 17 (not used)
- 18 (not used)
- 19 (not used)
- 20 P-value: average rating
- 21 Expected correlation
- 22 (not used)
- 23 (not used)
- 24 Name or Label
- 25 Include deleted
- 26 RMS Residual
- 27 (not used)
- 28 WMLE measure
- 29- (not used)

Example: Output into the PFILE= file only the Entry number and the WMLE measures.

```
POSFIELDS = NYNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNY
```

11.132 PRCOMP= residual type for principal components analyses in Tables 23, 24

Principal components analysis of item-response or person-response residuals can help identify structure in the misfit patterns across items or persons. The measures have been extracted from the residuals, so only uncorrelated noise would remain, if the data fit the Rasch model.

PRCOMP=S or Y Analyze the standardized residuals, $(\text{observed} - \text{expected})/(\text{model standard error})$.

Simulation studies indicate that PRCOMP=S gives the most accurate reflection of secondary dimensions in the items.

PRCOMP=R Analyze the raw score residuals, $(\text{observed} - \text{expected})$ for each observation. These report Wendy Yen's Q3 in [Table 23.99](#)

PRCOMP=L Analyze the logit residuals, $(\text{observed} - \text{expected})/(\text{model variance})$.

PRCOMP=O Analyze the observations themselves.

PRCOMP=K Observation probability

PRCOMP=H Observation log-probability

PRCOMP=G Observation logit-probability

PRCOMP=N Do not perform PCA analysis

Example 1: Perform a Rasch analysis, and then see if there is any meaningful other dimensions in the residuals:
PRCOMP=S Standardized residuals

Example 2: Analysis of the observations themselves is more familiar to statisticians.

PRCOMP=O Observations

In the PCA of th observations, Chien's (2012) Dimension Coefficient is $DC = (R12/R23)/(1 + (R12/R23))$ which simplifies to (eigenvalue of first component in the raw observations) / (first eigenvalue + third eigenvalue). Chien suggests values below 0.67 indicate a lack of unidimensionality.

Cronbach's Alpha with the Dimension Coefficient to Jointly Assess a Scale's Quality. Tsair-Wei Chien ... Rasch Measurement Transactions, 2012, 26:3 p. 1379 www.rasch.org/rmt/rmt263c.htm

11.133 PSELECT= person selection criterion

Persons to be selected may be specified by using the PSELECT= instruction to match characters within the person name. Persons deleted by [PDFILE=](#) etc. are never selected by PSELECT=.

This can be done before analysis in the control file or with "Extra specifications". It can also be done after the analysis using the "Specification" pull-down menu.

Persons to be selected may be specified by using the PSELECT= instruction to match characters within the person name. Persons deleted by [PDFILE=](#) etc. are never selected by PSELECT=.

This can be done before analysis in the control file or with "Extra specifications". It can also be done after the analysis using the "Specification" pull-down menu.

Selection rules:	
Control characters match label or name. They start at the first column of the label or name.	
?	matches any character
*	matches any string of characters - must be last selection character. If * is in the first column, then every available person or item is selected.
A	matches A in the person label, and similarly all other characters except { }
{..}	braces characters which can match a single character: {ABC} matches A or B or C.
{..-..}	matches single characters in a range. {0-9} matches digits in the range 0 to 9.
{..-..}	matches a single "-" {AB-} matches A or B or "-".
{~ABX}	omits persons or items which match A or B or X
@fieldname=	positions the next selection character at the start of the specified field

Each PSELECT= performed using the "Specification" pull-down menu selects from all those analyzed. For incremental selections from the [Specification menu](#), i.e., selecting from those already selected, specify +PSELECT=

Example 0: In [Example 0](#), select only Males:

PSELECT=@GENDER=M

In Example 0, select only Males with last names starting with R

PSELECT=@GENDER=M?R

Example 1: Select for analysis only persons with M in the 5th column of person name. Person name starts in column 6 of the data record:

```

NAME1=6 Person name field starts in col. 6
NAMELEN=8 Person name field is 8 characters long
PSELECT=????M* Column 5 of person name is sex
|
END NAMES
xxxxxBPL M J 01101000101001 selected

```

xxxxxMEL F S 01001000111100 omitted
-----1234 selection column

Example 2: Select for analysis all persons with code "A4" in columns 2-4 of their names. Person name starts in column 23, so target characters starts in column 24:

NAME1=23 person name starts in column 23
PSELECT="A 4*" quotes because a blank is included. A is in col. 2 etc.
selects ZA 4PQRS

Example 3: Select all Male (M in column 2) persons who are Asian or Hispanic (A or H in column 4):

PSELECT=?M?{AH}*
selects 1M3A456 MPQRS
selects 1M5H689 ABCDE
omits 1X2A123 QWERT

Example 4: Select Males (M in column 8) in School 23 (023 in column 14-16):

PSELECT= ??????M?????023*
Selects: 1234567MABCDE023XYZ

Example 5: Select codes 1,2,3,4,5,6,7, in column 2:

PSELECT=?{1-7}*

Example 6: Analyze only males (column 4 or person-id). Then report only School C (column 1). Then only report Grades 4 and 5 (column 2) in School C.

PSELECT=???M* in the Control file or at the Extra Specifications prompt.
PSELECT=C* using the Specification pull-down menu, after the analysis
+PSELECT=?{45}* using the Specification pull-down menu.

Example 7: There are 100000 examinees from 17 school districts in one data file. I want to process only 16 districts.

If the district code is one character, then you can select the 16 codes you want, for instance:

If the codes are letters in column 4 of the person label
PSELECT=???{ABDEFGHIJKLMNOPQR}
or
PSELECT=???{~C}; omits unwanted school district C

Example 8: If the field to be selected is already identified, then this simplifies counting the column:

@Dx = 16E17 ; \$C44W2
16E17 means that the field is in columns 16 and 17 of the person label. Look at a data record to see what is actually there.
PSELECT=????????????23 ; selects "23" in columns 16 and 17

Example 9: We want person group X (in column 4 of the person label) except for subgroups 11, 24 (in columns 6, 7 of the person label):

```
PSELECT = "???X?{12}{14}" ; this selects X and 11, 14, 21, 24
EDFILE=*
"?????14" "?" . ; convert 14 to missing data
"?????21" "?" . ; convert 21 to missing data
*
```

Example 10: Item statistics are needed for each person classification.

If you want PSELECT= to alter the item statistics, then please do separate Winsteps analyses for each PSELECT= value. Place the PSELECT= in your control file, or enter it at the "Extra Specifications (if any)" prompt. This PSELECT= will change what is shown for the persons and the items.

If you want to keep the item difficulties, person abilities and rating scale structure from the complete analysis, rather than re-estimating them. Then

Complete analysis -> [PFILE=pf.txt](#), [JFILE=if.txt](#), [SFILE=sf.txt](#),
[PAFILE=pf.txt](#), [IAFILE=if.txt](#), [SAFILE=sf.txt](#), PSELECT= -> selected analysis.

The "Specification" pull-down menu alter what is included in directly-relevant output tables.
 So, entering PSELECT= at the "Specification" pull-down menu will change what is shown in the person tables, but not what is shown in the item tables.

11.134 PSORT= column within person label for alphabetical sort in Table 19

Table 19 lists persons alphabetically. Table 1 and Table 16 list them alphabetically within lines. Ordinarily, the whole person name is used. Select the sorting columns in the person labels using the column [selection rules](#), e.g., starting in column Snn and ending in column Enn or of width Wnn.

Example 1: The person name is entered in the data file starting in column 20. It is a 6-digit student number followed by a blank and then gender identification in column 8 of the person name. Sort by gender identification for Table 19, then by student number.

```
NAME1=20
NAMLEN=8 ; student number + gender
PSORT=$S8W1+$S8W6 alphabetical sort on gender
TABLES=11111111111111111111111111111111
&END
|
END NAMES
xxxxxxxxxxxxxxxxxxxx123456 M 0010101101001002010102110011
xxxxxxxxxxxxxxxxxxxx229591 F 1102010020100100201002010021
      sort columns ^ ^
```

Example 2: The person name contains several important classifiers. Table 19 is needed for each one:

```
NAME1=14 Person name starts in column 14
ITEM1=24 Response start in column 24
TFILE=*
19 - - - 1 sort starts with column 1 of person name
19 - - - 8 sort starts with column 8 of person name
19 - - - 6 sort starts with column 6 of person name up to the end of the person name
      - entered as place-holders, see TFILE=
*
&END
|
END NAMES
xxxxxxxxxxxxxxxx1234 M 12 0010101101001002010102110011
xxxxxxxxxxxxxxxx2295 F 09 1102010020100100201002010021
```

Example 3: A version of Table 19, sorted on person name column 6, is to be specified on the DOS command line or on the Extra Specifications line. Commas are used as separators, and "-" as place-holders:

```
TFILE=* 19,-,-,-,6 *
```

11.135 PSUBTOTAL= columns within person label for Tables 28, 35

This specifies what part of the data record is to be used to classify persons for subtotal in Table 28.

PSUBTOTAL= file name	file containing details
PSUBTOTAL = *	in-line list
PSUBTOTAL = \$S1W1	field in person label

With tab-separated data and the subtotal indicator in a separate field from the Person label, specify the subtotal field as the person label field using [NAME1=](#), then PSUBTOTAL=\$S1W1

Format 1: PSUBTOTAL = \$S.W. or \$S.E. using the column [selection rules](#).
 \$S.W. e.g., \$S2W13 means that the label to be shown on the map starts in column 2 of the person label and is 13 columns wide.

\$\$..E.. e.g., \$\$S3E6 means that the label to be shown on the map starts in column 3 of the person label and ends in column 6.

These can be combined, and constants introduced, e.g,
 PSUBTOTAL=\$S3W2+"/".\$S7W2

If the person label is "KH323MXTR", the sub-grouping will be shown as "32/XT"

Format 2: PSUBTOTAL=*

This is followed by a list of sub-groupings, each on a new line:

```
PSUBTOTAL=*
$S1W1+$S7W2 ; Subtotals reported for person classifications according to these columns
$S3E5 ; Subtotals reported for person classifications according to these columns
*
```

Example: Subtotal by first letter of person name:

```
PSUBTOTAL=$S1W1
TFILE=*
27 ; produce the subtotal report
*
```

Here is a subtotal report (Table 28) for person beginning with "R"

```
"R" SUBTOTAL FOR 8 NON-EXTREME PUPILS
+-----+
|          RAW          MODEL      INFIT      OUTFIT      |
|          SCORE      COUNT      MEASURE      ERROR      MNSQ      ZSTD      MNSQ      ZSTD      |
+-----+-----+-----+-----+-----+-----+-----+-----+
| MEAN      28.1      25.0      4.04      3.48      .91      -.5      1.04      .0      |
| P.SD      5.5      .0      6.63      .14      .31      1.1      .54      1.4      |
| MAX.      38.0      25.0      16.30      3.82      1.61      2.0      2.37      3.4      |
| MIN.      19.0      25.0      -6.69      3.38      .64      -1.6      .60      -1.2      |
+-----+-----+-----+-----+-----+-----+-----+-----+
| REAL RMSE  3.63 TRUE SD  5.54 SEPARATION  1.52 PUPIL RELIABILITY .70 |
| MODEL RMSE  3.48 TRUE SD  5.64 SEPARATION  1.62 PUPIL RELIABILITY .72 |
| S.E. OF PUPIL MEAN = 2.50                                     |
| WITH 2 EXTREME = TOTAL 10 PUPILS MEAN = 3.05, P.SD = 28.19 |
| REAL RMSE  8.88 TRUE SD  26.75 SEPARATION  3.01 PUPIL RELIABILITY .90 |
| MODEL RMSE  8.83 TRUE SD  26.77 SEPARATION  3.03 PUPIL RELIABILITY .90 |
| S.E. OF PUPIL MEAN = 9.40                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+
| MAXIMUM EXTREME SCORE: 1 PUPILS                             |
| MINIMUM EXTREME SCORE: 1 PUPILS                             |
| LACKING RESPONSES: 1 PUPILS                                  |
| DELETED: 1 PUPILS                                           |
```

Example: You want an average score for each option on a multiple-choice test.

1. Include the item response string in the person label

```
Data record
Fred ABCDE

NI = 5
Item1=6
Name1=1
Name1length=10
```

2. Produce subtotals for every item

```
PSUBTOT=*
6 ; corresponds to Item 1
7
8
```

9
10
...
*

3. Specify Table 28

TFILE=*
28
*

4. Table 28 output is voluminous, but the numbers are there. Do a search for "| MEAN".

11.136 PTBISERIAL= point-biserial-type correlation coefficient

The point-biserial correlation (or the point-polyserial correlation) is the Pearson correlation between on the observations on an item (or by a person) and the person raw scores or measures (or item marginal scores or measures). These are crucial for evaluating whether the coding scheme and person responses accord with the requirement that "higher observations correspond to more of the latent variable" (and vice-versa). These are reported in Tables 14 for items and Table 18 for items persons. They correlate an item's (or person's) responses with the measures of the encountered persons (or items). In Rasch analysis, the point-biserial correlation, r_{pbis} , is a useful diagnostic indicator of data mis-coding or item mis keying: negative or zero values indicate items or persons with response strings that contradict the variable. Positive values are less informative than INFIT and OUTFIT statistics.

Point-Biserial and Point-Measure Correlations			
Control Instruction	Observed value	Explanation	Expected value (EXP.)
PTBISERIAL= Yes PTBISERIAL= Exclude	PBSX PTBISERL-EX	Point-biserial (or point-polyserial) correlation excluding the current observation from the raw score. Computes the point-biserial or point-polyserial correlation coefficients, r_{pbis} , for persons and items. This is the Pearson product-moment correlation between the scored responses (dichotomies and polytomies) and the "rest scores", the corresponding total (marginal) scores excluding the scored responses to be correlated. This is a point-biserial correlation for dichotomies, or a point-polyserial correlation for polytomies. Extreme (perfect, maximum possible and zero, minimum possible) scores are included in the computation, but missing observations are omitted pairwise. The Biserial correlation can be computed from the Point-biserial. This correlation loses its meaning when there are missing data or with CUTLO= or CUTHI= . Specify PTBISERIAL=X instead.	PBSX-E PBX-E
PTBISERIAL= All PTBISERIAL= Include	PBSA PTBISERL-AL	Point-biserial correlation for all observations including the current observation in the raw score. Computes the Pearson correlation between the total (marginal) scores including all responses and the responses to the targeted item and person. This is a point-biserial correlation for dichotomies, or a point-polyserial correlation for polytomies. This correlation loses its meaning when there are missing data or with CUTLO= or CUTHI= . Specify PTBISERIAL=N instead.	PBSA-E PBA-E
PTBISERIAL= No PTBISERIAL= RPM PTBISERIAL= Measure	PTMA PTMEASUR- AL	Point-measure correlation for all observations. Computes the Pearson point-measure correlation coefficients, r_{pm} between the observations and the measures, estimated from the raw scores including the current observation or the anchored values. Measures corresponding to extreme scores are included in the computation.	PTMA-E PMA-E

PTBISERIAL= X	PTMX PTMEASUR- EX	Point-measure correlation excluding the current observation from the estimation of the measure. Computes the Pearson point-measure correlation coefficients, r_{pm} between the observations and the measures or anchor values adjusted to exclude the current observation. Measures corresponding to extreme scores are included in the computation.	PTMX-E PMX-E
---------------	-------------------------	--	-----------------

Point-correlations are always reported for items. Point-correlations are reported for persons when (i) all the items are dichotomies, (ii) all the items have three categories, or (iii) all the items are in the same item group (Andrich Rating Scale Model). Otherwise, the observed and expected correlations are reported as zero in [PFILE=](#).

Here's how these correlations work:

Think of an item (or a person).

That item has a string of responses.

Each response ("point") is made by a person who has a raw score and a Rasch measure.

1. Correlate the raw scores with the responses. This is the point-biserial correlation (including the current response), PTBSA (PTBIS=All)
2. Correlate the raw scores (less the current response) with the responses. This is the point-biserial correlation corrected for auto-correlation, PTBSE. (PTBIS=Yes)
3. Correlate the Rasch measures (estimated including the current response) with the responses. This is the point-measure correlation, PTMEA. (PTBIS=No)
4. Correlate the Rasch measures (estimated without the current response) with the responses. This is the point-measure correlation: corrected for autocorrelation, PTMEX. (PTBIS=X)

Numerical example:

Person	Response to item	Measure	Raw Score	Raw score less current response	Measure (estimated without current response)
Jose	1	2.00	21	20	1.9
Mary	0	1.00	13	13	1.1
Robert	0	0.00	7	7	0.1
Point-measure correlation: PTBIS = N		0.87			
Point-biserial correlation (All responses): PTBIS = A			0.90		
Point-biserial correlation (Excluding current response): PTBIS= E				0.89	
Point-measure correlation (excluding current response from measure) : PTBIS = X					0.83

The "expected correlations" (the values of the correlations we would expect when the data fit the Rasch model perfectly) were introduced to avoid incorrect inferences based on the point-biserial correlations. For instance, eliminating items with low point-biserial correlations in situations where it is impossible for the point-biserial correlations to be high. In fact, without an "expected" reference point, it can be impossible to identify whether a reported correlation is too high, about right or too low.

For tests of different lengths including the same items we would first need to compare the items' "expected" correlation values, no matter which correlation we chose to report. This would provide the baseline for discussion about the idiosyncrasies of each item in each test.

Example 1: For [rank-order](#) or paired-comparison data, point-biserials are all -1. So specify Point-measure correlations.
PTBISERIAL=NO

Example 2: Winsteps results are to be compared to previous results for which the point-biserial correlation was based on the total marginal scores, including the responses to the targeted item.
PTBISERIAL=ALL

11.137 PVALUE= proportion correct or average rating

An important statistic in classical item analysis is the item proportion-correct (item-analysis p-value), the proportion of the sample succeeding on a dichotomous item. In Winsteps, this value is interpreted as the average of the responses to the item for both dichotomous and polytomous items.

This proportion-correct-value is not a hypothesis test probability-value (statistical p-value). For hypothesis tests of fit, use the ZSTD values, which are probability-values expressed as unit-normal deviates.

PVALUE = NO Do not report the item observed proportion-correct-value.

PVALUE=YES Report item observed proportion-correct-values in [Tables 6, 10](#), etc. and [IFILE=](#) (over-riding the existing [default](#) setting.)

Example: To parallel a classical analysis of an MCQ data set, it is desired to report the raw scores (including extreme persons and items), the point-biserials and the proportion-correct-values or average ratings.

TOTALSCORE= YES ; report original total raw scores

PTBISERIAL = YES ; report point-biserial, not point-measure correlations

PVALUE = YES ; report proportion-correct-values

```

+-----+
|ENTRY  TOTAL          | INFIT | OUTFIT |PTBSE|PROP-|
|NUMBER SCORE COUNT MEASURE ERROR|MNSQ  ZSTD|MNSQ  ZSTD|CORR.|VALUE| ITEM
+-----+
| 1    35    35   -6.59   1.85| MINIMUM ESTIMATED MEASURE | 1.00| 1= 1-4
| 4    32    35   -4.40   .81|.90   .0|.35   .8|.48|.91| 4= 1-3-4
| 12   6     35    2.24   .55|1.16  .6|1.06  .5|.26|.17| 12=1-3-2-4-3
| 18   0     35    6.13   1.84| MAXIMUM ESTIMATED MEASURE | .00| 18=4-1-3-4-2-1-4

```

11.138 PWEIGHT= person (case) weighting

PWEIGHT= allows for differential weighting of persons. The standard weights are 1 for all persons. To change the weighting of persons, specify [IWEIGHT=](#).

PWEIGHT of 2 has the same effect on item estimation as putting the person and responses into the analysis twice. It does not change a person's responses scored 0-1 into responses scored 0-2. When PWEIGHT is 0 for a person, the measure and fit statistics are reported for the person, but the person does not influence the measures or fit statistics of the other persons or items. PWEIGHT= applies to everything except the dimensionality computations ([Tables 23, 24](#)) where PWEIGHT= is set to 1 and IWEIGHT= is set to 1 for all persons and items.

Raw score, count, and standard error of measurement reflect the absolute size of weights as well as their relative sizes. Measure, infit and outfit and correlations are sensitive only to relative weights. If you want the standard error of the final weight-based measure to approximate the S.E. of the unweighted measure, then ratio-adjust case weights so that the total of the weights is equal to the total number of independent observations.

PWEIGHT= file name	weights are in a text file of format: (person entry number) (weight)
PWEIGHT = * (person entry number) (weight) *	in-line list of person-entry numbers and weights
PWEIGHT = \$S2W3 PWEIGHT = \$S2E4	weights in person labels which start at NAME1= using the column selection rules , e.g., starting in column S... with a width of W... or starting in column S and ending in column E. This can be expanded, e.g, PWEIGHT = \$S23W1+"."+\$S25W2 places the columns next to each other (not added to each other)
PWEIGHT = \$C2W3 PWEIGHT = \$C2E4	weights in data records using the column selection rules . column1 of the data recode is C1. The format is starting in column C... with a width of W... or starting in column C... and ending in column E. The columns can be anywhere in the data record

Example 1:

In a sample of 20 respondents, person 1 is to be given a weight of 2.5, all other persons have a weight of 1.

```
PWEIGHT=*
1 2.5
2-20 1
*
```

A better weighting, which would make the reported item standard errors more realistic by maintaining the original total sum of weights at 20, is:

```
PWEIGHT=*
1 2.33 ; 2.5 * 0.93
2-20 0.93 ; the sum of all weights is 20.0
*
```

ENTRY	RAW			MODEL	INFIT	OUTFIT	PTMEA				
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	WEIGH	KID
1	4	10	-1.20	.85	.65	-.9	.41	.3	.82	2.33	Richard M
2	7	10	1.70	1.17	.18	-1.2	.10	-.4	.94	.93	Tracie F
3	7	10	1.70	1.17	2.04	1.2	1.32	.7	.84	.93	Walter M

Example 2:

The data records contain the weights in columns 16-18 of the record. The person label starts in column 15, so this is also column 2 of the person label

```
NAME1 = 15 ; start of person label
NAMELEN = 20 ; length of person label
PWEIGHT= $S2W3 ; location in person label
&END
```

```
...
END NAMES
10110110011001 0.5 Person A
01001100011011 0.7 Person B
.....
```

or

```
NAME1 = 15 ; start of person label
NAMELEN = 20 ; length of person label
PWEIGHT= $C16W3 ; or $C16E18 ; column in data record also in person label
```

or

```
NAME1 = 20 ; start of person label in column 20 excluding the person weight
NAMELEN = 15 ; length of person label
PWEIGHT= $C16W3 ; or $C16E18 ; column in data record, not in person label
```

Example 3:

Person 4 is a dummy case, to be given weight 0. The measure and fit statistics are reported for person 4, but person 4 does not influence the measures or fit statistics of the other persons or items.

```
PWEIGHT=*
4 0 ; Person 4 has weight 0, other persons have standard weight of 1.
*
```

11.139 QCMLE = Quasi-CMLE estimates

QCMLE=Yes reports [quasi-CMLE estimates](#) in the measure tables.

For dichotomous data, CMLE item (delta) estimates can be deduced from the CMLE item probabilities for a person with score 1. Since JMLE probabilities are close to CMLE item probabilities, approximate item deltas, known as QCMLE estimates, can be calculated from JMLE probabilities. Similarly for person theta QCMLE estimates.

The theory of QCMLE is explained in a 2020 issue of [Rasch Measurement Transactions](#)

11.140 QUOTED= quote-marks around labels

Non-numeric values in the output files can be placed within quote-marks. This is required by some software in order to decode internal blanks within labels correctly. These apply to comma-separated and tab-separated output files.

QUOTED=Y "non-numeric values within quotation marks"

QUOTED=N non-numeric values without quotation marks.

Example: Produce an SFILE=

CSV=Y ; produce a comma-separated output file

QUOTED=Y ; with labels in quotation marks

```
" ;STRUCTURE MEASURE ANCHOR FILE"  
" ;CATEGORY", "Rasch-Andrich Threshold"  
0, .00  
1, -.86  
2, .86
```

QUOTED=N ; labels without quotation marks

```
;STRUCTURE MEASURE ANCHOR FILE  
;CATEGORY,Rasch-Andrich Threshold  
0, .00  
1, -.86  
2, .86
```

11.141 RARE= heading in Tables 1, 12, 16

Please see [MORE=](#)

11.142 RCONV= score residual at convergence

Scores increment in integers so that 0.1 is about as precise a recovery of observed data as can be hoped for.

Specifies what value the largest score residual, corresponding to any person measure or item calibration, must be less than in the iteration just completed for iteration to cease. The current largest value is listed in Table 0, and displayed on your screen. In large data sets, the smallest meaningful logit change in estimates may correspond to score residuals of several score points. See [convergence considerations](#).

The standard setting is **CONVERGE="E"**, so that iteration stops when either **LCONV=** or **RCONV=** is satisfied. (Note: this depends on Winsteps version - and may explain differences in converged values.)

Example: To set the maximum score residual, when convergence will be accepted, at 5 score points and maximum logit change in estimates of .01 logits. Your data consists of the responses of 5,000 students to a test of 250 items.

NI=250 ; 250 items

RCONV=5 ; score residual convergence at 5 score points

LCONV=.01 ; this is the standard.

CONVERGE=Both ;convergence when both RCONV= and LCONV= are met

11.143 REALSE= inflate S.E. for misfit

The modeled, REALSE=N, standard errors of measure estimates (abilities and difficulties) are the smallest possible errors. These always overstate the measurement precision.

Controls the reporting of standard errors of measures in all tables.

REALSE=N

Report modeled, asymptotic, standard errors (the standard).

REALSE=Y

Report the modeled standard errors inflated by the square root of the infit mean square, when it is greater than 1.0. This **inflates** the standard error to include uncertainty due to overall lack of fit of data to model.

See [Standard Errors: Model and Real](#) for more details.

11.144 RESCORE= response recoding

The responses in your data file may not be coded as you desire. The responses to some or all of the items can be rescored or keyed using RESCORE=. RESCORE= and [NEWSCORE=](#) are ignored when [KEYn=](#) is specified, except as below. If rescoring implies that the items have different rating (or partial credit) scale structures, [ISGROUPS=](#) may also be required.

RESCORE= *file name	file containing details
RESCORE = *	in-line list
RESCORE = codes	codes for item groups

RESCORE= has three forms: RESCORE=1101110 and RESCORE=* list * and RESCORE=*filename

RESCORE=" " or 2 or is omitted

All items are recoded using NEWSCORE=. RESCORE=2 is the standard when NEWSCORE= is specified.

RESCORE= some combination of 1's and 0's

Only items corresponding to 1's are recoded with NEWSCORE= or scored with KEYn=. When KEYn is specified, NEWSCORE= is ignored.

If some, but not all, items are to be recoded or keyed, assign a character string to RESCORE= in which "1" means "recode (key) the item", and "0" (or blank) means "do not recode (key) the item". The position of the "0" or "1" in the RESCORE= string must match the position of the item-response in the item-string.

When XWIDE=2 or more, then

either (a) Use one character per XWIDE and blanks,

NI=8

XWIDE=2

RESCORE=' 1 0 1 0 1 0 1 1'

or (b) Use one character per item with no blanks

NI=8

XWIDE=2

RESCORE='10101011'

Example 1: The original codes are "0" and "1". You want to reverse these codes, i.e., 1 0 and 0 1, for all items.

XWIDE=1 one character wide responses (the standard)

CODES =01 valid response codes are 0 and 1 (the standard)

NEWSCORE=10 desired response scoring

RESCORE=2 rescore all items - this line can be omitted

Example 2: Your data is coded "0" and "1". This is correct for all 10 items except for items 1 and 7 which have the reverse meaning, i.e. 1 0 and 0 1.

NI=10 ten items

CODES =01 standard, shown here for clarity

(a) old method - which still works:

NEWSCORE=10 revised scoring

RESCORE=1000001000 only for items 1 and 7

or

NEWSCORE=20

RESCORE=*

1 1 item 1 is to be rescored

7 1 item 7 is to be rescored

*

(b) new method - recommended:

IVALUE1 =10 revised scoring

IVALUE0 =01 scoring unchanged, so this line can be omitted.
IREFER =1000001000 only for items 1 and 7

If XWIDE=2, use one or two columns per RESCORE= code, e.g., " 1" or "1 " mean recode (key). " 0" or "0 " mean do not recode (key).

Example 3: The original codes are " 0" and " 1". You want to reverse these codes, i.e., 1 0 and 0 1, for items 1 and 7 of a ten item test.

NI =10 ; ten items
XWIDE =2 ; two characters wide
CODES =" 0 1" ; original codes
NEWSCORE=" 1 0" ; new values
RESCORE =" 1 0 0 0 0 0 1 0 0 0" ; rescore items 1 & 7

Example 4: The original codes are "0", "1", and "2". You want 0 0, 1 1, and 2 1 for all items

XWIDE=1 one character wide (standard)
CODES =012 valid codes
NEWSCORE=011 desired scoring

Example 5: The original codes are "0", "1", and "2". You want to make 0 2, 1 1, and 2 0, for even-numbered items in a twenty item test.

NI=20 twenty items
CODES =012 three valid codes
NEWSCORE=210 desired scoring
RESCORE=01010101010101010101 rescore "even" items

Example 6: The original codes are "0", "1", "2", "3" and some others. You want to make all non-specified codes into "0", but to treat codes of "2" as missing.

CODES = 0123 four valid codes
NEWSCORE= 01X3 response code 2 will be ignored
MISSCORE=0 treat all invalid codes as 0

Example 7: The original codes are "0", "1", "2", "3". You want to rescore some items selectively using KEY1= and KEY2= and to leave the others unchanged - their data codes will be their rating values. For items 5 and 6, 0 0, 1 0, 2 1, 3 2; for item 7, 0 0, 1 0, 2 0, 3 1. Responses to other items are already entered correctly as 0, 1, 2, or 3.

CODES =0123 valid codes
RESCORE=0000111000 rescore items 5,6,7
KEY1 =****223*** keyed for selected items
KEY2 =****33X*** the X will be ignored
^ read these columns vertically

Example 8: Multiple score key for items 1 to 10. Items 11 to 15 are on a rating scale of 1 to 5

CODES = abcd12345
KEY1 = bacdbaddcd*****
RESCORE= 111111111100000 ; RESCORE= signals when to apply KEY1=

11.145 RESFROM= location of RESCORE

Only use this if you have too many items to put conveniently on one line of the [RESCORE=](#) control variable.

RESFRM= instructs where to find the RESCORE= information.

RESFRM=N
RESCORE= is a control variable between before &END (the standard).

RESFRM=Y
RESCORE= information follows after &END but before the item names, if any, and is formatted exactly like a data record. It is helpful, for reference, to enter the label "RESCORE=" where the person name would go in your data record.

Example: KEY1= and KEY2= information follows the RESCORE= information, all are formatted like your data. No item names are provided,

```

NAME1 = 1 start of person-id
ITEM1 = 10 start of responses
NI = 10 ten items
INUMB = Y use item sequence numbers as names
CODES = ABCDE valid codes
RESFRM = Y rescore information in data record format
KEYFRM = 2 two keys in data record format
&END
RESCORE= 0000110000 RESCORE= looks like data
KEY1= ****AB**** KEY1= looks like data
KEY2= ****CA**** KEY2= looks like data record
George ABCDABCDAB first data record
| subsequent data records

```

11.146 RFILE= scored response file

Useful for reformatting data from a family of test forms, linked by a network of common items, into a single common structure suitable for one-step item banking.

RFILE=filename outputs a file which contains a scored/keyed copy of the input data. This file can be used as input for later analyses. Items and persons deleted by [PDFILE=](#) or the like are replaced by blank rows or columns in the scored response file. The file format is:

1. Person-id: 30 columns wide or maximum person-id length, whichever is larger.
2. Responses one per item:
 One column for each response if largest scored response is less than or equal to 9.
 Two columns for each responses if largest scored response is more than 9.
 Missing or unscored data are indicated by "."
 The width of the responses is not determined by XWIDE=.

RFILE=? opens a Browse window

```

M Rossner, Marc Daniel      1.1110.01...0.11..0.10.0
M Rossner, Lawrence F.     .....
M Rossner, Toby G.        ..1101101...1..1..1.1111
M Rossner, Michael T.     10100101..1..1111..0.1111
F Rossner, Rebecca A.     10101010011110011..111110
M Rossner, Tr Cat         1011.110111.10101..101.10
M Wright, Benjamin        ...00...0.....
M Lambert, Md., Ross W.   ..100.10....0.01..0..0..
M Schulz, Matthew         011010011.1..0011.11.0.11
M Hsieh, Daniel Seb       .1000101...11.11..0110.1

```

Example: I have 9 scored items. for example LP1, LP2, CS1, CS2, CS3, D1, R1, R2, TE1. I would like to combine the following into one subtest (LP1 and LP2) and (CS1,CS2,CS3) and (R1 and R2) using Winsteps. D1 and TE1 will remain unchanged. This means I will ultimately have 5 items LP, CS, D, R, and TE.

Winsteps cannot combine items automatically. So here is a suggestion if your data file is already in Winsteps format:

1. Launch Winsteps
2. Analyze your 9-item file
3. Winsteps menu bar, "Output Files", Output the RFILE= to Excel
4. In Excel, combine your items into subtests
5. Winsteps menu bar, "Data Setup", convert your Excel file into a new Winsteps control file
6. Analyze the new Winsteps control file.

11.147 RMSR= report root-mean-square residuals in measure tables

RMSR=Yes reports the summary root-mean-square residual (= observation - expectation) for each person or item in the measure tables.. Observations in extreme scores are excluded.

INFIT		OUTFIT		FIT-MEASURE		EXACT MATCH		RMSR	ITEM
MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%		
1.23	1.2	1.09	.5	.83	.84	42.9	50.9	.855	A. EATING
1.51	2.6	1.63	3.1	.79	.85	37.7	44.4	1.023	B. GROOMING
.85	-.9	.82	-1.0	.88	.87	48.6	42.8	.802	C. BATHING
.69	-2.0	.73	-1.7	.91	.86	42.9	44.0	.704	D. UPPER BODY DRESSING
.57	-3.0	.56	-2.9	.92	.87	64.3	43.3	.656	E. LOWER BODY DRESSING
.71	-1.9	.71	-1.8	.89	.86	54.3	44.1	.732	F. TOILETING
2.03	4.7	2.08	4.9	.76	.86	34.3	44.1	1.216	G. BLADDER
1.50	2.6	1.53	2.7	.78	.85	37.1	44.1	1.034	H. BOWEL
.52	-3.4	.60	-2.7	.92	.86	55.7	43.9	.616	I. BED TRANSFER
.49	-3.7	.53	-3.3	.91	.86	54.3	44.0	.608	J. TOILET TRANSFER
.85	-.9	.85	-.8	.88	.88	54.3	45.2	.796	K. TUB, SHOWER
.91	-.5	.80	-1.2	.88	.87	51.4	44.1	.830	L. WALK/WHEELCHAIR
1.02	.1	.83	-.8	.90	.88	44.3	51.3	.806	M. STAIRS

The blue column is the root-mean-square residual. The red box shows how it relates to item misfit [Table 14.1](#).

11.148 ROW1HEADING = heading in first row of output file

ROW1HEADING = Yes (default)

Most output files have a descriptive comment as the first line in the file. Column labels (if any) are in the second line.

Example: `IFILE=`

Test (Best Test Design p.31)

```
;ENTRY MEASURE ST COUNT SCORE MODLSE IN.MSQ
1 -60.86 -1 35.0 35.0 18.48 1.00
```

ROW1HEADING = No

Output files do have a descriptive comment as the first line in the file. Column labels (if any) are in the first line.

`HLINES=`No removes the descriptive comment and also any column labels.

Example: `IFILE=`

```
;ENTRY MEASURE ST COUNT SCORE MODLSE IN.MSQ
1 -60.86 -1 35.0 35.0 18.48 1.00
```

11.149 SAFILE= structure-threshold input anchor file

The `SAFILE=` (not `ISFILE=`, but see [PCM anchoring](#)) of one analysis may be used unedited as the `SAFILE=` of another.

The rating-scale structure parameter values (taus, Rasch-Andrich thresholds, steps) can be anchored (fixed) using `SAFILE=`. The anchoring option facilitates test form [equating](#). The structure in the rating (or partial credit) scales of two test forms, or in the item bank and in the current form, can be anchored at their other form or bank values. Then the common rating (or partial credit) scale calibrations are maintained. Other measures are [estimated](#) in the frame of reference defined by the anchor values. Use `IAFILE=` and `SAFILE=` if you need the polytomous item in one analysis to be identical in thresholds and overall difficulty to the same item in another analysis. Use only `SAFILE=` if you need the polytomous item in one analysis to be identical in thresholds to the same item in another analysis, but the overall item difficulties can differ.

<code>SAFILE=</code> file name	file containing details
<code>SAFILE = *</code>	in-line list
<code>SAFILE = ?</code>	opens a Browser window to find the file
No <code>ISGROUPS=</code> or all items in one group	
(bottom category) 0 example: 0 0	place holder for bottom category of rating scale in case it is not observed in the data

(category number) (anchor value) example: 2 1.5	Andrich threshold (step calibration) between categories 1 and 2 is anchored at 1.5 for all items (unless overridden)
ISGROUPS= specifies more than one group of items or PCM	
(item number) (bottom category) 0 example: 34 0 0 ; place holder for bottom category	For item 34 and all items in the same ISGROUPS= group, place holder for bottom category in case it is not observed in the data
(item number) (category number) (anchor value) example: 34 2 1.5 ; Andrich threshold	For item 34 and all items in the same ISGROUPS= group, Andrich threshold (step calibration) between categories 1 and 2 is anchored at 1.5
(item number-item number) (category number) (anchor value) example: 34-39 2 1.5	For items 34 to 39 and all items in the same ISGROUPS= group(s), Andrich threshold (step calibration) between categories 1 and 2 is anchored at 1.5
(1-NI=) (category number) (anchor value) example: 1-47 1 2.0	Specify the default value for a category threshold for all items (and so for all ISGROUPS= groups)
*	end of list

In order to anchor category structures, an anchor file must be created of the following form:

1. Use one line per category Rasch-Andrich threshold to be anchored.
2. If all items use the same rating scale (i.e. [ISGROUPS=](#) " ", the standard, or you assign all items to the same grouping, e.g. [ISGROUPS=222222..](#)), then type the category number, a blank, and the "structure measure" value (in logits or your user-rescaled units) at which to anchor the Rasch-Andrich threshold measure corresponding to that category (see [Table 3.2](#)). [Arithmetical expressions](#) are allowed.
3. If you wish to force category 0 to stay in an analysis, anchors its calibration at 0. Specify [SAITEM=Yes](#) to use the multiple ISGROUP= format
or
If items use different rating (or partial credit) scales (i.e. [ISGROUPS=0](#), or items are assigned to different groupings, e.g. [ISGROUPS=122113..](#)), then type the sequence number of any item belonging to the grouping, a blank, the category number, a blank, and the "structure measure" value (in logits if [USCALE=1](#), otherwise your user-rescaled units) at which to anchor the Rasch-Andrich threshold up to that category for that grouping. If you wish to force category 0 to stay in an analysis, anchor its calibration at 0.

This information may be entered directly in the control file using [SAFILE=*](#)

Anything after ";" is treated as a comment.

Example 1: Dichotomous: A score of, say, 438 means that you have 62% odds (and not 50% as it is default in Winsteps/Ministep!) of answering correctly to a dichotomous item of difficulty 438. How can I set this threshold from 50% to 62%?

In your control file, include:
 UASCALE=1 ; anchoring is in logits
 SAFILE=* ; anchors the response structure
 0 0 ; place holder for bottom category
 1 -0.489548225 ; $\ln((100\%-62\%)/62\%)$
 *

When you look at Table 1, you should see that the person abilities are now lower relative to the item difficulties.

Example 2: Polytomous: A score of, say, 438 means that you have 62% expectation of answering correctly to a polytomous item (0-1-2-3) of difficulty 438. How can I set the thresholds to 62%?

The default item difficulty for a polytomy is the point where the lowest and highest categories are equally probable. We need to make a logit adjustment to all the category thresholds equivalent to a change of difficulty corresponding to a rating of

.62*3 = 1.86.

This is intricate:

1. We need the current set of Rasch-Andrich thresholds (step calibrations) = F1, F2, F3.
2. We need to compute the measure (M) corresponding to a score of 1.86 on the rating scale
3. Then we need to anchor the rating scale at:

SAFILE=*

0 0 ; place holder for bottom category

1 F1 - M

2 F2 - M

3 F3 - M

*

An easy way to obtain M is to produce Winsteps "Output Files" menu, [GRFILE=](#) and then look up the Measure for the Score you want.

Example 3: A rating scale, common to all items, of three categories numbered 2, 4, and 6, is to be anchored at pre-set calibrations. The calibration of the Rasch-Andrich threshold from category 2 to category 4 is -1.5, and of the Rasch-Andrich threshold to category 6 is +1.5.

1. Create a file named, say, "STANC.FIL"

2. Enter the lines

2 0 ; place holder for bottom category of this rating scale

4 -1.5 ; Rasch-Andrich threshold from category 2 to category 4, anchor at -1.5 logits

6 1.5 ; Rasch-Andrich threshold from category 4 to category 6, anchor at +1.5 logits

Note: categories are calibrated pair-wise, so the Rasch-Andrich threshold values do not have to advance.

3. Specify, in the control file,

ISGROUPS="" (the standard)

SAFILE=STANC.FIL structure anchor file

or, enter directly in the control file,

SAFILE=*

4 -1.5

6 1.5

*

If you wish to use the multiple grouping format, i.e., specify an example item, e.g., 13

SAITEM=YES

SAFILE=*

13 4 -1.5

13 6 1.5

*

To check this: "A" after the Andrich threshold measure

```
+-----+
|CATEGORY  OBSERVED|OBSVD SAMPLE|INFIT  OUTFIT|| ANDRICH |CATEGORY|
|LABEL SCORE COUNT %|AVRGE EXPECT|  MNSQ  MNSQ||THRESHOLD| MEASURE|
+-----+-----+-----+-----+-----+-----+
|  4    4      620  34|   .14   .36|   .87   .72||  -1.50A|   .00 |
```

Example 4: A partial credit analysis (ISGROUPS=0) has a different rating scale for each item. Item 15 has four categories, 0,1,2,3 and this particular response structure is to be anchored at pre-set calibrations.

1. Create a file named, say, "PC.15"

2. Enter the lines

15 0 0 Bottom categories are always at logit 0

15 1 -2.0 item 15, Rasch-Andrich threshold to category 1, anchor at -2 logits

15 2 0.5

15 3 1.5

3. Specify, in the control file,

ISGROUPS=0

SAFILE=PC.15

IAFILE= file of item calibrations

Example 5: A grouped rating scale analysis (ISGROUPS=21134..) has a different rating scale for each grouping of items. Item 26 belongs to grouping 5 for which the response structure is three categories, 1,2,3 and this structure is to be anchored, but the difficulties of the individual items are to be re-estimated.

1. Create a file named, say, "GROUPING.ANC"
2. Enter the lines
 26 2 -3.3 for item 26, representing grouping 5, Rasch-Andrich threshold to category 2, anchored at -3.3
 26 3 3.3
3. Specify, in the control file,
 ISGROUPS =21134..
 SAFILE=GROUPING.ANC
 ; there is no IAFILE= because we want to re-estimate the item difficulties

Example 6: A partial-credit scale has an unobserved category last time, but we want to use those anchor values where possible.
 We have two choices.

- a) Treat the unobserved category as a structural zero, i.e., unobservable. If so...
 Rescore the item using IVALUE=, removing the unobserved category from the category hierarchy, and use a matching SAFILE=.

In the run generating the anchor values, which had STKEEP=NO,

```

+-----+
|CATEGORY  OBSERVED|OBSVD  SAMPLE|INFIT  OUTFIT|| ANDRICH |CATEGORY|
|LABEL SCORE COUNT %|AVRGE EXPECT|  MNSQ  MNSQ||THRESHOLD| MEASURE|
+-----+
| 1 1 33 0| -.23 -.15| .91 .93|| NONE |( -.85)| 1
| 2 2 23 0| .15 .05| .88 .78|| -1.12 | 1.44 | 2
| 4 3 2 0| .29 .17| .95 .89|| 1.12 |( 3.73)| 4
+-----+
  
```

In the anchored run:

```

IREFER=A..... ; item 1 is an "A" type item
CODES=1234 ; valid categories
IVALUEA=12*3 ; rescore "A" items from 1,2,4 to 1,2,3
SAFILE=*
1 1 .00 ; place holder for bottom category
1 2 -1.12
1 3 1.12
*
  
```

If the structural zeroes in the original and anchored runs are the same then, the same measures would result from:

```

STKEEP=NO
SAFILE=*
1 1 .00 ; place holder for bottom category
1 2 -1.12
1 4 1.12
*
  
```

- b) Treat the unobserved category as an incidental zero, i.e., very unlikely to be observed.

Here is Table 3.2 from the original run which produced the anchor values. The [NULL](#) indicates an incidental or sampling zero.

```

+-----+
|CATEGORY  OBSERVED|OBSVD  SAMPLE|INFIT  OUTFIT|| ANDRICH |CATEGORY|
|LABEL SCORE COUNT %|AVRGE EXPECT|  MNSQ  MNSQ||THRESHOLD| MEASURE|
+-----+
| 1 1 33 0| -.27 -.20| .91 .95|| NONE |( -.88)| 1
| 2 2 23 0| .08 -.02| .84 .68|| -.69 | .72 | 2
| 3 3 0 0| | .00 .00|| NULL | 1.52 | 3
| 4 4 2 0| .22 .16| .98 .87|| .69 |( 2.36)| 4
+-----+
  
```


Here is the matching SAFILE=

SAFILE=*

- 1 1 .00 ; place holder for bottom category
 - 1 2 -.69
 - 1 3 46.71 ; flag category 3 with a large positive value, i.e., unlikely to be observed.
 - 1 4 -46.02 ; maintain sum of Andrich thresholds (step calibrations) at zero.
- *

Example 7: Partial-credit Item difficulty is are to be set at an expected score of 1.3333 for an item scored 0,1,2 :

1. Do the standard analysis with [UPMEAN=0](#). Center the person abilities, so we can see the change in item difficulties later.
2. Output the [SFILE=](#) to Excel of Rasch-Andrich thresholds
3. Output the [GRFILE=](#) to Excel of the ICCs, item characteristic curves
4. From the GRFILE=, discover the measure for each item corresponding to 1.3333
5. Output the [IFILE=](#) to Excel showing the item difficulties
6. Subtract the item difficulty from the 1.3333 measure. This is the necessary shift in the item difficulty
7. Subtract this shift from every threshold for the item in the SFILE=
8. Copy-and-paste-text the shifted thresholds into a text-file SAFILE=
9. Reanalyze the data with the SAFILE= and UPMEAN=0. All the item difficulties should now be at their RP67 values, relative to the mean of the person abilities.

Example 8: Score-to-measure Table 20 is to be produced from known item and rating scale structure difficulties.

Specify:

IAFILE= ; the item anchor file

SAFILE= ; the structure/step anchor file (if not dichotomies)

CONVERGE=L ; only logit change is used for convergence

LCONV=0.005 ; logit change too small to appear on any report.

STBIAS=NO ; anchor values do not need estimation bias correction.

The data file comprises two dummy data records, so that every item has a non extreme score, e.g.,

For dichotomies:

Record 1: 10101010101

Record 2: 01010101010

For a rating scale from 1 to 5:

Record 1: 15151515151

Record 2: 51515151515

Redefining the Item Difficulty of Rating Scale items:

We want to define the difficulty of an item as 65% success on the item, instead of the usual approximately 50% success.

1. Suppose we have these Rasch-Andrich thresholds (step calibrations) from a standard rating-scale analysis:

Category	Rasch-Andrich Threshold
1	(0.00)
2	-.98
3	-.25
4	1.22

2. The item score range is 1-4, so

a) we need the relative measure corresponding to an expected score of 65% on the item = $1 + (4-1) \cdot 0.65 = 2.95$

3. We look at the [GRFILE=](#) and see that the measure corresponding to an expected score of 2.95 is about 0.58 (we can verify this by looking at the Graphs window, [Expected score ICC](#))

ITEM	MEAS	SCOR	INFO	0	1	2	3
1	.48	2.89	.67	.05	.23	.48	.23
1	.56	2.94	.65	.05	.22	.49	.25
1	.64	2.99	.63	.04	.20	.49	.27

4. We want the item difficulty to correspond to 65% success instead of its current approximately 50% correct. So we have raised the bar for the item. The item is to be reported as about 0.57 logits more difficult.
5. To force the item to be reported as 0.57 logits more difficult, we need the Andrich thresholds (step calibrations) to be 0.57 logits **easier** = -0.57 logits.

Category	Rasch-Andrich Threshold
1	(0.00)
2	$-.98 + -.57 = -1.55$
3	$-.25 + -.57 = -.82$
4	$1.22 + -.57 = .64$

6. Now, since the item mean remains 0, all the person measures will be reduced by 0.57 logits relative to their original values.

Dichotomies (MCQ, etc.) Mastery Levels:

Example 9: To set mastery levels at 75% on dichotomous items (so that maps line up at 75%, rather than 50%), we need to adjust the item difficulties by $\ln(75/(100-75)) = 1.1$ logits.

SAFILE=*

0 0 ; place holder for bottom category

1 -1.1 ; set the Rasch-Andrich threshold point 1.1 logits down, so that the person ability matches item difficulty at 75% success.

; If you are using [USCALE=](#), then the value is $-1.1 * USCALE=$

*

Similarly for 66.67% success or 66.67% master level: $\ln(66.67/(100-66.67)) = 0.693$ logits.

SAFILE=*

0 0

0 -0.6931 ; notice that this is negative

*

Similarly for 65% success or 65% master level: $\ln(65/(100-35)) = 0.691$ logits.

SAFILE=*

0 0 ; place holder for bottom category

0 -0.691 ; notice that this is negative

*

Polytomies (rating scales, partial credit, etc.):

When a variety of rating (or partial credit) scales are used in an instrument, their different formats perturb the item hierarchy. This can be remedied by choosing a point along each rating (or partial credit) scale that dichotomizes its meaning (not its scoring) in an equivalent manner. This is the pivot point. The effect of pivoting is to move the structure calibrations such that the item measure is defined at the pivot point on the rating (or partial credit) scale, rather than the standard point (at which the highest and lowest categories are equally probable).

Example 1. Anchoring polytomous items for the Rating Scale Model

CODES = 012 ; 3 category Rating Scale Model

IAFILE=*

0 0 ; place holder for bottom category

1 2.37 ; anchor item 1 at 2.37 logits

```
2 -1.23
*
```

```
SAFILE=*
```

```
0 0 ; the bottom category is always anchored at 0
1 -2.34 ; Andrich threshold (step difficulty) from category 0 to 1
2 2.34 ; Andrich threshold (step difficulty) from category 2 to 3
*
```

Example 2. Anchoring polytomous items for the Partial Credit and Grouped-Items models

```
CODES = 012 ; 3 category Rating Scale Model
```

```
ISGROUPS=0
```

```
IAFILE=*
```

```
0 0 ; place holder for bottom category
1 2.37 ; anchor item 1 at 2.37 logits
2 -1.23
*
```

```
SAFILE=*
```

```
; for item 1, relative to the difficulty of item 1
1 0 0 ; the bottom category is always anchored at 0
1 1 -2.34 ; Andrich threshold (step difficulty) from category 0 to 1
1 2 2.34 ; Andrich threshold (step difficulty) from category 2 to 3
; for item 2, relative to the difficulty of item 2
2 0 0 ; the bottom category is always anchored at 0
2 1 -1.54 ; Andrich threshold (step difficulty) from category 0 to 1
2 2 1.54 ; Andrich threshold (step difficulty) from category 2 to 3
*
```

Here is a general procedure.

Use [ISGROUPS=](#)

Do an unanchored run, make sure it all makes sense.

Write out an [SFILE=structure.txt](#) of the rating scale (partial credit) structures.

Calculate, for each item, the amount that you want the item difficulty to move. Looking at the [Graphs](#) menu or [Table 2](#) may help you decide.

Make this amount of adjustment to every value for the item in the SFILE=*

So, suppose you want item 3 to be shown as 1 logit more difficult on the item reports.

The SFILE=structure.txt is

```
3 0 0.0 ; place holder for bottom category
3 1 -2.5
3 2 -1.0
```

```
...
```

```
*
```

Change this to (add 1 to the values for 1 logit more difficult)

```
3 0 0 ; place holder for bottom category
3 1 -1.5
3 2 -0.0
```

```
...
```

```
*
```

This becomes the SAFILE=structure.txt of the pivoted analysis.

Example 10: Pivoting with ISGROUPS=. Positive (P) items pivot at an expected score of 2.5. Negative (N) items at an expected score of 2.0

```
ISGROUPS=PPPPPNNNNN
```

```
SAFILE=*
```

```
1 2 0.7 ; put in the values necessary to move the center to the desired spot
5 2 0.5 ; e.g., the "structure calibration" - "score-to-measure of pivot point"
```

```
*
```

Example 11: To set a rating (or partial credit) scale turning point: In the Liking for Science, with 0=Dislike, 1=Neutral, 2=Like, anything less than an expected score of 1.5 indicates some degree of lack of liking:

SAFILE=*

1 -2.22 ; put in the Andrich threshold (step calibration) necessary to move expected rating of 1.5 to the desired spot

```

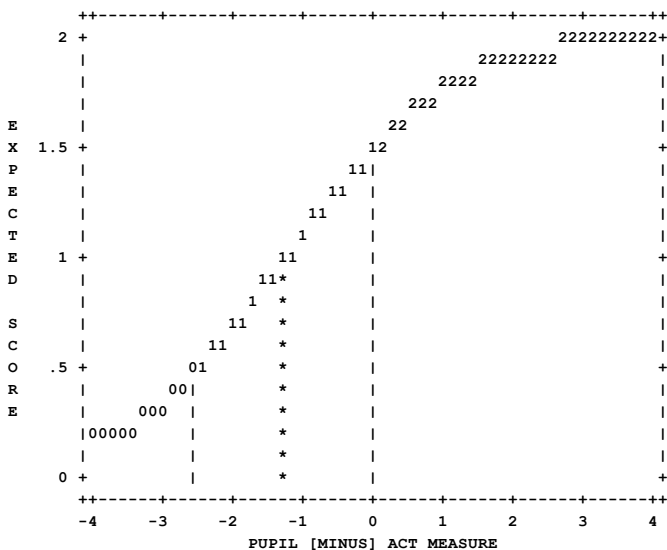
RATING SCALE PIVOTED AT 1.50
+-----+
|CATEGORY  OBSERVED|OBSVD  SAMPLE|INFIT  OUTFIT|| ANDRICH |CATEGORY|
|LABEL SCORE COUNT %|AVRGE  EXPECT|  MNSQ  MNSQ||THRESHOLD| MEASURE|
+-----+
|  0  0    197 22|-2.29 -2.42|  1.05  .99||  NONE  |(-3.42)| dislike
|  1  1    322 36|-1.17 -.99|   .90  .79|| -2.22  |-1.25| neutral
|  2  2    368 41|  .89  .80|   .98  1.29||  -.28  |(.92)| like
+-----+
|MISSING      1  0|  .04  |          ||          |
+-----+
AVERAGE MEASURE is mean of measures in category.

+-----+
|CATEGORY  STRUCTURE  | SCORE-TO-MEASURE |CUMULATIV| COHERENCE|
| LABEL   MEASURE   S.E. | AT CAT.  ---ZONE---|PROBABLTY| M->C C->M|
+-----+
|  0      NONE      | (-3.42) -INF  -2.50|          | 63% 44%| dislike
|  1    -2.22     .10 | -1.25 -2.50  .00| -2.34 | 55% 72%| neutral
|  2     -.28     .09 | (.92)  .00 +INF |  -.16 | 84% 76%| like
+-----+

```

Values of .00 for scores of 1.5 show effect of pivot anchoring on the rating (or partial credit) scale. The structure calibrations are offset.

TABLE 21.2 LIKING FOR SCIENCE (Wright & Masters p.18) sf.out Aug 1 21:31 2000
 EXPECTED SCORE OGIVE: MEANS



Example 12: A questionnaire includes several rating (or partial credit) scales, each with a pivotal transition-structure between two categories. The item measures are to be centered on those pivots.

1. Use ISGROUPS= to identify the item response-structure groupings.

2. Look at the response structures and identify the pivot point:

e.g., here are categories for "grouping A" items, after rescaling, etc.

Strongly Disagree 1

Disagree 2

Neutral 3

Agree 4

Strongly Agree 5

If agreement is wanted, pivot between 3 and 4, identified as transition 4.

If no disagreement is wanted, pivot between 2 and 3, identified as transition 3.

3. Anchor the transition corresponding to the pivot point at 0, e.g., for agreement:

e.g., for

```
ISGROUPS=AAAAAABBBBAACCC
```

```
SAFILE=*
```

6 4 0 6 is an item in grouping A, pivoted at agreement (Rasch-Andrich threshold from category 3 into category 4)

8 2 0 8 is an item in grouping B, pivoted at Rasch-Andrich threshold from category 2 into category 3

; no pivoting for grouping C, as these are dichotomous items

*

Example 13: Anchor files for dichotomous and partial credit items. Use the IAFILE= for anchoring the item difficulties, and SAFILE= to anchor partial credit structures. Winsteps decomposes the (delta) D_{ij} of partial credit items into $D_i + F_{ij}$.

The D_i for the partial credit and dichotomous items are in the IAFILE=

The F_{ij} for the partial credit files are in the SAFILE=

Suppose the data are A,B,C,D, and there are two partial credit items, scored 0,1,2, and two merely right-wrong. 0,1 then :

```
CODES=ABCD
```

```
KEY1=BCBC ; SCORE OF 1 ON THE 4 ITEMS
```

```
KEY2=DA** ; SCORE OF 2 ON THE PARTIAL CREDIT ITEMS
```

```
ISGROUPS=0
```

If the right-wrong MCQ items are to be scored 0,2, then

```
CODES=ABCD
```

```
KEY1=BC** ; SCORE OF 1 ON THE 4 ITEMS
```

```
KEY2=DABC ; SCORE OF 2 ON THE PARTIAL CREDIT ITEMS
```

```
ISGROUPS=0
```

but better psychometrically is:

```
CODES=ABCD
```

```
KEY1=BCBC ; SCORE OF 1 ON THE 4 ITEMS
```

```
KEY2=DA** ; SCORE OF 2 ON THE PARTIAL CREDIT ITEMS
```

```
IWEIGHT=*
```

```
3-4 2 ; items 3 and 4 have a weight of 2.
```

*

```
ISGROUPS=0
```

Then write out the item and partial credit structures

```
IFILE= items.txt
```

```
SFILE=pc.txt
```

In the anchored run:

```
CODES= ... etc.
```

```
IAFILE=items.txt
```

```
SAFILE=pc.txt
```

```
CONVERGE=L ; only logit change is used for convergence
```

```
LCONV=0.005 ; logit change too small to appear on any report.
```

Anchored values are marked by "A" in the Item Tables, and also Table 3.2

Anchoring with Partial-Credit Delta δ_{ij} (D_{ij}) values

Example:

```
Title = "Partial credit with anchored  $D_{ij}$  structures"
;-----
;          STRUCTURE MEASURE (Andrich threshold)
;          -----
;Item i delta_i1  delta_i2
;-----
;Item 1   -3.0    -2.0
;Item 2   -2.0     1.0
```

```

;Item 3    0.0    2.0
;Item 4    1.0    3.0
;Item 5    2.0    3.0
;-----

Item1 = 11 ; observations start in column 11
NI=5 ; 5 items
Name1 = 1 ; person label in column 1
CODES = 012 ; valid data values
ISGROUPS = 0 ; partial-credit model

IAFILE=*
1-5 0 ; item difficulties for all items set at 0
*

SAFILE=*
1 0 0 ; this is a placeholder for data code 0 for item 1
1 1 -3.0 ; value from ISFILE= I+THRESH column
1 2 -2.0
2 0 0
2 1 -2.0
2 2 1.0
3 0 0
3 1 0.0
3 2 2.0
4 0 0
4 1 1.0
4 2 3.0
5 0 0
5 1 2.0
5 2 3.0
*

&END
END LABELS
Person 1 22111
Person 2 21010

```

Equating with Partial Credit (PCM) items

Question: My data has three time-points, and I want to compare the item difficulty hierarchies across time-points, but the PCM thresholds are different at each time-point. What should I do?

Answer: Equating with PCM is always problematic. The threshold estimates are highly influenced by idiosyncrasies in the local dataset. Since Rasch findings are usually based on person estimates, and these are based on the item ICCs (expected scores on the items), then it really makes more sense to compare the ICCs than the thresholds.

Ben Wright's recommendation was to analyze all the data together to obtain the best compromise for the thresholds, and then anchor the thresholds at those values for the analysis of each time-point separately. See www.rasch.org/rmt/rmt101f.htm stage II. Accordingly the SFILE=sf.txt from the joint analysis becomes the SAFILE=sf.txt for all the separate time-point analyses.

Another approach is to treat each PCM item essentially as a dichotomy by choosing one Andrich threshold as the "pivot" threshold for each item, and then anchoring this threshold of each item at 0. Then the item difficulties are forced to conform with this threshold value at all time-points. Accordingly SAFILE= 0 for the pivot thresholds of all the items (other thresholds are not anchored) in all the time-point analyses. For example, in all the analyses:

```

ISGROUPS=0 ; Partial Credit Model
SAFILE=*
1 3 0 ; for item 1, anchor thresholds between categories 2 and 3 at 0
2 2 0 ; for item 2, anchor thresholds between categories 1 and 2 at 0
3 2 0 ; for item 3, anchor thresholds between categories 1 and 2 at 0
.....
*

```

11.150 SAITEM= item numbers in SAFILE with one grouping

Rating-scale Structure-Threshold Files [SFILE=](#) and Structure-Threshold Anchor Files [SAFILE=](#) (Rating (or partial credit) scale step anchoring files) have the format:

For only one active grouping in ISGROUPS= , or no ISGROUPS=		
scored-category number	Andrich threshold relative to item difficulty (step-calibration)	
For multiple groupings in ISGROUPS=, or SAITEM= Yes		
example-item-number in grouping	scored-category number	Andrich threshold relative to item difficulty (step-calibration)

Example 1: "Liking for Science" structure-threshold anchor file:

Standard format for the Rating Scale model:

ISGROUPS= ; no groupings specified

SAFILE=*

0 0 ; bottom category threshold is 0 as a place-holder

1 -1.8 ; Rasch-Andrich threshold from category 0 to category 1

2 1.8 ; Rasch-Andrich threshold from category 1 to category 2

*

Alternative format allowing an example item number to identify the grouping, say item 10

SAITEM=Yes

SAFILE=*

10 0 0 ; bottom category threshold is 0 as a place-holder

10 1 -1.8

10 2 1.8

*

11.151 SANCHQ= anchor category structure interactively

If your system is interactive, steps to be anchored can be entered interactively by setting SANCHQ=Y before the &END line. If you specify this, you will be asked if you want to anchor any steps. If you respond "yes", it will ask if you want to read these anchored items from a file; if you answer "yes" it will ask for the file name and process that file in the same manner as if [SAFILE=](#) had been specified. If you answer "no", you will be asked to enter the step measures (found in Table 3).

If there is only one rating (or partial credit) scale, enter the category numbers for which the Rasch-Andrich thresholds are to be anchored, one at a time, along with their logit (or user-rescaled by [USCALE=](#)) Andrich threshold (structure measure) calibrations. Bypass categories without measures. Enter 0 where there is a measure of "NONE". When you are finished, enter -1 in place of the category number.

If there are several rating (or partial credit) scales, enter one of the item numbers for each rating (or partial credit) scale, then the Andrich thresholds corresponding to its categories. Repeat this for each category of an item for each rating (or partial credit) scale. Enter 0 where there is a Andrich Threshold for a category of "NONE". Entering 0 as the item number completes anchoring.

Example 1: You are doing a number of analyses, anchoring the common rating (or partial credit) scale to different values each time. You want to enter the numbers at your PC:

SANCHQ=Y

You want to anchor items 4 and 8.

Winsteps asks you:

DO YOU WANT TO ANCHOR ANY STRUCTURES? respond YES(Enter)

DO YOU WISH TO READ THE ANCHORED STRUCTURES FROM A FILE?

respond NO(Enter)

INPUT STRUCTURE TO ANCHOR (-1 TO END):

respond 2(Enter) (the first category Rasch-Andrich threshold to be anchored)

```

INPUT VALUE AT WHICH TO ANCHOR Rasch-Andrich threshold:
respond 0(Enter) (the first anchor value)
INPUT Rasch-Andrich threshold TO ANCHOR (-1 TO END): 4(Enter)
INPUT VALUE AT WHICH TO ANCHOR Rasch-Andrich threshold:-1.5(Enter)
INPUT Rasch-Andrich threshold TO ANCHOR (-1 TO END): 6(Enter)
INPUT VALUE AT WHICH TO ANCHOR Rasch-Andrich threshold:1.5(Enter)
INPUT Rasch-Andrich threshold TO ANCHOR (-1 TO END): -1(Enter) (to end anchoring)

```

Example 2: You wish to enter the Andrich Thresholds for several rating scales, each comprising a grouping of items: SANCHQ=Y

```

Winsteps asks you:
DO YOU WANT TO ANCHOR ANY Rasch-Andrich thresholds? YES(Enter)
DO YOU WANT TO READ THE ANCHORED Rasch-Andrich thresholds FROM A FILE? NO
Item 1 represents the first grouping of items, sharing a common rating scale:
INPUT AN ITEM, REPRESENTING A GROUPING (0 TO END): 1
INPUT Rasch-Andrich threshold TO ANCHOR (-1 TO END): 0 bottom category
INPUT VALUE AT WHICH TO ANCHOR Rasch-Andrich threshold: 0 "NONE"
INPUT AN ITEM, REPRESENTING A GROUPING (0 TO END): 1
INPUT Rasch-Andrich threshold TO ANCHOR (-1 TO END): 1
INPUT VALUE AT WHICH TO ANCHOR Rasch-Andrich threshold: -0.5
INPUT AN ITEM, REPRESENTING A GROUPING (0 TO END): 1
INPUT Rasch-Andrich threshold TO ANCHOR (-1 TO END): 2
INPUT VALUE AT WHICH TO ANCHOR Rasch-Andrich threshold: 0.5
Item 8 represents the second grouping of items, sharing a common rating scale:
INPUT AN ITEM, REPRESENTING A GROUPING (0 TO END): 8
INPUT Rasch-Andrich threshold TO ANCHOR (-1 TO END): 0 bottom category
When all are anchored, enter 0 to end:
INPUT AN ITEM, REPRESENTING A GROUPING (0 TO END): 0

```

11.152 SCOREFILE= person score file

If SCOREFILE=filename is specified, a file is output which contains the measure and model standard error corresponding to every possible score on a test consisting of all the non-extreme items.

The SCOREFILE= and [Table 20](#) person ability estimates are estimated on the basis that the current item difficulty estimates are the "true" estimates. These are the person estimates if you anchored (fixed) the items at their reported estimates. [PFILE=](#), [Table 17](#) and the other Person Measure Tables show the person abilities that are the maximum likelihood estimates at the current stage of estimation. To make these two sets of estimates coincide, please tighten the convergence criteria in your Winsteps control file:

```

CONVERGE=L
LCONV=*.001 ; or tighter

```

If you want the score file for person measures including the extreme (zero, perfect) items, then

1. Run a standard analysis.
2. Output: [IFILE=if.txt](#) [SFILE=sf.txt](#)
3. Run the analysis again, with Extra specifications: [IAFILE=if.txt](#) [SAFILE=sf.txt](#)
4. The person measures will have altered somewhat to adjust for the imputed difficulties of the extreme items.
5. Output [Table 20](#) and SCOREFILE=

SCOREFILE=? opens a Browse window

This is also shown in [Table 20](#). It has 3 heading lines (unless [HLINES=N](#) or [ROW1HEADING=N](#)), and has the format:

```

; PERSON SCORE FILE FOR
; EXAMPLE ANALYSES
;Apr 28 22:37 2005  UMEAN=.000  USCALE=1.000
; TABLE OF SAMPLE NORMS (500/100) AND FREQUENCIES CORRESPONDING TO COMPLETE TEST
;SCORE MEASURE S.E. INFO NORMED S.E. FREQUENCY % CUM.FREQ. % PERCENTILE
...
    0 -5.2891  1.4230   .49  -76 125      0   .0      0   .0      0
...
    58  1.8144  .3351   8.90  546  29      15  2.9     363  71.2     70

```


59	1.9303	.3461	8.35	556	30	20	3.9	383	75.1	73
60	2.0544	.3588	7.77	567	31	21	4.1	404	79.2	77
...										
70	5.3232	1.4245	.49	853	125	3	.6	510	100.0	99

1. SCORE: Score on test of all items. [TOTALSCORE=](#)Yes includes extreme items. TOTALSCORE=No excludes extreme items (if any).

The score file shows integer raw scores, unless there are decimal weights for [IWEIGHT=](#). In which case, scores to 1 decimal place are shown. To obtain other decimal raw scores for short tests, go to the [Graphs pull-down menu](#). Select "Test Characteristic Curve". This displays the score-to-measure ogive. Click on "Copy data to clipboard". Open Excel. Paste. There will be to three columns. The second column is the measure, the third column is the raw score.

2. MEASURE: Measure (user-scaled by [USCALE=](#))

If this value is not the same as in [PFILE=](#)

1. Missing data. The measure in the PFILE= is based on the responses observed. The measure in the SCOREFILE= is based on all the items.

2. Convergence criteria. The convergence criteria: [CONVERGE=](#), [LCONV=](#), [RCONV=](#) are not tight enough for your data. Please try setting LCONV= and RCONV= to smaller values. SCFILE= measures are estimated from the item estimates reported in [Table 14](#) using the technique at <http://www.rasch.org/rmt/rmt102t.htm> converged at LCONV=*.001 logits and RCONV=*.001 score points. Suggested values:

3. S.E.: [Standard error](#) (user scaled by USCALE=) - model, because empirical future misfit is unknown.

4. INFO: Statistical information in measure (=1/Logit S.E.²) = observable points of the Test Information Function (TIF).

Measures locally-rescaled, so that sample mean=500, [standard deviation](#)=100

5. NORMED: Measure (rescaled)

6. S.E.: Standard error (rescaled)

Sample distribution:

7. FREQUENCY: Count of sample at this measure

8. %: Percent of sample at this measure

9. CUM.FREQ.: Count of sample at or below this measure

10. %: Percent of sample at or below this measure

11. PERCENTILE: Percentile of this sample lower than the current measure (range is 0-99).

If [CSV=Y](#), these values are separated by commas. When CSV=T, the commas are replaced by tab characters.

Example 1: You wish to write a file on disk called "MYDATA.SCF.txt" containing a score-to-measure table for the complete test.

```
SCOREFILE=MYDATA.SCF.txt
```

Example 2: You want a score-to-measure file for items with known difficulties.

```
ITEM1=1 ; start of response string
NAME1=1
NI=10 ; number of items
CODES=01 ; valid item codes
IAFILE=*
; known item difficulties here
1 0.53
.....
10 -0.34
*
SAFILE=*
; known structure "step" calibrations here, if rating scale or partial credit items
*
SCOREFILE=sm.txt ; the score-to-measure file - also see Table 20
&END
END LABELS
0101010101 ; two dummy data records
1010101010 ; give every item a non-extreme score
```

Example 3: Produce a cumulative percentage plot of raw scores:

- 1) Output the Scorefile= to Excel
- 2) Scatterplot with lines between points
- 3) x-axis: SCORES
- 4) y-axis: Cumulative Frequency %

For the inverse cumulative percentage plot (percentage point plot), switch the x-axis and the y-axis.

Example 4: more exact relationship between raw scores and Rasch measures.

- 1) "Output Files" menu, SCOREFILE= to Excel.
- 2) Excel scatterplot the measures against the scores
- 3) On the Excel scatterplot, a cubic trendline: display formula.

11.153 SCREEN= screen log file

SCREEN=filename (with no spaces) must be on the command line or in the short cut. Everything written to the Analysis window is also written to the screen file.

Example:

```
c:\> "C:\winsteps\winsteps.exe" controlfile.txt outputfile.txt SCREEN=screenfile.txt
```

The screen log file can be displayed from the Edit menu.

11.154 SDELQU= delete category structure interactively

This is better performed with [IREFER=](#) and [IVALUE=](#). SDELQU= is not supported in this version of Winsteps.

11.155 SDFILE= category structure deletion file

This is better performed with [IREFER=](#) and [IVALUE=](#). SDFILE= is not supported in this version of Winsteps.

11.156 SEPARATOR= or DELIMITER= data field delimiters

See [DELIMITER=](#)

11.157 SFILE= category structure-threshold output file

If SFILE=filename is specified, a file is output which contains the item and category information needed for anchoring structures. It has 4 heading lines (unless [HLINES=N](#) or [ROW1HEADING=N](#)), and has the format:

Example for dichotomous items:

```
; STRUCTURE-THRESHOLD MEASURE ANCHOR FILE
; CATEGORY Rasch-Andrich threshold
1 .00 ; this is a dummy threshold to tell Winsteps that 0 is the bottom category
2 .00
```

Example for the Andrich Rating Scale model:

```
; STRUCTURE-THRESHOLD MEASURE ANCHOR FILE
; CATEGORY Rasch-Andrich threshold
0 .00 ; this is a dummy threshold to tell Winsteps that 0 is the bottom category
1 -.86
2 .86
```

Example for the Partial Credit model:

```
STRUCTURE-THRESHOLD MEASURE ANCHOR FILE
; ACT CATEGORY Rasch-Andrich threshold MEASURE
1 0 .00 ; this is a dummy threshold to tell Winsteps that 0 is the bottom category
1 1 -1.64
1 2 1.64
2 0 .00 ; this is a dummy threshold to tell Winsteps that 0 is the bottom category
```

```

2 1 -.30
2 2 .30
3 0 .00 ; this is a dummy threshold to tell Winsteps that 0 is the bottom category
3 1 -1.13
3 2 1.13
....

```

Columns:

1. The item sequence number (I6), if required.
Shown only if there are multiple groupings in [ISGROUPS=](#) or [SAITEM=Yes](#), otherwise this is omitted.
2. The category number (I3) (STRU) - the higher numbered of the pair of categories on each side of the Andrich threshold.
3. Andrich threshold value (structure calibration) (F7.2) (user-rescaled by [USCALE=](#)) (number of decimals by [UDECIM=](#)) (MEASURE)
Use this for anchoring, together with [IFILE=](#)

If [CSV=Y](#), these values are separated by commas. When [CSV=T](#), the commas are replaced by tab characters.

[SFILE=?](#) opens a Browse window

Andrich Thresholds for Unobserved Categories

When [STKEEP=YES](#) and there are intermediate [null](#) categories, i.e., with no observations, then the Rasch-Andrich thresholds are infinite. To approximate infinite Andrich thresholds in computer arithmetic, the Andrich threshold into the unobserved category is set 40 logits above the highest threshold calibration. The Rasch-Andrich threshold out of the unobserved category, and into the next category, is set down by the same amount. Thus:

Category Structure Calibration Rasch-Andrich Threshold		
Category	Table 3.2	In SFILE
0	NULL	0
1	-3	-3
2	NULL	$40 + 2 = 42$
3	1	$1 - 40 - 2 = -41$
4	2	2
TOTAL:	0	0

This is also an attempt to model unobserved categories for the next time, when they may be observed.

If categories will not be observed next time, then please specify [STKEEP=NO](#), which automatically drops the category from the rating (or partial credit) scale.

If categories may be observed next time, then it is better to include a dummy data record in your data file which includes an observation of the missing category, and reasonable values for all the other item responses that accord with that missing category. This one data record will have minimal impact on the rest of the analysis.

11.158 SFUNCTION= function to model Andrich thresholds with polynomials

By default, the structure of a rating scale (polytomy) is estimated from the data, and expressed as Rasch-Andrich thresholds. When the data are sparse, or the rating scale is long, there may not be enough observations to estimate every threshold securely. [SFUNCTION=](#) defines a [smoothing function](#) for estimating all the rating scale thresholds. If all the categories are not observed in the data, [ISRANGE=](#) defines the range of the category numbers.

Alternative approaches are described at [Null or unobserved categories: structural and incidental zeroes](#), also at [Unobserved and dropped categories](#).

Modeling Rasch-Andrich Thresholds for unobserved and sparse categories

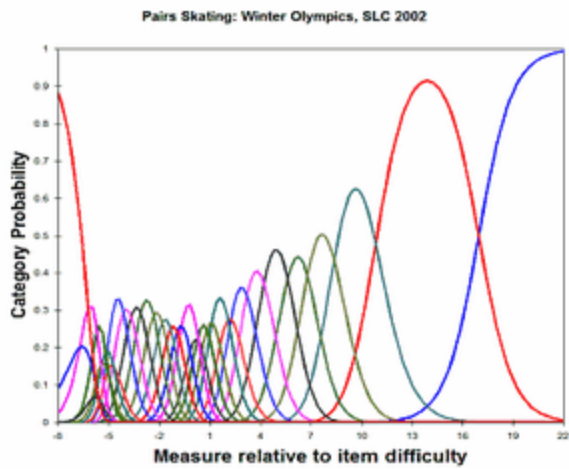
Example: Olympic Ice-Skating, [Exam 15.txt](#). The rating-scale is from 0.0 to 6.0, renumbered 00 to 60 for Rasch analysis. The empirical data are from the Pairs Figure-Skating at the Salt Lake City Olympics.

Category frequencies:

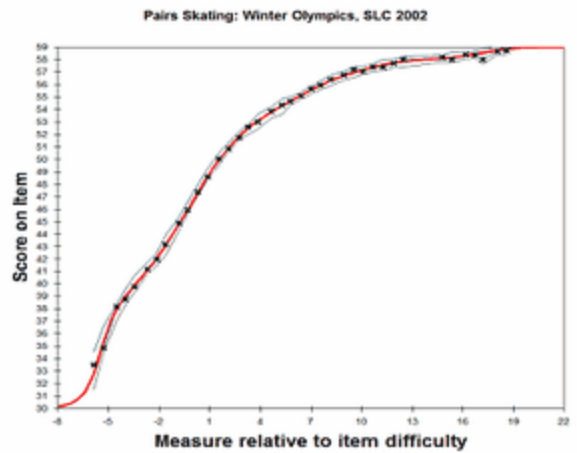
- Possible categories: 0-60 = 61 categories
- Observed categories: 29, 30-59 = 30 categories
- Unobserved categories: 00-28, 30, 60 = 31 categories

Analysis 1: Categories 29, 31-59 observed, renumbered 30-59. Estimated using the method of Wright, Masters (1982).

```
CODES= " 29 31 32 33 34 35 36 37 38 39 40 41 42 43 44+
        + 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59"
NEWSCORE = " 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44+
            + 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59"
STKEEP=NO
```



Category Probability Curves



Item Characteristic Curves

Red: Model, Blue: Empirical (with confidence bands)

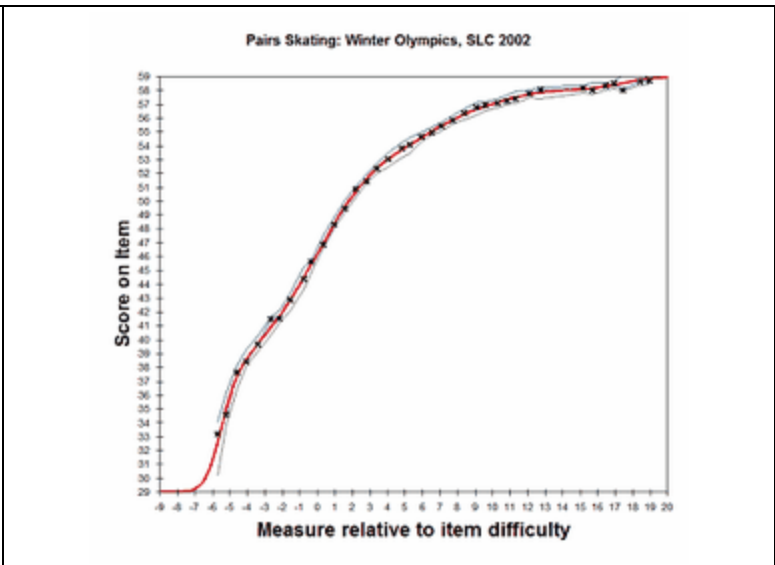
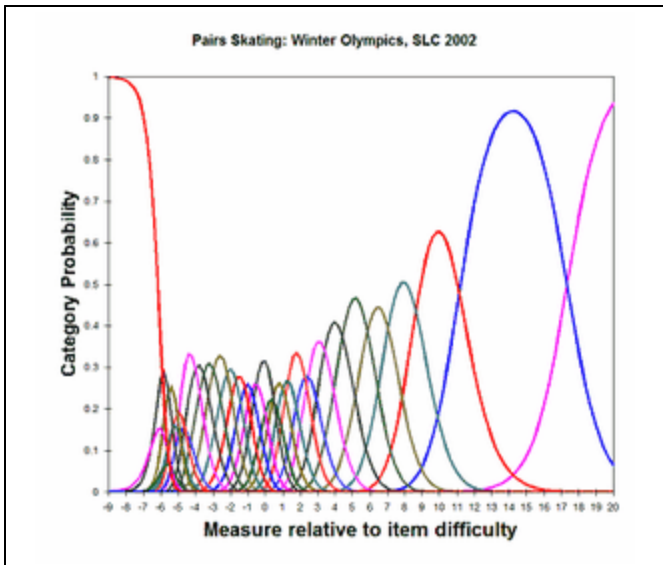
Analysis 2: Categories 29, 31-59 observed. Category 30 interpolated with Wilson (1991).

```
CODES= " 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44+
        + 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59"
```

or

```
CODERANGE= 29 - 59
```

```
STKEEP=YES
```



Analysis 3: 1st moment (thresholds at central location). Categories 0-60.

CODERANGE= 0 - 60

SFUNCTION=1 ; item thresholds are at item location. Threshold dispersion = 0.

ISRANGE=*

1 0 60 ; 1 is item in the item-group: category range 0-60

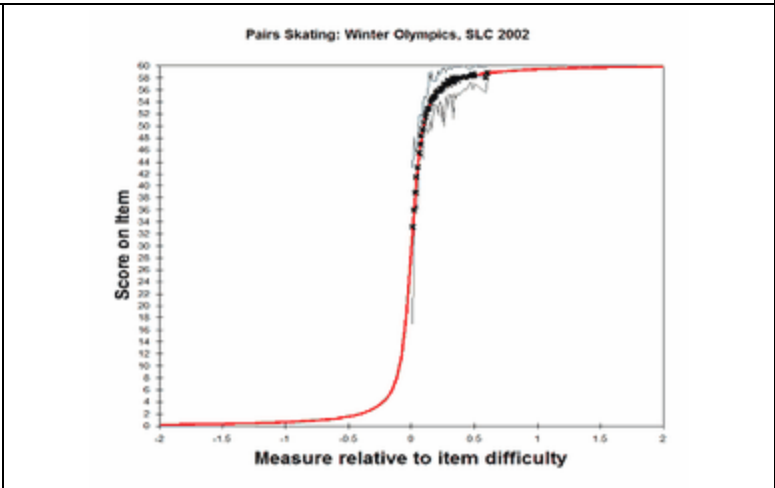
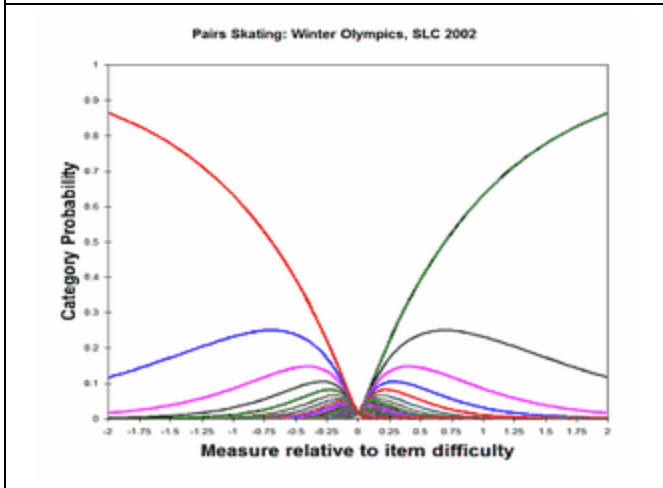
*

OR

SAFILE=*

0-60 0 ; all thresholds anchored at 0

*



Analysis 4: 2nd moment (thresholds with uniform dispersion) . Categories 0-60 with Andrich, Luo (2003), Pedler (1987)

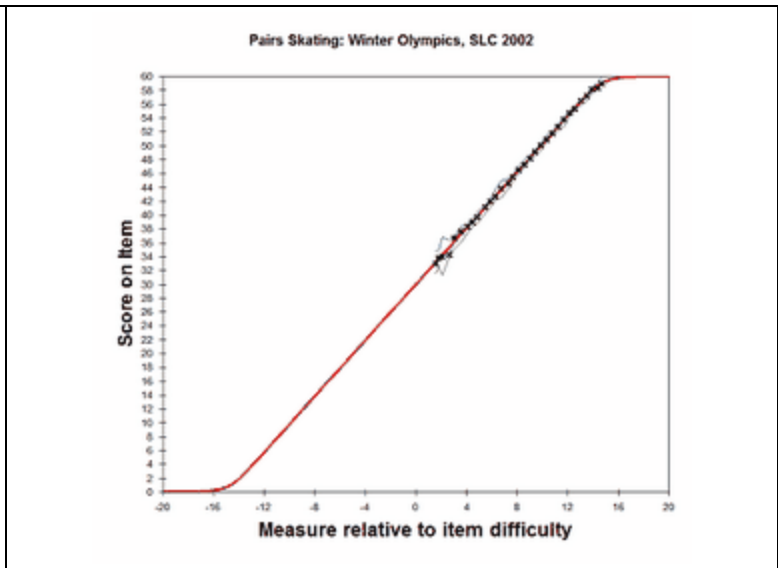
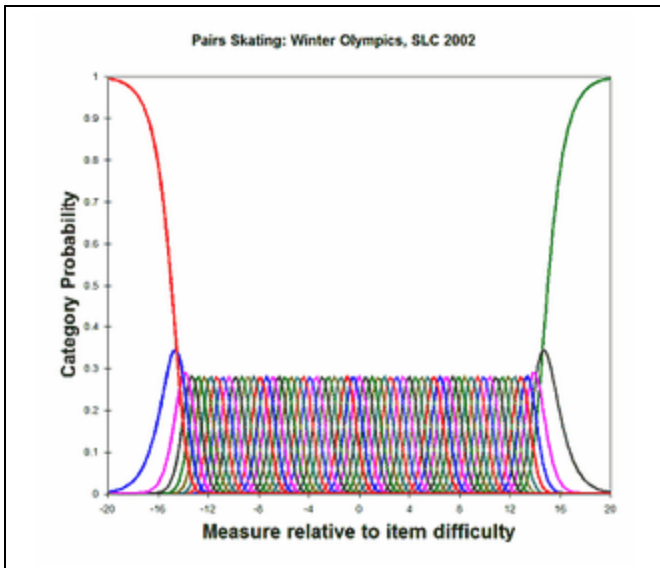
CODERANGE= 0 - 60

SFUNCTION=2

ISRANGE=*

1 0-60

*



Analysis 5: 3rd moment (thresholds with dispersion and skewness). Categories 0-60 with Andrich, Luo (2003), Pedler (1987)

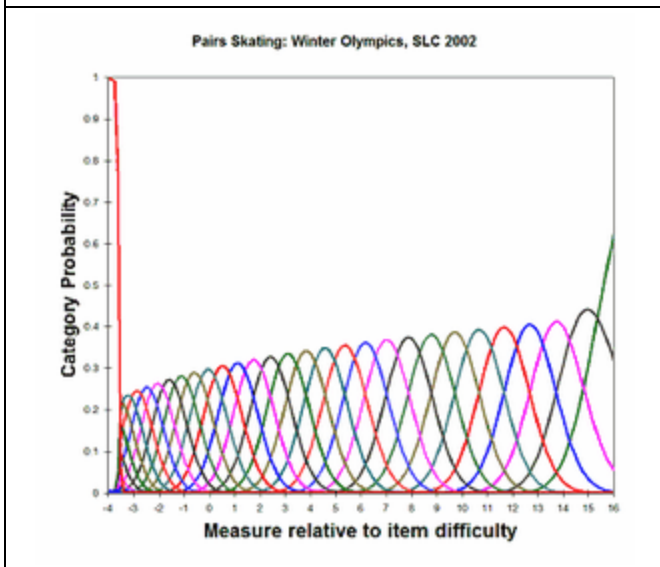
CODERANGE= 0 - 60

SFUNCTION=3

ISRANGE=*

1 0-60

*



Analysis 6: 4th moment (thresholds with dispersion, skewness and kurtosis). Categories 0-60 with Andrich, Luo (2003), Pedler (1987)

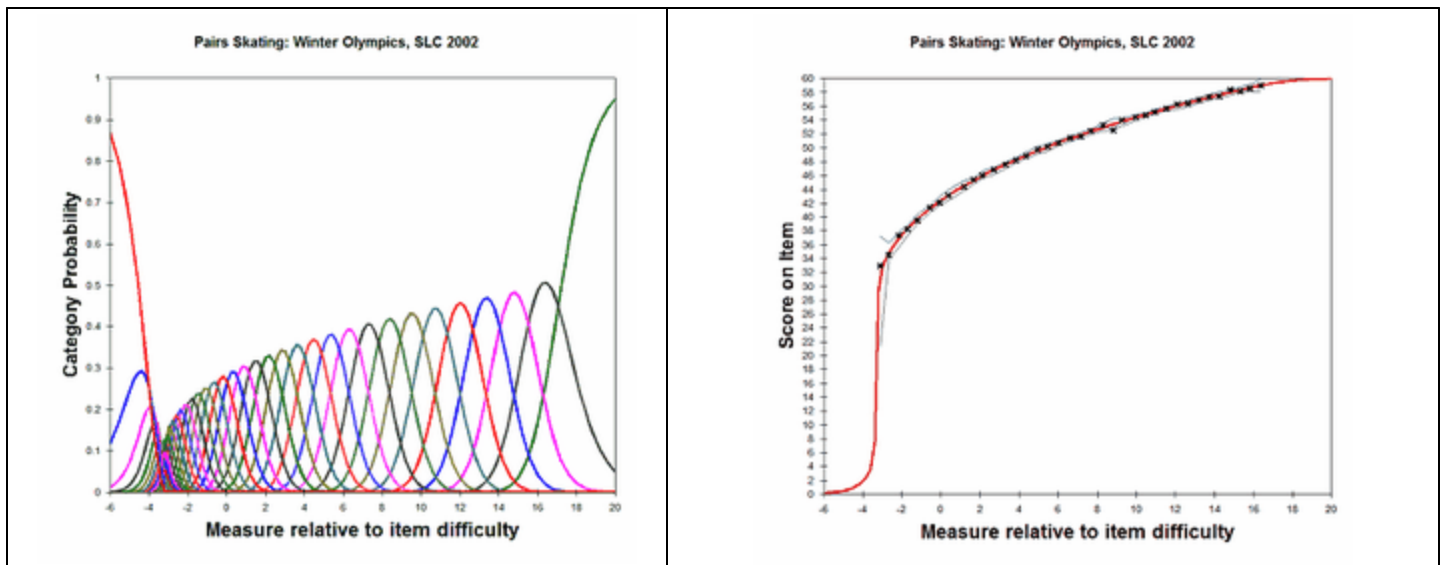
CODERANGE= 0 - 60

SFUNCTION=4

ISRANGE=*

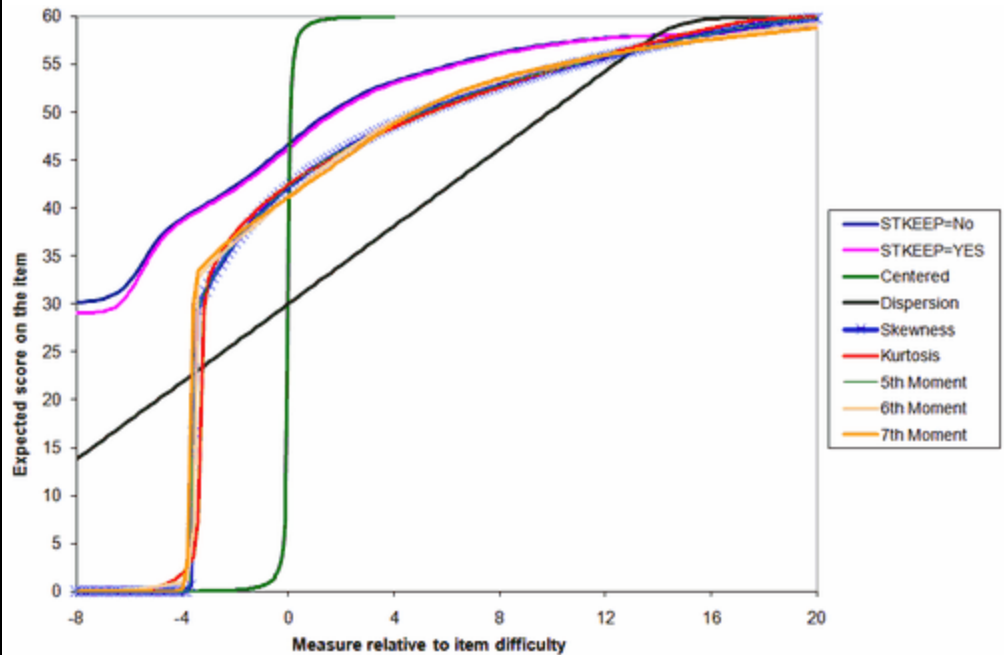
1 0-60

*



ICCs for these and higher moments

The 4th moment (kurtosis, red line) models the smoothest ICC for the full range of the rating scale 0-60. The higher moments (SFUNCTION=5,6,...) do not add meaningfully to the red line of the 4th moment (kurtosis, SFUNCTION=4).



Andrich, D. & Luo, G. (2003). Conditional Pairwise Estimation in the Rasch Model for Ordered Response Categories using Principal Components. *Journal of Applied Measurement*, 4(3), 205-221.

Pedler, P.J. (1987) Accounting for psychometric dependence with a class of latent trait models. Ph.D. dissertation. University of Western Australia.

Wilson M. (1991) Unobserved or Null Categories. *Rasch Measurement Transactions*, 5:1, 158 - www.rasch.org/rmt/rmt51b.htm

Wright B.D., Masters G.N. (1982) Rating Scale Analysis. MESA Press.

11.159 SICOMLETE= simulate complete data

Data simulated with [SIFILE=](#). See [simulated data file](#). When simulating a data file [SIFILE=](#) the missing-data pattern in the originating data may be copied or ignored.

SICOMLETE = Yes
generate a complete rectangular data set from the generating values.

SICOMLETE = No

generate a rectangular data set with missing values where the originating data set has missing values. This has no effect if the originating data set is complete, without missing responses.

Example: For a CAT dataset, we want to simulate responses only where responses were made:
SICOMPLETE = No

11.160 SIEXTREME= simulate extreme scores

Data are simulated with [SIFILE=](#). See [simulated data file](#). Simulated data files may contain extreme scores (zero, minimum possible and perfect, maximum possible scores) by chance.

SIEXTREME = Yes
allow extreme score to be part of the simulated data.

SIEXTREME = No
avoid generating extreme scores in the simulated data. This may not be possible. This option may also produce measures which are more central than the generating values.

Example: In a study of differential item functioning extreme scores are not wanted:
SIEXTREME = No

11.161 SIFILE= simulated data file

Data are simulated with [SIFILE=](#). See [simulated data file](#).

SIFILE = file name	name of output file to hold simulated data
SIFILE=? opens a Browse window	the format of the simulated file is controlled by the filename suffix . If there is more than one simulated file specified with SINUMBER= then the file name is automatically incremented by one for each simulated data file.

The simulated data file has the same format as the original data. It may be accessed from the original control file by using [DATA=](#) at the [Extra Specifications](#) prompt.

Other control variables for simulated data files are [SICOMPLETE=](#), [SIEXTREME=](#), [SIMEASURE=](#), [SINUMBER=](#), [SIRESAMPLE=](#), [SISEED=](#)

11.162 SIMEASURE= measure or data

Data simulated with [SIFILE=](#). See [simulated data file](#).

SIMEASURE = Yes	Simulate data from the Rasch measures
SIMEASURE = No	Simulate data by re-sampling from observed response strings.

11.163 SINUMBER= number of simulated data files

Data simulated with [SIFILE=](#). See [simulated data file](#). More than one simulated data file [SIFILE=](#) can be generated.

SINUMBER = 1 produce one simulated data file
SINUMBER > 1 produce more than one simulated data file.

For the first simulated data file, its name is given by SIFILE= (or a temporary file) and the random number seed is given by SISEED=.

For the second and subsequent simulated data files when SINUMBER>1, then the data file name increments from SIFILE= and the seed number increments from SISEED, unless the seed is based on the system clock.

Example. Produce 3 simulated files with seed starting at 101.
SIFILE = mysimulation.txt

SINUMBER = 3
SISEED = 101

The three output files will be:
mysimulation.txt with seed 101
mysimulation2.txt with seed 102
mysimulation3.txt with seed 103

11.164 SIRESAMPLE= number of persons resampled

Data can be simulated with [SIFILE=](#) . See [simulated data file](#).

SIRESAMPLE= 0	The simulated corresponds exactly with the original data
SIRESAMPLE= number	The simulated data file contains <i>number</i> person records. These are produced by: 1. SIMEASURE=Yes - re-sampling the person measures. 2. SIMEASURE=No - re-sampling the person response strings.

11.165 SISEED= simulated data seed

Data simulated with [SIFILE=](#) . See [simulated data file](#). Simulated data files are constructed using a random number generator for selecting the responses. The [random generator number](#) uses a seed value to initiate its pseudo-random sequence of numbers. The seed value can either be set by the user, or be the current value of the system clock.

SISEED = 0 or 1 means use the system clock to seed the random number generator.
SISEED = 2 or above (integer) means use this value to seed the random number generator.

If more than one simulated data file is produced with [SINUMBER=](#) then
if SISEED = 0 or 1, the new current setting of the system clock is used for each simulated file
if SISEED > 1, the user-specified seed is advanced by 1 for each simulated file

Example: You want to simulate the same data set several times, so you choose the seed:
SISEED = 237 ; 237 is any number.

11.166 SPFILE= supplementary control file

There is often a set of control instructions that you wish to make a permanent or temporary part of every control file. Such files can be specified with [SPFILE=](#). Multiple [SPFILE=](#) specifications can be included in one control file. Supplemental files called with [SPFILE=](#) can also include [SPFILE=](#) specifications.

SPFILE= file name	file containing details
SPFILE = *	in-line list (useful for testing)
SOFILE = ?	opens a Browser window to find the file

Example 1: The analyst has a standard set of convergence criteria and other control instructions to include in every control file.

a) Enter these into a standard DOS TEXT/ASCII file, e.g, SUPPL.TXT

The analyst's SUPPL.TXT contains:

```
LCONV=.01  
ITEM=TASK  
PERSON=CLIENT  
TABLES=101110011
```

b) Specify this supplemental file in the main control file, say, MAIN.TXT

```
TITLE='NEW YORK CLIENTS'  
SPFILE=SUPPL.TXT  
ITEM1=37  
NI=100
```

Example 2: The analyst has a set of control instructions that are used only for the final run. These are coded in a separate DOS TEXT file called FINAL.txt

```
C:>Winsteps CONTROL.txt OUTPUT.txt SPFILE=FINAL.txt
```

Example 3: KEYn= is a particularly useful application of SPFILE=.

Put the KEY1= instruction for each test form in its own DOS TEXT file, then reference that file rather than including the key directly in the control file.

Here is FORMA.KEY:

```
NI=23
```

```
CODES=ABCD
```

```
KEY1=ABCDDADBCDADDABBCAADBBA
```

Here is the control file:

```
TITLE='FORMA READING RASCH ANALYSIS'
```

```
ITEM1=20
```

```
SPFILE=FORMA.KEY
```

```
TABLES=111011011
```

```
|
```

11.167 STBIAS= correct for estimation bias

STBIAS=Y causes an approximate [correction](#) for estimation bias in [JMLE](#) estimates to be applied to measures and calibrations. This is only relevant if an exact probabilistic interpretation of logit differences is required for short tests or small samples. Set STBIAS=NO when using [IWEIGHT=](#), [PWEIGHT=](#), anchoring, [IAFILE=](#), [PAFILE=](#), [SAFILE=](#) or artificially lengthened tests or augmented samples, e.g., by replicating item or person response strings.

Fit statistics are computed without this estimation-bias correction. Estimation-bias correction makes the measures more central, generally giving a slight overfit condition to Outfit and Infit. Correct "unbiased" computation of INFIT and OUTFIT needs not only unbiased measures, but also probabilities adjusted for the possibility of extreme score vectors (which is the cause of the estimation bias).

STBIAS=YES instructs Winsteps to compute and apply statistical-bias-correction coefficients to the item difficulties and to the person measures - based on the current data. This becomes complicated for anchor values and scoring tables. With STBIAS=YES, the item anchor values are assumed to be bias-corrected. Consequently bias is applied to make them compatible with JMLE computations for the current data. The resulting person measures are JMLE person estimates, which are biased. So a person bias correction is applied to them.

For the special case of two items for each person, or two persons for each item, please use [PAIRED=Yes](#) to correct for bias.

With **STBIAS=No**, there is no statistical bias correction, so the internal and reported values are the same. The process is

For unanchored item values,

data + internal person estimates => internal item estimates => reported item estimates

For anchored item values,

anchored item values => internal item estimates => reported item estimates

For unanchored person values,

data + internal item estimates => internal person estimates => reported person estimates

For anchored person values,

anchored person values => internal person estimates => reported person estimates

For a scoring table

reported item estimates => internal item estimates => internal person estimates => reported person estimates

With **STBIAS= YES**, the internal and reported values differ. The process is

Compute bias correction coefficients for item estimates and for person estimates based on the current data.

For unanchored item values,
current data + internal person estimates => internal item estimates => item bias correction => reported item estimates
For anchored item values,
anchored item values => undo item bias correction => internal item estimates => item bias correction => reported item estimates

For unanchored person values,
data + internal item estimates => internal person estimates => person bias correction => reported person estimates
For anchored person values,
anchored person values => undo person bias correction => internal person estimates => person bias correction => reported person estimates

For a scoring table, the process is
reported item estimates => undo item bias correction => internal item estimates => internal person estimates => person bias correction => reported person estimates.

Note: it is seen that this process can result in estimates that are worse than uncorrected JMLE estimates. Consequently it may be advisable not to use STBIAS=YES unless the bias correction is clearly required.

Question: Are JMLE estimates always biased?

Answer: Yes, but the bias becomes inconsequential (less than the standard errors) for tests with more than 20 persons and more than 20 items - www.rasch.org/memo45.htm

Question: Are person estimates in JMLE biased as well as the item difficulty estimates?

Answer: The Rasch model does not know what is a person and what is an item. Smaller person samples for a given test length and shorter tests for a given person sample size generally make the estimation bias worse. Winsteps is constructed so that transposing the rows and columns of the data matrix (with dichotomous items or the Andrich rating scale model) produces statistically the same item and person measures (apart from a change of sign). [CMLE](#) and [MMLE](#) do not have this property. This transposition property is convenient in those types of analysis where it is not clear what is a "person" and what is an "item" - e.g., a matrix of street intersections and calendar dates with "0" for no traffic accident and "1" for traffic accident. It also enables analysis with a person-based "Partial Credit" model, where each person has a unique rating scale.

Question: Are person estimates obtain from known or anchored item difficulties statistically biased?

Answer: Under these circumstances, estimation is no longer "Joint" (persons and items), but becomes the MLE used to estimate person abilities from item difficulties in other estimation methods (CMLE, MMLE, PMLE, etc.)

Question: What is the bias correction used by Winsteps?

Assuming dichotomous, complete data:

Unbiased item estimate = biased estimate * (number of items - 1) / number of items

Unbiased person estimate = biased estimate * (number of persons - 1) / number of persons

Question: what about other data?

An empirical solution is to estimate the Rasch measures from your data. Note down the person S.D.

Then use the Winsteps "Output Files" "[simulate data](#)" option to simulate 10 datasets.

Analyze these, and average the person S.D.s

(This process can be automated using Winsteps "[BATCH=](#)" capability)

Compare the average with the original estimate person S.D. - this will tell you the size of the bias.

Then apply [USCALE=](#) bias correction to the original analysis to obtain unbiased estimates. The unbiased estimates will always be more central (smaller range) than the biased estimates.

Example 1: I have a well-behaved test of only a few items, but I want to correct for statistical estimation bias because I want to me to make exact probabilistic inferences based on differences between logit measures:

STBIAS=Y

Example 2: I have a set of item difficulties from RUMM (or ConQuest or OPLM or ...) and want to construct a score-table that matches the RUMM person measures.

```
IAFILE = (RUMM item difficulties)
STBIAS = No ; don't change the estimates - use the values RUMM uses
TFILE=*
20 ; Table 20 is the score table
*
```

```
...
END LABELS
01010101010101010 ; Dummy data records
101010101010101 ; so that Winsteps will run
```

11.168 STEPT3= include category structure summary in Table 3.2 or 21

The structure summary statistics usually appear in [Table 3](#). For grouped analysis this part of Table 3.2 can become long, in which case it can be moved to [Table 21](#).

Example: Don't output partial credit structure summaries in Table 3.2. Move them to Table 21:

```
ISGROUPS=0 ; each item has own rating scale
STEPT3=N ; report scale statistics in Table 21
```

11.169 STKEEP= keep unobserved intermediate categories in structure

Unobserved categories can be dropped from rating (or partial credit) scales and the remaining category recounted during estimation. For intermediate categories only, recounting can be prevented and unobserved categories retained in the analysis. This is useful when the unobserved categories are important to the rating (or partial credit) scale logic or are usually observed, even though they happen to have been unused this time. Category Rasch-Andrich thresholds for which anchor calibrations are supplied are always maintained wherever computationally possible, even when there are no observations of a category in the current data set.

Use STKEEP=YES when there may be intermediate categories in your rating (or partial credit) scale that aren't observed in this data set, i.e., incidental or sampling zeroes.

Use STKEEP=NO when your category numbering deliberately skips over intermediate categories, i.e., structural zeroes.

Use [IVALUE=](#) to rescore each item independently to keep or eliminate unobserved categories.

STKEEP=N Eliminate unused categories and close up the observed categories.

STKEEP=Y Retain unused non-extreme categories in the ordinal categorization.

When STKEEP=Y, missing categories are retained in the rating (or partial credit) scale, so maintaining the raw score ordering. But missing categories require locally indefinite structure calibrations. If these are to be used for anchoring later runs, compare these calibrations with the calibrations obtained by an unanchored analysis of the new data. This will assist you in determining what adjustments need to be made to the original calibrations in order to establish a set of anchor calibrations that maintain the same rating (or partial credit) scale structure.

To remind yourself, STKEEP=YES can be written as STRUCTUREKEEP=YES, STRKEEP=YES or STEPKEEP=YES and other similar abbreviations starting STK, STR and STEPK.

Example 1: Incidental unobserved categories. Keep the developmentally important rating (or partial credit) scale categories, observed or not. Your small Piaget rating scale goes from 1 to 6. But some levels may not have been observed in this data set.

```
STKEEP=Y
```

Example 2: Structural unobserved categories. Responses have been coded as "10", "20", "30", "40", but they really mean 1,2,3,4

CODES = "10203040"

XWIDE = 2

STKEEP=NO

; if STKEEP=YES, then data are analyzed as though categories 11, 12, 13, 14, etc. could exist, which would distort the measures.

; for reporting purposes, multiply Winsteps SCORE by 10 to return to the original 10, 20, 30 categorization.

Example 3: Some unobserved categories are structural and some incidental. Rescore the data and use STKEEP=YES. Possible categories are 2,4,6,8 but only 2,6,8 are observed this time.

(a) Rescore 2,4,6,8 to 1,2,3,4 using IVALUE= or NEWSCORE=

(b) Set STKEEP=YES, so that the observed 1,3,4 and unobserved 2 are treated as 1,2,3,4

(c) For reporting purposes, multiply the Winsteps SCORE by 2 using Excel or similar software.

CODES=2468

NEWSCORE=1234

STKEEP=YES

Example 4: There are both dichotomous items(0/1) and polytomous items(0/1/2) in the same test. Items 1-5 are dichotomies, and 6-10 are polytomies with the same rating scale:

ISGROUPS = DDDDDPPPPP ; items 1-5, the dichotomies, are in one item-group, items 6-10, the polytomies, in another.

CODES = 012 ; has all valid codes in the data file.

; For dichotomies, STKEEP= has no effect.

; For polytomies: If the data contain 0/1/2 then STKEEP= has no effect.

; If the data contain 0/2 (but no 1's), what is to be done?

STKEEP = No ; 0/2 is to be automatically recoded 0/1

STKEEP = Yes ; 0/2 is to be analyzed as 0/1/2

Incidental and Structural Zeroes: Extreme and Intermediate

For missing intermediate categories, there are two options.

If the categories are missing because they cannot be observed, then they are "structural zeroes". Specify "STKEEP=NO". This effectively recounts the observed categories starting from the bottom category, so that 1,3,5,7 becomes 1,2,3,4.

If they are missing because they just do not happen to have been observed this time, then they are "incidental or sampling zeros". Specify "STKEEP=YES". Then 1,3,5,7 is treated as 1,2,3,4,5,6,7.

Categories outside the observed range are always treated as structural zeroes.

When STKEEP=Y, unobserved intermediate categories are imputed using a mathematical device noticed by Mark Wilson. This device can be extended to runs of unobserved categories.

Idiosyncratic rating scales: the Barthel Index

Winsteps has two ways of analyzing rating scales such as 0,2,5,8,10 in the Barthel Index.

1. STKEEP=No. Unobserved rating-scale categories are structural zeroes. This converts 0,2,5,8,10 into 0,1,2,3,4

2. STKEEP=Yes (the default). Unobserved rating-scale categories are sampling zeroes. This converts 0,2,5,8,10 into 0,1,2,3,4,5,6,7,8,9,10

It appears that the Barthel Index is using weighted categories, so its scoring implies:

0,2,5,8,10 = 0, 2x1, 2.5x2, 2.66x3, 2.5x4

Winsteps does not support different weighting of different categories. This can be done in *Facets*, but *Facets* does not output KeyForms.

An approximation in Winsteps could be:

0,2,5,8,10 ≈ 0, 2.5x1, 2.5x2, 2.5x3, 2.5x4 = (STKEEP=No) 0,1,2,3,4 with an item weight of [WEIGHT= 2.5](#)

then edit the [KeyForm](#) from 0,1,2,3,4 to become 0,2,5,8,10

11.170 SUBSETS= perform subset detection

When there are missing data, subsets of the observations may not be measurable with other subsets. For instance, Person sample A responds to Test Form A, and Person sample B responds to Test Form B. If there are no common items nor common persons, combining these two datasets into one analysis will produce two disconnected subsets of responses.

SUBSETS= Yes performs [subset and connectivity detection](#).

SUBSETS = No bypasses subset detection.

When subset detection is activated, "Probing Data Connection" displays in the Winsteps analysis window.

```
PROBING DATA CONNECTION: to skip out: Ctrl+F - to bypass: subset=no
>=====<
Consolidating 8 potential subsets pairwise ...
>=====<
Consolidating 8 potential subsets indirectly pairwise ...
>=====<
Warning: Data are ambiguously connected into 8 subsets. Measures may not be comparable across subsets.
Subsets details are in Table 0.4
```

Subsets are reported In [Table 0.4](#) if any are discovered.

TABLE 0.4 Example of subset reporting

```
SUBSET DETAILS

SUBSET 1 OF 2 ITEM AND 2 PERSON
ITEM: 2-3
PERSON: 2-3
SUBSET 2 OF 2 ITEM AND 2 PERSON
ITEM: 4-5
PERSON: 4-5
SUBSET 3 OF 1 PERSON
PERSON: 6
SUBSET 4 OF 2 ITEM AND 2 PERSON
ITEM: 7-8
PERSON: 7-8
SUBSET 5 OF 2 ITEM AND 2 PERSON
ITEM: 9-10
PERSON: 9-10
SUBSET 6 OF 2 PERSON
PERSON: 11-12
SUBSET 7 OF 3 PERSON
PERSON: 13-15
SUBSET 8 OF 1 ITEM
ITEM: 6
```

Subsets are also shown to the right of the measure tables, [Table 14.1](#), [Table 18.1](#), etc.

```
-----
TCH|      |
XP%| PERSON|
-----+-----
.0.1| Ivan  | SUBSET 5
.0.1| Jack  | SUBSET 5
5.7| George| SUBSET 4
5.7| Henry | SUBSET 4
8.7| David | SUBSET 2
8.7| Edward| SUBSET 2
6.4| Ben   | SUBSET 1
6.4| Carl  | SUBSET 1
2.0| Frank | SUBSET 3
-----+-----
```

11.171 SVDEPOCHS= maximum epochs for SVD estimation

SVDEPOCHS= sets the maximum number of epochs (iterations through the data) for estimating each SVD factor.

See [SVDFILE=](#).

11.172 SVDFACTORS= singular-value decomposition factors

SVDFACTORS= sets the number of factors to estimate. Factor processing can be forced to conclude with Ctrl+F, "finish factoring".

See [SVDFILE=](#).

11.173 SVDFILE= singular-value decomposition file

Singular-value decomposition (SVD) is a technique for factoring a data matrix. Here it is used to look for patterns in the Rasch residuals which may indicate multidimensionality. SVD values are estimated by iterating through the residual data matrix. Each residual is explained, as much as possible, by a value attributed to the person (row) and a value attributed to the item (column). These values are multiplied.

The SVD estimation method implemented in Winsteps uses randomization for speed and flexibility, so the results reported for each estimation run are slightly different. For a description of the algorithm, see "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. Nathan Halko, Per-Gunnar Martinsson, Joel A. Tropp. SIAM Rev., Survey and Review section, Vol. 53, num. 2, pp. 217-288, June 2011. <https://arxiv.org/pdf/0909.4061.pdf>

SVDFILE=? opens a Browse window

The general form of the Rasch-SVD model is:

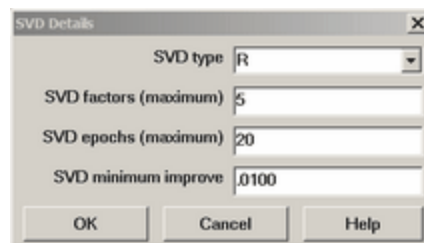
$$X_{ni} = E_{ni} + V_n * U_i \pm \varepsilon$$

where X_{ni} is the observation when person n meets item i . E_{ni} is the Rasch-expected value. V_n is the SVD value for the person. U_i is the SVD value for the item. ε is the unexplained part of the residual.

After the first set of SVD values have been estimated, the process can be continued deeper down to further levels:

$$X_{ni} = E_{ni} + V_{1n} * U_{1i} + V_{2n} * U_{2i} \dots \pm \varepsilon$$

SVD control settings. These can be in the control file, or from the [Output file menu. SVDFILE option.](#)



SVDFACTORS=	is the number of SVD terms (factors) to be estimated for each residual. SVD computation can be concluded with Ctrl+F.
SVDEPOCHS=	is the maximum number of epochs (iterations through the data) for estimating each SVD factor
SVDMIN=	is the minimum improvement of the root-mean-square residuals based on the current SVD estimates at the end of an SVD iteration (epoch) through the data matrix of residuals in order for SVD estimation to finish for that SVD factor.
SVDMETHOD=	is the type of residual to be decomposed. R = the raw residual or S = the standardized residual.
SVDFILE=	name of the output file for the SVD values.

SVD file format:

```

; SVD FILE FOR LIKING FOR SCIENCE (Wright & Masters p.18) SVDTYPE= S SVDFACTORS= 4 SVDMIN= .0010 KID= 75 ACT= 25
; RMSR .5422 .4793 .4595 .4244 .4027
; RMSS 1.0404 .8715 .8133 .7424 .6983
; KID MEASURE 1 2 3 4 ACT MEASURE 1 2 3
1 .6099 -.0607 .2212 -.4504 -.1882 1 -.3960 -.2002 -.9116 .1108
2 6.0751 0 0 0 0 2 -.7086 -.7039 .2789 -1.1431
3 1.0973 .0243 -.1957 -.2856 .0619 3 .4174 -.3841 .0428 .0975
4 .2611 .0999 .1702 .2774 .2440 4 1.7504 .3857 -.9536 .9610
5 -.6791 .8925 -.0656 -.2772 -.3505 5 2.4183 2.5782 3.2365 -1.1817

```

The first four lines are heading lines. To omit them, specify [HLINES=NO](#) or in the [Output Files specifications dialog box](#). [ROW1HEADING=N](#) to omit the first line only.

RMSR = Root-mean-square-residual corresponding to the column of Rasch measures or SVD factors.
 RMSS = Root-mean-square-standardized-residual corresponding to the column of Rasch measures or SVD factors.
 MEASURE = Rasch measure in [Table 14](#) etc.
 1 2 3 4 ... are the columns of SVD values. For each SVD factor, every person and every item has an SVD value.

SVD factors corresponding to extreme scores have 0 values.

11.174 SVDMIN= singular-value decomposition minimum improvement

SVDMIN= sets the change in root-mean-square (RMS) at which an SVD factor is deemed to have converged. If the change in the RMS at the end of an epoch (iteration through the data) is less than SVDMIN=, then estimation advances to the next SVD factor.

Factor convergence can be forced with Ctrl+E, "end epochs".

See [SVDFILE=](#).

11.175 SVDTYPE= singular-value decomposition residual type

SVDTYPE= sets the type of residual to factor.
 SVDTYPE=R is the raw residual: (observation - Rasch expectation)
 SVDTYPE=S is the standardized residual: (observation - Rasch expectation) / $\sqrt{\text{(model variance)}}$

See [SVDFILE=](#).

11.176 T1CAT= show category number

In Tables 1.5, 1.6, 1.7, 1.8, 1.9, 12.5, 12.6, 12.7, 12.8, 12.9

T1CAT=	Category numbers in column headings are
(blank)	as in earlier versions of Winsteps
Scored	Category value after specified scoring with NEWSCORE= , etc. (if any)
Recounted	Category score counting up from lowest observed value
Zero	Category score counting up from zero

11.177 T1I#= number of items summarized by "#" symbol in Table 1

For ease in comparing the outputs from multiple runs, force consistent x-axis scaling by using [MRANGE=](#), [T1I#=](#) and [T1P#](#). Choose T1I# to be the largest number of items summarized by one "#" from any of the separate runs.

Example: In one run, the bottom of Table 1 (or Table 16) states that
 EACH "#" IN THE ITEM COLUMN IS 20 ITEMS
 In another run: EACH "#" IN THE ITEM COLUMN IS 15 ITEMS
 To make the runs visually comparable, specify the bigger value:
 T1P#=20

11.178 T1P#= number of persons summarized by "#" symbol in Table 1

For ease in comparing the outputs from multiple runs, and to force consistent x-axis scaling by using [MRANGE=](#), [T1P#](#) and T1P#=. Choose T1P# to be the largest number of persons summarized by one "#" from any of the separate runs.

T1P# can increase the number of persons for each # above the value computed by Winsteps.

Example: In one run, the bottom of Table 1 (or Table 12) states that
 EACH "#" IN THE PERSON COLUMN IS 250 PERSONS
 In another run: EACH "#" IN THE PERSON COLUMN IS 300 PERSON
 To make the runs visually comparable, specify the bigger value:
 T1P#=300

11.179 T1WEIGHT= show by weight in Table 1

Weighting	T1WEIGHT= No	T1WEIGHT= Yes
IWEIGHT=0, PWEIGHT=0	Items, persons not shown in Tables 1, 12, 16	Not shown in Tables 1, 12, 16
IWEIGHT=, PWEIGHT=	Items, persons shown in Tables 1, 12, 16	Items, persons shown once each in Tables 1, 12, 16
IWEIGHT=weighting , PWEIGHT=weighting	Items, persons shown once in Tables 1, 12, 16	Items, persons shown weight-tiems each in Tables 1, 12, 16

11.180 T2SELECT= selects Table 2 subtables

T2SELECT= specifies which sub-tables in Tables 2 are to display. "1" to display the subtable. "0" to omit the subtable. Formerly included with [CURVES=](#)

T2SELECT=1010111 displays subtables 2.1, 2.3, 2.5, 2.6, 2.7, when Table 2 is selected.

T2SELECT=	Table 2 displays	
10000000	2.1.2.11	Modal categories (most probable)
01000000	2.2.2.12	Mean categories (average or expected: Rasch-half-point thresholds)
00100000	2.3.2.13	Median categories (cumulative probabilities: Rasch-Thurstone thresholds)
0001000	2.4.2.14	Structure calibrations (Rasch model parameters: rating scale, partial credit, "restricted", "unrestricted": Rasch-Andrich thresholds)
0000100	2.5.2.15	Observed average measures of persons for scored categories (empirical averages)
0000010	2.6.2.16	Observed average measures of persons (empirical averages)
0000001	2.7.2.17	Expected average measures of persons

11.181 T45OPTIONS= field choices for Table 45

See [Table 45](#)

11.182 T7OPTIONS = selects Table 7.1 and 11.1 detail lines

T7OPTIONS= specifies which detail lines are shown in [Table 7.1](#) and [Table 11.1](#) . The default is T7OPTIONS=OZ, display *Observations* and *Z-standardized* residuals

T7OPTIONS= some or all of	Table 7.1 and Table 11.1 display:
O	Observations
E	Expected values of the observations
R	Residual = Observation - Expectation
Z	Z-standardized residual = Residual / square-root(Model variance of observation around its expectation)

Example: Table 7.1 with T7OPTIONS = OERZ in an analysis of [Example0.txt](#):

```
TABLE OF POORLY FITTING KID      (ACT IN ENTRY ORDER)
NUMBER - NAME -- POSITION ----- MEASURE - INFIT (MNSQ)  OUTFIT

      72  M Jackson, Solomon      -1.32    2.0  A    5.2
OBSERVED:      1:  1  0  0  1  0  0  0  1  0  0  0  <- T7OPTIONS= O
EXPECTED:      0.6 0.7 0.3 0.1 0.1  0.4 0.2 0.1 0.3 1.1 <- T7OPTIONS= E
RESIDUAL:      0.4 -0.7 -0.3 0.9 -0.1 -0.4 -0.2 0.9 -0.3 -1.1 <- T7OPTIONS= R
Z-RESIDUAL:      2                2                <- T7OPTIONS= Z
```

11.183 TABLES= output tables

For more flexibility, use the [Output Tables](#) pull-down menu.

Tables= causes the specified Tables to be written into the report output file. It has no effect on the pull-down menus.

A character string that tells Winsteps which output tables to prepare for printing. The sequence number of the "1" or "0" in the TABLES= string matches the table number. For more elaborate table selection, use [TFILE=](#).

"1" means prepare the corresponding table.

"0" or anything else means do not prepare the corresponding table.

TABLES= is processed after [TFILE=](#) so that control variables in TFIL= are active for TABLES=

[TABLE 0](#) is appended to the end of the TABLES= output except if TFIL= is also specified.

Example 1: You want only Tables 2,4,6,8,10 and 20 to be prepared

```
TABLES=01010101010000000001000
```

This is the same as specifying:

```
TFIL=*
```

```
2
```

```
4
```

```
6
```

```
8
```

```
10
```

```
20
```

```
*
```

Example 2: You want only Tables 1-4.

```
TABLES=1111
```

Example 3. To save all the tables at one time, here are three ways to do this:

= probability of 1 * probability of 0 for dichotomies,
 so maximum value is 0.25 when person ability measure = item difficulty measure

b) for targeting:

weighted observation = variance * observation

weighted expected score = variance * expected score

c) sum these across persons and items (and structures)

d) required "targeted" estimates are obtained when, for each person, item, structure sum (weighted observations) = sum (weighted expected scores)

e) for calculation of fit statistics and [displacement](#), weights of 1 are used but with the targeted parameter estimates. Displacement size and excessive misfit indicate how much "off-target" aberrant behavior exists in the data.

For targeting, there are many patterns of responses that can cause infinite measures, e.g. all items correct except for the easiest one. The convergence criteria limit how extreme the reported measures will be.

Polytomous items: look at the [Item Information function](#) from the [Graphs menu](#) to see the weighting. With TARGET=Y, responses are weighted by their information.

11.185 TCCFILE= Test Characteristic Curve and Test Information Function

TCCFILE= specifies the file to hold the coordinates for the Test Characteristic Curve and the Test Information Function. See [Test Characteristic Curve](#) and the [Test Information Function](#) in the [Graph Window](#), also [Table 20](#). Also the accumulated probabilities across rating-scale categories for all items.

; TCC + TIF CURVES							The contents of the TCCFILE=
MEASURE	SCORE	INFO	SEM	0	1	2	
-8.00	.05	.05	4.31	24.95	.05	.00	
-7.98	.06	.05	4.27	24.94	.05	.00	
-7.97	.06	.06	4.24	24.94	.06	.00	
-7.95	.06	.06	4.20	24.94	.06	.00	
-7.94	.06	.06	4.17	24.94	.06	.00	
-7.92	.06	.06	4.14	24.94	.06	.00	
-7.90	.06	.06	4.11	24.94	.06	.00	
MEASURE							The person ability measure (theta in logits or user-scaled USCALE= units)
SCORE							Test Characteristic Curve, TCC: the expected score on the test for the MEASURE
INFO							Test Information Function, TIF: the statistical information in the score
SEM							Standard Error of Measurement for the MEASURE when estimated from the SCORE = $1 / \sqrt{\text{INFO}}$
0							Cumulative probability of lowest category (with lowest category numbered 0)
...							Cumulative probability of intermediate categories
Highest							Cumulative probability of highest category

Example 1: www.rasch.org/memo1.htm Figure 2 shows the TCC for two person groups on the same plot. How can I do this?

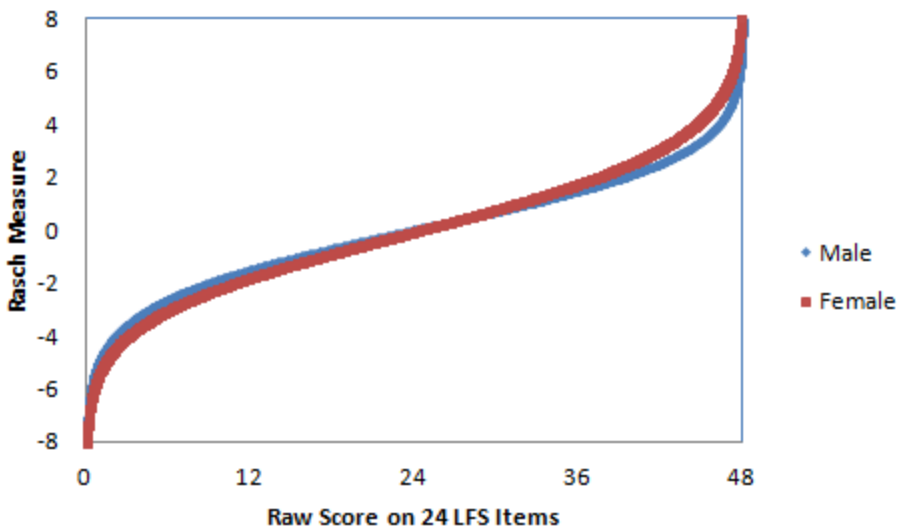
These needs two analyses, but only one control file and data file. Put the group code in column 1 of the person label. Then, here is the procedure with [Example0.txt](#) - since item 19 is extreme for one group, it is omitted with IDELETE=19 from both groups

analysis 1: at the Extra Specification prompt, type the person group code for the first group:
PSELECT=F
do the analysis, then
Output Files menu:
Output TCCFILE= to Excel temporary file

analysis 2: at the Extra Specification prompt, type the person group code for the second group:
PSELECT=M
do the analysis, then
Output Files menu:
Output TCCFILE= to Excel

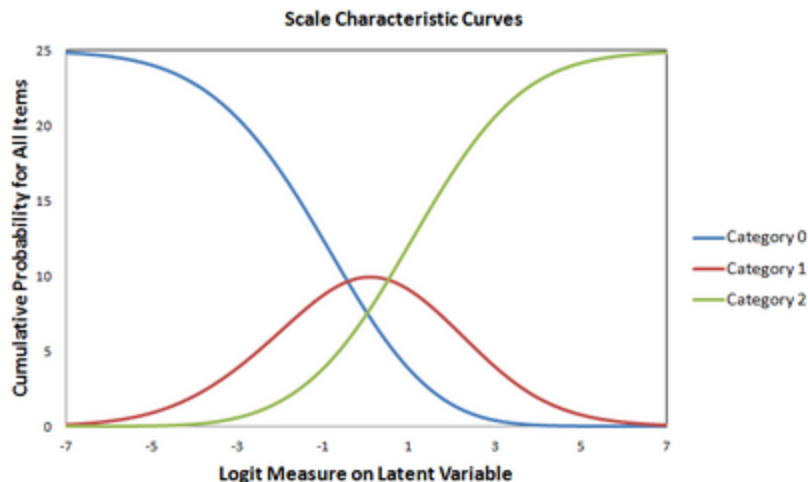
Excel plot:
copy the first two columns of both Excel files into one worksheet.
scatterplot the MEASURE columns against the matching SCORE columns

Here is what I see:



Example 2: The Scale Characteristic Curves in Dogan (2018) "An Application of the Partial Credit IRT Model in Identifying Benchmarks for Polytomous Rating Scale Instruments", PARE, 23:7.

Output the TCCFILE= to Excel and plot the cumulative probability curves against the measures. Here they are for [Example0.txt](#) using the Rating Scale Model:



11.186 TFILE= input file listing tables to be output

TFILE= causes the specified Tables to be written into the report output file. It has no effect on the [Output Tables](#) menus.

TFILE= file name	file containing details
TFILE = *	in-line list which ends with *
TFILE = ?	opens a Browser window to find the TFILE= file

TFILE= contents	
Table number, such as 22	All standard subtables for the Table are output. Table 0 is the iteration report
Table and subtable number, such as 3.1	Only the specified subtable is output. Every subtable can be reported individually.
Table and subtable range, such as 23.11-33 (which means 23.11 - 23.33)	All subtables in the specified range of subtables are output
Table and subtable open-ended range, such as 23.11+ 21.0+	All subtables numbered equal to, or higher than the specified subtable are output
Control instruction = value, such as OFILE=report1.txt PSELECT={MF}	All instructions active in the Specification menu dialog box can be included. These instructions become active when Winsteps reaches them in processing TFILE= after estimation has completed.
<p><i>Example 1:</i> In this example, 3 report output files will be written. These include two versions of Table 2.2.</p> <pre>TFILE=* OFILE=table1report.txt 1 ; output Table 1 OFILE=table2.2report.txt 2.2 ; output Table 2,2 OFILE=table2.2breport.txt MRANGE=5 2.2 ; output Table 2.2 MRANGE= ; to cancel the MRANGE=5 instruction *</pre>	
<p><i>Example 2:</i> A DIF subtable is output in a format more suitable for processing by other software</p> <pre>TFILE=* DIF=\$S1W2 BOXSHOW=No HEADER=No OFILE=difoutput.txt 30.1 *</pre>	

[TABLES=](#) selects the tables in a fixed sequence, and prints only one copy. TFILE= allows the analyst to print multiple copies of tables to different specifications. TFILE= specifies the name of an input ASCII file. Each line of the file contains a table number or table number.sub-table and other control parameters, separated by blanks or commas. Unused control values are specified with "-". The list may be entered directly into the control file with TFILE=* (see Example 2).

TFILE= Parameters:	
Control variable = Value	The Control Variables allowed in the Specification Menu dialog box are can be entered here for immediate action.

Example: TITLE="My Report"							
Output Table selection: (enter unused parameters with "-" or ".")							
1: Table number. subtable	2	3	4	5	6	7	8
Distribution map 1.0, 1.1	Lowest measure	Highest measure	Rows per unit	-	Persons per '#'	Items per '#'	
Response plot 2, 7.2	Lowest measure	Highest measure	Columns per marked division	Reference category for sorting	High rating adjustment	Low rating adjustment	Unexpected only
Person fit plots 4, 5	Lowest measure	Highest measure	Columns per marked division				
Person/Item list: 6, 10, 13, 14, 15, 17, 18, 19	Low fit bar	High fit bar					
Item fit plots 8, 9, 23	Lowest measure	Highest measure	Columns per marked division				
Item map 12, 1.2	Lowest measure	Highest measure	Rows per unit	Sort column within item name	-	Items per '#'	
Item list alphabetical 15	-	-	-	Sort column within item name			
Person map 16, 1.3	Lowest measure	Highest measure	Rows per unit	Sort column within person name	Persons per '#'		
Person list alphabetical 19	-	-	-	Sort column within person name			
Score table 20	-	-	Columns per marked division				
Category curves 21	Lowest measure	Highest measure	Columns per marked division				

TFILE= Table number selection		
Table number, for example 15	selects the Table and all its Subtables	

Table number.Subtable number For example: 6.4	selects only the Table's Subtable	For each Subtable request, Winsteps regenerates all the required Table information. For example, TFILE=* 23.3 ; performs Table 23 PCA analysis then reports Table 23.3 23.4 ; performs Table 23 PCA analysis then reports Table 23.4
Table number.Subtable number+ For example: 18.3+	select the specified Subtable and all following Subtables, if any.	
6.2, 10.2, 13.2, 14.2, 15.2, 17.2, 18.2, 19.2, 25.2, 26.2	selects the specified Subtable with CHART=Yes	
10.3, 13.3, 14.3, 15.3, 25.3, 26.3	selects the specified Subtable with DISTRATORS=YES	
3.3	STEPT3.NE.'Y'	

Example 1: The analyst wishes to select and print several tables:

TFILE=TABLES.TF

TABLES.TF is a DOS (ASCII) file with the following lines:					
; Table number	Low Range	High Range	Columns per unit	Sort on column	
2					; print Tables 2.1, 2.2, 2.3
10.2	0.5	1.5			; print Table 10.2 with fit bars at 0.5 and 1.5
8	-5	5			; print Table 8 with range -5 to +5 logits
9	-2	7			; range -2 to +7 logits, 10 columns per logit
9	-5	5	10		; print Table 9 again, different range
15	-	-	10	4	; print Table 15, sorted on column 4 of item name

or enter directly into the control file,

```
TFILE=*
2
10.2
8 -5 5
9 -2 7 10
9 -5 5 10
15 - - - 4
*
```

Example 2: Analyst wishes to specify on the DOS control line, Table 15 sorted on item name column 4. Values are separated by commas, because blanks act as end-of-line separators.

C:>Winsteps SF.TXT SFO.TXT TFILE=* 15,-,-,4 *

Example 3: To output the empirical Keyforms in Table 17.3, etc.:

```
TFILE=*
17.3
*
```


11.187 TITLE= title for output listing

Use TITLE= to label output distinctly and uniquely.

Up to 60 characters of title. This title will be printed at the top of each page of output.

Example: You want the title to be: Analysis of Math Test
TITLE="Analysis of Math Test"

Quote marks "" or '' are required if the title contains any blanks.

11.188 TOTALSCORE= show total scores with extreme observations

TOTALSCORE=No

Winsteps uses an adjusted raw score for estimation, from which observations that form part of extreme scores or [CUTLO=](#) or [CUTHI=](#) have been dropped. Scored responses are transformed (re-counted) so that the lowest response is zero. This is displayed in [Table 13.1](#), [PFILE=](#), [IFILE=](#), [TABLE 20](#), [SCOREFILE=](#) etc.

TOTALSCORE=Yes

The total raw score from the data file, after any recoding and weighting, is shown. This usually matches the numbers used in raw-score analysis and the counts in [Table 13.3](#), [TABLE 20](#), [SCOREFILE=](#) etc.

This can be changed from the *Specification* pull-down menu.

Example: KCT.txt with TOTALSCORE=No. This shows scores on 15 non-extreme items.

```

+-----+
|ENTRY  RAW          | INFIT | OUTFIT | SCORE| | | |
|NUMBER SCORE COUNT MEASURE ERROR|MNSQ  ZSTD|MNSQ  ZSTD|CORR.|WEIGH| KID  |
+-----+
|  24   12    15    3.50   .91|1.81  1.4|.79   .0|.52|1.00| Rick  M|
|   7    11    15    2.68   .90|1.64  1.1|1.49  .1|.59|1.00| Susan  F|
|  15    11    15    2.68   .90|.35  -1.7|.15  -.4|.76|1.00| Frank  M|

```

With TOTALSCORE=Y. This shows scores on all 18 items.

```

+-----+
|ENTRY  TOTAL        | INFIT | OUTFIT | SCORE| | | |
|NUMBER SCORE COUNT MEASURE ERROR|MNSQ  ZSTD|MNSQ  ZSTD|CORR.|WEIGH| KID  |
+-----+
|  24   15    18    3.50   .91|1.81  1.4|.79   .0|.52|1.00| Rick  M|
|   7   14    18    2.68   .90|1.64  1.1|1.49  .1|.59|1.00| Susan  F|
|  15   14    18    2.68   .90|.35  -1.7|.15  -.4|.76|1.00| Frank  M|

```

11.189 TRPOFILE= transposing the data matrix

To transpose the rows and columns of the data matrix, select Transpose on the [Output Files menu](#) or include TRPOFILE= in your control file. Details are given at [Transpose file specifications](#).

TRPOFILE= file name. This is the name of the transposed file.

This operates in association with TRPOTYPE= control characters. This controls what is written to the transposed file.

TRPOFILE=? opens a Browse window

Example 1: You want to "rack" the data in exam12lo.txt and exam12hi.txt, so that each person record contains the ratings from both files:

1. Analyze exam12.txt with data=exam12lo.txt
Transpose the data file only to exam12lod.txt
The "lo" items are now rows

2. Analyze exam12.txt with data=exam12hi.txt
Transpose the data only to exam12hid.txt
The "hi" items are now rows

Transpose the control file only to exam12tran.txt
The person labels are now shown as item labels

3. Analyze exam12tran.txt with data=exam12lod.txt+exam12hid.txt
The columns are the original persons. The rows are all the original items, "lo" and "hi".
Transpose the control and data file to exam12rack.txt
The persons return to the rows.
The columns are now all the original items, "lo" and "hi".

4. Analyze exam12rack.txt with Winsteps. This is the desired control and data file.

Example 2: We want to see if one group of persons have a different response "style" than another group of persons on a 6-category rating scale. To do this,

1. Transpose the Winsteps control-and-data file so that the rows become columns: Winsteps output file menu: Transpose.
2. In the transposed Winsteps control-and-data file,
GROUPS = 12112111.... to match the person groups.
3. Save the transposed file and analyze it with Winsteps.
4. Look at Table 3.2. There will be one rating scale for Group 1 and another rating scale for Group 2.

11.190 TRPOTYPE= transposing specifications

TRPOTYPE= controls the format of the transposed file written to [TRPOFILE=](#). Details are given at [Transpose file specifications](#).

Transposing switches around the rows and the columns in the Winsteps control and data file.

[TRPOTYPE=](#) has three characters:

First Character: The response-level data are:

O = Original data: data codes are those in the data file.

S = Scored data: data after applying [NEWSCORE=](#), [IVALUE=](#), [KEY=](#), etc.

R = Recounted data: the data recounted upwards from 0 after applying [STKEEP=No](#), etc.

Second Character: Deleted and deselect persons and items are: ([IDFILE=](#), [IDELETE=](#), [ISELECT=](#), [PDFILE=](#), [PDELETE=](#), [PSELECT=](#))

S = Squeezed out: deleted and deselected items and persons are omitted from the transposed file.

M = Missing data: deleted and deselected items and persons remain in the transposed file, but all their responses are indicated as missing data, "."

Third Character: Output:

B = Both Control and Data file: the transposed file can be analyzed. Additional control instructions may be required in the control section in order for the analysis to proceed correctly.

C = Control file only: Only the control file for the transposed data is output. The transposed data are not output.

D = Data file only: Only the transposed data are output. The control instructions are not.

Example: Transpose the multiple choice data after rescaling.

TRPOFILE = transposed-file.txt

TRPOTYPE = SMD ; Scored data, Missing data codes for deletions, Data only

11.191 UAMOVE= move all anchor values a pre-set amount

Anchor (fixed) values for the persons are specified by [PAFILE=](#) and for the items by [IAFILE=](#). All the anchor values can be increased or decreased by the same amount using UAMOVE=. The value is applied before any rescaling with [UASCALE=](#) or [USCALE=](#). [UJMEAN=](#) and [UPMEAN=](#) are ignored.

Reported values = USCALE= * (UAMOVE= + anchor value)/UASCALE=

Example 1: All anchor values are to be increased by 1.5 units for this analysis:

IAFILE = *

12 2.5 ; the anchor value for item 12 is 2.5 units

*

UAMOVE = 1.5 ; the amount by which to increase the anchor values
Item 12 will be reported in Table 14 with an anchor value "A" of $2.5 + 1.5 = 4.0$ logits.

Example 2: Anchor values in logits centered on 0. Report values to be user-scaled 100 units per logit, centered on 500.

UASCALE=1
USCALE=100
UAMOVE = $500/100 = 5$
IAFILE=*
1 0.5 ; 0.5 logits
*

Reported value for item 1 anchor value = 0.5 logits = $100 * (5 + 0.5) = 550$ user-units

11.192 UANCHOR= anchor values supplied in user-scaled units

For backwards compatibility only.

UANCHOR= simplifies conversion from previously computed logit measures to user-scaled measures.

If UASCALE<>0.0 then UANCHOR=A is forced.

If UASCALE=0.0 then UANCHOR=N is forced.

UANCHOR=A or N or L specifies that the anchor values are in [UASCALE=](#) units per logit. Reported measures will be user-rescaled by [UMEAN=](#) (or [UIMEAN=](#) or [UPMEAN=](#)) and [USCALE=](#).

Logit measure = anchored measure/UASCALE - UMEAN/USCALE

UANCHOR=Y specifies that anchor values are in USCALE= units per logit.

Example 1: Your item bank calibrations are user-scaled with 10 units per logits, but you want to report person measures in CHIPS (BTD p.201):

UASCALE=10 ; user-scaling of anchor values
UANCHOR=A
USCALE=4.55

Example 2: Your previous test was in logits, but now you want the user-scaling 100 per logit.

UASCALE=1 ; user-scaling of anchor values: logits
UANCHOR=Logits
USCALE=100

11.193 UASCALE= anchor user-scale value of 1 logit

Specifies the number of units per logit of the anchor values.

UASCALE=0 - the anchor values are in the current output scale set with [USCALE=](#)

UASCALE= other value - this is the units per logit of the anchor values.

[UIMEAN=](#) and [UPMEAN=](#) are ignored when there are anchor values.

Example 1: The anchor values are user-scaled such that 1 logit is 45.5 units, so that differences of -100, -50, 0, +50, +100 correspond to success rates of 10%, 25%, 50%, 75%, 90%:

UASCALE = 45.5 ; 1 logit = 45.4 user-units

Example 2: The anchor values are on one scaling, but you want the reported values to be on another scaling.

UASCALE= 5 ; there are 5 units per logit in the scaled values
USCALE=10 ; you want 10 units per logit on the reported values

Example 3: The anchor measures are in logit scaling, the output measures are to be in user-scaling.

UASCALE= 1 ; anchor values in logits

USCALE=30 ; 30 user-units per logit

Example 4: Anchor measures are in same scaling as the output measures

UASCALE= 0 ; anchor values in USCALE= scaling
USCALE=user- ; also used for anchor values
value

Example 5: Anchor values are scaled in Probits, not Logits

UASCALE = 0.588 ; [1 logit = 0.588 probits](#)

Example 6: Anchor values are to be rescaled with a new range

IAFILE= (values) the anchor values in anchor-value scaling
UASCALE= value ; anchor value scaling: on input anchor values converted to logits
TFILE=* in control file: done after analysis, but before reporting
USCALE=value user-rescaling with new range
UIMEAN=value user-rescaling with new offset
..... any other instructions for TFIL=

*

11.194 UCOUNT= number of unexpected responses: Tables 6, 10

This sets the maximum number of "most unexpected responses" to report in [Tables 6.6, 10.6](#). Also the maximum number of persons and items to report in the anti-Guttman matrices in [Tables 6.4, 6.5, 10.4, 10.5](#)

MOST UNEXPECTED RESPONSES

DATA	OBSERVED	EXPECTED	RESIDUAL	ST. RES.	ACT	KID	ACT	KID
0	0	1.93	-1.93	-7.66	18	73	GO ON PICNIC	SANDBERG, RYNE
2	2	.07	1.93	7.57	23	72	WATCH A RAT	JACKSON, SOLOMON
2	2	.07	1.93	7.57	23	29	WATCH A RAT	LANDMAN, ALAN
0	0	1.93	-1.93	-7.40	19	71	GO TO ZOO	STOLLER, DAVE

Example: Show 100 "Most Unexpected Responses" in Table 10.6:

UCOUNT=100

11.195 UDECIMALS= number of decimal places reported

This is useful for presenting your output measures and calibrations in a clear manner by removing meaningless decimal places from the output. Range is 0 (12345.) to 4 (1.2345).

How small is meaningless? Look at the Standard Error columns. Any value clearly less than a standard error has little statistical meaning.

Use the "Specification" pull-down menu to alter the value of UDECIMALS= for individual reports.

Example 1: You want to report measures and calibrations to the nearest integer:

UDECIMALS = 0

Example 2: You want to report measures and calibrations to 4 decimal places because of a highly precise, though arbitrary, pass-fail criterion level:

UDECIMALS = 4

11.196 UEXTREME= include extreme scores for UIMEAN= or UPMEAN=

In unanchored analyses, the item difficulties are centered by [UIMEAN=](#) or the person abilities are centered by [UPMEAN=](#).

UEXTREME = No (the default). Do **not** include items or persons with extreme scores (minimum possible or maximum possible) when computing the centering.

UEXTREME = Yes. Do include items or persons with extreme scores (minimum possible or maximum possible) when computing the centering.

Example: using [Exam1.txt](#)

UIMEAN=0 and UEXTREME=No (items with extreme scores **excluded** when setting the local origin at zero)

SUMMARY OF 14 MEASURED (NON-EXTREME) TAP								
	TOTAL			MODEL	INFIT		OUTFIT	
	SCORE	COUNT	MEASURE	ERROR	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	16.9	35.0	.00	.71	.96	.0	.68	-.1
P.SD	12.9	.0	3.48	.21	.28	.7	.58	.5
MAX.	32.0	35.0	4.80	1.07	1.56	1.2	2.21	1.1
MIN.	1.0	35.0	-4.40	.45	.59	-1.3	.11	-.6

SUMMARY OF 18 MEASURED (EXTREME AND NON-EXTREME) TAP								
	TOTAL			MODEL	INFIT		OUTFIT	
	SCORE	COUNT	MEASURE	ERROR	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	18.9	35.0	-.76	.96				
P.SD	14.0	.0	4.26	.51				
MAX.	35.0	35.0	6.13	1.85				
MIN.	.0	35.0	-6.59	.45	.59	-1.3	.11	-.6

UIMEAN=0 and UEXTREME=YES (items with extreme scores **included** when setting the local origin at zero)

SUMMARY OF 14 MEASURED (NON-EXTREME) TAP								
	TOTAL			MODEL	INFIT		OUTFIT	
	SCORE	COUNT	MEASURE	ERROR	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	16.9	35.0	.75	.71	.96	.0	.68	-.1
P.SD	12.9	.0	3.46	.21	.27	.7	.57	.5
MAX.	32.0	35.0	5.53	1.07	1.55	1.2	2.16	1.1
MIN.	1.0	35.0	-3.63	.45	.59	-1.4	.11	-.6
REAL RMSE	.77	TRUE SD	3.38	SEPARATION	4.40	TAP	RELIABILITY	.95
MODEL RMSE	.74	TRUE SD	3.38	SEPARATION	4.57	TAP	RELIABILITY	.95
S.E. OF TAP MEAN = .96								

SUMMARY OF 18 MEASURED (EXTREME AND NON-EXTREME) TAP								
	TOTAL			MODEL	INFIT		OUTFIT	
	SCORE	COUNT	MEASURE	ERROR	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	18.9	35.0	.00	.96				
P.SD	14.0	.0	4.24	.50				
MAX.	35.0	35.0	6.86	1.84				
MIN.	.0	35.0	-5.80	.45	.59	-1.4	.11	-.6

11.197 UIMEAN=or UMEAN= the mean or center of the item difficulties

Assigns your chosen numerical value to the average measure of the non-extreme items ([UEXTREME=No](#)) or all items ([UEXTREME=Yes](#)), i.e., a criterion-referenced measure. UMEAN= and UIMEAN= are the same specification. The last UIMEAN= or [UPMEAN=](#) values in the control file is actioned.

[UIMEAN=](#) and [UPMEAN=](#) are ignored when there are anchor values, [IAFILE=](#) or [PAFILE=](#). Anchor values are treated according to [UASCALE=](#)

Table 20 gives the UMEAN= and [USCALE=](#) values for a conversion that gives the measures a range of 0-100.

Example 1: You want to recenter the item measures at 10 logits, and so add 10 logits to all reported measures, to avoid reporting negative measures for low achievers:
UIMEAN = 10

Example 2: You want to recenter and user-rescale the item measures, so that the range of observable measures goes from 0 to 100.

Look at Table 20.1. Beneath the Table are shown the requisite values, e.g.,
UMEAN = 48.3 ; this is the same as UIMEAN=48.3
USCALE = 9.7

For more examples, and how to compute this by hand, see [User-friendly rescaling](#)

Rescaling with anchor values:

When you do the analysis, you need the anchored values and their scaling:

UASCALE= anchor values per logit

IAFILE=*

1 20

2 34

...

*

Then you have options:

A. Use the "Help" menu scaling calculator to discover the new values you want:

UIMEAN= new mean

USCALE= new scaling

B. If you know the rescaling you want, include in the control file:

TFILE=*

UIMEAN= new mean

USCALE= new scaling

*

In both options, use the Output Tables menu or Output Files menu to produce the output you need.

11.198 UMEAN=or UIMEAN= the mean or center of the item difficulties

See [UIMEAN=](#)

11.199 UPMEAN= the mean or center of the person abilities

Assigns your chosen numerical value to the average of the non-extreme abilities for persons ([UEXTREME=No](#)) or all persons ([UEXTREME=Yes](#)), i.e., this provides a norm-referenced user-scaling.

The last UIMEAN= or [UPMEAN=](#) values in the control file is actioned.

[UIMEAN=](#) and [UPMEAN=](#) are ignored when there are anchor values, [IAFILE=](#) or [PAFILE=](#). Anchor values are treated according to [UASCALE=](#)

Example 1: You want to used conventional IRT norm-referenced user-scaling with person mean of 0 and person S.D. of 1.

UPMEAN = 0

USCALE = 1 / (person S.D. in logits) ; find this from Table 3.1

If there are extreme person scores, see [User-friendly rescaling](#)

Example 2: I want to compare the mean performance of random samples of examinees from my database. Will UPMEAN= help in this?

UPMEAN=0 sets the local mean of the persons to zero (excluding extreme scores) regardless of the sampling.

If you wish to investigate the behavior of the person mean for different person samples, then

(1) analyze all persons and items: set UPMEAN=0, for convenience, and write [IFILE=](#). For better comparison, set [STBIAS=NO](#).

(2) anchor the items using [IAFILE=](#)

(3) analyze samples of persons with the anchored items.

The person means reported in [Table 3.1](#) now show person means (with or without extreme scores) in the one frame of reference across all analyses defined by the anchored items.

For more examples, and how to compute this by hand, see [User-friendly rescaling](#)

Example 3: We want the mean of a target group in our analysis to be 50.

1. Put a code in the person label for the target group (or sort them to be adjacent rows) in the data file.
2. Do a free analysis of everything.
3. In the Specification menu, [PSELECT=](#) the target group on the code or [PDELETE=+](#) the target range of rows.
4. Output [Table 3.1](#) - this will give you the mean of the target group
5. Winsteps Specification menu: UPMEAN = - mean of target group
6. Output [Table 3.1](#) - this will give you the mean of the target group which should be 0.
7. If not mean=0, then UPMEAN = - mean of target group from step 4, adjusted by value in step 6.
8. Repeat 5. and 6. if needed.
9. Write down the UPMEAN= value for later, if needed. You can see it in the [Control Variables](#) list.
10. Reinstate everyone: PSELECT= or PDELETE=
11. Run your reports.

Item Easiness = Reversed Item Difficulties

The reversal formula for the items on the maps in [Tables 1.10, 1.11, 1.12, 1.13, 12.12, and 16.13](#) is:

Item Easiness = 2*Mean Person Ability - Item Difficulty

For a full set of item easiness statistics, set the Mean Person Ability as the reference point, then reverse the user-scaling . Use the [Scaling Calculator](#) or:

For the person ability statistics:

UEXTREME = Yes

UPMEAN = 0 or preferred reference value

USCALE = current value of USCALE=

For the matching item easiness statistics:

UEXTREME = Yes

UPMEAN = 0 or preferred reference value

USCALE = - current value of USCALE=

11.200 USCALE= the user-scaled value of 1 logit

Specifies the number of reported user-scaled units per logit. When USCALE=1 (or USCALE= is omitted) then all measures are in [logits](#).

[Table 20](#) gives the [UMEAN=](#) and [USCALE=](#) values for a conversion that gives the measures a range of 0-100.

Example 1: You want to user-rescale 1 logit into 45.5 units, so that differences of -100, -50, 0, +50, +100 correspond to success rates of 10%, 25%, 50%, 75%, 90%:

USCALE = 45.5

Example 2: You want to reverse the measurement directions, since the data matrix is transposed so that the 'items' are examinees and the 'persons' are test questions:

USCALE = -1

KEYn=, RESCORE=, ISGROUPS= will still apply to the columns, not the rows, of the data matrix. Centering will still be on the column measures.

Example 3: You want to approximate the "probit" measures used in many statistical procedures.

UPMEAN = 0 ; set the person sample mean to zero

USCALE = 0.59 ; probits * 1.7 = logits

but if you want your person sample to approximate N(0,1):

1. first analysis:

UPMEAN=0

USCALE=1

look at the person S.D. in Table 3.1

2. second analysis:

UPMEAN=0

USCALE=1/person S.D. from 1.

Example 4: You want to report measures in "log-odds units to the base 10" (lods) instead of the standard "logits to the base e".

USCALE=0.434294482 ; the conversion between "natural" and base-10 logarithms.

For more examples, and how to compute this by hand, see [User-friendly rescaling](#)

11.201 UTF8ENC= UTF-8 encoding

UTF8ENC encoding = Yes	This control and data file are encoded with UTF-8
UTF8ENC encoding = No	This control and data file are not encoded with UTF-8
UTF8ENC encoding = ? or anything else	Winsteps determines whether the file is encoded with UTF-8

If the file is encoded with UTF-8, but the UTF-8 code cannot be displayed, then the [UTF8SUB=](#) character is displayed

11.202 UTF8SUB= substitute character for multibyte UTF-8

Winsteps replaces multibyte UTF-8 codes with a substitute ASCII code when needed to maintain column-alignment in Tables.

The freeware text-editor NotePad++ <http://notepad-plus-plus.org/> displays UTF-8 codes.

Example: UTF8CODE= ~

Table 6.6

```

-----
| DATA |OBSERVED|EXPECTED|RESIDUAL|ST. RES.|MEASDIFF| ITEM | PERSON | ITEM |
|-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 2 | 2 | .39 | 1.61 | 2.04 | -.71 | 4 | 6 | ~~~~~~ | 馬
| 蒔 撰 |
| 初 編 |
|-----+-----+-----+-----+-----+-----+-----+-----+-----+
substituting for 四部叢刊

```


11.203 W300= Output files in Winsteps 3.00 format

New or revised statistics are added as extra or changed columns in IFILE= and PFILE= when they are introduced into Winsteps.

To revert to an earlier format of [IFILE=](#) and [PFILE=](#), specify W300=Yes. This produces the PFILE= and IFILE= in the format of Winsteps 3.00 1/1/2000.

Example: IFILE= in current format:

```
;ACT      FILE FOR
;ENTRY MEASURE STTS COUNT  SCORE  ERROR IN.MSQ IN.ZSTD OUT.MS OUT.ZSTD DISPL PTME WEIGHT  DISCR G M NAME
  1   -.89  1   75.0   109.0   .23   .74  -1.97   .67  -1.89   .00   .64  1.00   1.06  0 R WATCH BIRDS
  2   -.61  1   75.0   116.0   .20   .76  -1.54   .56  -1.55   .00   .58  1.00   1.07  0 R READ BOOKS ON ANIMALS
```

IFILE= in W300=Yes format:

```
;ACT      FILE FOR
;ENTRY MEASURE ST COUNT  SCORE  ERROR IN.MSQ IN.ZSTD OUT.MS OUT.ZSTD DISPL CORR G M NAME
  1   -.89  1   75   109   .23   .74  -1.97   .67  -1.89   .00   .64  0 R WATCH BIRDS
  2   -.61  1   75   116   .20   .76  -1.54   .56  -1.55   .00   .58  0 R READ BOOKS ON ANIMALS
```

PFILE= in current format:

```
;PUPIL      FILE FOR
;ENTRY MEASURE STTS COUNT  SCORE  ERROR IN.MSQ IN.ZSTD OUT.MS OUT.ZSTD DISPL PTME WEIGHT  DISCR NAME
  1   .49  1   25.0   30.0   .35   .96  -.15   .84  -.43   .00   .69  1.00   .00 ROSSNER, MARC DANIEL
;  2   5.99  0   25.0   50.0   1.84  1.00   .00  1.00   .00   .00   .00  1.00   .00 ROSSNER, LAWRENCE F.
```

PFILE= in W300=Yes format:

```
;PUPIL      FILE FOR
;ENTRY MEASURE ST COUNT  SCORE  ERROR IN.MSQ IN.ZSTD OUT.MS OUT.ZSTD DISPL CORR NAME
  1   .49  1   25   30   .35   .96  -.15   .84  -.43   .00   .69 ROSSNER, MARC DANIEL
  2   5.99  0   25   50   1.84  1.00   .00  1.00   .00   .00   .00 ROSSNER, LAWRENCE F.
```

Notes:

TOTAL=YES is active for both current and old formats.

;" shown for extreme scores, such as person 2, in current format, but not in old format.

COUNT and SCORE are shown rounded to nearest integer in old format.

11.204 WEBFONT= font of webpages

WEBFONT= is used to specify the font for webpage display with [ASCII=](#)Webpage.

A standard font instruction is:

```
WEBFONT="Lucida Console, Andale Mono, Courier New"
```

See [ASCII=](#) for how this works.

11.205 WHEXACT= Wilson-Hilferty exact normalization

Some early versions of Winsteps have the standard WHEXACT=NO.

ZSTD INFIT is the "t standardized Weighted Mean Square" shown at the bottom of RSA p. 100. ZSTD (standardized as a z-score) is used of a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's [t-statistic](#) distribution value has been adjusted to a unit normal value.

ZSTD OUTFIT is the "t standardized Unweighted Mean Square" based on the terms on RSA p. 100.

The Wilson-Hilferty transformation converts mean-square values to their equivalent "t standardized" normal deviates. See RSA p. 101

$$t_i = (v_i^{1/3} - 1)(3/q_i) + q_i / 3$$

The degrees of freedom are $2/(q_i * q_i)$. The Wilson-Hilferty transformation is accurate for $d.f. > 25$. With very small $d.f.$, it becomes misleading, consequently the $d.f.$ are not allowed to fall below 1, so that $q_i \leq 1.4142$.

Under certain circumstances, it can report the paradoxical finding that the mean-square apparently reports an overfit condition, but the normal deviate an underfit.

To allow this possibility, specify `WHEXACT=Y`

To suppress it, specify `WHEXACT=N`. The final $q/3$ term is omitted from the transformation.

Does not apply if `NORMAL=Yes..`

Example: A person takes a test of 20 dichotomous items and obtains an unweighted chi-square value of 19.5.

`WHEXACT=Y`

The OUTFIT mean-square is 0.975, i.e., apparently slightly overfitting. The exact normal deviate is .03, i.e., very slightly underfitting.

`WHEXACT=N`

The OUTFIT mean-square is 0.975, i.e., apparently slightly overfitting. The reported normal deviate is -.08, i.e., slightly overfitting.

11.206 WMLE = Warm's Mean Likelihood Estimation

`WMLE=Yes` reports [Warm's Mean Likelihood](#) estimates in the measure tables.

11.207 XFILE= analyzed response file

If `XFILE=filename` is specified in the control file, a file is output which enables a detailed analysis of individual response anomalies.

`XFILE=?` opens a Browse window

This file contains 4 heading lines (unless `HLINES=N` or `ROW1HEADING=N`) followed by one line for each person-by-item response used in the estimation. Each line contains:

Field number	Description	Number Format	Abbreviation
1	Person number	7.0	PERSON
2	Item number	7.0	ITEM
3	Response value (after scoring with <code>KEY=</code> , <code>IVALUE=</code> , etc.)	4.0	OBS
4	Response value (after scoring and recounting. This usually only happens for rating scales with unobserved intermediate categories and <code>STKEEP=NO</code> .)	4.0	ORD
5	Expected response value. For dichotomous items, probability of success. This is computed from the measures without correction by <code>STBIAS=</code> . If you want to see the values with <code>STBIAS=YES</code> , then: 1. Perform the analysis with <code>STBIAS=YES</code> 2. Output <code>IFILE=if.txt</code> <code>PFILE=PF.txt</code> <code>SFILE=sf.txt</code> 3. Perform the analysis again with <code>STBIAS=NO</code> <code>IAFILE=if.txt</code> <code>PAFILE=pf.txt</code> <code>SAFILE=sf.txt</code>	7.3	EXPECT
6	Modeled Variance of observed values around the expected value This is also the statistical information in the observation. Square root (modeled variance) is the observation's raw-score standard deviation	7.3	VARIAN
7	Standardized residual: (Observed - Expected)/Square root (Variance). This approximates a unit-normal deviate. Values outside ± 2 are unexpected.	7.3	ZSCORE
8	Score residual: (Observed - Expected)	7.3	RESIDL

9	Person measure in <u>USCALE=</u> units	7.2*	PERMEA
10	Item measure in USCALE= units	7.2*	ITMMEA
11	Measure difference (Person measure - Item measure) in USCALE= units	7.2*	MEASDF
12	Log-Probability of observed response. These can be summed for the Log-Likelihood Chi-Square, Akaike Information Criterion (<u>AIC</u>), and Bayesian Information Criterion (<u>BIC</u>), etc.	7.3	L-PROB
13	Predicted person measure from this response alone in USCALE= units	7.2*	PPMEAS
14	Predicted item measure from this response alone in USCALE= units	7.2*	PIMEAS
15	Modeled Kurtosis of the observed values around their expectation	7.3	KURT
16	More Probable of the Responses to be observed	4.0	MPR
17	Response weight (person weight x item weight)	7.3	WEIGHT
18	Response status: 0 = Not scored, 1 = Standard, 2 = in Extreme person score, 3 = in Extreme item score, 4 = in Extreme person and item scores	2.0	ST
19	Response code in data file	A	CODE
20	Person label	A	PLABEL
21	Item label	A	ILABEL
7.3 means "7 columns with 3 decimal places". 2* means Decimal Places set by the XFILE= dialog box from the Output Files menu or <u>UDECIM=</u> if not set. A means alphanumeric character field			

```

: RESIDUAL FILE FOR LIKING FOR SCIENCE (Wright & Masters p.18) May 22 17:34 2013
KID  ACT  OBS  DES  EXPECT  VARIAN  ZSCORE  RESID  PERMEA  ITMMEA  MEASDF  L-PROB  PPMEAS  PIMEAS  KURTOS  MPR  WEIGHT  C  PLABEL  ILABEL
1  1  1  1  1.414  .989  -.868  -.414  .489  -.489  .959  -.919  -.489  .489  .388  2  1.000  1  ROSSNER, MARC DANIEL WATCH BIRDS
1  2  2  2  1.510  .941  .889  .472  .489  -.791  1.264  -.555  1.289  -1.587  .388  2  1.000  2  ROSSNER, MARC DANIEL READ BOOKS ON ANIMALS
1  3  1  1  1.885  .459  -.896  -.865  .489  .942  1.141  -.821  .942  .489  .457  1  1.000  1  ROSSNER, MARC DANIEL READ BOOKS ON PLANTS
1  4  1  1  .499  .959  .843  .581  .489  1.889  -1.188  -.928  1.889  .489  .321  0  1.000  1  ROSSNER, MARC DANIEL WATCH SEAS CHANGE
1  5  1  1  .381  1.246  1.418  .899  .489  2.351  -1.848  -1.327  2.351  .489  .218  0  1.000  1  ROSSNER, MARC DANIEL FIND BOTTLES AND CANS
1  6  0  0  1.115  .455  -1.854  -1.115  .489  .292  .251  -1.735  -1.838  2.555  .451  1  1.000  0  ROSSNER, MARC DANIEL LOOK UP STRANGE ANIMAL OR PLANT
1  7  2  2  .755  .495  1.897  1.245  .489  1.825  -.541  -2.878  3.895  -1.587  .423  1  1.000  2  ROSSNER, MARC DANIEL WATCH ANIMAL MOVE
1  8  0  0  .510  .965  -.873  -.510  .489  1.589  -1.188  -.835  -.481  2.555  .394  0  1.000  0  ROSSNER, MARC DANIEL LOOK IN SIDEWALK CRACKS

```

Fields can be selected interactively and the default field selection changed at [XFILE= dialog box](#).

If CSV=Y, the values are separated by commas. When CSV=T, the commas are replaced by tab characters. For "non-numeric values in quotation marks", specify QUOTED=Y.

This file enables a detailed analysis of individual response anomalies. The response residual can be analyzed in three forms:

- 1) in response-level score units, from [(observed value - expected value)].
- 2) in logits, from [(observed value - expected value)/variance].
- 3) in standard units, [(observed value - expected value)/(square root of variance)].

Predicted person measure: Imagine that this observation was the only observation made for the person ... this value is the measure we would predict for that person given the item measure.

Predicted item measure: Imagine that this observation is the only observation made for this item ... this value is the measure we would predict for that item given the person measure.

The formulas are the same as for a response string of more than 1 observation. For dichotomies, see www.rasch.org/rmt/rmt102t.htm and for polytomies www.rasch.org/rmt/rmt122q.htm

Example 1: You wish to write a file on disk called "MYDATA.XF" containing response-level information for use in examining particularly response patterns:

XFILE=MYDATA.XF

Example 2: You wish to compute differential item functioning, DIF, for a specific classification group of people:

If Table 30 is not suitable, here is a simple approximation:

Since one item does not have enough information to measure a person, for item bias we have to do it on the basis of a classification group of people.

From the XFILE,

add the "score residuals" (not standardized) for everyone in classification "A" on a particular item.

Add the "modeled variance" for everyone in the classification.

Divide the residual sum by the variance sum. This gives an estimate of the DIF for classification "A" relative to the grand mean measure.

Do the same for classification "B" on the same item.

To contrast classification "A" with classification "B" then

DIF size "AB" = DIF estimate for "A" - DIF estimate for "B"

A significance t-test is $t = \text{DIF size "AB"} / \sqrt{(1/\text{variance sum for classification A} + 1/\text{variance sum for classification B})}$

Example 3: You want to convert lucky guesses into missing data for some items.

Click on Output Files menu

Click on XFILE=

In the [XFILE= Fields](#) dialog box, type the Item numbers you want.

Click on OK

Output to Excel (Temporary file)

Sort the Excel file on Residual

The largest positive residuals are the lucky guesses

Rectangular- Copy the person entry numbers and the item entry numbers into your Winsteps control file:

[EDFILE=*](#)

person item . ; the . is to indicate "make this response missing data"

person item . ; the . is to indicate "make this response missing data"

person item . ; the . is to indicate "make this response missing data"

*

Save your Winsteps control file.

In the next analysis these unexpected correct responses will be scored as missing. The raw scores will also change.

11.208 XWIDE= columns per response

The number of columns taken up by each response in your data file (1 or 2 or more). If possible, enter your data one column per response. If there are two columns per response, make XWIDE=2. If your data includes responses entered in both 1 and 2 character-width formats, use [FORMAT=](#) to convert all to XWIDE=2 format. When XWIDE=2, these control variables require two columns per item or per response code: [CODES=](#), [KEYn=](#), [KEYSCR=](#), [NEWSCORE=](#), [IVALUE=](#). Either 1 or XWIDE= columns can be used for [RESCORE=](#), [ISGROUPS=](#), [RESCORE=](#) and [IREFER=](#)

See [Example 11](#), Item responses two characters wide.

Example 1: The responses are scanned into adjacent columns in the data records,

```
XWIDE=1 ; Observations 1 column wide
CODES=01
```

Example 2: Each response is a rating on a rating scale from 1 to 10, and so requires two columns in the data record,

```
XWIDE=2 ; 2 columns per datum
CODES=010203040506070801011 ; 2 CHARACTERS FOR EACH OBSERVATION
or
CODES=" 1 2 3 4 5 6 7 8 91011" ; 2 CHARACTERS FOR EACH OBSERVATION
```

Example 3: Some responses take one column, and some two columns in the data record. Five items of 1-character width, code "a", "b", "c", or "d", then ten items of 2-character width, coded "AA", "BB", "CC", "DD". These are preceded by person-id of 30 characters.

```
XWIDE=2 Format to two columns per response
FORMAT=(30A1,5A1,10A2) Name 30 characters, 5 1-chars, 10 2-chars
CODES ="a b c d AABBCDD" "a" becomes "a "
NEWSCORE="1 2 3 4 1 2 3 4 " response values
RESCORE=2 rescore all items
NAME1=1 person id starts in column 1
ITEM1=31 item responses start in column 31
NI=15 15 items all now XWIDE=2
```

Example 4: Scores are entered as hundreds, but are to be analyzed as incrementing by 1

```
XWIDE = 4
```

```
CODES = " 100 200 300 400 500 600 700 800 9001000"  
STKEEP = No ; analyze the ascending scores as ordinal levels
```

For more examples, see [CODES=](#) and [Decimal, percentage and continuous data](#).

11.209 #1#= user-defined replaceable token

#1#= to #9#= are 9 user-defined tokens that can be used to replace values in other control specifications. Each can be up to 32 characters long. Tokens can reference tokens that are already defined. Undefined tokens are removed from control specifications.

Example 1:

```
&INST  
#1# = myfile  
#2# = 3  
IFILE=#1##2#.txt ; this is processed as IFILE=myfile3.txt
```

Example 2:

```
Extra Specifications? #3#=analysis2  
Control file:  
PFILE = #3#pf.txt ; this is processed as PFILE = analysis2pf.txt
```

Example 3:

```
At the DOS prompt, or in a Batch file:  
c:\myfolder>c:\Winsteps\Winsteps.exe mycontrol.txt #3#=analysis2  
Control file:  
OUTFILE = #3#out.txt ; this is processed as PFILE = analysis2out.txt  
PFILE = #3#pf.txt ; this is processed as PFILE = analysis2pf.txt
```

11.210 &END end of control variables

The first section in a control file contains the control variables, one per line. Its end is indicated by &END.

```
TITLE = "Example control file"  
ITEM1 = 1  
NI = 10  
NAME1 = 12  
&END  
..... ; Item labels here  
END LABELS
```

11.211 &INST start of control instructions

&INST is ignored by current versions of Winsteps. It is maintained for backward compatibility with earlier versions, where it was required to be the first control instruction. It is still present in some example files, again for backwards compatibility.

```
&INST ; this is allowed for compatibility  
TITLE = "Old control file"  
....
```

11.212 ; starts a comment

; starts comment for control variables or item labels. A different character can be substituted using [Edit Initial Settings](#)

Example:

```
NI = 24 ; this test has 24 items  
&END  
2+2 ; the first item tests simple addition
```

....
END LABELS

11.213 @Field= name for location in label

@Fieldname= allows for user-defined names for locations with the person or item labels to be specified with the [column selection](#) rules.

@Fieldname = value

Field name: a user-specified name which can include letters and numbers, but not = signs. Field names are converted to capital letters, and must be referenced in full. Fieldname is 1 to 9 characters, e.g., @X or @123456789 or @FIRSTNAME

Value: a user-specified values which must accord with the column selection rules.

@ILABEL= is the length of the longest item label

@PLABEL= is the length of the longest person label

Example 1: The gender of persons is in column 14 of the person label. A DIF report on gender is wanted.

@GENDER = 14 ; gender indicator in column 14 of the person label

DIF = @GENDER ; DIF classification is by Gender column

or

DIF = @gender ; lower case letters in field names are allowed

but not

DIF = @GEN ; abbreviations of field names are not allowed

TFILE=*

30 ; produce the DIF Table

*

This can also be done by the pull-down menus

Specification menu box: @GENDER = 14

Output Tables menu: 30. Items: DIF

Right-click on DIF selection box: @GENDER

Click on OK box

12 Examples of Control and Data Files

12.1 Example 0: Rating scale data: The Liking for Science data

Rather than attempting to construct a control file from scratch, it is usually easier to find one of these examples that is similar to your analysis, and modify it.

Control and data file, Example0.txt, for the Liking for Science data (see RSA) contains the responses of 75 children to 25 rating-scale items. The responses are 0-dislike, 1-neutral, 2-like. To analyze these data, start Winsteps, then:

Control file name?:

"Files" pull-down menu

"Control file name"

"Examples" folder

Example0.txt

Open

Report output file: (Enter)	<i>press Enter for a temporary file.</i>
Extra specifications:(Enter)	<i>press Enter because there are no extra specifications</i>

```
; This is file "example0.txt" - ";" starts a comment
TITLE= 'LIKING FOR SCIENCE (Wright & Masters p.18)'
NI= 25 ; 25 items
ITEM1= 1 ; responses start in column 1 of the data
NAME1= 30 ; person-label starts in column 30 of the data
```

```

ITEM= ACT ; items are called "activities"
PERSON= KID ; persons are called "kids"
CODES= 012 ; valid response codes (ratings) are 0, 1, 2
CLFILE= * ; label the response categories
0 Dislike ; names of the response categories
1 Neutral
2 Like
* ; "*" means the end of a list
<END ; this ends the control specifications
; These are brief descriptions of the 25 items
WATCH BIRDS
READ BOOKS ON ANIMALS
READ BOOKS ON PLANTS
WATCH GRASS CHANGE
FIND BOTTLES AND CANS
LOOK UP STRANGE ANIMAL OR PLANT
WATCH ANIMAL MOVE
LOOK IN SIDEWALK CRACKS
LEARN WEED NAMES
LISTEN TO BIRD SING
FIND WHERE ANIMAL LIVES
GO TO MUSEUM
GROW GARDEN
LOOK AT PICTURES OF PLANTS
READ ANIMAL STORIES
MAKE A MAP
WATCH WHAT ANIMALS EAT
GO ON PICNIC
GO TO ZOO
WATCH BUGS
WATCH BIRD MAKE NEST
FIND OUT WHAT ANIMALS EAT
WATCH A RAT
FIND OUT WHAT FLOWERS LIVE ON
TALK W/FRIENDS ABOUT PLANTS
END NAMES ;this follows the item names: - the data follow:
1211102012222021122021020 ROSSNER, MARC DANIEL
222222222222222222222222 ROSSNER, LAWRENCE F.
2211011012222122122121111 ROSSNER, TOBY G.
1010010122122111122021111 ROSSNER, MICHAEL T.
1010101001111001122111110 ROSSNER, REBECCA A.
1011211011121010122101210 ROSSNER, TR CAT
22200222022222222222222222 WRIGHT, BENJAMIN
221002102222020122022022 LAMBERT, MD., ROSS W.
0110100112122001121120211 SCHULZ, MATTHEW
2100010122221121122011021 HSIEH, DANIEL SEB
2220011102222202202202222 HSIEH, PAUL FRED
0100220101210011021000101 LIEBERMAN, DANIEL
1211000102121121122111000 LIEBERMAN, BENJAMIN
2110020212022100022000120 HWA, NANCY MARIE
111111111111111111122111111 DYSON, STEPHIE NINA
22211211222222222222122022 BUFF, MARGE BABY
2222022222222222222222122 SCHATTNER, GAIL
222202102222221222122022 ERNST, RICHARD MAX
2221011002222211122020022 FONTANILLA, HAMES
1100010002122212122020022 ANGUIANO, ROB
1111011112122111122021011 EISEN, NORM L.
121111112222121122121121 HOGAN, KATHLEEN
221112112222121222122122 VROOM, JEFF
1221012102222110122021022 TOZER, AMY ELIZABETH
222112022222221222122022 SEILER, KAREN
111111111111111111111111111 NEIMAN, RAYMOND
1110011101122111022120221 DENNY, DON
2211102002122121022012011 ALLEN, PETER
1000101001110000022110200 LANDMAN, ALAN
100000002122000022012100 NORDGREN, JAN SWEDE
2211011122210111221211111 SABILE, JACK
1210010011222110121022021 ROSSNER, JACK
2212021212222202202202021 ROSSNER, BESS
2222122222222222222222122 PASTER, RUTH
1110211011122111122011111 RINZLER, JAMES
11102010011221111220111111 AMIRAULT, ZIPPY
1210010012122111122012021 AIREHEAD, JOHN
2211021112222121122222121 MOOSE, BULLWINKLE
2221022112222122222122022 SQUIRREL, ROCKY J.
22210110222221222122022 BADENOV, BORIS
2222222222222222222222122 FATALE, NATASHA
2121011112122221122011021 LEADER, FEARLESS
2220112022221212220202021 MAN, SPIDER
1210021012212211022011021 CIANCI, BUDDY
2221022222222222222222022 MCLOUGHLIN, BILLY

```

```

2200020002222000222122010 MULLER, JEFF
1110220112212111022010000 VAN DAM, ANDY
2221022222222221222122022 CHAZELLE, BERNIE
1100110101110112111111210 BAUDET, GERARD
2222121222222222222122122 DOEPPNER, TOM
1210021022222211220210222 REISS, STEVE
1111111111111111122111111 STULTZ, NEWELL
0100101010110011021010000 SABOL, ANDREW
1111021121122112122021011 BABBOO, BABOO
1010010011122112222111110 DRISKEL, MICHAEL
2221012012222222222122022 EASTWOOD, CLINT
22200220222222222022022 CLAPP, CHARLIE
122101012222221222021021 CLAPP, DOCENT
222102112222221222022222 CLAPP, LB
122111112222121122121021 SQUILLY, MICHAEL
2111010022122111122011022 SQUILLY, BAY OF
22200000222222022020000 BEISER, ED
12110101222212222122010 BECKER, SELWIN
2210022112222121222122011 CORLEONE, VITO
121121100222012122121200 CORLEONE, MICHAEL
1201010012221202202202021 PINHEAD, ZIPPY
1211001001212121222121012 MALAPROP, MRS.
1200010102120120022001010 BOND, JAMES
12100101222212222021021 BLOFELD, VILLAIN
22220220122212222022021 KENT, CLARK
220122200122002222022201 STOLLER, DAVE
1001000100011010021200201 JACKSON, SOLOMON
2012010122210201102202022 SANDBERG, RYNE
22200220022222222020212 PATRIARCA, RAY
1200110112122221022020010 PAULING, LINUS

```

When analysis completes, use the "Output Tables" pull-down menu to examine the measures and quality-control fit statistics. Start with Table 1, which gives you a quick summary. Here are the Science Activity items in measure order, Table 13. The hierarchy of item labels, in the right-hand column, defines the construct that has been measured. "Go on picnic" is easiest to like. "Find bottles" is hardest to like.

ACT STATISTICS: MEASURE ORDER

ENTRY	RAW	MODEL	INFI	OUTFIT	PTMEA					
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	ACT
5	35	74	2.42	.22	2.30	5.6	3.62	7.3	.05	FIND BOTTLES AND CANS
23	40	74	2.18	.21	2.41	6.3	4.11	9.0	.00	WATCH A RAT
20	48	74	1.83	.20	1.33	2.0	1.82	3.7	.42	WATCH BUGS
4	50	74	1.75	.20	.89	-.7	.91	-.4	.60	WATCH GRASS CHANGE
8	52	74	1.67	.20	1.10	.7	1.21	1.2	.51	LOOK IN SIDEWALK CRACKS
7	67	74	1.10	.19	.97	-.1	1.01	.1	.59	WATCH ANIMAL MOVE
9	78	74	.71	.19	1.18	1.3	1.17	1.0	.53	LEARN WEED NAMES
16	81	74	.60	.19	.97	-.2	.95	-.3	.51	MAKE A MAP
25	83	74	.53	.19	.80	-1.5	.74	-1.6	.66	TALK W/FRIENDS ABOUT PLANTS
3	86	74	.42	.19	.57	-3.5	.54	-3.0	.72	READ BOOKS ON PLANTS
14	86	74	.42	.19	.82	-1.3	.75	-1.4	.62	LOOK AT PICTURES OF PLANTS
6	89	74	.31	.19	.81	-1.4	.76	-1.4	.61	LOOK UP STRANGE ANIMAL OR PLANT
17	93	74	.16	.19	.65	-2.7	.59	-2.4	.70	WATCH WHAT ANIMALS EAT
22	95	74	.08	.19	.83	-1.2	.74	-1.4	.63	FIND OUT WHAT ANIMALS EAT
24	105	74	-.31	.20	.90	-.6	.79	-.9	.60	FIND OUT WHAT FLOWERS LIVE ON
1	107	74	-.40	.21	.55	-3.5	.49	-2.5	.64	WATCH BIRDS
15	109	74	-.48	.21	.78	-1.5	.64	-1.6	.61	READ ANIMAL STORIES
2	114	74	-.71	.22	.93	-.4	.72	-1.0	.58	READ BOOKS ON ANIMALS
21	117	74	-.85	.22	.84	-.9	.65	-1.3	.58	WATCH BIRD MAKE NEST
11	119	74	-.96	.23	.63	-2.4	.49	-1.9	.59	FIND WHERE ANIMAL LIVES
13	125	74	-1.29	.25	1.22	1.1	.94	.0	.47	GROW GARDEN
10	128	74	-1.49	.26	.78	-1.1	.57	-1.1	.50	LISTEN TO BIRD SING
12	135	74	-2.04	.31	.70	-1.2	.51	-1.0	.45	GO TO MUSEUM
19	139	74	-2.48	.36	1.08	.4	1.10	.4	.30	GO TO ZOO
18	143	74	-3.15	.47	1.50	1.2	1.23	.5	.14	GO ON PICNIC
MEAN	93.0	74.0	.00	.23	1.02	-.2	1.08	.0		
P.SD	30.9	.0	1.41	.06	.45	2.3	.87	2.8		

12.2 Example 1: Dichotomous data: Simple control file with data included

A control file, EXAM1.TXT, for an analysis of the Knox Cube Test (see BTB) a test containing 18 items, each item is already scored dichotomously as 0,1. The person-id data begins in column 1 and the item string begins in column 11. No items will be deleted, recoded, or anchored. The number of data lines is counted to determine how many children took the test. use the pull-down "Diagnosis" and "Output Table" menus to see the output tables. For an explanation of the output obtained, see later in the manual. Run this example with:

Control file: **EXAM1.TXT**
 Report output file: **(Enter)**
 Extra specifications: **(Enter)**

```

; This file is EXAM1.TXT - (";" starts a comment)
TITLE='KNOX CUBE TEST' ; Report title
NAME1=1 ; First column of person label in data file
ITEM1=11 ; First column of responses in data file
NI=18 ; Number of items
CODES=01 ; Valid response codes in the data file
CLFILE=* ; Labels the observations
0 Wrong ; 0 in data is "wrong"
1 Right ; 1 in data is "right"
* ; "*" is the end of a list
PERSON=KID ; Person title: KID means "child"
ITEM=TAP ; Item title: TAP means "tapping pattern"
&END ; Item labels for 18 items follow
1-4 ; tapping pattern of first item: cube 1 then cube 4 are tapped.
2-3
1-2-4
1-3-4
2-1-4
3-4-1
1-4-3-2
1-4-2-3
1-3-2-4
2-4-3-1
1-3-1-2-4
1-3-2-4-3
1-4-3-2-4
1-4-2-3-4-1
1-3-2-4-1-3
1-4-2-3-1-4
1-4-3-1-2-4
4-1-3-4-2-1-4 ; last tapping pattern: 7 actions to remember!
END NAMES ; END NAMES or END LABELS must come at end of list
Richard M 11111110000000000 ; Here are the 35 person response strings
Tracie F 11111111100000000
Walter M 11111111100100000
Blaise M 11110010100000000
Ron M 11111111110000000
William M 11111111110000000
Susan F 11111111111101000
Linda F 11111111110000000
Kim F 11111111110000000
Carol F 11111111110000000
Pete M 11101111100000000
Brenda F 11111010110000000
Mike M 11111001111100000
Zula F 11111111110000000
Frank M 11111111111100000
Dorothy F 11111111101000000
Rod M 11110111110000000
Britton F 11111111100100000
Janet F 11111111100000000
David M 11111111110010000
Thomas M 11111111111010000
Betty F 11111111111000000
Bert M 111111111100110000
Rick M 111111111110100110 ; best performance
Don M 11101100000000000
Barbara F 11111111110000000
Adam M 11111100000000000
Audrey F 11111111101000000
Anne F 111111001110010000
Lisa F 11111111100000000
James M 11111111110000000

```

```

Joe      M 11111111111100000000
Martha   F 111100100100000000
Elsie    F 111111111101010000
Helen    F 111000000000000000    ; worst performance: last data line - just stop!

```

12.3 Example 2: Control and anchor files

A control file, EXAM2.TXT, for the analysis of a test containing 18 items, each item already scored dichotomously as 0,1. The person-id data begins in column 1 and the item-response string begins in column 11. The standard tables will be appear in the printout. There is user scaling. Items 2, 4, 6 and 8 are anchored at 400, 450, 550 and 600 units respectively, supplied in file EXAM2IAF.TXT Your data is in file EXAM2DAT.TXT:

```

; This file is EXAM2.TXT
TITLE='KNOX CUBE TEST - ANCHORED' ; the title for output
NI=18 ; the number of items
ITEM1=11 ; position of first response in data record
NAME1=1 ; first column of person-id in data record
PERSON=KID
ITEM=TAP
DATA=EXAM2DAT.TXT ; name of data file
IAFILE=EXAM2IAF.TXT ; this is item anchor (input) file: it is the IFILE= of an earlier analysis
CONVERGE=L ; use only Logits for convergence criterion
LCONV=.005 ; converged when biggest change is too small to show on any report.
; What follows is equivalent to the IAFILE= above
; IAFILE=* ; item anchor file list
; 2 400 ; item 2 anchored at 400 units
; 4 450 ; item 4 anchored at 450 units
; 6 550 ; item 6 anchored at 550 units
; 8 600 ; item 8 anchored at 600 units
; *

UIMEAN=500 ; user scaling - item mean
USCALE=100 ; user scaling - 1 logit = 100 user units
UDECIM=0 ; print measures without decimals
&END
1-4 ; item labels, starting with the first item
.
4-1-3-4-2-1-4
END NAMES ; End of this file

```

The anchoring information is contained in file EXAM2IAF.TXT and contains the following lines, starting in column 1:

```

2 400 ; item 2 anchored at 400 units:
; if logits are user-rescaled, then anchor values are also expected to be user-rescaled.
; for logit anchor values, specify UASCALE=1
4 450 ; item 4 anchored at 450 units
6 550 ; item 6 anchored at 550 units
8 600 ; item 8 anchored at 600 units

```

Item calibration files, IFILE=, from prior runs can be used as item anchor files, IAFILE=, of later runs.

Your data is in the separate file, EXAM2DAT.TXT, with person-id starting in column 1, and item responses starting in column 11:

```

Richard M 111111100000000000
Tracie F 111111111100000000
.
Elsie F 111111111101010000
Helen F 111000000000000000 End of this file

```

12.4 Example 3: Item recoding and item deletion

The test has 25 items, specified in EXAM3.TXT. The item response string starts in column 12. Person-id's start in column 1 (the standard value). Original item codes are "0", "1", "2" and "X". All items are to be recoded and the original-to-new-code assignments will be 0 0, 1 2, 2 1 and X 3. Items 5, 8, and 20 through 25 are to be deleted from the analysis, and are specified in the control. The misfit criterion for person or item behavior is 3.0. Tables 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 13, 15, 17, 19, 20 and 21 are to appear in your report output file EXAM3OUT.TXT. Sequence numbers are used as item names. Data are in file EXAM3DAT.TXT.

```

; This file is EXAM3.TXT
TITLE="A test with 25 recoded items" ; informative title
NI=25 ; 25 items

```

```

ITEM1=12          ; item responses start in column 12
CODES =012X      ; valid codes
NEWSCORE=0213    ; corresponding response score
RESCORE=2        ; specifies rescore all items
TABLES=111111110101010101110 ; selected tables to go in output file
FITP=3.0         ; person misfit cut-off for reporting
FITI=3.0         ; item misfit cut-off for reporting
DATA=EXAM3DAT.TXT ; name of data file
IDFILE=*         ; list of items to delete: or use IDFILE=file name
5                ; delete item entry number 5
8                ; delete item entry number 8
20-25           ; delete item entry numbers 20 through 25
*               ; end of list
INUMB=Y         ; use item sequence numbers as names
&END            ; end of control specifications
first-item-label
second-item-label
....
END LABELS

```

The data is in file EXAM3DAT.TXT

```

101F20FJDP 21XX2XXXXX111X1200001X2X
102M20PFNP X221012221122211222120X2
.
175F  FBDP 1X00X00000200012X02220100
176F23FEDP 21121022012002121 2202000      person id's contain demographic information

```

12.5 Example 4: Selective item recoding

The test has 18 items, specified in file EXAM4.TXT. The response string starts in column 1. Person-id's start in column 41. Original codes are 0,1 in data file EXAM4DAT.TXT. Items 2, 3, and 4 are to be recoded as 1,0. All tables are to appear in report file EXAM4OUT.TXT, in a standardized form.

```

TITLE="An example of selective item recoding" ; page title
NI=18          ; 18 items
ITEM1=1        ; item responses start in column 1
NAME1=41       ; person-id starts in column 41
NAMLEN=9       ; person-id has 9 characters: Richard M
PSELECT= ??????M ; select M in position 9 of person name
NAMLMP=7       ; 7 characters to appear on maps: "Richard"
CODES=01       ; the observed response codes
CLFILE=*       ; the code label list
0 Incorrect    ; the observed response codes
1 Correct
*              ; end of list
IREFER=ABBBAAAAAAAAAAAAA ; can be letters or numbers
IVALUEA=01     ; A-type items keep original scoring (this line can be omitted)
IVALUEB=10     ; B-type items have reverse scoring

DATA=EXAM4DAT.TXT ; name of data file
MRANGE=5        ; plots and maps have half-range of 5 logits
LINELENGTH=0    ; allow maximum line length
MAXPAGE=0       ; no maximum page length
TABLES=11111111111111111111111111111111 ; output all Tables to the Output file
&END
1-4
R2-3           ; R reminds us item coding was reversed
R1-2-4
2-1-4
.
4-1-3-4-2-1-4
END NAMES

```

The data file, EXAM4DAT.TXT, is

```

100011100000000000      Richard M
100011111100000000      Tracie F
.
100100000000000000      Helen F

```

12.6 Example 5: Scoring key for items, also CAT responses

A multiple-choice-question (MCQ) computer-adaptive test (CAT), in file EXAM5.TXT with responses "a", "b", "c", "d" and a scoring key for 69 items. Your data are in the control file. This was administered as a CAT test, then the response file formatted into a "flat" file with one row per person and one column per item.

```
; This file is EXAM5.TXT
TITLE="An MCQ Test" ; the title
NI=69 ; 69 items
ITEM1=10 ; response string starts in column 10
NAME1=1 ; person-id starts in column 1
CODES=abcd ; valid response codes
MISSCORE=-1 ; standard scoring of missing data, means that blanks are ignored
KEY1 = dcbbbbadbdcacacddabdbaaaccbdddcaadccccdbdcccbbdbcccdbdcbacacbcdddb
; scoring key of correct answers
ITEM=TOPIC ; items are topics
PERSON=STDNT ; respondents are students
NAMLMP=2 ; first 2 characters on maps, e.g., nl
PFILE=EXAM5PF.TXT ; write out person measures
CSV=Y ; separate values by commas in PFILE=
HLINES=N ; write out no heading lines in PFILE=
; Many spreadsheets and statistics programs expect a file of numbers separated by commas.
; Use IFILE= or PFILE= with CSV=Y and HLINES=N.
MJMLE=0 ; allow as many JMLE iterations as necessary
EXTRSC=0.5 ; most conservative (central) extreme measures wanted
XFILE=EXAM5XF.TXT ; write out individual response residual file
&END
nl01 Month
nl02 Sign
.
sb02 newspaper
sb03 newspaper
END NAMES
IM CAT a dcacc ccabbcaa
NM KAT b badad accaaba aa c dd ab c
.
NH RIC ddb b dbdcbcaadba ba acd bad db c d cc
IL HOL a a da d d ccbddd bcd dc ca
```

12.7 Example 6: Keys in data record FORMAT

Do not use `FORMAT=` unless there is no other way. It is tricky to set up correctly. `MFORMS=` is much easier.

A test of 165 multiple-choice items with multiple data lines per data record. The scoring key is formatted in the same way as the data lines:

```
; This file is EXAM6.TXT
TITLE='Demonstration of KEY1 record' title
FORMAT=(1X,10A,T23,50A,/,T23,50A,/,T23,50A,/,T23,15A)
; The first character is ignored,
; then 10 characters in first record are person id,
; then starting in column 23,
; 50 columns in first 3 records,
; and 15 responses in fourth record.

; Using MFORMS= to reformat the same data record
; MFORMS=*
; Data= filename ; put the input data in a separate file
; L=4 ; 4 data input lines per output record
; P1-10=2 ; person label characters 1-10 start in column 2 of line 1 of input data
; I1-50=23 ; responses to items 1-50 start in column 23 of line 1 of input data
; I51-100=2:23 ; responses to items 51-100 start in column 23 of line 2 of input data
```

```

; I101-150=3:23 ; responses to items 101-150 start in column 23 of line 3 of input data
; I151-165=4:23 ; responses to items 151-165 start in column 23 of line 4 of input data
; * ; end of MFORMS=
; ; Note this does not reformat the Keyfrm=, so use KEY1=

; In the reformatted record
NAME1=1 ; Person-id starts in column 1
ITEM1=11 ; Item responses start in column 11 of reformatted record
NI=165 ; There are 165 items
CODES="ABCD " ; The raw responses are ABCD and BLANK.
; Put character strings in " " if blanks are to be included.
MISSCORE=0 ; Blanks and invalid codes scored wrong=0
KEYFRM=1 ; There is a KEY1 record after &END which is formatted
; exactly like a data record specifying the correct responses.
RFILE=exam6rf.txt ; this shows effect of reformatting and scoring
PTBISERIAL=YES ; Raw score point-biserial excluding extremes
TFILE=* ; List of desired tables
1.0 ; Subtable 0 of Table 1
3 ; Table 3.1, 3.2, etc.
6
10
20
* ; end of list

```

```

&END ; KEY1= formatted like your data follows:
Key 1 Record CDABCDBDABCADCBDBCADBABDDCDABCABDCACBADACBADBAACD
after &END CCBDACABDADCBDCABBCACDBAABCDADCDACBCABCDCADABACDA
in FORMAT= format BADCDBADCBADCBACBADBCAADBCCBDCBACDBACBADCDADBACDB
before item names ABDACDCDBADBCAB
A1 First item name
A2
.
A164
A165
END NAMES
090111000102 10001 BDABADCDACDDCCADBCBDBCDACADDCAACCCBCCADBDABADCAADD
ABDDDCABDADCBDCBACADABCDCCDCBDBCCABBBCDCADDDCDA
BDCDBABCDCDDCBADCBACBDCBDBACBCCADBBABAADDCBABAAC
DCBCCACABCDCCB
090111000202 10002 BDCDCDCDADCBCCBDBDCABCBDACDABCAABCAACBBBACAADDAACA
ACBCACBBADADCBDBCCBDCDDACCBCADCACCAACDBCCDADDBACDA
BADDCBDBDCDCBACCCBBAABDBCBCCBAADBABBADBDDBDCAA
DCDBCDDBADBCCB
.
090111008402 10084 CDABCDAADDCCDDDCBDCBCCDACDBBCCACDBCCCBDDACBADCAACD
ACBDCDBDADCCBDDBCABACCBDBDCBCCACCBACDCADABACDA
BABBDCADBDDBCDADDDCDBACBDCBCCCAACDBACBDDDBDCCAAACB
DBACDBCBADDCBC
090111008502 10085 CDABADADABCADCDDBDABBCBACDABCCABACCCDAAACBADAAACD
ACBCDCBBDADCDDBCCBDBADDBBBDCAACAABCBDDDCADABACDA
BADABBADBBADCADACDABBAACACAABDCBACDBADBACCDBACBADA
BCABCBCDBADDCB

```

or, using [continuation lines](#) for the key:

```

CODES="ABCD " ; The raw responses are ABCD and BLANK
;KEYFRM= ; omit this, not needed
KEY1 = CDABCDBDABCADCBDBCADBABDDCDABCABDCACBADACBADBAACD+
+CCBDACABDADCBDCABBCACDBAABCDADCDACBCABCDCADABACDA+
+BADCDBADCBADCBACBADBCAADBCCBDCBACDBACBADCDADBACDB+
+ABDACDCDBADBCAB

```

; "+" are continuation characters

12.8 Example 7: A partial credit analysis

A 30 item MCQ Arithmetic Test is to be analyzed in which credit is given for partial solutions to the questions. Each item is conceptualized to have its own response structure, as in the Masters' Partial Credit model. Estimation for the Partial Credit model is described in RSA, p. 87.

In this example, item 1 has 3 scoring levels. "C" is correct, worth 2 points. "A" and "D" are partially correct, worth 1 point. "B" is incorrect, worth 0 points. [CODES=](#) identifies all possible valid responses. In this example, [KEY1=](#) identifies responses worth 2 points. KEY2= and KEY3= identify responses worth 1 point. The values of KEY1=, KEY2= and KEY3= are set by [KEYSCR=](#). So for item 1, KEY1=C..., KEY2=A..., and KEY3=D. Response B is not in a KEY= and so is scored 0. Here, invalid responses are treated as not-administered. If invalid responses are to be treated as "wrong", specify [MISSCORE=0](#).

```
; This file is EXAM7.TXT
TITLE="A Partial Credit Analysis" page heading
NAME1=1 ; Person-id starts in column 1
ITEM1=23 ; Item responses start in column 23
NI=30 ; There are 30 items

; USING KEY1= (an alternative is to use IREFER=)
CODES=ABCD ; Scores entered as A through D
; 0 1 2 3
; 123456789012345678901234567890
KEY1=CDABCDBDABCADCBDBCADBBDDCDABC ; Fully correct
KEY2=ABCDABCCBADDABACABBBACCCCDABAB ; Partially correct
KEY3=DABABAAACC*CCABAC***** ; Some partially correct
; if no matching response, use a character not in CODES=, e.g., *
; the keys are matched in sequence, "B" for item 15 matches Key1=, and Key3= is ignored
KEYSCR=211 ; KEY1 fully correct (2 points),
; KEY2, KEY3 partially correct (1 point each)

ISGROUPS=0 ; Each item is its own grouping, i.e., the Partial Credit model
MODELS=R ; Each item has its own Andrich rating scale
STKEEP=Y ; Keep all intermediate categories in analysis, even if never observed
CURVES=111 ; Print all 3 item curves in Tables 2 and 21
CATREF=2 ; Use category 2 for ordering in Table 2
&END
0.7 A1 ; item labels
1.1 A2
.
3.1 A29
2.7 A30
END NAMES
090111000102 10001 BDABADCDACCDCCADBCBDCDDACADDC
090111000202 10002 BDCDCDCDADCBCCBDEDCABCBDACDABC
.
090111005302 10053 BDABCDDCCDDCDBACBABABCMDBDC
090111005402 10054 BDADCDCDADCBCABBDADBBBCBDDDDDC
```

12.9 Example 8: Items with various rating scale models

A 4 item test in which each item has a 9-level scoring rubric. We suspect that there are not really 9 levels of competence. After several Winsteps analyses, we choose to recode the rating scale or partial credit items by collapsing categories in order to increase measurement effectiveness ([separation](#)) and increase parameter stability. An objection to collapsing is that it violates the Rasch model. This is only true if the uncollapsed data strictly accord with the model. In fact, the collapsed data may fit the Rasch model better than the uncollapsed. We have to compare the collapsed and uncollapsed analyses to decide.

```
; This file is EXAM8.TXT
TITLE="Success and Failure Items"
NAME1=6
```

```

ITEM1=1
NI=4 ; 4 items
ISGROUPS=1234 ; one item per grouping, same as ISGROUPS=0
IREFER=ABCD ; the 4 items are to be recoded differently. Item 1 is type "A", etc.
CODES=123456789 ; the codes in the data file
IVALUEA=333456666 ; the recoding for A-type items in IREFER=, i.e., Item 1
IVALUEB=333455555
IVALUEC=333444444
IVALUED=444456777
&END
Maze
Passengers
Blocks
Egg race
END NAMES
5536 M Richard
4545 F Tracie
4345 M Walter
3454 M Blaise
4435 M Ron
....

```

Table 14.3 shows the recoding:

ITEMS CATEGORY/OPTION/Distractor FREQUENCIES: ENTRY ORDER										
ENTRY	DATA	SCORE	DATA		USED		AVERAGE		OUTF	
NUMBER	CODE	VALUE	COUNT	%	COUNT	%	MEASURE	MNSQ	ITEM	
1	2	3	1	3	1	3	-2.39	.3	Maze	
	3	3	7	22	6	19	-1.57	.7		
	4	4	10	31	10	32	-.54	.5		
	5	5	10	31	10	32	1.23	.6		
	6	6	3	9	3	10	2.42	.7		
	7	6	1	3	1	3	2.78	.5		

12.10 Example 9: Grouping and modeling items

A 20 item test. Items 1, 2 are dichotomous items, coded "Y", "N". Item 3 is a "Partial Credit" item. Items 4-10 are all ratings on one Andrich scale or test-level partial-credit scale (Grouping 1), and items 11-20 are all ratings on another Andrich scale or test-level partial-credit scale (Grouping 2). These are grouped with `ISGROUPS=`. Winsteps discovers from the data what the item structure is. Items 3-20 have response codes "A", "B", "C", "D", "E" or "a", "b", "c", "d", "e".

```

; This file is EXAM9.TXT
TITLE="Grouping and Modeling"
ITEM1=11 ; Item responses start in column 11
NI=20 ; There are 20 items
ISGROUPS=DD011111112222222222 ; The groupings in item order
IREFER=DDAAAAAAAAAAAAAAAAAAAA ; for recoding
CODES = 10ABCDEabcde ; Response codes to all items
IVALUED= 10***** ; for Items 1 & 2
IVALUEA= **1234512345 ; for Items 3-20
DATA=EXAM9DAT.TXT ; Location of data file
IWEIGHT=* ; Item weighting file list
3 2.5 ; Item 3 has weight of 2.5
*
&END
RD Prompt 1 ; Item id's remind us of MODELS= and ISGROUPS=
RD Prompt 2 ; MODELS= and ISGROUPS= are shown in item measure Tables 10, 13, 14, 15
S0 Logic
R1 Grammar 1
R1 Grammar 2
R1 Grammar 3

```

```

R1 Grammar 4
R1 Grammar 5
R1 Grammar 6
R1 Grammar 7
R2 Meaning 1
R2 Meaning 2
R2 Meaning 3
R2 Meaning 4
R2 Meaning 5
R2 Meaning 6
R2 Meaning 7
R2 Meaning 8
R2 Meaning 9
R2 Meaning 10
END NAMES

```

The data is in file EXAM9DAT.TXT:

```

Richard M 00bCDCDddCDdddCDccE
Tracie F 00BcBABBccbBbbBbBBBb
.
James M 00ccaBbabBAcbacbaBbb
Joe M 10BdBBBBccBccbcbcbC

```

12.11 Example 10: Combining tests with common items

This uses [MFORMS=](#), but it can also be done, more awkwardly, with [FORMAT=](#)

Test A, in file EXAM10A.TXT, and TEST B, in EXAM10B.TXT, are both 20 item tests. They have 5 items in common, but the distractors are not necessarily in the same order. The responses must be scored on an individual test basis. Also the validity of each test is to be examined separately. Then one combined analysis is wanted to [equate](#) the tests and obtain bankable item measures. For each file of original test responses, the person information is in columns 1-11, the item responses in 41-60.

The combined data file specified in EXAM10C.TXT, is to be in RFILE= format. It contains

Person information of 11 characters: Columns 1-30 (always), but only 1-11 are active.
Item responses to 35 items: Columns 31-64

The identification of the common items is:

Test Item Number (=Location in item string)							
Bank	1	2	3	4	5	6-20	21-35
A:	3	1	7	8	9	2, 4-6, 10-20	
B:	4	5	6	2	11		1, 3, 7-10, 12-20

I. From Test A, make a response (RFILE=) file..

```

; This file is EXAM10A.TXT
TITLE="Analysis of Test A"
RFILE=EXAM10AR.TXT ; The constructed response file for Test A
NI=20 ; 20 items
ITEM1=41 ; Items start in column 41 of data record
NAME1=1 ; Start of person label
NAMELEN=11 ; Length of person label
CODES="ABCD# " ; Beware of blanks and # meaning wrong!
; Blanks are included in CODES=, but they are scored incorrect, because never keyed correct
KEY1=CCBDACABDADCBDCAABCA ; the MCQ key
&END

```



```

BANK 2 TEST A 1 ; first item name
BANK 6 TEST A 2
BANK 1 TEST A 3

```

```

.
BANK 20 TEST A 20
END NAMES

```

```

Person 01 A          BDABCEBDDACDBCACBDBA
.
Person 12 A          BADCACADCDABDDDCBACA

```

The RFILE= file, EXAM10AR.TXT, is:

```

0          1          2          3          4          5
12345678901234567890123456789012345678901234567890
Person 01 A          00000000110010001001
.....
Person 12 A          00001110000001001011

```

II. From Test B, make a response (RFILE=) file.

```

; This file is EXAM10B.TXT
TITLE="Analysis of Test B"
RFILE=EXAM10BR.TXT ; The constructed response file for Test B
NI=20
ITEM1=41 ; Items start in column 26 of reformatted record
NAME1=1 ; Start of person label
NAMELEN=11 ; Length of person label
CODES="ABCD# " ; Beware of blanks meaning wrong!
KEY1=CDABCEBDDACDBCACBDBA ; Key in data record format
&END
BANK 21 TEST B 1
BANK 4 TEST B 2
BANK 22 TEST B 3

```

```

.
BANK 35 TEST B 20
END NAMES

```

```

Person 01 B          BDABDDCDBBCCCCDAACBC
.
Person 12 B          BADABBADCBADDBBBBBB

```

The RFILE= file, EXAM10BR.TXT, is:

```

Person 01 B          01110101011001000100
...
Person 12 B          00000001010000101000

```

III. Analyze Test A's and Test B's RFILE='s together:

```

; This file is EXAM10C.TXT
TITLE="Analysis of Tests A & B (already scored)"
NI=35 ; 35 items in total
ITEM1=31 ; reformatted data record
CODES=01 ; scored right-wrong.
; Blanks ignored as "not administered"
MFORMS=* ; multiple data files in different formats
DATA=EXAM10AR.txt ; first file
L=1 ; one line per data record
P1-11 = 1 ; person id starts in column 1 of input file
I1=33 ; item 1 is Test A's item 3 in Test A's column 33
I2=31 ; item 2 is Test A's item 1 in column 31
I3-5=37 ; items 3-5 are Test A's items 7-9 starting in column 37
I6=32
I7-9=34

```

```

I10-20=40
#
DATA=EXAM10BR.txt ; second data file
L=1
P1-11 = 1      ; person id starts in column 1 of input file
I1-3 = 34      ; items 1-3 are Test B's items 4-6 starting in Test B's column 34
I4 = 32        ; item 4 is Test B's item 2 in column 32
I5 = 41        ; items 5 in Test B's item 11 in column 41
I21 = 31
I22 = 33
I23-26 = 37
I27-35 = 42
*
&END
BANK 1  TEST A 3 B 4
...
BANK 35 TEST B 20
END NAMES

```

The combined data file (which can be accessed from the **Edit** pull-down menu) is:

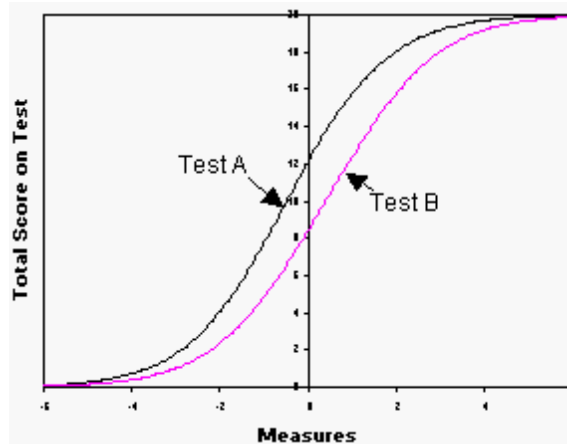
```

Person 01 A      00001000010010001001
...
Person 12 A      00100001100001001011
Person 01 B      10111          010101001000100
...
Person 12 B      00000          000101000101000

```

After running EXAM10C, I want to see two ICCs: One for test A and another Test B. How do I do this?

This graph is not produced directly by Winsteps, but can be produced in Excel.



After the Exam10C analysis,

use the "Specification" pull-down menu to delete items not in Test A: Use [IDELETE=21-35](#)

Display the [Test Characteristic Curve](#)

Select "Copy data to clipboard".

Paste into an Excel worksheet

Use the "Specification" pull-down menu to reinstate all items: IDELETE=+1-35

Use the "Specification" pull-down menu to delete items not in Test B: Use IDELETE=6-20

Display the Test Characteristic Curve

Select "Copy data to clipboard".

Paste into an Excel worksheet

In Excel, scatterplot the pasted columns.

12.12 Example 11: Item responses two characters wide

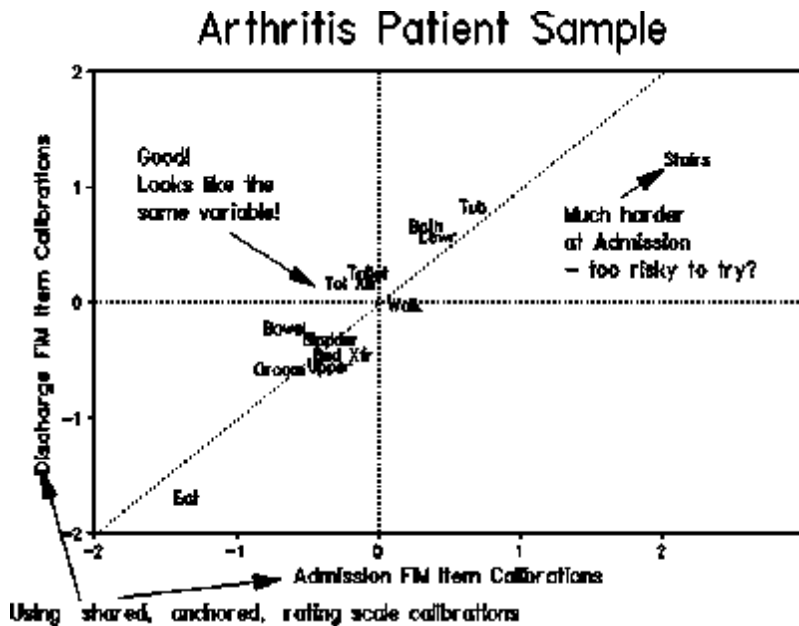
The "Liking for Science" data (see RSA) is in file EXAM11.TXT. Each observation is on a rating scale where 0 means "dislike", 1 means "don't care/don't know" and 2 means "like". The data has been recorded in two columns as 00, 01 and 02. XWIDE= is used.

```
; This file is EXAM11.TXT
TITLE='LIKING FOR SCIENCE (Wright & Masters p.18)'
XWIDE=2 ; Responses are 2 columns wide
CODES=000102 ; Codes are 00 01 and 02
CLFILE=* ; Category label filelist
00 Dislike ; Category 00 in data file means "Dislike"
01 Neutral
02 Like
* ; End of category list
ITEM1=1 ; Items start in column 1
NI=25 ; 25 Items
NAME1=51 ; Person id starts in column 51
NAMLMP=20 ; Show 20 characters of id on maps
TABLES=11111111111111111111111111111111 ; All Tables
CURVES=111 ; Print all curves in Tables 2 and 21
IFILE = EXAM11IF.TXT ; Output item measure file
PFILE = EXAM11PF.TXT ; Output person measure file
SFILE = EXAM11SF.TXT ; Output structure calibration file
RFILE = EXAM11RF.TXT ; Output reformatted response file
XFILE = EXAM11XF.TXT ; Output observation and residual file
UIMEAN = 455 ; User scaling: mean 455
USCALE = 94 ; 94 user units per logit
LINELENGTH = 0 ; Print with minimum of split lines
MAXPAGE = 0 ; Print with no page breaks in long tables
&END
WATCH BIRDS
READ BOOKS ON ANIMALS
.
FIND OUT WHAT FLOWERS LIVE ON
TALK W/FRIENDS ABOUT PLANTS
END NAMES
01020101010002000102020202000201010202000201000200ROSSNER, MARC DANIEL
02020202020202020202020202020202020202020202020202020202ROSSNER, LAWRENCE F.
.
020202000002020000020202020202020202000202000102PATRIARCA, RAY
010200000101000101020102020201000202000200000100PAULING, LINUS
BLANK RECORD
```

12.13 Example 12: Comparing high and low samples with rating scales

Rasch estimates are constructed to be as sample independent as is statistically possible, but you must still take care to maintain comparability of measures across analyses. For instance, if a rating scale or partial credit structure is used, and a high-low measure split is made, then the low rating scale (or partial credit) categories may not appear in the data for the high measure sample and vice versa. To compare item calibrations for the two samples requires the response structure to be calibrated on both samples together, and then the response structure calibrations to be anchored for each sample separately. Comparison of patient measures from separate analyses requires both the response structure calibrations and the item calibrations to share anchor calibrations. 35 arthritis patients have been through rehabilitation therapy. Their admission to therapy and discharge from therapy measures are to be compared. They have been rated on the 13 mobility items of the Functional Independence Measure (FIM™). Each item has seven levels. At admission, the patients could not perform at the higher levels. At discharge, all patients had surpassed the lower levels (Data courtesy of C.V. Granger & B. Hamilton, ADS). A generic control file is in EXAM12.TXT. The admission ratings are in EXAM12LO.TXT and the discharge ratings in EXAM12HI.TXT. Three analyses are performed: 1) joint analysis of the admission (low) and discharge (high) data

to obtain response structure calibrations, 2 & 3) separate runs for the admission (low) and discharge (high) data to obtain item calibrations. For a more complex situation, see [Example 17](#).



```

; This common control file is EXAM12.TXT
TITLE='GENERIC ARTHRITIS FIM CONTROL FILE'
ITEM1=7 ; Responses start in column 7
NI=13 ; 13 mobility items
CODES=1234567 ; 7 level rating scale
CLFILE=* ; Defines the rating scale
1 0% Independent
2 25% Independent
3 50% Independent
4 75% Independent
5 Supervision
6 Device
7 Independent
*
&END
A. EATING
B. GROOMING
C. BATHING
D. UPPER BODY DRESSING
E. LOWER BODY DRESSING
F. TOILETING
G. BLADDER
H. BOWEL
I. BED TRANSFER
J. TOILET TRANSFER
K. TUB, SHOWER
L. WALK/WHEELCHAIR
M. STAIRS
END NAMES

```

The admission data is in file EXAM12LO.TXT:

```

21101 5523133322121 Patient number in cols 1-5, ratings in 7-19
21170 4433443345454
|
22618 4433255542141
22693 3524233421111

```

The discharge data is in file EXAM12HI.TXT:

```
21101 5734366655453 Ratings generally higher than at admission
21170 6466677777676
|
22618 7667656666565
22693 7776677676677
```

The batch file to run this is (see [BATCH=](#)), under Windows XP, 2000, EXAM12.CMD:

```
REM COMPARISON OF ITEM CALIBRATIONS FOR HIGH AND LOW SAMPLES
START /WAIT ..\Winsteps EXAM12.TXT EXAM12OU.TXT DATA=EXAM12LO.TXT+EXAM12HI.TXT
      TITLE=ADMIT+DISCHARGE SFILE=EXAM12SF.TXT BATCH=Y
START /WAIT ..\Winsteps EXAM12.TXT EXAM12LU.TXT DATA=EXAM12LOF.TXT TITLE=ADMIT
      SAFILE=EXAM12SF.TXT IFILE=EXAM12LI.TXT BATCH=Y
START /WAIT ..\Winsteps EXAM12.TXT EXAM12HU.TXT DATA=EXAM12HIF.TXT TITLE=DISCHARGE
      SAFILE=EXAM12SF.TXT IFILE=EXAM12HI.TXT BATCH=Y
```

under WINDOWS-98, EXAM12BAT.BAT: (not Windows-9 nor 3.1)

```
REM COMPARISON OF ITEM CALIBRATIONS FOR HIGH AND LOW SAMPLES
START /w ..\Winsteps EXAM12.TXT EXAM12OU.TXT DATA=EXAM12LO.TXT+EXAM12HI.TXT
      TITLE=ADMIT+DISCHARGE SFILE=EXAM12SF.TXT BATCH=Y
START /w ..\Winsteps EXAM12.TXT EXAM12LU.TXT DATA=EXAM12LO.TXT TITLE=ADMIT SAFILE=EXAM12SF.TXT
      IFILE=EXAM12LIF.TXT BATCH=Y
START /w ..\Winsteps EXAM12.TXT EXAM12HU.TXT DATA=EXAM12HI.TXT TITLE=DISCHARGE
      SAFILE=EXAM12SF.TXT IFILE=EXAM12HIF.TXT BATCH=Y
```

Under WINDOWS-NT (early versions), EXAM12NT.BAT:

```
REM COMPARISON OF ITEM CALIBRATIONS FOR HIGH AND LOW SAMPLES
..\Winsteps EXAM12.TXT EXAM12OU.TXT DATA=EXAM12LO.TXT+EXAM12HI.TXT TITLE=ADMIT&DISCHARGE
      SFILE=EXAM12SF.TXT BATCH=Y
..\Winsteps EXAM12.TXT EXAM12LU.TXT DATA=EXAM12LO.TXT TITLE=ADMIT SAFILE=EXAM12SF.TXT
      IFILE=EXAM12LIF.TXT BATCH=Y
..\Winsteps EXAM12.TXT EXAM12HU.TXT DATA=EXAM12HI.TXT TITLE=DISCHARGE SAFILE=EXAM12SF.TXT
      IFILE=EXAM12HIF.TXT BATCH=Y
```

To run this, select "Run batch file" from "Batch" pull-down menu, and right-click on "Exam12bat.bat" or "Exam12cmd.cmd" in the dialog box, then left-click on "open".

The shared structure calibration (Andrich threshold) anchor file is EXAM12SF.TXT:

```
; structure measure FILE FOR
; ADMIT&DISCHARGE
; May 23 13:56 1993
; CATEGORY structure measure
1 .00
2 -2.11
3 -1.61
4 -1.25
5 .06
6 1.92
7 2.99
```

The item calibrations measures for admission and discharge are written into [IFILE=](#) files, with comma-separated values ([CSV=Y](#)), so that they can easily be imported into a spreadsheet.

12.14 Example 13: Paired comparisons as the basis for measurement

Paired comparisons can be modeled directly with the Facets computer program. For Winsteps a dummy facet of "occasion" or "pairing" must be introduced. On each occasion (in this example, each column), there is a winner '1', a loser '0', or a draw 'D' recorded for the two players. The data can be set up with the paired objects as rows and the pairings as columns, or the paired objects as columns or the pairings as rows. Analytically it makes no differences, so do it the way that is easier for you.

In this example, the paired objects are the rows, and the pairings are the columns. In column 1 of the response data in this example, Browne (1) defeated Mariotti (0). In column 2, Browne (D) drew with Tatai (D). Specifying [PAIRED=YES](#) adjusts the measures for the statistical bias introduced by this stratagem. Each player receives a measure and fit statistics. Occasion measures are the average of the two players participating. Misfitting occasions are unexpected outcomes. Point-biserial correlations have little meaning. Check the occasion summary statistics in Table 3.1 to verify that all occasions have the same raw score.

```
; This common control file is EXAM13.TXT
TITLE = 'Chess Matches at the Venice Tournament, 1971'
Name1 = 1 ; Player's name
Item1 = 11 ; First match results
PERSON = PLAYER
ITEM = MATCH ; Example of paired comparison
CODES = 0D1 ; 0 = loss, D = draw (non-numeric), 1 = win

; if you wish to just consider won-loss, and ignore the draws, omit the following line:
NEWSCORE = 012 ; 0 = loss, 1 = draw, 2 = win

CLFILE=*
0 Loss
D Draw
1 Win
*
NI = 66 ; 66 matches (columns) in total
PAIRED = YES ; specify the paired comparison adjustment
INUMBER = YES ; number the matches in the Output
&END
Browne 1D 0 1 1 1 1 1 D D 1 1
Mariotti 0 1 D 0 1 1 1 D 1 D 1
Tatai D 0 1 D D 1 1 1 1 1 D
Hort 1D1 D D 1 D D D 1 0
Kavalek 010D D D 1 D 1 1 D
Damjanovic 00DDD D D D 1 D 1
Gligoric 00D0DD D 1 1 1 0
Radulov 000D0DD D 1 D 1
Bobotsov DD0DDD0D 0 0 1
Cosulich D00D00001 1 1
Westerinen 0D000D0D10 1
Zichichi 00D1D010000
```

Part of the output is:

PLAYER STATISTICS: MEASURE ORDER										
ENTRY	RAW									
NUMBR	SCORE	COUNT	MEASURE	ERROR	MNSQ	ZSTD	MNSQ	ZSTD	PLAYER	
1	17	11	1.09	.35	1.10	.2	1.02	.1	Browne	
2	15	11	.68	.32	1.02	.0	.96	-.1	Mariotti	
3	14	11	.50	.31	.86	-.4	.83	-.5	Tatai	
4	14	11	.50	.31	1.34	.9	1.54	1.3	Hort	
5	13	11	.33	.31	.81	-.6	.80	-.6	Kavalek	

6	11	11	.00	.30	.35	-2.8	.37	-2.6	Damjanovic
7	10	11	-.17	.30	.90	-.3	.91	-.3	Gligoric
8	9	11	-.34	.31	.52	-1.8	.52	-1.7	Radulov
9	8	11	-.51	.31	1.00	.0	1.00	.0	Bobotsov
10	8	11	-.51	.31	1.18	.5	1.15	.4	Cosulich
11	7	11	-.69	.32	.95	-.1	.89	-.3	Westerinen
12	6	11	-.88	.33	1.86	1.8	1.90	1.7	Zichichi
+-----+-----+-----+-----+-----+									
MEAN	11.	11.	.00	.32	.99	-.2	.99	-.2	
+-----+-----+-----+-----+-----+									

12.15 Example 14: Multiple rescorings, response structures and widths

Introductory example: A test has 30 multiple choice question (keyed A, B, C, D) and 5 essay questions, each has its own rating scale definition (scored 0, 1, or 2), i.e., they accord with the partial credit model. [ISGROUPS=](#) is used to identify the partial credit items. [IREFER=](#) and [IVALUE=](#) are used for scoring the questions

```
; This control file is EXAM14.TXT
TITLE = "AN EXAMPLE OF MULTIPLE ITEM TYPES"
; THIS PART DEFINES THE ANALYTICAL MODEL
; MCQ ARE ONE GROUPING, WHICH WE WILL CALL "M" - PARTIAL CREDIT ITEMS ARE GROUPED AS "0"
; "0" MEANS EACH ITEM IS IN A GROUPING BY ITSELF, i.e., partial credit
ISGROUPS = MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM00000

; THIS PART DEFINES THE RESPONSE SCORING
; The MCQ key is: ABCAACBAAABABABBBAAABBBAAABAB
IREFER = ABCAACBAAABABABBBAAABBBAAABAB00000 ; REFERENCE THE MCQ ITEMS BY THEIR SCORING KEY
CODES = ABCD012; ALL VALID CODES IN THE DATA FILE
IVALUE0= 0000012; THESE ARE THE ESSAY ITEMS
IVALUEA= 1000000 ; MCQ ITEMS WITH A AS THE CORRECT ANSWER
IVALUEB= 0100000 ; B IS CORRECT
IVALUEC= 0010000 ; C IS CORRECT
IVALUED= 0001000 ; D IS CORRECT
MISSING-VALUES-SCORED = 0 ; SO ALL 9'S (OR ANYTHING ELSE NOT IN CODES=) ARE SCORED 0

ITEM1 = 1 ; START OF ITEM STRING
NI = 35 ; NUMBER OF ITEM RESPONSES
NAME1 = 37 ; START OF PERSON IDENTIFICATION
&END
; ITEM LABELS HERE
END LABELS
ACACBAA9DC9AB9C9DCC9ABBA9AACB00102 1 ; DATA ARE HERE
A9D9AAAB9DA999AA9BB9999999B9AA12100 2
A9C9AACD9AB9AB9B99D9DB9ACBD9AB10120 3
A9C9AACB9AD99999CA9ABD9999999900120 4
A9C9AAAB9AA99999CD9ABB9999999910120 5
A9C9ADCBBAA9CDABD9DACCABBA9ABC21212 6
B9D9AAAB9CB99999DB9ABC9999999900100 7
A9D9BAAB9AA99999BA9ABB9999999900101 8
A9D9BACB9DC99999CA9DBB9999999921201 9
A9A9AACD9BA99999AD9ABB9999999910120 10
A9C9AACB9AA99999DC9ABD9999999900120 11
A9C9AAAB9CA99999BA9CBB9999999921201 12
A9C9CACB9AC99999CB9ABD9999999910120 13
A9D9AACB9AA99999AA9ABD999999D900000 14
A9C9DDCB9AA99999CB9ABD999999C9D21201 15
A9C9AABB9AD99999CD9ABC99999ABDAB11110 16
A9CAACB9ABBC9ADBB9ABDABBA9ADCB00120 17
CBBAB9CAAC9BBBC9BCBACDD9ADDCAB10120 18
A9C9BAAB9AD99CB9BBBA9ABDACCDD9A00120 19
C9D9BDD9B9BBB9ACBADBC9AADBBCBC21201 20
```

```

A9C9AABB9DABA9ABDDABCABBA9AACB00120 21
A9D9BDCB9DCAC9DBBADBBACBA9ABAC00001 22
A9C9BACB9DADA9DBBDABBACBA9ABBB21201 23
A9D9BACC9AAD9DBBAABBADBC9ABCB10012 24
A9D9BAAB9ABCC9ABBDABBACBB9ABBB21201 25

```

Example 14: A test comprises multiple-choice questions (MCQ), true-false, and two different response structure formats. These are combined in one data file. The MCQ scoring key, and also the True-False scoring key are used as item cluster references in the IREFER= and IVALUE= specifications to simplify key checking.

EXAM14DT.TXT data format is:	
Cols 1-5	Person id.
Cols 6-10	5 MCQ items (A,B,C,D, # = missing, wrong. Some items may be miskeyed
Cols 11-15	5 True-False items, responses (S, N). For some of these S="True" is correct, for others "N"=False is correct
Cols 16-25	10 Rating scale items (N=0,P=1,M=2,S=3). Some items may be reversed.
Cols 26-29	2 Evaluation items (0-12). - (See exam14b.txt) - <i>second analysis only.</i>

First analysis: all responses one column wide

The control file, EXAM14.TXT, is:

```

; This file is EXAM14.TXT
TITLE="Multiple Response Formats, Same Response Width"
DATA=exam14dt.txt
;EXAM14DT.TXT data format is
;Cols 1-5 Person id.
;Cols 6-10 5 MCQ items (A,B,C,D, # = missing, wrong)
; Some items may be miskeyed
;Cols 11-15 5 True-False items, responses (S, N)
; For some of these S="True" is correct, for others "N"=False is correct
;Cols 16-25 10 Rating scale items (N=0,P=1,M=2,S=3)
; Some items may be reversed.
;Cols 26-29 2 Evaluation items (0-12). - (See exam14b.con)
NAME1=1
ITEM1=6
NI=20

; THESE CODES LINE UP WITH THE ITEM COLUMNS
;      0      1      2
;      12345678901234567890

; TO DEFINE RESPONSE STRUCTURE CLUSTERS
ISGROUPS = 11111222223333333333

; TO DEFINE RESCORING CLUSTERS
IREFER = BACDCSNSNSRRRRRRRRRR
; IREFER = X MATCHES IVALUEX=

; IVALUE?= MATCHES WITH CODES=
CODES = ABCD#SNPM ; CODES IN ORIGINAL DATA FILE
IVALUEA = 10000**** ; MCQ RESPONSE A IS CORRECT
IVALUEB = 01000**** ; MCQ RESPONSE B IS CORRECT
IVALUEC = 00100**** ; MCQ RESPONSE C IS CORRECT
IVALUED = 00010**** ; MCQ RESPONSE D IS CORRECT
IVALUES = *****10** ; "S" IS THE CORRECT ANSWER
IVALUEN = *****01** ; "N" IS THE CORRECT ANSWER
IVALUER = *****3012 ; "NPMS" RATING SCALE
STKEEP=YES ; KEEP UNUSED INTERMEDIATE CATEGORIES IN RATING SCALES
INUMBER=YES ; NO ITEM INFORMATION AVAILABLE
&END

```


Second analysis: responses one and two columns wide

Including the last two items with long numeric response structures and 2-column format, the control file becomes EXAM14B.TXT. Since some responses are in 2 column format, XWIDE=2. FORMAT= is used to transform all responses into XWIDE=2 format.

EXAM14B.TXT is:

```
&INST
TITLE="Multiple Response Formats, Same Response Width"
DATA=EXAM14DT.TXT
;EXAM14DT.TXT data format is
;Cols 1-5 Person id.
;Cols 6-10 5 MCQ items (A,B,C,D, # = missing, wrong)
; Some items may be miskeyed
;Cols 11-15 5 True-False items, responses (S, N)
; For some of these S="True" is correct, for others "N"=False is correct
;Cols 16-25 10 Rating scale items (N=0,P=1,M=2,S=3)
; Some items may be reversed.
;Cols 26-29 2 Evaluation items (0-12)
NAME1=1
ITEM1=6
NI=22 ;20-1 COLUMN + 2 2-COLUMN ITEMS
XWIDE=2 ;XWIDE FOR WIDEST FIELD
FORMAT=(5A1, 20A1,2A2) ; PERSON LABEL & FIRST 20 ITEMS 1 CHARACTER COLUMNS
; ; LAST 2 ITEMS ARE 2 CHARACTER COLUMNS
; THESE ARE SET UP FOR XWIDE=2

; FOR RESPONSE STRUCTURE DEFINITIONS
; TO DEFINE RATING SCALE CLUSTERS
ISGROUPS = 1111122222333333333344
IREFER = BACDCSNSNSRRRRRRRRRREE
; IREFER = X MATCHES IVALUEX=

; IVALUE?= MATCHES WITH CODES=
; CODES ARE 2 CHARACTERS WIDE
CODES = "A B C D # S N P M 121110 9 8 7 6 5 4 3 2 1 0"; CODES IN DATA
IVALUEA = "1 0 0 0 * * * * * * * * * * * * * * * *"; MCQ A IS CORRECT
IVALUEB = "0 1 0 0 0 * * * * * * * * * * * * * * * *"; MCQ B IS CORRECT
IVALUEC = "0 0 1 0 0 * * * * * * * * * * * * * * * *"; MCQ C IS CORRECT
IVALUED = "0 0 0 1 0 * * * * * * * * * * * * * * * *"; MCQ D IS CORRECT
IVALUES = "* * * * * 1 0 * * * * * * * * * * * * * * *"; "S" IS CORRECT
IVALUEN = "* * * * * 0 1 * * * * * * * * * * * * * * *"; "N" IS CORRECT
IVALUER = "* * * * * 3 0 1 2 * * * * * * * * * * * * * *"; "NPMS" RATING SCALE
IVALUEE = "* * * * * * * * * * 121110 9 8 7 6 5 4 3 2 1 0"; 0-12 RATING SCALE
STKEEP=YES ; KEEP UNUSED INTERMEDIATE CATEGORIES IN RATING SCALES
INUMBER=YES ; NO ITEM INFORMATION AVAILABLE
&END
```

This can also be done with [MFORMS=](#)

EXAM14C.TXT is:

```
; This file is EXAM14C.TXT
&INST
TITLE="Multiple Response Formats, Same Response Width"
;EXAM14DT.TXT data format is
;Cols 1-5 Person id.
;Cols 6-10 5 MCQ items (A,B,C,D, # = missing, wrong)
; Some items may be miskeyed
;Cols 11-15 5 True-False items, responses (S, N)
; For some of these S="True" is correct, for others "N"=False is correct
```


12.16 Example 15: Figure skating: Multidimensionality, DIF or Bias

The Pairs Skating competition at the 2002 Winter Olympics in Salt Lake City was contentious. It resulted in the awarding of Gold Medals to both a Russian and a Canadian pair, after the French judge admitted to awarding biased scores. Multidimensionality, differential item functioning, and item bias are all manifestations of disparate subdimensions within the data. In judged competitions, judge behavior can introduce unwanted subdimensions.

The data comprise 4 facets: skaters + program + skill + judges → rating

For this analysis, each pair is allowed to have a different skill level, i.e., different measure, on each skill of each performance. The judges are modeled to maintain their leniencies across all performances.

In this judge-focused rectangular 2-facet analysis: (skaters + program + skill = rows) + (judges = columns) → rating

The rating scale is very long, 0-60. Alternative methods of analysis are shown in [SFUNCTION=](#).

The control file and data are in **exam15.txt**.

```
; This file is EXAM15.TXT
Title = "Pairs Skating: Winter Olympics, SLC 2002"
Item = Judge
Person = Pair
NI = 9 ' the judges
Item1 = 14 ' the leading blank of the first rating
Xwide = 3 ' Observations are 3 CHARACTERS WIDE for convenience
NAME1 = 1 ' start of person identification
NAMELENGTH = 13 ' 13 characters identifiers

; CODES NEXT LINE HAS ALL OBSERVED RATING SCORES
CODES= " 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44+
      + 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60"

STEPKEEP=YES ; maintain missing intermediate scores in the hierarchy

@order = 1-2 ; pair order number at finish of competition in person label columns 1-2
@program = 11 ; program in person label column 11: Short or Free
@skil = 13 ; skill in person label column 13: Technical or Artistic
PSUBTOT = @order
DIF = @order ; judge "DIF" across skating pairs
tfile=*
30 ; produce Table 30 for judge-by-pairs "DIF"
28 ; produce Table 28 for skater-pair summary statistics
*
&END
1 Rus ;Mrs. Marina SANIAIA : RUSSIA
2 Chn ;Mr. Jiasheng YANG : CHINA
3 USA ;Mrs. Lucy BRENNAN : USA
4 Fra ;Miss Marie Reine LE GOUGNE : FRANCE
5 Pol ;Mrs. Anna SIEROCKA : POLAND
6 Can ;Mr. Benoit LAVOIE : CANADA
7 Ukr ;Mr. Vladislav PETUKHOV : UKRAINE
8 Ger ;Mrs. Sissy KRICK : GERMANY
9 Jap ;Mr. Hideo SUGITA : JAPAN

; Description of Person Identifiers
; Cols. Description
; 1-2 Order immediately after competition (@order)
; 4-5 Skaters' initials
; 7-9 Nationality
; 11 Program: S=Short F=Free
; 13 Skill: T=Technical Merit, A=Artistic Impression
END LABELS
```

```

1 BS-Rus S T 58 58 57 58 58 58 58 57 ; 1 BEREZHNAVAYA Elena / SIKHARULIDZE Anton : RUS
1 BS-Rus S A 58 58 58 58 59 58 58 58 ; 2 BEREZHNAVAYA Elena / SIKHARULIDZE Anton : RUS
1 BS-Rus F T 58 58 57 58 57 57 58 58 57 ; 3 BEREZHNAVAYA Elena / SIKHARULIDZE Anton : RUS
1 BS-Rus F A 59 59 59 59 59 58 59 58 59 ; 4 BEREZHNAVAYA Elena / SIKHARULIDZE Anton : RUS
2 SP-Can S T 57 57 56 57 58 58 57 58 56 ; 5 SALE Jamie / PELLETIER David : CAN
2 SP-Can S A 58 59 58 58 58 59 58 59 58 ; 6 SALE Jamie / PELLETIER David : CAN
2 SP-Can F T 58 59 58 58 58 59 58 59 58 ; 7 SALE Jamie / PELLETIER David : CAN
2 SP-Can F A 58 58 59 58 58 59 58 59 59 ; 8 SALE Jamie / PELLETIER David : CAN
3 SZ-Chn S T 57 58 56 57 57 57 56 57 56 ; 9 SHEN Xue / ZHAO Hongbo : CHN
.....

```

From this data file, estimate judge severity. In my run this took 738 iterations, because the data are so thin, and the rating scale is so long.

Here is some of the output of **Table 30**, for Judge DIF, i.e., Judge Bias by skater pair order number, `@order = $S1W2`.

```

+-----+
| Pair   DIF   DIF   Pair   DIF   DIF   DIF   JOINT   Judge   |
| CLASS ADDED S.E. CLASS ADDED S.E. CONTRAST S.E. t   Number Name |
+-----+
| 13     -.93  .40 18     1.50  .39   -2.43  .56 -4.35   9 9 Jap |
| 14     -1.08 .36 18     1.50  .39   -2.58  .53 -4.83   9 9 Jap |
+-----+

```

The most significant statistical bias is by the Japanese judge on skater pairs 13 and 14 vs. 18. These pairs are low in the final order, and so of little interest.

Table 23, the [principal components/contrast](#) analysis of Judge residuals is more interesting. Note that Judge 4, the French judge, is at the top with the largest contrast loading. The actual distortion in the measurement framework is small, but crucial to the awarding of the Gold Medal!

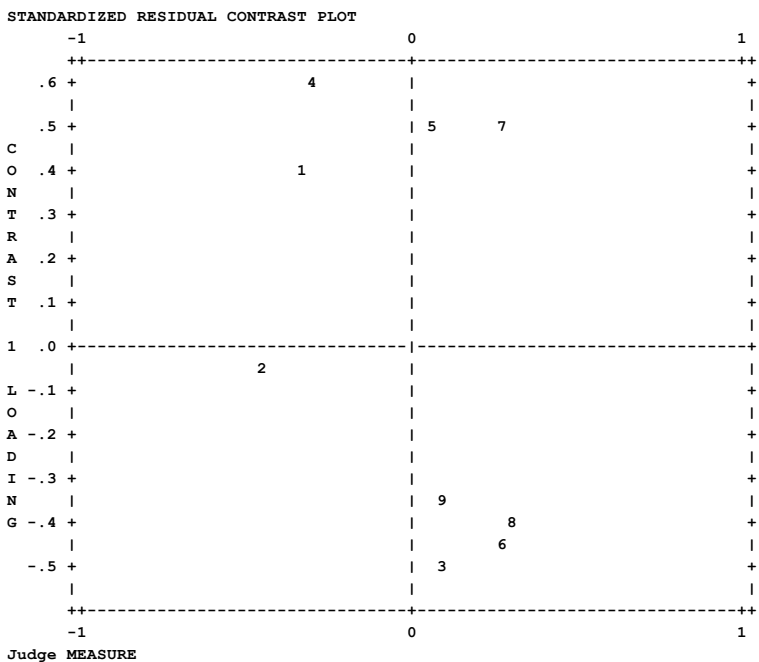


Table 23 variance table also shows a very high explained variance by the measures, 97%, and the estimates require many iterations to converge. Why is this?

The Olympic Ice-skating data is problematic. This is because the judges' ratings are edited prior to display to the public. The ISU (International Skating Union) feel that disagreement among the judges about a skater's performance would look bad. So the head judge instructs disagreeing judges to redo their ratings. All this goes on while we are waiting for the

judges' ratings to display. Sometimes there is quite a long wait! The result is that the data are too Guttman-like. Hence the large explained variance :-)

In fact, the large explained variance indicates that we may have lost measurement accuracy and precision. Rasch uses the randomness in the data to construct the variable. There is little randomness, particularly among high-performing skaters, so the variable definition is weak.

Another flaw in the ratings is that smart judges can game the system. They know that the data will be almost Guttman, so if they make very small tweaks to bias the ratings, the head judge won't catch them, but their effects can be profound - even altering the Gold Medal winners.

Analysts have pointed out these and other flaws to the ISU. Their response has been to make the rating process even more obscure :-)

12.17 Example 16: Disordered categories - thresholds and person anchoring

This example illustrates two types of disordering in rating scale structures: category disordering and Rasch-Andrich [threshold](#) disordering. For further discussion, see [disordering](#). It also illustrates anchor values in the person labels and category labeling.

The control file and data are in **exam16.txt**.

```
; This control file is EXAM16.TXT
title = "Attitude Survey illustrating two types of disordering"
ni      = 3 ; three items
item1 = 1 ; item responses start in column 1
xwide = 2 ; data codes are two columns wide
codes = "1 2 3 " ; valid two character codes
name1 = 9 ; person label starts in column 9
namelength = 3 ; length of person labels

; pafile = $slw3 ; person anchor values in columns 1-3 of person labels
@panchors = 1-3 ; person anchor field is in columns 1 to 3 of person labels
pafile = @panchors ; person anchors in field @panchors

ISGROUPS = 0 ; allow each item to have its own rating scale (the partial credit model)
clfile = * ; category labels: item+category
1+1 Never ; item 1 category 1 is "Never"
1+2 Sometimes ; well-behaved rating scale
1+3 Often
2+1 Car ; categories as ordered in frequency in 1930
2+2 Ship ; now these categories are disordered
2+3 Plane ; ship travel now rarer than planes
3+1 No
3+2 Neutral ; very few in this narrow intermediate category
3+3 Yes
*
&END ; item labels follow
Smooth advance everywhere - probability curves a "range of hills"
Disordered categories - disordered empirical average measures for categories
Low middle frequency - high discrimination - disordered Rasch-Andrich thresholds
END LABELS ; end of item labels, data follow ...
1 1 1 0.0 ; person anchor logits in person labels
1 1 1 0.1 ; advancing anchored person measures
2 3 1 0.2 ; but not advancing categories for item 2
2 3 1 0.3
2 2 2 0.4 ; only one observation of category 2 for item 3, but in the correct place = high
discrimination
2 2 3 0.5
2 2 3 0.6
```

3 2 3 0.7
 3 2 3 0.8 ; last data record

On the [Diagnosis Menu](#): Empirical Item-Category Measures:

TABLE 2.5 Attitude Survey

OBSERVED AVERAGE MEASURES FOR PERSONS (BY OBSERVED CATEGORY)

0	1	2	3	4	5	NUM	ITEM
	1	2	3			1	Smooth advance everywhere
	1	3	2			2	Disordered categories
		1	2	3		3	Low middle frequency - high discrimination

On the Diagnosis Menu: Category Function:

TABLE 3.2 Attitude Survey

FOR GROUPING "0" ITEM NUMBER: 1 Smooth advance everywhere - probability curves a "range of hills"

ITEM DIFFICULTY MEASURE OF 2.00 ADDED TO MEASURES

CATEGORY	OBSERVED	OBSVD	SAMPLE	INFINIT	OUTFIT	ANDRICH	CATEGORY
LABEL	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ
						THRESHOLD	MEASURE
1	1	2	22	1.00	1.62	.49	.52
2	2	5	56	2.00	2.00	.00	.00
Rasch-Andrich thresholds							
3	3	2	22	3.00	2.38	.49	.52
categories							

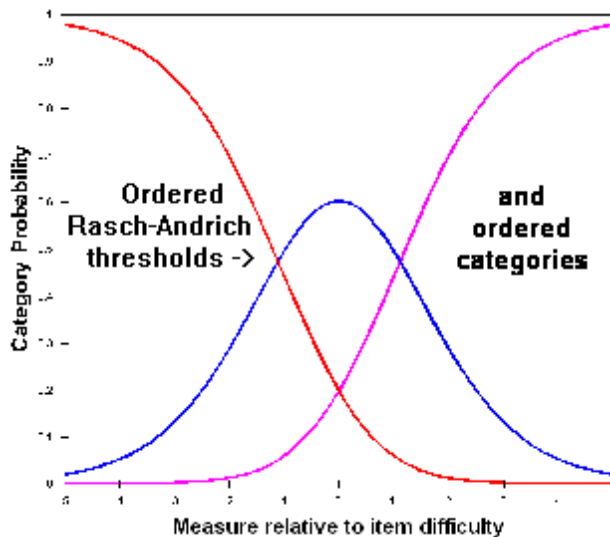


TABLE 3.3 Attitude Survey

FOR GROUPING "0" ITEM NUMBER: 2 Disordered categories - disordered empirical [average measures](#) for categories

ITEM DIFFICULTY MEASURE OF 2.00 ADDED TO MEASURES

CATEGORY	OBSERVED	OBSVD	SAMPLE	INFINIT	OUTFIT	ANDRICH	CATEGORY
----------	----------	-------	--------	---------	--------	---------	----------

ITEM	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	THRESHOLD	MEASURE	DESCRIPTION
1	1	2	22	1.00	1.62	.49	.52	NONE	(-.29)	1 1 Car
2	2	5	56	2.40	2.00	1.01	1.01	-1.11	2.00	2 2 Ship <- Ordered Rasch-Andrich thresholds
3	3	2	22	2.00*	2.38	1.28	1.23	1.11	(4.29)	3 3 Plane <- Disordered categories

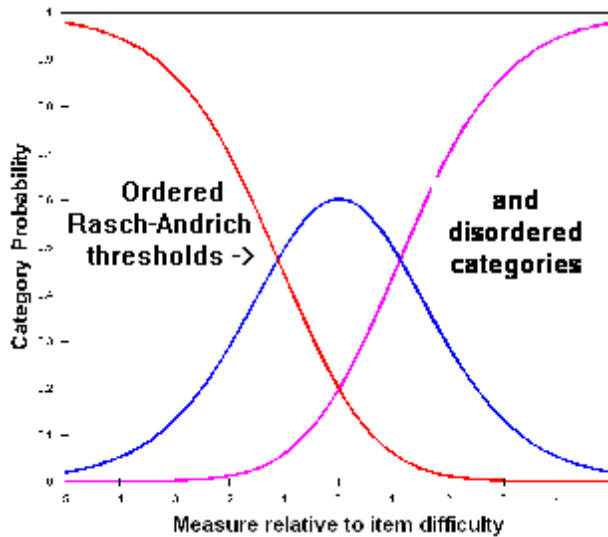
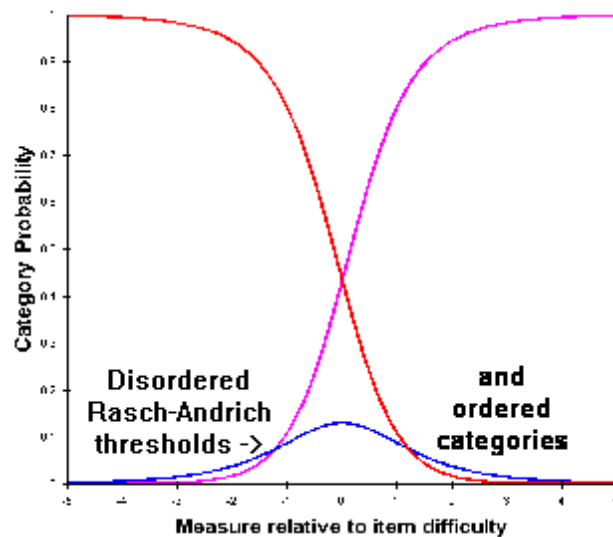


TABLE 3.4 Attitude Survey

FOR GROUPING "0" ITEM NUMBER: 3 Low middle frequency - high discrimination - disordered Rasch-Andrich thresholds

ITEM DIFFICULTY MEASURE OF 2.00 ADDED TO MEASURES

ITEM	SCORE	COUNT	%	AVRGE	EXPECT	MNSQ	MNSQ	THRESHOLD	MEASURE	DESCRIPTION
1	1	4	44	1.50	1.65	.74	.64	NONE	(.86)	1 1 No
2	2	1	11	2.00	2.00	.00	.00	1.21	2.00	2 2 Neutral <- Disordered Rasch-Andrich thresholds
3	3	4	44	2.50	2.35	.74	.64	-1.21	(3.14)	3 3 Yes <- Ordered categories



12.18 Example 17: Rack, stack, principal components

Comparisons of measures at two time-points or in two situations can become complicated. [Example 12](#) illustrates a straightforward situation. It uses the 13 motor items of the FIM®. This example uses the 18 items of the FIM, 13 motor and 5 cognitive, at two time points. There is a 7-level rating scale that is intended to operate in the same way across all items.

In exam17s.txt, the data have been **stacked**. The items are modeled to maintain their difficulties across the two time points, and the 52 patients are entered twice, once at admission to rehabilitation and once at discharge from it, so there 104 data records. Changes in patient independence can be identified by cross-plotting the admission and discharge measures for each patient, as in [Example 12](#). This can be done by using the [Plots](#) menu to plot the measures against themselves, and then, in Excel, pasting the discharge measures over the top of the admission measures for the y-axis. Here is the procedure:

1. Do the analysis with all time 1 data above all time 2 data, and the persons in the same order at both time-points.
2. [Output Files](#) menu: PFILE=pf.txt
3. Edit pf.txt
4. Place time 1 persons in pf1.txt
5. Place time 2 persons in pf2.txt
6. Make sure that the person labels match (put in blank rows to make them align)
7. Winsteps [Plots menu](#): Scatterplot
8. x-axis file is pf1.txt
9. y-axis file is pf2.txt
10. Plot the measures
11. [Plot](#) is produced in Excel

In exam17r.txt, the data have been **racked**. The persons are modeled to maintain their abilities across the two time points, and the 18 FIM items are entered twice, once at admission to rehabilitation and once at discharge from it, so there 36 items. Average changes in patient performance on individual items can be identified by cross-plotting the admission and discharge measures for each item. This can be done by using the [Plots](#) menu to plot the measures against themselves, and then, in Excel, pasting the discharge measures over the top of the admission measures for the y-axis.

A further feature is the contrasts in the Principal Components Analysis of Residuals, [Tables 23 and 24](#). The 13 motor items and 5 cognitive items are probing different aspects of patient independence. Do they function as one variable in this sample? See [Principal components/contrast](#). Patients also have different impairments. Are their measures comparable? See [Differential Item Function](#).

```
; This control file is EXAM17S.TXT
TITLE='Example 17S: 18-item FIM control file: stacked admission then discharge'
ITEM1=1 ; Responses start in column 7
NI=18 ; 13 motor items + 5 cognitive items
```



```

NAME1=20 ; start of person label
; variables in person label
@SETTING=$S1W1 ; setting: A=Admission, D=Discharge
@SEX=$S3W1 ; Gender: 1=Male, 2=Female
@IGC=$S5W2 ; Impairment Group Code: 8=Orthopedic, 9=Cardiac, 10=Pulmonary, 11=Burns,
12=Congenital, 13=Other
@ID=$S8W2 ; Patient sequence number

; variables in item label
@SUBTEST=$S1W1 ; subtest: M=Motor, C=Cognitive
@ILETTER=$S3W1 ; code letter of item

CODES=1234567 ; 7 level rating scale
CLFILE=* ; Defines the rating scale
1 0% Independent
2 25% Independent
3 50% Independent
4 75% Independent
5 Supervision
6 Device
7 Independent
*
&END
M A. EATING
....
C R. MEMORY
END NAMES
334412312331123112 A 1 8 1
432322442223134444 A 2 8 2
....

; This control file is EXAM17R.TXT
TITLE='Example 17: 18-item FIM control file: racked admission and discharge'
ITEM1=1 ; Responses start in column 7
NI=37 ; 13 motor items + 5 cognitive items: admission and discharge
NAME1=39 ; start of person label

; variables in person label
@SEX=$S1W1 ; Gender: 1=Male, 2=Female
@IGC=$S3W2 ; Impairment Group Code: 8=Orthopedic, 9=Cardiac, 10=Pulmonary, 11=Burns,
12=Congenital, 13=Other
@ID=$S5W2 ; Patient sequ. number

; variable in item label
@SETTING=$S1W1 ; setting: A=Admission, D=Discharge
@SUBTEST=$S3W1 ; subtest: M=Motor, C=Cognitive
@ILETTER=$S5W1 ; code letter of item

CODES=1234567 ; 7 level rating scale
CLFILE=* ; Defines the rating scale
1 0% Independent
2 25% Independent
3 50% Independent
4 75% Independent
5 Supervision
6 Device
7 Independent
*
&END
A M A. EATING
....
A C R. MEMORY

```

```

-      Blank for ease of seeing
D M A. EATING
....
D C R. MEMORY
END NAMES
334412312331123112 554535546665345545 1 8 1
432322442223134444 777677777666567777 2 8 2
....

```

12.19 Example of subset reporting

You see: **Warning:** Data are ambiguously connected into 6 subsets. Measures may not be comparable across subsets.

Quick (but arbitrary) solution: add to the data file two dummy person records so that all persons and items become directly comparable.

Dichotomous data:

```

Dummy person 1: responses: 010101010...
Dummy person 2: responses: 101010101...

```

This says: "the middle level of performance for all subsets of persons is the same."

Rating scale data, where "1" is the lowest category, and "5" is the highest category:

```

Dummy person 1: responses: 121212121...
Dummy person 2: responses: 212121212...

```

This says: "the bottom level of performance for all subsets of persons is the same."

Explanation: Connectivity (or subsetting) is a concern in any data analysis involving missing data. In general, nested data are not connected.

fully-crossed data (also called "complete data") are connected.

partially-crossed data may or may not be connected.

Winsteps examines the responses strings for all the persons. It verifies that every non-extreme response string is linked into one network of success and failure on the items. Similarly, the strings of responses to the items are linked into one network of success and failure by the persons.

If person response string A has a success on item 1 and a failure on item 2, and response string B has a failure on item 1 and a success on item 2, then A and B are connected. This examination is repeated for all pairs of response strings and all pairs of items. Gradually all the persons are connected with all the other persons, and all the items are connected with all the other items. But if some persons or some items cannot be connected in this way, then Winsteps reports a "connectivity" problem, and reports which subsets of items and persons are connected.

Mathematics: connectivity is part of [Graph Theory](#). The person/item/judge/... parameters of the Rasch model are the vertices and the observations are the edges. In an undirected graph, we need every vertex to be connected directly or indirectly to every other vertex. A connection is established between two vertices when one vertex is observed to have both a higher observation and a lower observation than another vertex in the same context, or when both both vertices have the same intermediate category of a rating scale in the same context.

Thus there are two situation for failure to connect:

1) there is no direct or indirect link between two vertices, e.g., two different datasets analyzed together with no common parameters. This is detected by the Winsteps/Facets subset routine.

2) the vertices are connected by observations, but the observations do not meet the requirements, e.g., all the person respond to all the items, but half the persons score in the upper half of the rating scale on every item, and the other half of the persons score in the lower half of the rating scale on every item. This is called a "Guttman split" in the data. This is usually obvious in the reported estimates as a big gap on the Wright maps between the two halves of the person distribution.

Example 1: Connection problems and subsets in the data are shown in this dataset. It is **Examsubs.txt**.

```

Title = "Example of subset reporting"
Name1 = 1
Namelength = 24 ; include response string in person label
Item1 = 13
NI = 12
CODES = 0123 ; x is missing data
ISGROUPS = DDDDDDDDDRRR ; items 1-10 are dichotomies; items 11-12 share a rating scale
MUCON = 3 ; Subsetting can cause very slow convergence
TFILE=*
18.1
14.1
0.4
*
&End
01 Subset 1
02 Subset 1
03 Subset 2
04 Subset 2
05 Subset 7
06 Subset 4
07 Subset 4
08 Subset 5
09 Subset 5
10 Subset 5
11 Subset 6
12 Subset 6
END LABELS
01 Extreme 11111
02 Subset 1 01111
03 Subset 1 10111
04 Subset 2 00101
05 Subset 2 00011
06 Subset 3 011
07 Subset 3 011
08 Subset 4 001
09 Subset 4 010
10 Subset 5 0x1
11 Subset 5 10x
12 Subset 5 x10
13 Subset 6 01
14 Subset 6 10
15 Subset 6 23
16 Subset 6 32

```

The Iteration Screen reports:

```

CONVERGENCE TABLE
-Control: \HOLDW95\examples\examsubs.txt Output: \examples\ZOU571WS.TXT
| PROX ACTIVE COUNT EXTREME 5 RANGE MAX LOGIT CHANGE |
| ITERATION PERSON ITEM CATS PERSON ITEM MEASURES STRUCTURE |
>=====|
| 1 15 12 8 2.00 1.06 -2.0794 |
>=====|
| 2 15 12 6 2.38 1.84 2.6539 -1.6094 |
>=====|

```

```

|      3      14      12      6      2.67      1.60      2.7231      .0000 |
>=====<
|      4      14      12      6      2.68      2.33      -2.3912      .0000 |
>=====<
|      5      14      12      6      2.97      1.77      2.3246      |
>=====<
|      6      14      12      6      2.97      2.54      -2.2191      |
>=====<
|      7      14      12      6      3.22      2.10      2.0372      |
Probing data connection: to skip out: Ctrl+F - to bypass: subset=no
Processing unanchored persons ...
>=====<
Consolidating 9 potential subsets pairwise ...
>=====
Consolidating 9 potential subsets indirectly pairwise ...
>=====<
Consolidating 8 potential subsets pairwise ...
>=====
Consolidating 7 potential subsets pairwise ...
>=====
Consolidating 7 potential subsets indirectly pairwise ...
>=====<
Warning: Data are ambiguously connected into 7 subsets. Measures may not be comparable
        across subsets.
        Subsets details are in Table 0.4

```

[Table 18.1](#)

PERSON STATISTICS: ENTRY ORDER

```

-----
|ENTRY  | |
|NUMBER | | PERSON
|-----+-----|
|   1   | | 01 Extreme 11111 | MAXIMUM MEASURE
|       | |                   | < Guttman split here >
|   2   | | 02 Subset 1 01111 | SUBSET 1
|   3   | | 03 Subset 1 10111 | SUBSET 1
|       | |                   | < Guttman split here >
|   4   | | 04 Subset 2 00101 | SUBSET 2
|   5   | | 05 Subset 2 00011 | SUBSET 2
|       | |                   | < Guttman split here >
|   6   | | 06 Subset 3   011 | SUBSET 3
|   7   | | 07 Subset 3   011 | SUBSET 3
|       | |                   | < Guttman split here >
|   8   | | 07 Subset 4   001 | SUBSET 4
|   9   | | 09 Subset 4   010 | SUBSET 4
|       | |                   | < Subset split here >
|  10   | | 10 Subset 5     0x1 | SUBSET 5
|  11   | | 11 Subset 5     10x | SUBSET 5 < Indirect connection >
|  12   | | 12 Subset 5     x10 | SUBSET 5
|       | |                   | < Subset split here>
|  13   | | 13 Subset 6     01 | SUBSET 6
|  14   | | 14 Subset 6     10 | SUBSET 6
|       | |                   | < undetected Guttman split here:
|       | | Winsteps failed! >
|  15   | | 15 Subset 6     23 | SUBSET 6
|  16   | | 16 Subset 6     32 | SUBSET 6
|-----+-----|

```

<i>In Tables and Notes:</i>	<i>Explanation:</i>
< Guttman split here >	The persons above the split performed an unknowable amount different from the persons below the split. There is no item on which this subset succeeded

	and another subset failed, and also this subset failed and the other subset succeeded. The data are not "well-conditioned" (Fischer G.H., Molenaar, I.W. (eds.) (1995) Rasch models: foundations, recent developments, and applications. New York: Springer-Verlag. p. 41-43).
< Subset split here >	The persons in this subset responded to different items than persons in other subsets. We don't know if these items are easier or harder than items in other subsets.
< Indirect connection >	The persons responded to different items, but they are connected by a loop of successes and failures.
< undetected Guttman split here >	Winsteps subset-detection did not report that persons 13 and 14 always score lower than persons 15 and 16, causing a Guttman split. We do not know how much better persons 15 and 16 are than persons 14 and 15. Winsteps subset-detection may fail to report subsets. Unreported subsets usually cause big jumps in the reported measures.
Data are ambiguously connected	Measures for persons in different subsets are not comparable. Winsteps always reports measures, but these are only valid within subsets. We do not know how the measures for persons in one subset compare with the measures for persons in another subset. Reliability coefficients are accidental and so is Table 20 , the score-to-measure Table. Fit statistics and standard errors are approximately correct.
Measures may not be comparable across subsets	Please always investigate when Winsteps reports subsets, even if you think that all your measures are comparable.
MAXIMUM MEASURE, MINIMUM MEASURE, DROPPED, INESTIMABLE	Persons and items with special features are not included in subsets. Extreme scores (zero, minimum possible and perfect, maximum possible scores) imply measures that are beyond the current frame of reference. Winsteps uses Bayesian logic to provide measures corresponding to those scores.
SUBSET 1, 2, 4	These are directly connected subsets. Within each subset, a person has succeeded on an item and failed on an item, and <i>vice-versa</i> . The person performances are directly pairwise comparable within the subset. The persons in this subset have either succeeded on items in other subsets, or failed on items in other subsets, or have missing data on items in other subsets.
SUBSET 3	These two persons have the same responses, so they are in the same subset. No one succeeded on their failed items item, and also failed on their successful item.
SUBSET 5	This is an indirectly connected subset. There is a loop of successes and failures so that the performances of all three persons are connected indirectly pairwise.
SUBSET 6	Persons 13 and 14 are directly comparable using categories 0 and 1 of the rating scale. Persons 15 and 16 are directly comparable using categories 2 and 3 of the rating scale. Winsteps has not detected that persons 13 and 14 always rate lower than persons 15 and 16, causing a Guttman split.
SUBSET 7 (Table 14.1)	No person is in the same subset as this item. There is no subset in which persons both succeeded and failed on this item.
Connecting SUBSETS	Here are approaches: 1. Collect more data that links items across subsets. Please start Winsteps analysis as soon as you start data collection. Then subset problems can be remedied before data collection ends. 2. Dummy data. Include data for imaginary people in the data file that connects the subsets. 3. Anchor persons or items. Anchor equivalent items (or equivalent persons) in the different subsets to the same values - or juggle the anchor values to make the mean of each subset the same (or whatever)

	<p>4. Analyze each subset of persons and items separately. In Table 0.4, Winsteps reports entry numbers for each person and each item in each subset, so that you can compare their response strings. To analyze only the items and persons in a particular subset, such as subset 4 above, specify the items and persons in the subset:</p> <p>IDELETE= +9-10 PDELETE= +10-11</p>
Memory was not allocatable to probe connectivity	If the data are complete, ignore this message. If the data are sparse, add dummy data records. They will have little influence on connected data, but will connected up data with subsets. See also Memory

Table 14.1

```

-----
|ENTRY      |          |
|NUMBER     | ITEM      | G |
|-----+-----|
|    1     | 01 Subset 1 D | SUBSET 1
|    2     | 02 Subset 1 D | SUBSET 1
|          |          |
|    3     | 03 Subset 2 D | SUBSET 2
|    4     | 04 Subset 2 D | SUBSET 2
|          |          |
|    5     | 05 Subset 7 D | SUBSET 7
|          |          |
|    6     | 06 Subset 4 D | SUBSET 4
|    7     | 07 Subset 4 D | SUBSET 4
|          |          |
|    8     | 08 Subset 5 D | SUBSET 5
|    9     | 09 Subset 5 D | SUBSET 5
|   10     | 10 Subset 5 D | SUBSET 5
|          |          |
|   11     | 11 Subset 6 R | SUBSET 6
|   12     | 12 Subset 6 R | SUBSET 6
|-----+-----|

```

< Guttman split here >

< Guttman split here >

< Guttman split here >

< Guttman split here >

< Guttman split here >

Table 0.4 reports

```

SUBSET DETAILS

Subset 1 of 2 ITEM and 2 PERSON
ITEM: 1-2
PERSON: 2-3
Subset 2 of 2 ITEM and 2 PERSON
ITEM: 3-4
PERSON: 4-5
Subset 3 of 2 PERSON
PERSON: 6-7
Subset 4 of 2 ITEM and 2 PERSON
ITEM: 6-7
PERSON: 8-9
Subset 5 of 3 ITEM and 3 PERSON
ITEM: 8-10
PERSON: 10-12
Subset 6 of 2 ITEM and 4 PERSON
ITEM: 11-12
PERSON: 13-16
Subset 7 of 1 ITEM
ITEM: 5

```

Example 2: Analyzing two separate datasets together.

Dataset 1. The Russian students take the Russian items. This is connected. All the data are in one subset.

Dataset 2. The American students take the American items. This is connected. All the data are in one subset.

Dataset 3. Datasets 1 and 2 are put into one analysis. This is not connected. The data form two subsets: the Russian one and the American one. The raw scores or Rasch measures of the Russian students cannot be compared to those of the American students. For instance, if the Russian students score higher than the American students, are the Russian students more able or are the Russian items easier? The data cannot tell us which is true.

Winsteps attempts to estimate an individual measure for each person and item within one frame of reference. Usually this happens. But there are exceptions.

12.20 Examples on www.winsteps.com

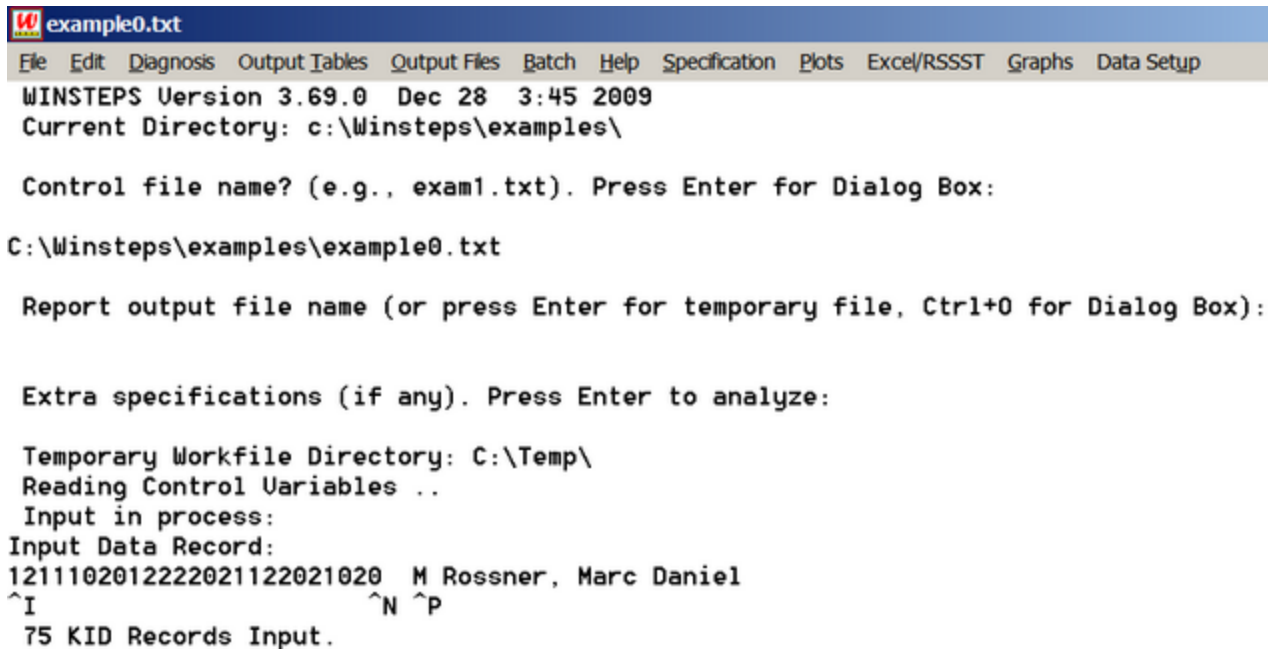
www.winsteps.com has other examples of Winsteps control and data files. For example:

1. The Languages of Love - www.winsteps.com/languages.htm
2. Masterchef Australia 2017 - www.winsteps.com/masterchef.htm
3. Masterchef Australia 2018 - www.winsteps.com/masterchef-2018.htm

13 Analysis Window

13.1 The Analysis Window

While Winsteps is running, information about the analysis is displayed on the screen. The iterative estimation process is by [logistic curve-fitting](#). Here is an example based on the "Liking for Science" data, [example0.txt](#)



```
example0.txt
File Edit Diagnosis Output Tables Output Files Batch Help Specification Plots Excel/RSSST Graphs Data Setup
WINSTEPS Version 3.69.0 Dec 28 3:45 2009
Current Directory: c:\Winsteps\examples\

Control file name? (e.g., exam1.txt). Press Enter for Dialog Box:

C:\Winsteps\examples\example0.txt

Report output file name (or press Enter for temporary file, Ctrl+O for Dialog Box):

Extra specifications (if any). Press Enter to analyze:

Temporary Workfile Directory: C:\Temp\
Reading Control Variables ..
Input in process:
Input Data Record:
1211102012222021122021020 M Rossner, Marc Daniel
^I ^N ^P
75 KID Records Input.
```

The Winsteps analysis window can display:

```
WINSTEPS Version: 4.5.2
program running - shows version number
```

```
Reading Control Variables .. Processing your control variables
Input in process ..... Reading in your data: each . is 1,000 persons
```

```
1211102012222021122021020 M Rossner, Marc Daniel
^I ^N ^P
```

^I where item response string starts, ITEM1=

^N last item response, NI=

^P where person label starts, NAME1=

75 person records input Total person records found

CONVERGENCE TABLE								
-Control: \examples\example0.txt					Output: \examples\ZOU706WS.TXT			
PROX	ACTIVE COUNT			EXTREME 5 RANGE		MAX LOGIT CHANGE		
ITERATION	KID	ACT	CATS	KID	ACT	MEASURES	STRUCTURE	
1	75	25	3	3.78	3.20	3.8918	.0740	
2	74	25	3	4.59	3.71	.8258	-.6158	
3	74	25	3	4.83	3.92	.2511	-.1074	
PROBING DATA CONNECTION: to skip: subset=no								
Control: \examples\example0.txt					Output: \examples\ZOU706WS.TXT			
JMLE	MAX SCORE	MAX LOGIT	LEAST CONVERGED		CATEGORY		STRUCTURE	
ITERATION	RESIDUAL*	CHANGE	KID	ACT	CAT	RESIDUAL	CHANGE	
1	2.84	-.1955	60	22*	2	.01	.0076	
2	.71	-.0335	53	15*	0	.01	.0144	
3	-.43	.0297	53	5*	1	.01	.0101	
4	.32	.0237	18	11*	1	.01	.0079	
5	.24	.0184	18	11*	0	.01	.0060	
6	.19	.0141	18	11*	0	.00	.0045	
7	.14	.0108	18	11*	0	.00	.0035	
8	.11	.0082	18	11*	0	.00	.0026	
9	.08	.0062	18	11*	0	.00	.0020	
10	.06	.0048	18	11*	0	.00	.0015	

To stop iterations: press Ctrl+S

To cancel the analysis: press Ctrl+F

CONVERGENCE TABLE: See [Table 0.2](#)

PROBING DATA CONNECTION: See [SUBSETS=](#)

PROX and JMLE: See [Estimation methods](#)

Check that values diminish, and become near 0. See [convergence considerations](#)


```

Calculating Fit Statistics
>=====<
Standardized Residuals N(0,1) Mean: .06 S.D.: 1.04
Time for estimation: 0:0:0.750
LIKING FOR SCIENCE (Wright & Masters p.18)
-----
| KID      75 INPUT      75 MEASURED      INFIT      OUTFIT  |
|          TOTAL      COUNT      MEASURE REALSE      IMNSQ  ZSTD  OMNSQ  ZSTD|
| MEAN     31.7      25.0      .97      .43      .99   -.2   1.08   .1|
| S.D.     8.6       .0       1.35     .20     .50   1.6   1.04   1.9|
| REAL RMSE .47 TRUE SD  1.26 SEPARATION  2.66 KID  RELIABILITY .88|
|-----|
| ACT      25 INPUT      25 MEASURED      INFIT      OUTFIT  |
|          TOTAL      COUNT      MEASURE REALSE      IMNSQ  ZSTD  OMNSQ  ZSTD|
| MEAN     95.0      75.0      .00      .25     1.02  -.2   1.08   .0|
| S.D.     30.9      .0       1.41     .08     .45   2.3   .87   2.8|
| REAL RMSE .26 TRUE SD  1.38 SEPARATION  5.32 ACT  RELIABILITY .97|
|-----|
Output written to C:\Winsteps\examples\ZOU280WS.TXT
CODES= 012
Measures constructed: use "Diagnosis" and "Output Tables" menus

```

CALCULATING FIT STATISTICS

During this iteration through the data, the person and item fit statistics are computed. The exact formulas are in the book Rating Scale Analysis, p. 100 - also at [Fit Equations](#).

Summary statistics are extracted from [Table 3.1](#) with [REALSE=Yes](#) and including extreme scores.

14 Output Tables

14.1 Table 0 Analysis Report

14.1.1 Table 0.1 Analysis identification

- [Table 0.1](#) Analysis Identification
- [Table 0.2](#) Convergence report
- [Table 0.3](#) Control file
- [Table 0.4](#) Subset details

Table 0 is in the Report Output File in the [Edit menu](#). In the [Output Table menu](#), request it from [Subtable](#), 0.

```

Table Heading omitted here.
*****
*
*          * * * * W I N S T E P S * * * *
*          -----
*
*          - RASCH ANALYSIS FOR TWO-FACET MODELS -
*
*          PERSON, ITEM & RESPONSE STRUCTURE MEASUREMENT AND FIT ANALYSIS
*
*          WINSTEPS®(R)   www.winsteps.com
*
*          Copyright © 2020 by John M. Linacre. All rights reserved.
*          VERSION 4.5.2
*****

TITLE= LIKING FOR SCIENCE (Wright & Masters p.18)
CONTROL FILE: C:\Winsteps\sف.txt
OUTPUT FILE: C:\Winsteps\ZOU042ws.txt
DATE: Apr 22 2015 14:45

ACT DELETIONS: 2-3 10-15 17-19 21 24
76 PUPIL Records Input.

```

This shows the Title, Control file, Output report file, date and time of this analysis. This page contains the authorship and version information.

Winsteps is updated frequently. Please refer to the web page for current version numbers and recent enhancements at www.winsteps.com

If the output file name has the form ZOU???.ws.txt, then it is a temporary file which will be erased when Winsteps terminates. You can "Save As" as a permanent file. This output file is on the [Edit menu](#).

14.1.2 Table 0.2 Convergence report

(controlled by [LCONV=](#), [RCONV=](#), [CONVERGE=](#), [MPROX=](#), [MJMLE=](#), [CUTLO=](#), [CUTHI=](#))

[Table 0.1](#) Analysis Identification

[Table 0.2](#) Convergence report

[Table 0.3](#) Control file

[Table 0.4](#) Subset details

Table 0 is in the Report Output File in the [Edit menu](#). In the [Output Table menu](#), request it from [Subtable](#), 0.

```

CONVERGENCE TABLE
+-----+
| PROX          ACTIVE COUNT          EXTREME 5 RANGE          MAX LOGIT CHANGE |
| ITERATION    PUPILS  ACTS    CATS    PUPILS  ACTS    MEASURES  STRUCTURE |
+-----+
|      1       76    25    3      3.59   1.62      3.1355   -.1229 |
|      2       74    12    3      4.03   1.90      .3862    -.5328 |
|      3       74    12    3      4.19   1.96      .1356    -.0783 |
WARNING: DATA ARE AMBIGUOUSLY CONNECTED INTO 6 SUBSETS. MEASURES ACROSS SUBSETS ARE NOT COMPARABLE
see Connection Ambiguities
+-----+
| JMLE         MAX SCORE   MAX LOGIT   LEAST CONVERGED   CATEGORY STRUCTURE |
| ITERATION    RESIDUAL*   CHANGE     PUPIL  ACT    CAT  RESIDUAL  CHANGE |
+-----+
|      1       -2.04      .2562      7     5*    2     -.72     .0003 |
+-----+
Time for estimation: 0:0:0.166

```

The meanings of the columns are:

PROX normal approximation algorithm - for quick initial estimates

ITERATION number of times through your data to calculate estimates

ACTIVE COUNT number of parameters participating in the estimation process after elimination of deletions, and perfect (maximum possible) and zero (minimum possible) scores:

PERSONS person parameters

ITEMS item parameters

CATS rating scale categories - shows 2 for dichotomies

These counts may reduce because persons, items or categories may have been

1. deleted
2. dropped because they have no responses
3. dropped from standard estimation because they are unanchored and have extreme scores. These will be reported with Bayesian estimates.

EXTREME 5 RANGE

PERSONS The current estimate of the spread between the average measure of the top 5 persons and the average measure of the bottom 5 persons.

ITEMS The current estimate of the spread between the average measure of the top 5 items and the average measure of the bottom 5 items.

MAX LOGIT CHANGE

MEASURES maximum logit change in any person or item estimate. This is expected to decrease gradually until convergence, i.e., less than LCONV=.

STRUCTURE maximum logit change in any Andrich Threshold estimate - for your information - need not be as small as MEASURES.

JMLE [JMLE](#) joint maximum likelihood estimation - for precise estimates

ITERATION number of times through your data to calculate estimates

It is unusual for more than 100 iterations to be required

MAX SCORE RESIDUAL maximum score residual (difference between integral observed score and decimal expected score) for any person or item estimate - used to compare with [RCONV=](#). This number is expected to decrease gradually until convergence acceptable.

* indicates to which person or item the residual applies.

MAX LOGIT CHANGE maximum logit change in any person or item estimate - used to compare with [LCONV=](#). This number is expected to decrease gradually until convergence is acceptable.

LEAST CONVERGED element numbers are reported for the person, item and category farthest from meeting the convergence criteria.

* indicates whether the person or the item is farthest from convergence.

the CAT (category) may not be related to the ITEM to its left. See [Table 3.2](#) for details of unconverted categories.

CATEGORY RESIDUAL maximum count residual (difference between integral observed count and decimal expected count) for any response structure category - for your information. This number is expected to decrease gradually. Values less than 0.5 have no substantive meaning.

STRUCTURE CHANGE maximum logit change in any structure calibration (usually Rasch-Andrich Threshold). Not used to decide convergence, but only for your information. This number is expected to decrease gradually.

Look for scores and residuals in the last iteration to be close to 0,

14.1.3 Table 0.3 Control file

[Table 0.1](#) Analysis Identification

[Table 0.2](#) Convergence report

[Table 0.3](#) Control file

[Table 0.4](#) Subset details

Table 0 is in the Report Output File in the [Edit menu](#). In the [Output Table menu](#), request it from [Subtable](#), 0.

Table 0.3 shows the control file used in the analysis. It includes the extra specifications and expands [SPFILE=](#) commands. This is also appended to the [LOGFILE=](#), when specified.

A complete listing of control variables is obtained using "Control variable file=" from the [Output Files](#) pull-down menu, which is also [Table 32](#).

"Extra Specifications" are listed after [&END](#).

```
&INST
TITLE='LIKING FOR SCIENCE +
      +(Wright & Masters p.18)' ;demonstrates continuation line
ITEMS=ACT
PERSONS=PUPIL
ITEM1=1
NI=25
NAMLMP=20
XWIDE=2
NAME1=51
isubtot=$s1W1
psubtot=$s2W1
pweight=$s2w1
iweight=$s3w1
dif=$s3W1
dpf=$s4W1
CODES=000102
; ISGROUPS=0
CFILE=*
00 dislike
01 neutral
02 like
*
idfile=*
13 GROW GARDEN |
10 LISTEN TO BIRD SING |
2 READ BOOKS ON ANIMALS |
```


Table 1.0 Wright map - person and item labels on one page

```

MEASURE                                KID - MAP - TAP
<more>|<rare>
 5                                     + 1-3- 1-4- 1-4- 4-1-
 4                                     |
                                     T+
 3                                     Mik Tra |S 1-4-
                                     Bla +
 2                                     Dav Rod Wal Zul S+ 1-3- 1-4-
                                     |
 1                                     Aud Jam Kim Ron Wil + 1-3-
                                     |
 0                                     M+M
Ann Bre Don Dor Hel Jan Joe Lis Mar Pet Ric Tho |
-1                                     +
-2                                     Bet Els Sus | 2-4-
-3                                     Ber Bri S+
                                     | 1-4-
-4                                     Ada Ric +
-5                                     Bar Car |S 1-3- 3-4-
                                     T+ 1-4- 2-1-
                                     Lin | 1-3-
                                     Fra + 1-2- 1-4 2-3
<less>|<freq>
    
```

Column Labels for Tables 1.0, 1.1, 1.2, 1.4

PERSON= MAP ITEM=
MORE= T RARE=
 |
LESS= ⊥ FREQ= frequent

Table 1.0 is printed, if person and item maps can be squeezed into one page. MAXPAGE= controls height of table (number of lines for each logit). LINELENGTH= controls width of table.

Left-hand column locates the person ability measures along the variable. The persons often have a normal distribution.

Right-hand column locates the item difficulty measures along the variable. [Table 12](#) has the full item labels. Items arranged by measure: look for the hierarchy of item names to spell out a meaningful construct from easiest at the bottom to hardest at the top.

For dichotomous items, look for an even spread of items along the variable (the y-axis) with no gaps. Gaps can indicate poorly defined or poorly tested regions of the variable. Good tests usually have the items targeted (lined up with) the persons.

For polytomous items, [Table 2.2](#) shows the operational range of the item with its rating scale.

You can use IMAP= and PMP= to choose parts of labels to show. NAMLMP= limits the number of characters of each name reported. LINELENGTH= changes the available space on the line. MAXPAGE= controls the length of the Table. MRANGE= controls the displayed range of measures. ISORT= and PSORT= control the sort order within rows.

For the full person labels, please use [Table 16](#).

For the full item labels, please use [Table 12](#).

Table 1.1 Wright map - person and item distributions

```

MAP OF KID AND TAP
MEASURE                                |                                MEASURE
<more> ----- KID    +- TAP    ----- <rare>
 5                                     + XXXX    5
 4                                     |
                                     T+
 3                                     XX |S X    3
                                     X +
 2                                     XXXX S+ XX    2
                                     |
 1                                     XXXXX + X    1
                                     |
 0                                     M+M    0
XXXXXXXXXXXXXXXXX |
-1                                     +    -1
                                     XXX | X
-2                                     XX S+    -2
                                     | X
-3                                     XX +    -3
                                     XX |S XX
    
```

This table shows the distribution of the persons and items. The variable is laid out vertically with the most able persons, and most difficult items at the top.

Left-hand column locates the person ability measures along the variable.

Right-hand column locates the item difficulty measures along the variable.

-4		T+ XX	-4
		X X	
-5		X + XXX	-5
<less>	----- KID	-- TAP	----- <freq>

Table 1.2 Wright map - person distribution + item labels

```

MEASURE      KID - MAP - ACT
              <more>|<rare>
5             . +
              . |
              . |
4             . +
              . |
              . T|
3             . +
              . |T
              ## | FIND BOTTLES AND  WATCH A RAT
2             .## S+
              ## | LOOK IN SIDEWALK  WATCH BUGS          WATCH GRASS CHANG
              .### |S
1             .#### M+ WATCH ANIMAL MOVE
              .## | LEARN WEED NAMES  MAKE A MAP          TALK W/FRIENDS AB
##### | LOOK AT PICTURES  LOOK UP STRANGE A  READ BOOKS ON PLA
0             .##### +M FIND OUT WHAT ANI  WATCH WHAT ANIMAL
              # S| FIND OUT WHAT FLO  READ ANIMAL STORI  WATCH BIRDS
              # | READ BOOKS ON ANI
-1            # + FIND WHERE ANIMAL  WATCH BIRD MAKE N
              # |S GROW GARDEN        LISTEN TO BIRD SI
              . T|
-2            + GO TO MUSEUM
              | GO TO ZOO
              |T
-3            + GO ON PICNIC
              <less>|<freq>
EACH "#" IS 2: EACH "." IS 1

```

Left-hand column locates the person ability measures along the variable. The persons often have a normal distribution.

Right-hand column locates the item difficulty measures along the variable. The item labels are abbreviated to fit on one page. [Table 12](#) has the full item labels.

Effective operational range of an instrument with dichotomous items.

Table 1.2 is enough if the data are dichotomies. Then each item has an operational range of about 1.5 logits up and down from the item difficulty. For an ability at 1.5 logits from the difficulty of a dichotomous item, the response to the item contain about 60% of the statistical information of a response for an ability equal to the item difficulty.

In Table 1.2 the X's are located at the point where there is a 50% chance of being observed in the bottom category and a 50% chance of being observed in the top category. For dichotomies, there are only 2 categories, so this is the whole story.

Table 1.3 Wright map - item distribution + person labels

```

MEASURE ACT - MAP - KID
              <rare>|<more>
5             + M
              | F
              |
4             + F
              | F
              |T M
3             + F
              T| M
              XX | F F F M
2             +S F M M M M
              XXX | F M M M
              S| F F M M M M M
1             X +M F F M M M M M M
              XXX | M M M M M
              XXX | F F M M M M M M M M M M M
0             XX M+ M M M M M M M M M M M
              XXX |S F M
              X | F M
-1            XX + F M
              XX S| M M
              |T M
-2            X +
              X |
              T|
-3            X +
              <freq>|<less>

```

Left-hand column shows the item difficulty distribution measures along the variable.

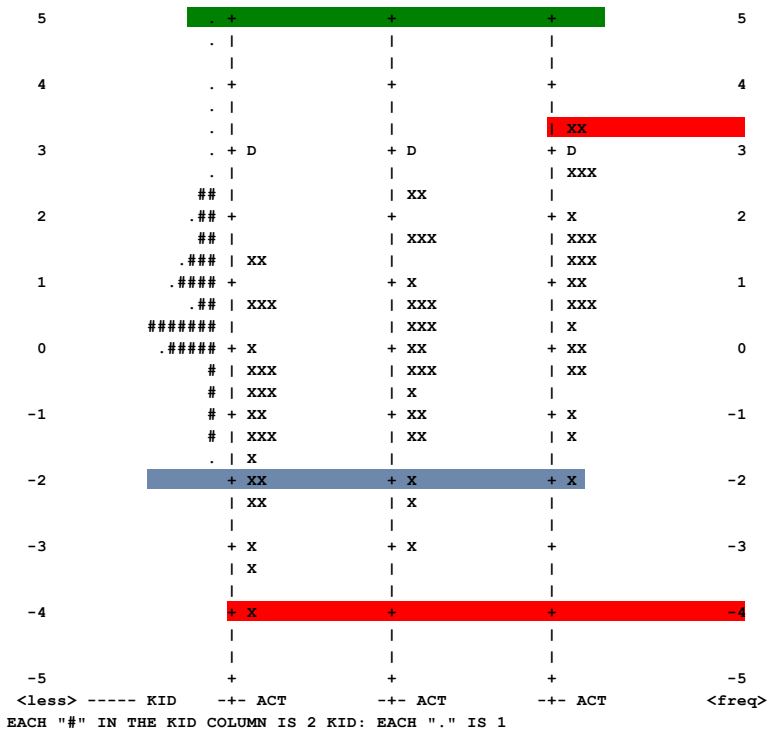
Right-hand column locates the person ability measures along the variable. The persons often have a normal distribution. The person labels are abbreviated to fit on one page. [Table 16.3](#) shows the full person labels

Column Labels		
ITEM=	MAP	PERSON=
RARE=	T	MORE=
FREQ=	↓	LESS=

Table 1.4 Wright map - person and item distributions + polytomous item range

MEASURE		BOTTOM P=50%	MEASURE	TOP P=50%	MEASURE
<more>	----- KID	-- ACT	-- ACT	-- ACT	<rare>

Table 1.4 is only produced when there are polytomous items. Table 1.4 shows the distribution of the person MEASURES and items. The variable is laid out vertically with



the most able persons, and most difficult items at the top. Each person or polytomous item is indicated by an "X", or, when there are too many "X"s to display on one line, several persons or items are represented by a "#". Less than that number by ".". So that if "#" represents 4, then "." represents 1 to 3. Each item is shown three times in Table 1.4. Dichotomous items, "D", have only one location, the same in all three item columns.

In the center item column, "MEASURE", each item is placed at its item difficulty measure (in [Table 13.1](#), etc).

In the left-hand item column, "BOTTOM", the item is shown at the measure level (Rasch-Thurstone threshold) corresponding to a [probability of .5](#) of exceeding (or being rated in) the bottom rating (or partial credit) scale category.

In the right-hand item column, "TOP", the item is shown at the measure level corresponding to a probability of .5 of being rated in (or falling below) the top rating (or partial credit) scale category.

The plotted values are in the [ISFILE=](#). These locations are also shown in [Table 2.3](#).

The left-hand column locates the person ability measures along the variable. Observe that the top pupil (top . in left column and [green band](#)) is well above the top category of the most difficult act items (top XX in right-most column and top [red band](#)), but that all pupils are above the top category of the easiest item (bottom X in right-most column and [blue band](#)). "Above" here means with greater than 50% chance of exceeding.

Effective operational range of an instrument with polytomous items.

In Table 1.4, everyone between the extreme Rasch-Thurstone thresholds = 50% cumulative thresholds ([red bands](#) in the Figure) has a more than 50% chance of being rated *above* the bottom category of at least one item and *below* the top category of at least one item. Here we have polytomies (rating scale, partial credit), so each item has a range from the bottom category of the rating scale (2nd column in Table 1.4) through the overall item difficulty (3rd column in Table 1.4) to the top category of the rating scale (4th column Table 1.4).

	<p>In Table 1.4, the X's in the 2nd column are located at the points on the latent variable where a person would have a 50% chance of being observed in the bottom category and a 50% chance of being observed in a higher category. So this is like the "item difficulty" of the bottom category.</p> <p>In Table 1.4, the X's in the 4th column are located at the points on the latent variable where a person would have a 50% chance of being observed in the top category and a 50% chance of being observed in a lower category. So this is like the "item difficulty" of the top category.</p> <p>So the item difficulty of each item covers the whole range in Table 1.4 from its X in the 2nd column to its X in the 4th column.</p>
--	--

Polytomies: Tables 1.5 - 1.8 - which to report? Think: what is my message to my audience? See [Category Boundaries and Thresholds](#)

Table 1.5 - Rasch-half-point thresholds work best when you need to show average ratings at each point on the rating scale. This works well with people familiar with conventional statistical analysis.

Table 1.6 - Rasch-Thurstone thresholds are best when you need the equivalent of dichotomizing the rating scale at each category boundary = is the person above or below this boundary? - www.rasch.org/rmt/rmt233e.htm - This is used in reporting to parents in Australia.

Table 1.7 - Andrich-Thresholds are best when you need to show which individual category is most probable at each point along the latent variable. They become confusing when thresholds are [disordered](#). This is often used in clinical situations when the user needs to know what behavior is more likely to be seen.

Table 1.8 - Category centers. The locations on the latent variable at which each category is most probable = points at which the expected score on the item is the category value = a center of each category. This is useful when you need to report how the rating scale categories "step up" the latent variable.

<p>Table 1.5 Item map with expected score zones</p> <pre> MEASURE KID - MAP - ACT - Expected score zones (Rasch-half-point thresholds) <more> <rare> Neutral Like 5 # + 4 . + . T 3 # +T .## .### S+ ### S FIND BOTTLES AND CANS .05 1 ##### M+ WATCH A RAT .05 ##### WATCH BUGS .05 ##### LOOK IN SIDEWALK CRACKS .05 ##### WATCH GRASS CHANGE .05 0 ##### +M WATCH ANIMAL MOVE .05 </pre>	<p>Table 1.5 is only produced when there are polytomous items.</p> <p>The right-hand column shows the items positioned at the lower edge of each expected score zone (the half-point thresholds). The half-point threshold for category c is indicated by .c5 when ASCII=Yes or .c½ when ASCII=No or ASCII=Web. The expected score zone above item-label .05 extends from expected category score 0.5 to 1.5 at item-label .15. Lower than item-label .05 is the zone from 0 to 0.5. These correspond to the thresholds on the Expected Score ICC.</p>
--	--

<pre> WATCH BIRD MAKE NEST .15 LEARN WEED NAMES .05 GROW GARDEN .15 LOOK AT PICTURES OF PLANTS .05 LISTEN TO BIRD SING .15 LOOK UP STRANGE ANIMAL OR PLAN.05 MAKE A MAP .05 READ BOOKS ON PLANTS .05 TALK W/FRIENDS ABOUT PLANTS .05 -1 # + FIND OUT WHAT ANIMALS EAT .05 GO TO MUSEUM .15 WATCH WHAT ANIMALS EAT .05 .# T S FIND OUT WHAT FLOWERS LIVE ON .05 GO TO ZOO .15 READ ANIMAL STORIES .05 READ BOOKS ON ANIMALS .05 WATCH BIRDS .05 -2 + FIND WHERE ANIMAL LIVES .05 GO ON PICNIC .15 WATCH BIRD MAKE NEST .05 GROW GARDEN .05 LISTEN TO BIRD SING .05 -3 +T GO TO MUSEUM .05 GO TO ZOO .05 -4 + GO ON PICNIC .05 -5 + <less> <freq> Dislike Neutral </pre>	<p>If you put the item number at the start of the item labels after &END, you can show only the item numbers on this plot by using NAMLMP= or IMAP=. Column headings are the category labels that match the (rescored) category numbers in CFILE= or CLFILE=.</p> <p>Where there are more items than can be shown on one line, the extra items are printed on subsequent lines, but the latent variable " " does not advance and is left blank. The plotted values are in the ISFILE=.</p> <p>Left-hand column shows the distribution of person ability measures along the variable.</p> <p>Column labels in Tables 1.5, 1.6, 1.7, 1.8</p> <pre> PERSON= MAP ITE MORE= T RA LESS= ↓ FR </pre>
--	--

Table 1.6 Item map with 50% cumulative probabilities (Rasch-Thurstone thresholds)

MEASURE	KID - MAP - ACT - 50% Cumulative probabilities (Rasch-Thurstone thresholds)		
	<more> <rare>	Neutral	Like
5	# +		
4	. +		
	. T		FIND BOTTLES AND CANS .2
3	# +T		WATCH A RAT .2
			WATCH BUGS .2
			WATCH GRASS CHANGE .2
	##		LOOK IN SIDEWALK CRACKS .2
2	### S+		WATCH ANIMAL MOVE .2
	### S	FIND BOTTLES AND CANS .1	LEARN WEED NAMES .2
			LOOK AT PICTURES OF PLANTS .2
			LOOK UP STRANGE ANIMAL OR PLANT.2
			MAKE A MAP .2
			READ BOOKS ON PLANTS .2
			TALK W/FRIENDS ABOUT PLANTS .2
1	##### M+	WATCH A RAT .1	FIND OUT WHAT ANIMALS EAT .2
		WATCH BUGS .1	WATCH WHAT ANIMALS EAT .2
	.#####	LOOK IN SIDEWALK CRACKS .1	FIND OUT WHAT FLOWERS LIVE ON .2
		WATCH GRASS CHANGE .1	READ ANIMAL STORIES .2
			READ BOOKS ON ANIMALS .2
			WATCH BIRDS .2
0	##### +M	WATCH ANIMAL MOVE .1	FIND WHERE ANIMAL LIVES .2
			WATCH BIRD MAKE NEST .2
	.# S	LEARN WEED NAMES .1	GROW GARDEN .2
		LOOK AT PICTURES OF PLANTS .1	LISTEN TO BIRD SING .2
		LOOK UP STRANGE ANIMAL OR PLANT.1	
		MAKE A MAP .1	
		READ BOOKS ON PLANTS .1	
		TALK W/FRIENDS ABOUT PLANTS .1	
-1	# +	FIND OUT WHAT ANIMALS EAT .1	GO TO MUSEUM .2
		WATCH WHAT ANIMALS EAT .1	
	.# T S	FIND OUT WHAT FLOWERS LIVE ON .1	GO TO ZOO .2
		READ ANIMAL STORIES .1	

Table 1.6 is only produced when there are polytomous items.

The right-hand column shows the items positioned at median [50% cumulative probability](#) (the Rasch-Thurstone thresholds) at the lower edge of each rating probability zone. Above label.1, categories 1 and above are more probable. Below label.2, categories 1 and below are most probable. Between label.1 and label.2 is the zone which can be thought of as corresponding to a rating of 1. These correspond to the [Cumulative Probability](#) thresholds.

If you put the item number at the start of the item labels after &END, you can show only the item numbers on this plot by using NAMLMP= or IMAP=. Columns are headed by the (rescored) categories in CFILE=.

<pre> READ BOOKS ON ANIMALS .1 WATCH BIRDS .1 -2 + FIND WHERE ANIMAL LIVES .1 GO ON PICNIC .2 WATCH BIRD MAKE NEST .1 GROW GARDEN .1 LISTEN TO BIRD SING .1 -3 +T GO TO MUSEUM .1 GO TO ZOO .1 -4 + GO ON PICNIC .1 -5 + <less> <freq> Dislike Neutral EACH "#" IS 2: EACH "." IS 1 </pre>	<p>Where there are more items than can be shown on one line, the extra items are printed on subsequent lines, but the latent variable " " does not advance and is left blank. The plotted values are in the ISFILE=.</p> <p>Left-hand column shows the distribution of person ability measures along the variable.</p>
---	--

<p>Table 1.7 Item map with Andrich thresholds (modal categories if ordered)</p> <pre> MEASURE KID - MAP - ACT - Andrich thresholds (modal categories if ordered) <more> <rare> Neutral Like 5 . + . 4 . + . . T 3 . + . T FIND BOTTLES AND CANS .2 WATCH A RAT .2 LOOK IN SIDEWALK CRACKS .2 WATCH BUGS .2 WATCH GRASS CHANGE .2 ## .## S+ ## FIND BOTTLES AND CANS .1 LEARN WEED NAMES .2 .### S WATCH A RAT .1 LOOK AT PICTURES OF PLANTS .2 MAKE A MAP .2 READ BOOKS ON PLANTS .2 TALK W/FRIENDS ABOUT PLANTS .2 1 .#### M+ WATCH BUGS .1 FIND OUT WHAT ANIMALS EAT .2 WATCH GRASS CHANGE .1 LOOK UP STRANGE ANIMAL OR PLANT .2 WATCH WHAT ANIMALS EAT .2 .## LOOK IN SIDEWALK CRACKS .1 FIND OUT WHAT FLOWERS LIVE ON .2 ##### WATCH ANIMAL MOVE .1 READ ANIMAL STORIES .2 WATCH BIRDS .2 0 .##### +M LEARN WEED NAMES .1 FIND WHERE ANIMAL LIVES .2 READ BOOKS ON ANIMALS .2 WATCH BIRD MAKE NEST .2 # S LOOK AT PICTURES OF PLANTS .1 GROW GARDEN .2 MAKE A MAP .1 READ BOOKS ON PLANTS .1 TALK W/FRIENDS ABOUT PLANTS .1 # FIND OUT WHAT ANIMALS EAT .1 LISTEN TO BIRD SING .2 LOOK UP STRANGE ANIMAL OR PLANT .1 WATCH WHAT ANIMALS EAT .1 -1 # + # S FIND OUT WHAT FLOWERS LIVE ON .1 GO TO MUSEUM .2 READ ANIMAL STORIES .1 WATCH BIRDS .1 . T FIND WHERE ANIMAL LIVES .1 GO TO ZOO .2 READ BOOKS ON ANIMALS .1 WATCH BIRD MAKE NEST .1 -2 + GROW GARDEN .1 LISTEN TO BIRD SING .1 GO ON PICNIC .2 T -3 + GO TO MUSEUM .1 GO TO ZOO .1 -4 + GO ON PICNIC .1 -5 + <less> <freq> Dislike Neutral EACH "#" IS 2: EACH "." IS 1 </pre>	<p>Table 1.7 is only produced when there are polytomous items.</p> <p>The right-hand column shows the items positioned at their Andrich thresholds (points of equal probability of adjacent categories) at the lower edge of each rating probability zone. When the thresholds are ordered, above label.2, the most probable category is 2. Below label.1, the most probable category is 0. Between label.1 and label.2 is the zone which can be thought of as corresponding to a rating of 1.</p> <p>If you put the item number at the start of the item labels after &END, you can show only the item numbers on this plot by using NAMLMP= or IMAP=. Columns are headed by the (rescored) categories in CFILE=.</p> <p>Where there are more items than can be shown on one line, the extra items are printed on subsequent lines, but the latent variable " " does not advance and is left blank. The plotted values are in the ISFILE=.</p> <p>Left-hand column shows the distribution of person ability measures along the variable.</p>
---	---

<p>Table 1.8 Item map of measures for category scores (maximum probability of observing a category)</p>	<p>Table 1.8 is only produced when there are polytomous items.</p>
--	--

```

MEASURE      KID - MAP - ACT - Measures for category scores
              (maximum probability of observing a
              category)
              <more>|<rare> Dislike   Neutral   Like
5            X +
            X |
            |
            | Find bottles .2
            | Watch a rat .2
4            X +
            |
            | Watch bugs .2
            | Look in side .2
            | Watch grass .2
            X |
            T|
            X | Watch animal .2
3            X +
            X |T
            |
            | Learn weed n .2
            | Make a map .2
            | Talk w/frien .2
            XX | Find bottles .1 Look at pict .2
            | Look up stra .2
            | Read books o .2
            XXXX S| Watch a rat .1 Find out wha .2
            | Watch what a .2
2            XXX +
            XX | Watch bugs .1 Find out wha .2
            | Watch grass .1
            XX | Look in side .1 Read animal .2
            | Watch birds .2
            XXXX |S Read books o .2
            XXX | Watch animal .1 Find where a .2
            | Watch bird m .2
1            XXXXXX +
            XXXX M| Learn weed n .1 Grow garden .2
            XXX | Make a map .1 Listen to bi .2
            | Talk w/frien .1
            XXXXXXXX | Find bottles .0 Look at pict .1
            | Look up stra .1
            | Read books o .1
            XXXXXXXXXXXX | Watch a rat .0 Watch what a .1
0            XXXXXXXX +M Find out wha .1 Go to museum .2
            X | Watch bugs .0
            X S| Look in side .0 Find out wha .1 Go to zoo .2
            | Watch grass .0 Read animal .1
            | Watch birds .1
            XX |
            | Read books o .1
            | Watch bird m .1
-1           XX + Watch animal .0 Find where a .1 Go on picnic .2
            | Grow garden .1
            XX |S Learn weed n .0 Listen to bi .1
            | Make a map .0
            X T| Look at pict .0
            | Read books o .0
            | Talk w/frien .0
            | Look up stra .0
-2           + Find out wha .0 Go to museum .1
            | Watch what a .0
            |
            | Find out wha .0 Go to zoo .1
            | Watch birds .0
            | Read animal .0
            |T Read books o .0
-3           + Find where a .0
            | Watch bird m .0
            | Go on picnic .1
            | Grow garden .0
            | Listen to bi .0
            |
-4           +
            | Go to museum .0
            |
            | Go to zoo .0
            |
-5           +
            | Go on picnic .0
            |
            |
-6           +

```

The right-hand column shows the items positioned at the measures where the expected score on the item = the category number. This is also the measure at which the category has its highest probability. Extreme categories are shown at measures corresponding to the category score adjusted by .25 score points (or LOWADJ= and HIADJ=).

If you put the item number at the start of the item labels after &END, you can show only the item numbers on this plot by using [NAMLMP=](#) or [IMAP=](#). Columns are headed by the (rescored) categories in [CFILE=](#).

Where there are more items than can be shown on one line, the extra items are printed on subsequent lines, but the latent variable "|" does not advance and is left blank. The plotted values are in the [ISFILE=](#).

Left-hand column shows the distribution of person ability measures along the variable.

<less> <freq> Dislike Neutral Like

Table 1.9 Item map of average person measure for each category score

MEASURE	PUPIL - MAP - PUPIL - Average Measures for category scores	like	3	4
5	X + X 			
4	X + X T X X +			
3	X T XX XXXX S		WATCH GRASS CHAN.4 LOOK IN SIDEWALK.4	
2	XXX + XX XX XXXX S XXX		WATCH BUGS.4 WATCH ANIMAL MOV.4 READ BOOKS ON PL.4 TALK W/FRIENDS A.4 LOOK AT PICTURES.4 MAKE A MAP.4 WATCH WHAT ANIMA.4 LEARN WEED NAMES.4 LOOK UP STRANGE.4 FIND OUT WHAT AN.4 WATCH BIRDS.4 READ ANIMAL STOR.4 FIND OUT WHAT FL.4 FIND WHERE ANIMA.4 READ BOOKS ON AN.4 WATCH BIRD MAKE.4 LISTEN TO BIRD S.4 FIND BOTTLES AND.4 GROW GARDEN.4 GO TO MUSEUM.4 GO TO ZOO.4	
1	XXXXXX + FIND BOTTLES AND.2 XXXXX M XXX XXXXXXX XXXXXXXXX XXXXXXXXXX	GO TO ZOO.2 WATCH A RAT.2 WATCH A RAT.2 WATCH BUGS.3 FIND BOTTLES AND.3 WATCH ANIMAL MOV.3 MAKE A MAP.3 LOOK IN SIDEWALK.3 LOOK UP STRANGE.3 LEARN WEED NAMES.3 LOOK AT PICTURES.3 TALK W/FRIENDS A.3 READ BOOKS ON PL.3 WATCH WHAT ANIMA.3 FIND OUT WHAT AN.3 FIND OUT WHAT FL.3 WATCH GRASS CHAN.2 WATCH BIRDS.3 LEARN WEED NAMES.2 READ ANIMAL STOR.3 WATCH ANIMAL MOV.2	WATCH GRASS CHAN.3 WATCH A RAT.3 WATCH BUGS.3 WATCH A RAT.4 WATCH ANIMAL MOV.3 MAKE A MAP.3 LOOK IN SIDEWALK.3 LOOK UP STRANGE.3 LEARN WEED NAMES.3 LOOK AT PICTURES.3 TALK W/FRIENDS A.3 READ BOOKS ON PL.3 WATCH WHAT ANIMA.3 FIND OUT WHAT AN.3 FIND OUT WHAT FL.3 WATCH GRASS CHAN.2 WATCH BIRDS.3 LEARN WEED NAMES.2 READ ANIMAL STOR.3 WATCH ANIMAL MOV.2	GO ON PICNIC.4
0	XXXXXXX +M X X S XX	MAKE A MAP.2 TALK W/FRIENDS A.2 LOOK AT PICTURES.2 FIND OUT WHAT AN.2 LOOK UP STRANGE.2 READ BOOKS ON PL.2 READ ANIMAL STOR.2 FIND OUT WHAT FL.2 WATCH BIRD MAKE.2 READ BOOKS ON AN.2 WATCH WHAT ANIMA.2 GROW GARDEN.2 FIND WHERE ANIMA.2	READ BOOKS ON AN.3 LISTEN TO BIRD S.3 GO ON PICNIC.3 FIND WHERE ANIMA.3 GROW GARDEN.3 WATCH BIRD MAKE.3 GO TO MUSEUM.3 GO TO ZOO.3	

Table 1.9 is only produced when there are polytomous items.

The right-hand column shows the categories of the items positioned by the average of the measures of the persons who responded in each category.

If you put the item number at the start of the item labels after &END, you can show only the item numbers on this plot by using [NAMLMP=](#) or [IMAP=](#). Columns are headed by the (rescored) categories in [CFILE=](#).

Where there are more items than can be shown on one line, the extra items are printed on subsequent lines, but the latent variable "|" does not advance and is left blank. The plotted values are in the [ISFILE=](#).

Left-hand column shows the distribution of person ability measures along the variable.

<pre> WATCH BIRDS.2 -1 XX + XX S LISTEN TO BIRD S.2 X T -2 + <less> <freq> Dislike Neutral Like </pre>	
--	--

Table 1.10 Wright map - person and item (by easiness) labels on one page

MEASURE	KID - MAP - ACT
	<more> <freq>
5	M R ++ F F
4	F P ++ F S M D T
3	F M ++ Go M W T
	F B F C F S M C Go
2	F C M E M E M S M V S++ Go F R M B M K M M
	F C F H M C M H M M M P M S S Gro Lis
1	F L F T M B M B M F M L M R M R M S M++ Fin Wat M B M C M P M R M S Rea
	F D F M M A M A M B M C M D M E M H M R M R M S M S M S Fin Rea Wat
0	M A M A M D M L M M M N M P M R M R M S M V ++M Fin Wat F H M B S Loo Loo Rea
-1	F R M B Lea Mak Tal F N M L ++ Wat M J M L S
	M S T Loo Wat Wat
-2	++ Fin Wat T
-3	++ <less> <rare>

Left-hand column locates the person labels according to ability measures along the latent variable.

Right-hand column locates the item labels by their easiness measures along the latent variable. These are the difficulty measures [reversed](#) around the mean person difficulty "M". The item labels are abbreviated to fit on one page. [Table 12.12](#) has the full item labels.

Column Labels

PERSON=	MAP	ITEM=
MORE=	T	FREQ=
LESS=	↓	RARE=

Table 1.11 Wright map - person and item (by easiness) distributions

MEASURE	KID	TAP	MEASURE
<more>	----	----	<freq>
5	++ XXX	X	5
		X	
4	T++ XX	S XX	4
		X ++	
3	X ++	X	3
		XXXX S++	
2	XXXX S++	X	2
		XXXXX ++	
1	XXXXX ++		1
		M+++M	
0	XXXXXXXXXXXXX		0
		++ X	
-1	XXX		-1
		XX S++ XX	
-2	XX S++ XX		-2
		XX ++	
-3	XX ++	S X	-3
		T++	
-4	XX S X	X	-4
		X ++ XXXX	
-5	X ++ XXXX	----	-5
<less>	----	----	<rare>

This table shows the distribution of the persons and items. The variable is laid out vertically with the most able persons, and most difficult items at the top.

Left-hand column shows the distribution of person ability measures along the variable.

Right-hand column locates the item labels by their easiness measures along the latent variable. These are the difficulty measures reversed around the mean person difficulty "M".

Column Labels

PERSON=	MAP	ITEM=
MORE=	T	FREQ=
LESS=	↓	RARE=

Table 1.12 Wright map - person distribution + item (by easiness) labels

MEASURE	KID - MAP - ACT
	<more> <freq>
5	. ++ .

Left-hand column displays the distribution of person ability measures along the latent variable. The persons often have a normal distribution.

<pre> 4 . ++ . . T 3 . ++ Go on picnic . T ## Go to zoo 2 .## S++ Go to museum ## .### S Grow garden Listen to bird si 1 .#### M++ Find where animal Watch bird make n .## Read books on ani ##### Find out what flo Read animal stori Watch birds 0 .##### ++M Find out what ani Watch what animal # S Look at pictures Look up strange a Read books on pla # Learn weed names Make a map Talk w/friends ab -1 # ++ Watch animal move # S . T Look in sidewalk Watch bugs Watch grass chang -2 ++ Find bottles and Watch a rat T -3 ++ <less> <rare> EACH "#" IS 2. EACH "." IS 1. </pre>	<p>Right-hand column locates the item labels by their easiness measures along the latent variable. These are the difficulty measures reversed around the mean person difficulty "M". The item labels are abbreviated to fit on one page. Table 12.12 has the full item labels.</p> <p style="text-align: center;">Column Labels</p> <table style="margin-left: auto; margin-right: auto;"> <tr> <td>PERSON=</td> <td>MAP</td> <td>ITEM=</td> </tr> <tr> <td>MORE=</td> <td>T</td> <td>FREQ=</td> </tr> <tr> <td></td> <td> </td> <td></td> </tr> <tr> <td>LESS=</td> <td>↓</td> <td>RARE=</td> </tr> </table>	PERSON=	MAP	ITEM=	MORE=	T	FREQ=				LESS=	↓	RARE=
PERSON=	MAP	ITEM=											
MORE=	T	FREQ=											
LESS=	↓	RARE=											

<p>Table 1.13 Wright map - item (by easiness) distribution + person labels</p> <pre> MEASURE TAP - MAP - KID <freq> <more> 5 XXX ++ X 4 XX ++T XX S Ric Sus 3 ++ Fra X 2 ++S Ber Bet Els Tho X 1 ++ Bri Car Dav Joe Zul 0 M++M Ann Aud Bar Dor Jam Kim Lin Mik Ron Tra Wal Wil -1 X ++ Jan Lis Rod -2 XX ++S Bre Pet -3 ++ Ada Ric X S Bla Mar -4 ++T Don -5 XXXX ++ Hel <rare> <less> </pre>	<p>Left-hand column locates the item labels by their easiness measures along the latent variable. These are the difficulty measures reversed around the mean person difficulty "M".</p> <p>Right-hand column locates the person ability measures along the variable. The persons often have a normal distribution. The person labels are abbreviated to fit on one page. Table 16 has the full person labels</p> <p style="text-align: center;">Column Labels</p> <table style="margin-left: auto; margin-right: auto;"> <tr> <td>ITEM=</td> <td>MAP</td> <td>PERSON=</td> </tr> <tr> <td>FREQ=</td> <td>T</td> <td>MORE=</td> </tr> <tr> <td></td> <td> </td> <td></td> </tr> <tr> <td>RARE=</td> <td>↓</td> <td>LESS=</td> </tr> </table>	ITEM=	MAP	PERSON=	FREQ=	T	MORE=				RARE=	↓	LESS=
ITEM=	MAP	PERSON=											
FREQ=	T	MORE=											
RARE=	↓	LESS=											

MEASURE	Location of person or item on the unidimensional latent variable, in logits or USCALE= units
<more>	higher person "ability" = higher raw score with complete data
<less>	lower person "ability" = lower raw score with complete data
<rare>	higher item "difficulty" = lower item "easiness" = lower score on the item with complete data: PVALUE=
<frequ> <frequent>	lower item "difficulty" = higher item "easiness" = higher score on the item with complete data: PVALUE=
M	Mean of person or item distribution
S	One standard deviation from the person or item mean
T	Two standard deviations from the person or item mean
X	one person or item
#	several persons or items, e.g., EACH "#" = 4. The value for # can be changed with T1I# for items, T1P# for persons.
.	between 1 and (# - 1) persons or items, e.g., if # = 4, then "." = 1 to 3
	Variable advances. Lines without " " have the same measure as the nearest " " above.

	The double line indicates the two sides have opposite orientations: person ability and item easiness. The item difficulties are reversed . This is useful if the items and persons are being compared to the responses.
--	--

<i>On plot:</i>	<i>Description:</i>
MEASURE	Location on the unidimensional latent variable, in logits or USCALE= units
<more>	higher person "ability" = higher raw score with complete data
<less>	lower person "ability" = lower raw score with complete data
<rare>	higher item "difficulty" = lower score on the item with complete data: PVALUE=
<frequent>	lower item "difficulty" = higher score on the item with complete data: PVALUE=
M	Mean of person or item distribution
S	One standard deviation from the person or item mean
T	Two standard deviations from the person or item mean
X	one person or item
#	several persons or items, e.g., "#" = 4. The value for # can be changed with T1#=# , T1P#=# .
.	between 1 and #-1 persons or items, e.g., "." = 1 to 3
or +	Variable advances. Lines without " " or "+" have the same measure as the nearest " " or "+" above.

Example: How can we use a Wright item-person map to investigate a test of multiple-choice (MCQ) items?

Please look at your item map. Rasch item maps show the distribution of the item difficulties in logits (equivalent to item p-values) aligned with the distribution of person abilities in logits (equivalent to raw scores). The easiest items are at the bottom (or left-hand side) of the item map. The hardest items are at the top (or right-hand side) of the item map. The most able students are at the top (or right-hand side) of the item map. The least able students are at the bottom (or left-hand side) of the item map.

1. We should have "content" validity for our items. Please put the content-area into the item label for each item, so that you can see the content-areas on the item map. Do the content areas for the items on the item map match your intention such as the curriculum? For example, are there "geography" items in your item map of "arithmetic" items?

2. We should have "construct" validity for our items. When you look at the ordering (hierarchy) of item difficulties, do their content-areas match your intention? For example, in an item map of "arithmetic" items, we expect to see that the "addition" items are the easiest items at the bottom of the map, and the "division" items are the hardest items at the top of the map, with "subtraction" and "multiplication" generally between "addition" and "division".

3. For most sets of items used in school situations, we want the success of most students to be between about 50% and 90%. This is from about 0 logits on the item map (which is the mean difficulty, average difficulty, of the items) up to about +2 logits on the item map.

If much of the person ability distribution is below 0 logits, then the set of items is too hard. You need some easier items if you want to measure the persons accurately. If much of the person ability distribution is about +2 logits, then the set of items is too easy. You need some harder items if you want to measure the persons accurately.

4. We want all the items to contribute to useful measurement of the persons. If some items on the item map are more than 1 logit below almost all of the persons, then those items may be too easy. If some items on the map are above almost all of the persons, then those items may be too hard. Items that are too easy or too hard do not contribute to useful measurement, and may provoke the students into guessing or carelessness.

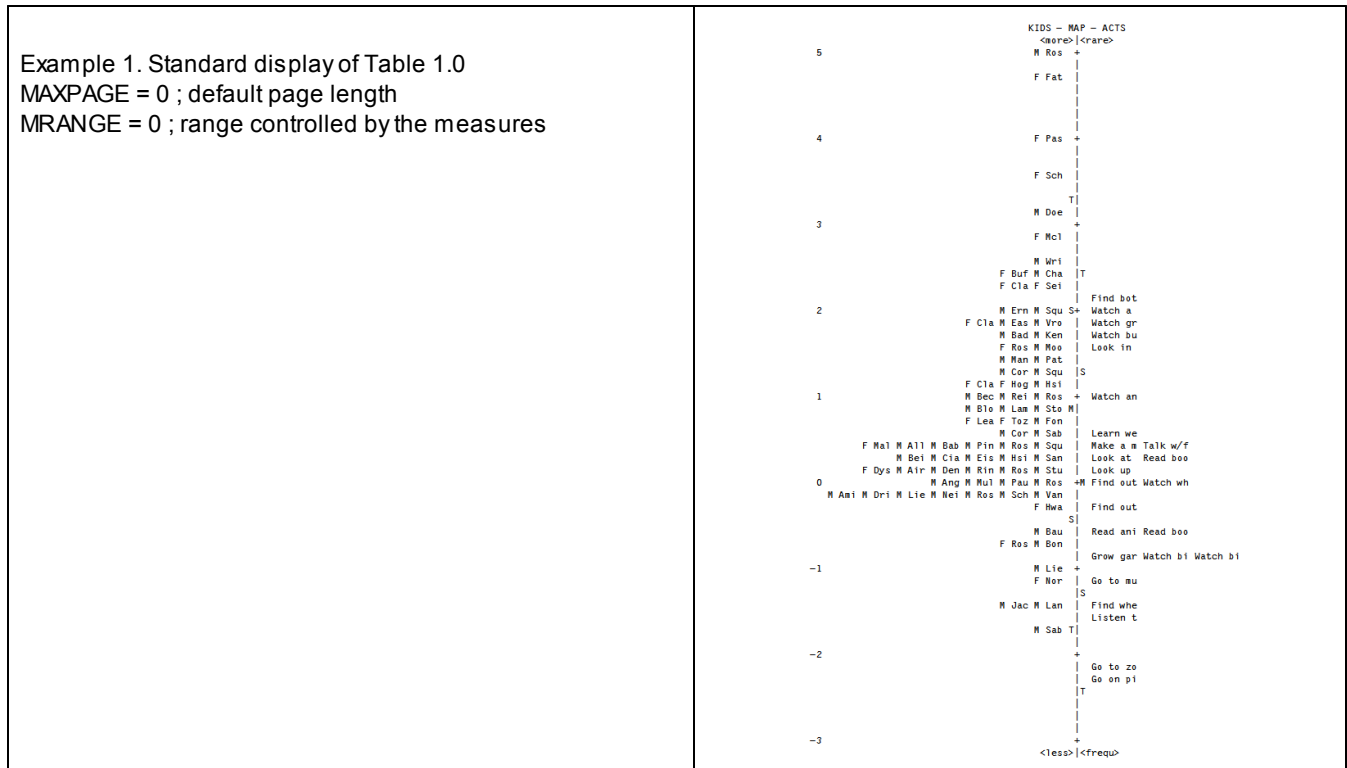
5. We want to measure all the students with good precision. For this we need the item difficulties to have a roughly uniform distribution. If there are gaps of more than 0.5 logits in the distribution of item difficulties, then items are needed to fill these gaps. If there are rows on the item maps with many items, then there are many items with the about the same difficulty. These items may be duplicative.

Prettifying the Wright item-person map

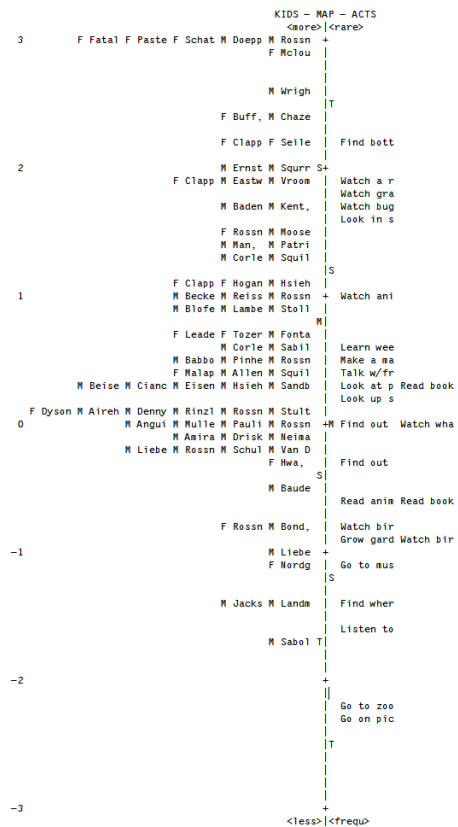
Get the map to look how you want on your screen by adjusting font sizes, type faces, [MRANGE=](#), [MAXPAGE=](#), etc., then use your PrintScreen key to make a copy of the screen. Paste into Word, PowerPoint, etc. These Tables can be copied-and-pasted into Excel spreadsheets, and then Excel "Text to columns" used to beautify the output. A more elaborate procedure is at [Excel item map](#). Or get the map to look how you want on your screen by adjusting font sizes, type faces, [MRANGE=](#), [MAXPAGE=](#), etc., then use your PrintScreen key to make a copy of the screen. Paste into Word, PowerPoint, etc.

Winsteps can plot one group of persons on a Wright map. To plot two groups of persons onto one Wright map,

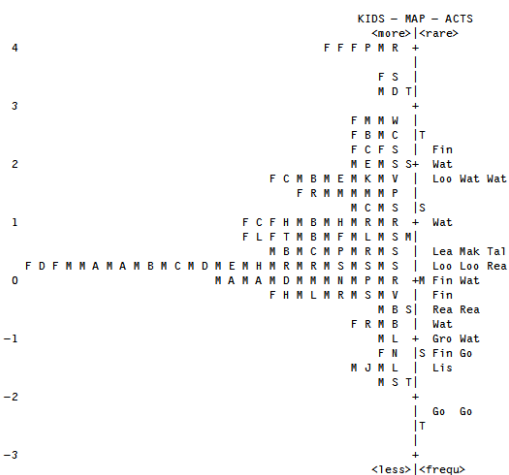
1. Plot one group of persons on the Wright map.
2. Copy the Wright map into a Word document (courier font)
3. Plot the second group of person on the Wright map.
4. Copy the Wright map into a Word document (courier font)
5. Rectangular-copy (alt+mouse) the person distribution of the second group into the Wright map of the first group.



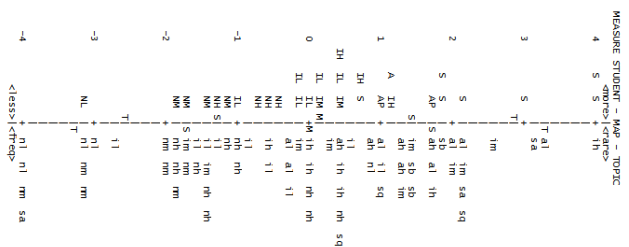
Example 2. Symmetric display around the item mean of Table 1.0
 MAXPAGE = 0 ; default page length
 MRANGE = -3 ; range away from item mean



Example 3. Custom display of Table 1.0
 MAXPAGE=50 ; pre-set page length
 TFILE=* ; in the control file
 1 -3 4 ; Table 1 with display from -3 to +4 logits.
 *



Example 4. A horizontal Wright map.
 1) Output Table 1.0
 2) Copy-and-Paste into a Word document
 3) Document "New Section" before and after map
 4) Highlight the map:
 5) Fixed space font: Consolas, Lucida Console
 6) Small font size
 7) Alt-F11 to launch VBA
 8) in Immediate Window:
 Selection.Orientation = wdTextOrientationVerticalFarEast
 9) Press Enter
 Text is rotated!



14.3 Table 2 Category plots

14.3.1 Table 2 Multiple-choice distractor plot

(controlled by [T2SELECT=](#), [MRANGE=](#))

The codes for the response options (distractors) are located according to the measures corresponding to them. Each subtable is presented two ways: with the response code itself (or one of them if several would be in the same place), e.g., Table 2.1, and with the score corresponding to the option, e.g. Table 2.11 (numbered 10 subtables higher).

[Table 2](#) for polytomous items.

[Table 2.1](#): shows the most probable response on the latent variable. In this example, for item "a107", "a" (or any other incorrect option) is most probable up to 3.2 logits, when "d", the correct response, becomes most probable according to the Rasch model.

TABLE 2.1: MOST PROBABLE RESPONSE: MODE (BETWEEN "0" AND "1" IS "0", ETC.) (ILLUSTRATED BY AN OBSERVED CATEGORY)

-4	-3	-2	-1	0	1	2	3	4	NUM	TOPIC

a								d	55	a107 newspaper
a								c	64	sa01 magazine
.....										
b	a							a	12	nm07 sign on wall
a	d							d	10	nm05 public place

-4	-3	-2	-1	0	1	2	3	4	NUM	TOPIC
		1	11 1111	111 212	3 2 12	12 1	1	2	STUDENTS	
			S	M	S				T	
		0	10 20	30 40	50 60	70 80	90	99	PERCENTILE	

M = Mean, the average of the person measures, S = One Standard Deviation from the mean, T = Two P.SDs. from the mean. Percentile is percentage below the specified position.

[Table 2.11](#) is the same as Table 2.1, but the options are shown by their scored values, not by their codes in the data.

TABLE 2.11: MOST PROBABLE RESPONSE: MODE (BETWEEN "0" AND "1" IS "0", ETC.) (BY CATEGORY SCORE)

-4	-3	-2	-1	0	1	2	3	4	NUM	TOPIC

0								1	1	55 a107 newspaper
0								1	1	64 sa01 magazine

[Table 2.2](#): shows the predicted average response on the latent variable. In this example, for item "a107", "a" (or any other incorrect option) is the predicted average response up to 3.2 logits, then "d", the correct response, becomes the average predictions. The ":" is at the transition from an average expected wrong response to an average expected "right" response, i.e., where the predicted average score on the item is 0.5, the Rasch-half-point thresholds. The "a" below "2" is positions where the expected average score on the item is 0.25. Similarly "d" would be repeated where the expected average score on the item is 0.75, according to the Rasch model.

TABLE 2.2 EXPECTED SCORE: MEAN (":" INDICATES HALF-POINT THRESHOLD) (ILLUSTRATED BY AN OBSERVED CATEGORY)

-4	-3	-2	-1	0	1	2	3	4	NUM	TOPIC

a					a	:	d	55	a107 newspaper	
a					a	:	c	64	sa01 magazine	

[Table 2.12](#) is the same as Table 2.2, but the options are shown by their scored values, not by their codes in the data.

TABLE 2.12 EXPECTED SCORE: MEAN (":" INDICATES HALF-POINT THRESHOLD) (BY CATEGORY SCORE)

-4	-3	-2	-1	0	1	2	3	4	NUM	TOPIC

0						0	:	1	55	a107 newspaper
0						0	:	1	64	sa01 magazine

[Table 2.3](#) shows the Rasch-Thurstone thresholds = 50% cumulative probability points. The lower category ("a" and other wrong answers) has a greater than 50% probability of being observed up to 3.2 logits, when "d", the correct answer, has a higher than 50% probability.

TABLE 2.3 50% CUMULATIVE PROBABILITY (RASCH-THURSTONE THRESHOLD): MEDIAN (ILLUSTRATED BY AN OBSERVED CATEGORY)

	-4	-3	-2	-1	0	1	2	3	4			
	-----+-----+-----+-----+-----+-----+-----+-----+-----									NUM	TOPIC	
a									d	d	55	al07 newspaper
a								c	c	64	sa01 magazine	

[Table 2.13](#) is the same as Table 2.3, but the options are shown by their scored values, not by their codes in the data.

TABLE 2.13 50% CUMULATIVE PROBABILITY (RASCH-THURSTONE THRESHOLD): MEDIAN (BY CATEGORY SCORE)

	-4	-3	-2	-1	0	1	2	3	4			
	-----+-----+-----+-----+-----+-----+-----+-----+-----									NUM	TOPIC	
0									1	1	55	al07 newspaper
0								1	1	64	sa01 magazine	

[Table 2.4](#) shows the item difficulties (or more generally the Rasch-Andrich thresholds) coded by the option of the higher category. For item "al07" this is "d", the correct option.

TABLE 2.4 STRUCTURE MEASURES (RASCH-ANDRICH THRESHOLDS: equal-adjacent-probability thresholds) (ILLUSTRATED BY AN OBSERVED CATEGORY)

	-4	-3	-2	-1	0	1	2	3	4			
	-----+-----+-----+-----+-----+-----+-----+-----+-----									NUM	TOPIC	
									d		55	al07 newspaper
								c		64	sa01 magazine	

[Table 2.14](#) is the same as Table 2.4, the Rasch-Andrich thresholds, but the options are shown by their scored values, not by their codes in the data.

TABLE 2.14 STRUCTURE MEASURES (RASCH-ANDRICH THRESHOLDS: equal-adjacent-probability thresholds) (BY CATEGORY SCORE)

	-4	-3	-2	-1	0	1	2	3	4			
	-----+-----+-----+-----+-----+-----+-----+-----+-----									NUM	TOPIC	
									1		55	al07 newspaper
								1		64	sa01 magazine	

[Table 2.5](#) shows the [average measures](#) of persons choosing wrong distractors (illustrated by one of the wrong distractors, "a") and the average measures of persons choosing a correct distractor (illustrated by one of the correct distractors, "d").

TABLE 2.5 OBSERVED AVERAGE MEASURES FOR STUDENTS (scored) (ILLUSTRATED BY AN OBSERVED CATEGORY)

	-4	-3	-2	-1	0	1	2	3	4			
	-----+-----+-----+-----+-----+-----+-----+-----+-----									NUM	TOPIC	
					a				d	55	al07 newspaper	
						a		c		64	sa01 magazine	

[Table 2.15](#) is the same as Table 2.5, but the options are shown by their scored values, not by their codes in the data.

TABLE 2.15 OBSERVED AVERAGE MEASURES FOR STUDENTS (scored) (BY CATEGORY SCORE)

	-4	-3	-2	-1	0	1	2	3	4			
	-----+-----+-----+-----+-----+-----+-----+-----+-----									NUM	TOPIC	
					0				1	55	al07 newspaper	
						0		1		64	sa01 magazine	

[Table 2.6](#), shown first from the Diagnosis menu, shows the average measures from [Table 14.3](#) of the persons choosing each distractor. "m" usually indicates the average measure of persons with missing data. Table 2.6 shows the average ability of the group of examinees who chose each option. Table 2.15 (above) shows the average ability of the people who got an item right and wrong. Table 2.5 (above again) has the right and wrong scoring illustrated with specific MCQ options. If there are more than one correct or incorrect option, only one of each is shown.

TABLE 2.6 OBSERVED AVERAGE MEASURES FOR STUDENTS (unscored) (BY OBSERVED CATEGORY)

	-4	-3	-2	-1	0	1	2	3	4			
	-----+-----+-----+-----+-----+-----+-----+-----+-----									NUM	TOPIC	
				m	ab	c			d	55	al07 newspaper	
				m			d		c	64	sa01 magazine	

Code for unidentified missing data: m

[Table 2.16](#) is the same as Table 2.6, but the options are shown by their scored values, not by their codes in the data.

TABLE 2.16 OBSERVED AVERAGE MEASURES FOR STUDENTS (unscored) (BY CATEGORY SCORE)

-4	-3	-2	-1	0	1	2	3	4			
-----+-----+-----+-----+-----+-----+-----+-----+-----									NUM	TOPIC	
				m	00	0			1	55	a107 newspaper
				m		0			1	64	sa01 magazine

[Table 2.7](#) shows the measures that would be predicted to be observed for incorrect and correct responses if the persons responded exactly as the Rasch model predicts. "a" (an incorrect distractor) shows the average measure for persons in the sample who would be predicted to fail the item, and "d" (a correct distractor) shows the average measure for persons in the sample who would be predicted to succeed on the item.

TABLE 2.7 EXPECTED AVERAGE MEASURES FOR STUDENTS (scored) (ILLUSTRATED BY AN OBSERVED CATEGORY)

-4	-3	-2	-1	0	1	2	3	4			
-----+-----+-----+-----+-----+-----+-----+-----+-----									NUM	TOPIC	
					a				d	55	a107 newspaper
						a			c	64	sa01 magazine

[Table 2.17](#) is the same as Table 2.7, but the options are shown by their scored values, not by their codes in the data.

TABLE 2.17 EXPECTED AVERAGE MEASURES FOR STUDENTS (scored) (BY CATEGORY SCORE)

-4	-3	-2	-1	0	1	2	3	4			
-----+-----+-----+-----+-----+-----+-----+-----+-----									NUM	TOPIC	
				0					1	55	a107 newspaper
						0			1	64	sa01 magazine

14.3.2 Table 2 Polytomies: Most probable, expected, cumulative, structure, average measures

(controlled by [T2SELECT=](#), [MRANGE=](#), [CATREF=](#))

[Table 2](#) for multiple-choice items.

Each plot answers a different question:

What category is most likely? The maximum probability (mode) plot.

What is the average or expected category value? The expected score (mean) plot.

What part of the variable corresponds to the category? The cumulative probability (median) plot.

The numeric information for these plots is in [ISFILE=](#)

Which Table should be used for a standard setting procedure?

Most standard setting is based on "average" or "frequency" considerations. For instance,

"If we observed 1000 candidates whose measures are known to be exactly at the pass-fail point, ...

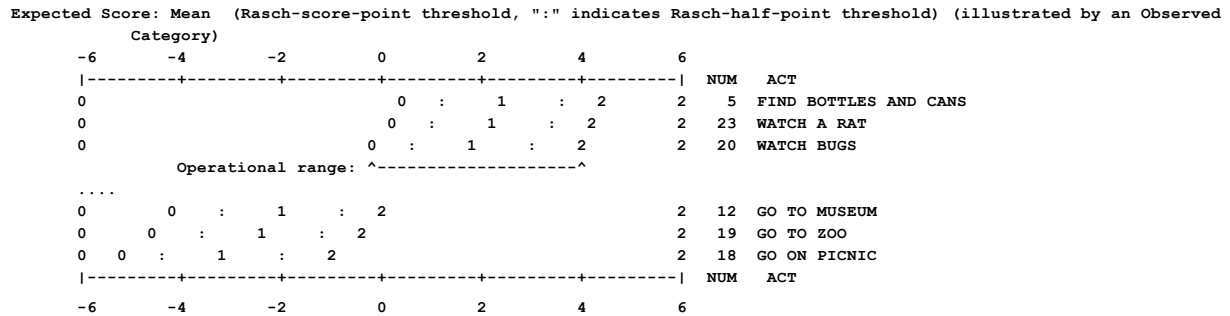
..., we would expect their average score to be the pass-fail score." If this is how you think, then the Table you want is Table 2.2 (matches Table 12.5)

..., we would expect 50% to pass and 50% to fail the pass-fail score." If this is how you think, then the Table you want is Table 2.3 (matches 12.6)

..., we would expect more to be in the criterion pass-fail category of each item than any other category." If this is how you think, then the Table you want is Table 2.1 (no matching 12.)

"Our current sample is definitive, ...

Table 2.2 & Table 2.12: In the "Expected Score" Table, the standard output (or selected with `CURVES=010`) answers the question "what is the average rating that we expect to observe for persons of a particular measure?" This rating information is expressed in terms of expected scores (with ":" at the half-point thresholds). Extreme scores are located at expected scores .25 score points away from the extremes. This plot also indicates the operational range of each item.



In this example of a 5 category rating scale, the operation range of the top item is indicated in red.



Table 2.3 & Table 2.13: The "Cumulative Probability" Table: Rasch-Thurstone thresholds, selected with `CURVES=001`, answers the question "whereabouts in the category ordering is a person of a particular measure located?" This information is expressed in terms of median cumulative probabilities (the point at which the probability of scoring in or above the category is .5).

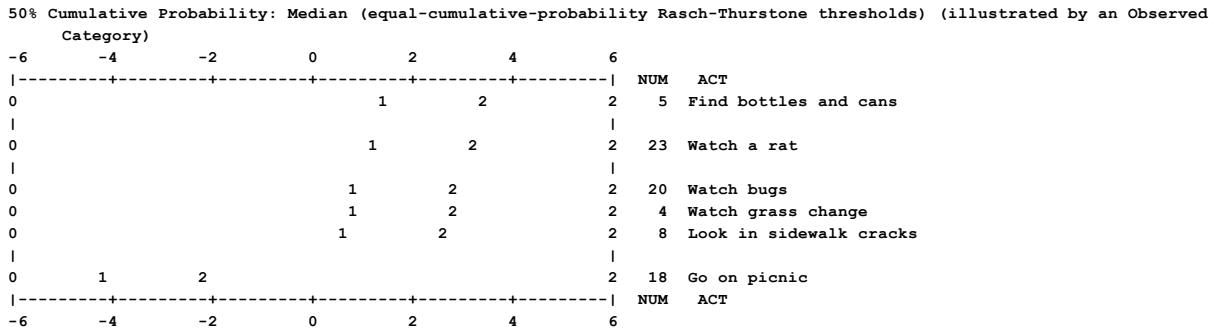
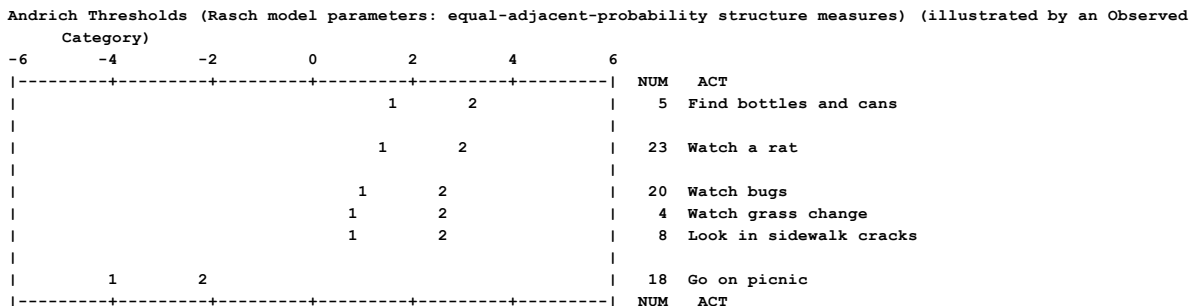
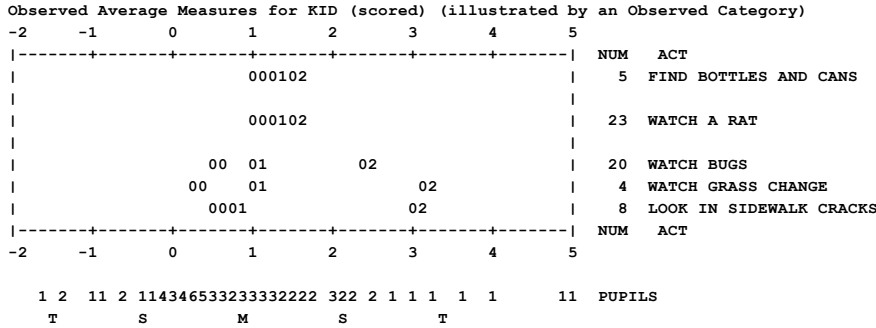


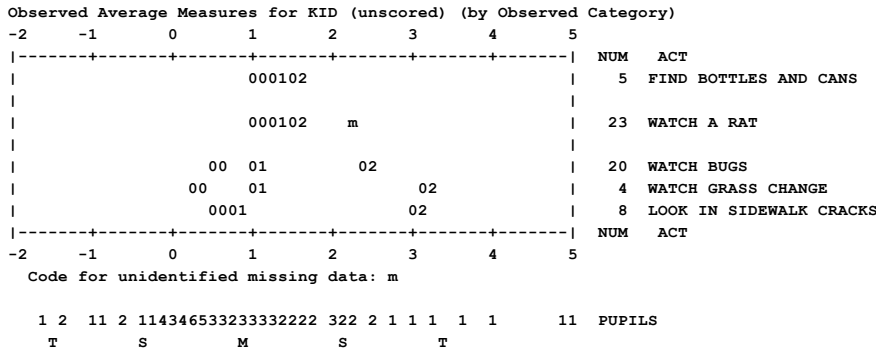
Table 2.4 & Table 2.14 show Rasch structure calibrations: Rasch-Andrich thresholds (step parameters, step measures, step difficulties, rating (or partial credit) scale calibrations). These are the relationships between adjacent categories, and correspond to the points where adjacent category probability curves cross, i.e., are equally probable of being observed according to a Rasch model.



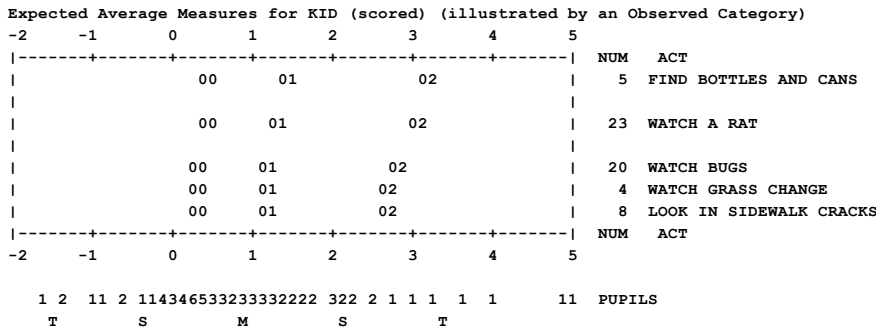
[Table 2.5](#) & [Table 2.15](#) plot the observed average person measures for each scored category. It reflects how this sample used these categories. The plotted values cannot fall outside the range of the sample.



[Table 2.6](#) & [Table 2.16](#) plot the observed average person measures for each observed category. It reflects how this sample used these categories. The plotted values cannot fall outside the range of the sample. "m" in the plot indicates the average measure of those for whom their observation is missing on this item. This Table is shown first from the Diagnosis pull-down menu.



[Table 2.7](#) & [Table 2.17](#) plot the expected average person measures for each category score. It reflects how this sample were expected to use these categories. The plotted values cannot fall outside the range of the sample. This Table applies the empirical person distribution to Table 2.2.



14.3.3 Table 2.1 Modal observed categories (most probable categories)

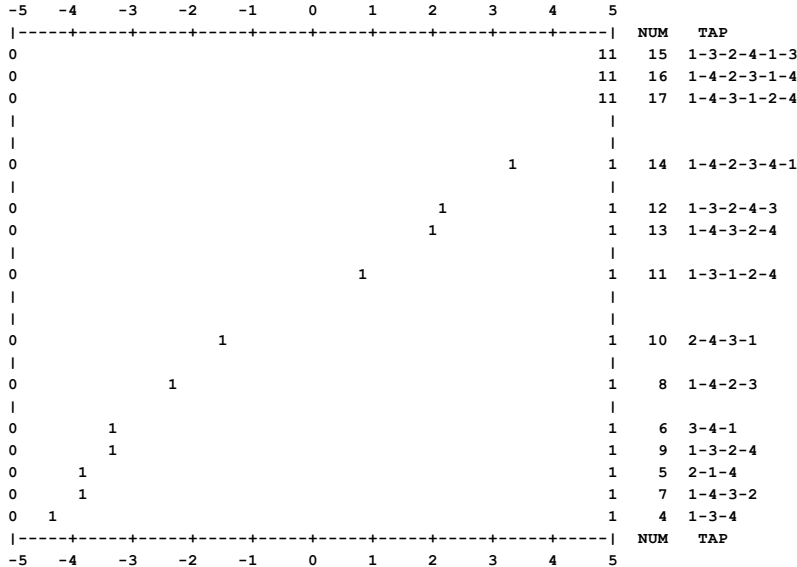
(controlled by [T2SELECT=](#), [MRANGE=](#), [MTICK=](#), [CATREF=](#))

Table 2 for multiple-choice items.

Table 2 for polytomous items.

Table 2.1: shows the most probable response on the latent variable, relative to each of the other responses. The most probable response for an item (row) is the category number shown to the left of the desired location on the x-axis. The items are shown vertically by difficulty distribution. The person distribution is shown below the plot.

MOST PROBABLE RESPONSE: MODE (BETWEEN "0" AND "1" IS "0", ETC.) (ILLUSTRATED BY AN OBSERVED CATEGORY)

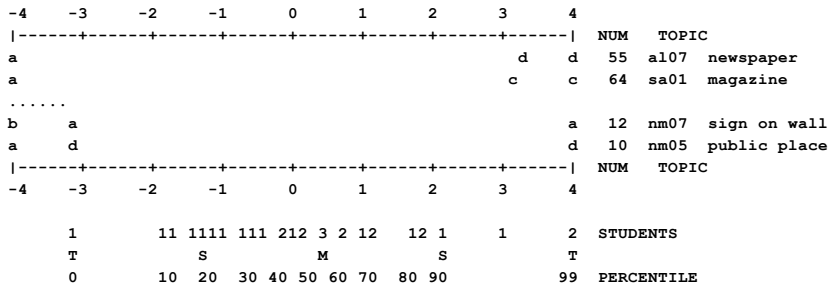


NUM	TAP
1	KID
10	PERCENTILE
20	
30	
40	
50	
60	
70	
80	
90	
99	

Multiple-choice items

this example, for item "al07", "a" (or any other incorrect option) is most probable up to 3.2 logits, when "d", the correct response, becomes most probable according to the Rasch model.

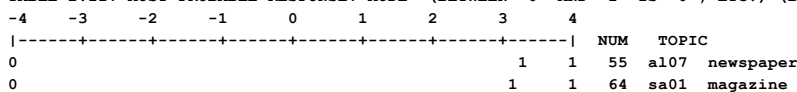
TABLE 2.1: MOST PROBABLE RESPONSE: MODE (BETWEEN "0" AND "1" IS "0", ETC.) (ILLUSTRATED BY AN OBSERVED CATEGORY)



M= Mean, the average of the person measures, S = One Standard Deviation from the mean, T = Two P.SDs. from the mean. Percentile is percentage below the specified position.

Table 2.11 is the same as Table 1, but the options are shown by their scored values, not by their codes in the data.

TABLE 2.11: MOST PROBABLE RESPONSE: MODE (BETWEEN "0" AND "1" IS "0", ETC.) (BY CATEGORY SCORE)



Polytomous items

Tables 2.1, The "Most Probable Response" Table, selected with CURVES=001, answers the question "which category is a person of a particular measure most likely to choose?" This is the most likely category with which the persons of logit (or user-rescaled) measure shown below would respond to the item shown on the left. The area to the extreme left is all "0"; the area to the extreme right is at the top category. Each category number is shown to the left of its modal area. If a category is not shown, it is never a most likely response. An item with an extreme, perfect (maximum possible) or zero (minimum possible), score is not strictly estimable, and is omitted here. Blank lines are used to help approximate the placement of the items on the latent variable.

This table presents in one picture the results of this analysis in a form suitable for inference. We can predict for people of any particular measure what responses they would probably make. "M" depicts an "average" person. The left "T" a low performer. The right "T" a high performer. Look straight up from those letters to read off the expected response profiles.

Table 2.1 to 2.7 reports with observed categories, i.e., those in the CODES= statement.

Table 2.11 to 2.17 report with scored categories, i.e., after IVALUE=, RESCORE=, KEY1=, etc., if any.

```

MOST PROBABLE RESPONSE: MODE (BETWEEN "0" AND "1" IS "0", ETC.)
NUM  ITEM          -5    -4    -3    -2    -1    0    1    2    3    4    5
-----
  5  FIND BOTTLES  0                1        2        2
 20  WATCH BUGS   0                1        2        2
  8  LOOK IN SIDE  0                1        2        2
  7  WATCH ANIMAL  0                1        2        2
 17  WATCH WHAT A  0                1        2        2
deliberate space |
 21  WATCH BIRD M  0                1        2        2
 10  LISTEN TO BI  0                1        2        2
 12  GO TO MUSEUM  0                1        2        2
 18  GO ON PICNIC  0                1        2        2
-----
-5    -4    -3    -2    -1    0    1    2    3    4    5
1
PERSON          1  2 1121 17413533622232221 11 1 1 11
                T  S  M  S  T

```

14.3.4 Table 2.2 Mean observed categories (average or expected categories: Rasch-half-point thresholds)

(controlled by T2SELECT=, MRANGE=, CATREF=)

Table 2 for multiple-choice items.

Table 2 for polytomous items.

```

EXPECTED SCORE: MEAN (Rasch-score-point threshold, ":" indicates Rasch-half-point threshold) (ILLUSTRATED BY AN OBSERVED CATEGORY)
-5 -4 -3 -2 -1 0 1 2 3 4 5
|-----| NUM TAP
0                0 :1 15 1-3-2-4-1-3
0                0 :1 16 1-4-2-3-1-4
0                0 :1 17 1-4-3-1-2-4
|
|
0                0 : 1 1 14 1-4-2-3-4-1
|
0                0 : 1 1 12 1-3-2-4-3
0                0 : 1 1 13 1-4-3-2-4
|
0                0 : 1 1 11 1-3-1-2-4
|
|
0                0 : 1 1 10 2-4-3-1
|
0                0 : 1 1 8 1-4-2-3

```

-----+-----+-----+-----+-----+-----+-----+-----+-----+-----										NUM	TAP	
-5	-4	-3	-2	-1	0	1	2	3	4	5		
1	1	2	2	2	3	2	5	4	1	2	KID	
T		S			M		S		T			
0	10	20	30	60	80	90	99				PERCENTILE	

Table 2.2 is used for inference:

Past-sample-dependent inference: If you expect the future person to be from a person sample exactly like the past sample (distribution, ability level, misfit, etc.) , then base your decisions on [Table 2.7](#).

Past-sample-distribution-dependent inference: If you expect the future person to be from a person sample similar to the past sample (distribution, ability level, but not misfit) , then base your decisions on [Table 2.5](#).

Past-sample-independent-as-possible: If you have no information about the future person, except that the latent variable will be the same as for the past sample, then base your decisions on Table 2.2.

Does the future repeat the past? Or is the future something entirely new? The truth is usually somewhere between these extremes.

Multiple-choice items

Table 2.2: shows the predicted average response on the latent variable. In this example, for item "a107", "a" (or any other incorrect option) is the predicted average response up to 3.2 logits, then "d", the correct response, becomes the average predictions. The ":" is at the transition from an average expected wrong response to an average expected "right" response, i.e., where the predicted average score on the item is 0.5, the Rasch-half-point thresholds. The "a" below "2" is positions where the expected average score on the item is 0.25. Similarly "d" would be repeated where the expected average score on the item is 0.75, according to the Rasch model.

TABLE 2.2 EXPECTED SCORE: MEAN (":" INDICATES HALF-POINT THRESHOLD) (ILLUSTRATED BY AN OBSERVED CATEGORY)

-4	-3	-2	-1	0	1	2	3	4	NUM	TOPIC
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----										
a						a	:	d	55	a107 newspaper
a						a	:	c	64	sa01 magazine

Table 2.12 is the same as Table 2, but the options are shown by their scored values, not by their codes in the data.

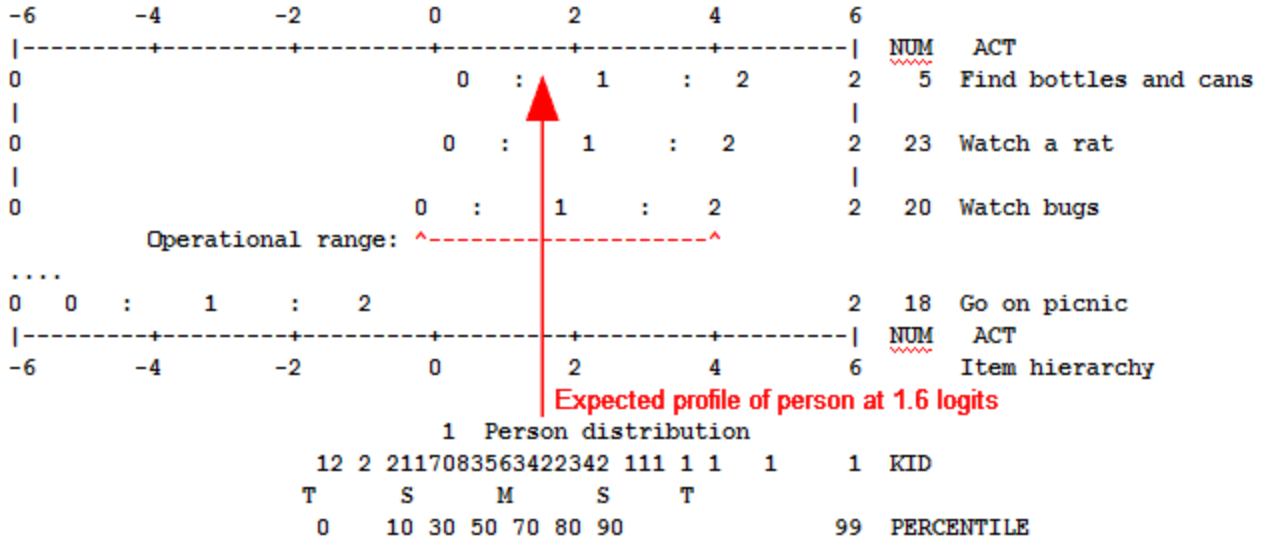
TABLE 2.12 EXPECTED SCORE: MEAN (":" INDICATES HALF-POINT THRESHOLD) (BY CATEGORY SCORE)

-4	-3	-2	-1	0	1	2	3	4	NUM	TOPIC
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----										
0						0	:	1	55	a107 newspaper
0						0	:	1	64	sa01 magazine

Polytomous items

Tables 2.2 & 2.12: In the "Expected Score" Table, the standard output (or selected with [CURVES=010](#)) answers the question "what is the average rating that we expect to observe for persons of a particular measure?" This rating information is expressed in terms of expected scores (with ":" at the half-point thresholds). Extreme scores are located at expected scores .25 score points away from the extremes. This plot also indicates the operational range of each item.

EXPECTED SCORE: MEAN (Rasch-score-point threshold, ":" indicates Rasch-half-point threshold)
 (ILLUSTRATED BY AN OBSERVED CATEGORY)



In this example of a 5 category rating scale, the operation range of the top item is indicated in red.



Example: From a paper by B.D. Wright - www.rasch.org/rmt/rmt141a.htm

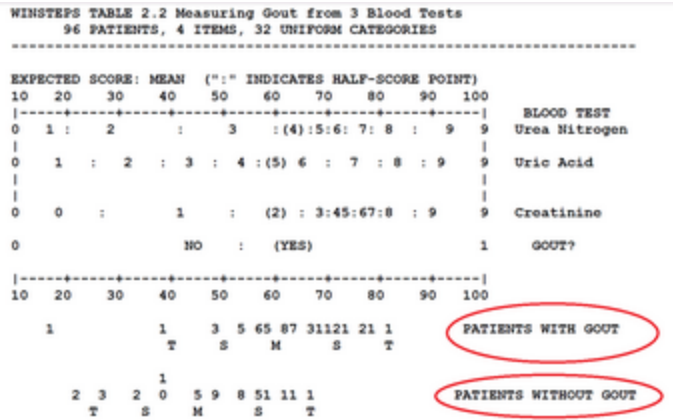


Figure 2. The Complete Score on One Pair

Here is what Ben Wright did:

1. Put a code for Gout/No-Gout in every person record
2. Do the Winsteps analysis for everyone.
3. In the Specification menu box, **PSELECT=** Gout produce Table 2.2
4. In the Specification menu box, **PSELECT=** to restore everyone
5. In the Specification menu box, **PSELECT=** No-Gout produce Table 2.2
6. Copy-and-paste the person distribution from 5. onto the bottom of 3.

14.3.5 Table 2.3 Median observed categories (cumulative category probabilities: Rasch-Thurstone thresholds) (controlled by T2SELECT=, MRANGE=, CATREF=)

[Table 2](#) for multiple-choice items.
[Table 2](#) for polytomous items.

Multiple-choice items

Table 2.3 shows the 50% cumulative probability points, the Rasch-Thurstone thresholds. The lower category ("a" and other wrong answers) has a greater than 50% probability of being observed up to 3.2 logits, when "d", the correct answer, has a higher than 50% probability.

TABLE 2.3 50% CUMULATIVE PROBABILITY (RASCH-THURSTONE THRESHOLD) (ILLUSTRATED BY AN OBSERVED CATEGORY)

	-4	-3	-2	-1	0	1	2	3	4			
	-----+-----+-----+-----+-----+-----+-----+-----+-----									NUM	TOPIC	
a									d	d	55	a107 newspaper
a								c	c	64	sa01 magazine	

Table 2.13 is the same as Table 3, but the options are shown by their scored values, not by their codes in the data.

TABLE 2.13 50% CUMULATIVE PROBABILITY (RASCH-THURSTONE THRESHOLD) (BY CATEGORY SCORE)

	-4	-3	-2	-1	0	1	2	3	4			
	-----+-----+-----+-----+-----+-----+-----+-----+-----									NUM	TOPIC	
0								1	1	55	a107 newspaper	
0								1	1	64	sa01 magazine	

Polytomous items

Tables 2.3 & 2.13: The "Cumulative Probability" Table: Rasch-Thurstone thresholds, selected with CURVES=001, answers the question "whereabouts in the category ordering is a person of a particular measure located?" This information is expressed in terms of median cumulative probabilities (the point at which the probability of scoring in or above the category is .5).

50% CUMULATIVE PROBABILITY: MEDIAN (equal-cumulative-probability Rasch-Thurstone thresholds) (ILLUSTRATED BY AN OBSERVED CATEGORY)

	-6	-4	-2	0	2	4	6				
	-----+-----+-----+-----+-----+-----+-----							NUM	ACT		
0					1	2	2	5	FIND BOTTLES AND CANS		
0					1	2	2	23	WATCH A RAT		
0					1	2	2	20	WATCH BUGS		
0					1	2	2	4	WATCH GRASS CHANGE		
0					1	2	2	8	LOOK IN SIDEWALK CRACKS		
	-----+-----+-----+-----+-----+-----+-----							NUM	ACT		
	-6	-4	-2	0	2	4	6				
					1						
				12	2	2117083563422342	111	1	1	1	KID
				T	S	M	S	T			
				0	10	30	50	70	80	90	99 PERCENTILE

14.3.6 Table 2.4 Andrich thresholds (structure calibrations)

(controlled by T2SELECT=, MRANGE=, CATREF=)

[Table 2](#) for multiple-choice items.
[Table 2](#) for polytomous items.

Multiple-choice items

Table 2.4 shows the item difficulties (or more generally the Rasch-Andrich thresholds) coded by the option of the higher category. For item "a107" this is "d", the correct option.

```

Andrich Thresholds (Rasch model parameters: equal-adjacent-probability structure measures) (ILLUSTRATED BY AN OBSERVED CATEGORY)
-4  -3  -2  -1  0  1  2  3  4
|-----+-----+-----+-----+-----+-----+-----+-----+-----| NUM  TOPIC
|                                     | d   | 55  al07  newspaper
|                                     | c   | 64  sa01  magazine

```

Polytomous items

Table 2.4 shows the Rasch structure calibrations which are the Rasch-Andrich thresholds, also called step parameters, step measures, step difficulties, rating (or partial credit) scale calibrations. These are the relationships between adjacent categories, and correspond to the points where adjacent category probability curves cross, i.e., are equally probable of being observed according to a Rasch model.

```

Andrich Thresholds (Rasch model parameters: equal-adjacent-probability structure measures) (ILLUSTRATED BY AN OBSERVED CATEGORY)
-6  -4  -2  0  2  4  6
|-----+-----+-----+-----+-----+-----+-----+-----| NUM  ACT
|                                     | 5   | FIND BOTTLES AND CANS
|                                     | 23  | WATCH A RAT
|                                     | 18  | GO ON PICNIC
|-----+-----+-----+-----+-----+-----+-----+-----| NUM  ACT
-6  -4  -2  0  2  4  6
|                                     | 1   | KID
|                                     | 12  | 2117083563422342
|                                     | T   | S   M   S   T
|                                     | 0   | 10 30 50 70 80 90
|                                     | 99  | PERCENTILE

```

14.3.7 Table 2.5 Observed average measures of persons by scored category (empirical averages)

(controlled by [T2SELECT=](#), [MRANGE=](#), [CATREF=](#))

[Table 2](#) for multiple-choice items.

[Table 2](#) for polytomous items.

```

OBSERVED AVERAGE MEASURES FOR KID (scored) (ILLUSTRATED BY AN OBSERVED CATEGORY)
-5  -4  -3  -2  -1  0  1  2  3  4  5
|-----+-----+-----+-----+-----+-----+-----+-----| NUM  TAP
|                                     | 15  | 1-3-2-4-1-3
|                                     | 16  | 1-4-2-3-1-4
|                                     | 17  | 1-4-3-1-2-4
|                                     | 14  | 1-4-2-3-4-1
|                                     | 12  | 1-3-2-4-3
|                                     | 13  | 1-4-3-2-4
|                                     | 11  | 1-3-1-2-4
|                                     | 10  | 2-4-3-1
|                                     | 8   | 1-4-2-3
|                                     | 6   | 3-4-1
|                                     | 9   | 1-3-2-4
|                                     | 5   | 2-1-4
|                                     | 7   | 1-4-3-2
|                                     | 4   | 1-3-4
|-----+-----+-----+-----+-----+-----+-----+-----| NUM  TAP
-5  -4  -3  -2  -1  0  1  2  3  4  5
|                                     | 1   | KID
|                                     | T   | S   M   S   T
|                                     | 0   | 10 20 30 60 80 90 99
|                                     | 99  | PERCENTILE

```

Table 2.5 is used for inference:

Past-sample-dependent inference: If you expect the future person to be from a person sample exactly like the past sample (distribution, ability level, misfit, etc.) , then base your decisions on [Table 2.7](#).

Past-sample-distribution-dependent inference: If you expect the future person to be from a person sample similar to the past sample (distribution, ability level, but not misfit) , then base your decisions on [Table 2.5](#).

Past-sample-independent-as-possible: If you have no information about the future person, except that the latent variable will be the same as for the past sample, then base your decisions on [Table 2.2](#).

Does the future repeat the past? Or is the future something entirely new? The truth is usually somewhere between these extremes.

To plot Table 2.5 in Excel:

The numbers for the Y-axis are in `IFILE=` (MEASURE)

The number for the X-axis are in `DISFILE=` (AVGE MEAS)

Multiple-choice items

Table 2.5 shows the [average measures](#) of persons choosing wrong distractors (illustrated by one of the wrong distractors, "a") and the average measures of persons choosing a correct distractor (illustrated by one of the correct distractors, "d").

```

TABLE 2.5 OBSERVED AVERAGE MEASURES FOR STUDENTS (scored) (ILLUSTRATED BY AN OBSERVED CATEGORY)
-4 -3 -2 -1 0 1 2 3 4
|-----|
|          a          d 55 a107 newspaper
|          a          c | 64 sa01 magazine
|

```

Table 2.15 is the same as Table 5, but the options are shown by their scored values, not by their codes in the data.

```

TABLE 2.15 OBSERVED AVERAGE MEASURES FOR STUDENTS (scored) (BY CATEGORY SCORE)
-4 -3 -2 -1 0 1 2 3 4
|-----|
|          0          1 55 a107 newspaper
|          0          1 | 64 sa01 magazine
|

```

Polytomous items

Tables 2.5 & 2.15 plot the observed average person measures for each scored category. It reflects how this sample used these categories. The plotted values cannot fall outside the range of the sample.

The rating-scale categories, such as 0, 1 and 2, are plotted for each item (y-axis) at the average measure of the persons who responded in that category (x-axis).

```

OBSERVED AVERAGE MEASURES BY SCORED CATEGORY FOR PUPILS
-2 -1 0 1 2 3 4 5
|-----|
|          000102          | 5 FIND BOTTLES AND CANS
|          000102          | 23 WATCH A RAT
|          00 01          02 | 20 WATCH BUGS
|          00 01          02 | 4 WATCH GRASS CHANGE
|          0001          02 | 8 LOOK IN SIDEWALK CRACKS
|-----|
-2 -1 0 1 2 3 4 5
1 2 11 2 114346533233332222 322 2 1 1 1 1 1 11 PUPILS
T S M S T

```

14.3.8 Table 2.6 Observed average measures of persons (empirical averages)

(controlled by `T2SELECT=`, `MRANGE=`, `CATREF=`)

[Table 2](#) for multiple-choice items.

[Table 2](#) for polytomous items.

OBSERVED AVERAGE MEASURES FOR KID (unscored) (BY OBSERVED CATEGORY)												
-5	-4	-3	-2	-1	0	1	2	3	4	5		
					0				1		NUM	TAP
					0				1		15	1-3-2-4-1-3
					0				1		16	1-4-2-3-1-4
					0				1		17	1-4-3-1-2-4
					0		1				14	1-4-2-3-4-1
					0		1				12	1-3-2-4-3
					0			1			13	1-4-3-2-4
					0		1				11	1-3-1-2-4
					0		1				10	2-4-3-1
							1				8	1-4-2-3
							1				6	3-4-1
							1				9	1-3-2-4
	0						1				5	2-1-4
	0						1				7	1-4-3-2
		0					1				4	1-3-4
	0						1				4	1-3-4
											NUM	TAP
-5	-4	-3	-2	-1	0	1	2	3	4	5		
1	1	2	2	2	3	5	4	1	2		KID	
T		S		M		S		T			PERCENTILE	
0	10	20	30	60	80	90	99					

Multiple-choice items

Table 2.6, shown first from the Diagnosis menu, shows the average measures from [Table 14.3](#) of the persons choosing each distractor. "m" usually indicates the average measure of persons with missing data. Table 2.6 shows the average ability of the group of examinees who chose each option. Table 2.15 (above) shows the average ability of the people who got an item right and wrong. Table 2.5 (above again) has the right and wrong scoring illustrated with specific MCQ options. If there are more than one correct or incorrect option, only one of each is shown.

TABLE 2.6 OBSERVED AVERAGE MEASURES FOR STUDENTS (unscored) (BY OBSERVED CATEGORY)											
-4	-3	-2	-1	0	1	2	3	4			
										NUM	TOPIC
					m	ab	c		d	55	al07 newspaper
					m		d	c		64	sa01 magazine

Code for unidentified missing data: m

Table 2.16 is the same as Table 6, but the options are shown by their scored values, not by their codes in the data.

TABLE 2.16 OBSERVED AVERAGE MEASURES FOR STUDENTS (unscored) (BY CATEGORY SCORE)											
-4	-3	-2	-1	0	1	2	3	4			
					m	00	0		1	55	al07 newspaper
					m		0	1		64	sa01 magazine

Polytomous items

Tables 2.6 & Table 2.16 plot the observed [average person measures](#) from [Table 14.3](#) for each observed category. It reflects how this sample used these categories. The plotted values cannot fall outside the range of the sample. "m" in the plot indicates the average measure of those for whom their observation is missing on this item. This Table is shown first from the Diagnosis pull-down menu.

OBSERVED AVERAGE MEASURES BY OBSERVED CATEGORY FOR PUPILS											
-2	-1	0	1	2	3	4	5				
										NUM	ACT
				000102						5	FIND BOTTLES AND CANS
				000102	m					23	WATCH A RAT
				00	01	02				20	WATCH BUGS

Multiple-choice items

Table 2.11 shows the most probable response on the latent variable. Table 2.11 is the same as [Table 2.1](#), but the options are shown by their scored values, not by their codes in the data.

TABLE 2.11: MOST PROBABLE RESPONSE: MODE (BETWEEN "0" AND "1" IS "0", ETC.) (BY CATEGORY SCORE)

-4	-3	-2	-1	0	1	2	3	4					
-----+-----+-----+-----+-----+-----+-----+-----+-----									NUM	TOPIC			
0								1	1	55 al07 newspaper			
0								1	1	64 sa01 magazine			
-----+-----+-----+-----+-----+-----+-----+-----+-----									NUM	TOPIC			
-4	-3	-2	-1	0	1	2	3	4					
		1	11	1111	111	212	3	2	12	12	1	2	STUDENTS
		T		S		M		S		T			
		0		10	20	30	40	50	60	70	80	90	99 PERCENTILE

M= Mean, the average of the person measures, S = One Standard Deviation from the mean, T = Two P.SDs. from the mean. Percentile is percentage below the specified position.

Polytomous items

Tables 2.1 & 2.11: The "Most Probable Response" Table, selected with CURVES=001, answers the question "which category is a person of a particular measure most likely to choose?" This is the most likely category with which the persons of logit (or user-rescaled) measure shown below would respond to the item shown on the left. The area to the extreme left is all "0"; the area to the extreme right is at the top category. Each category number is shown to the left of its modal area. If a category is not shown, it is never a most likely response. An item with an extreme, perfect (maximum possible) or zero (minimum possible), score is not strictly estimable, and is omitted here. Blank lines are used to help approximate the placement of the items on the latent variable.

This table presents in one picture the results of this analysis in a form suitable for inference. We can predict for people of any particular measure what responses they would probably make. "M" depicts an "average" person. The left "T" a low performer. The right "T" a high performer. Look straight up from those letters to read off the expected response profiles.

Table 2.1 to 2.7 reports with observed categories, i.e., those in the CODES= statement.

Table 2.11 to 2.17 report with scored categories, i.e., after IVALUE=, RESCORE=, KEY1=, etc., if any.

MOST PROBABLE RESPONSE: MODE (BETWEEN "0" AND "1" IS "0", ETC.) (BY CATEGORY SCORE)

-6	-4	-2	0	2	4	6							
-----+-----+-----+-----+-----+-----+-----							NUM	ACT					
0				1	2		2	5 FIND BOTTLES AND CANS					
0				1	2		2	23 WATCH A RAT					
0				1	2		2	20 WATCH BUGS					
0				1	2		2	4 WATCH GRASS CHANGE					
0				1	2		2	8 LOOK IN SIDEWALK CRACKS					
0	1	2					2	18 GO ON PICNIC					
-----+-----+-----+-----+-----+-----+-----							NUM	ACT					
-6	-4	-2	0	2	4	6							
				1									
				12	2	2117083563422342	111	1	1	1	1	1	KID
				T		S		M		S		T	
				0	10	30	50	70	80	90			99 PERCENTILE

14.3.11 Table 2.12 Mean categories by category score (average or expected: Rasch-half-point thresholds)

(controlled by [T2SELECT=](#), [MRANGE=](#), [CATREF=](#))

[Table 2](#) for multiple-choice items.

[Table 2](#) for polytomous items.

EXPECTED SCORE: MEAN (Rasch-score-point threshold, ":" indicates Rasch-half-point threshold) (BY CATEGORY SCORE)

-5	-4	-3	-2	-1	0	1	2	3	4	5	NUM	TAP	
	-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----												
0									0	:	1	15	1-3-2-4-1-3
0								0		:	1	16	1-4-2-3-1-4
0								0		:	1	17	1-4-3-1-2-4
0								0	:	1	1	14	1-4-2-3-4-1
0							0	:	1		1	12	1-3-2-4-3
0						0	:	1			1	13	1-4-3-2-4
0					0	:	1				1	11	1-3-1-2-4
0			0	:	1						1	10	2-4-3-1
0		0	:	1							1	8	1-4-2-3
0 0		:	1								1	6	3-4-1
0 0		:	1								1	9	1-3-2-4
0	:	1									1	5	2-1-4
0	:	1									1	7	1-4-3-2
0	:	1									1	4	1-3-4
	-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----												
-5	-4	-3	-2	-1	0	1	2	3	4	5	NUM	TAP	
						1							
1	1	2	2	2	3	2	5	4	1	2			KID
T			S		M		S		T				
0		10		20	30	60	80	90		99			PERCENTILE

Table 2.12 is used for inference:

Past-sample-dependent inference: If you expect the future person to be from a person sample exactly like the past sample (distribution, ability level, misfit, etc.) , then base your decisions on [Table 2.7](#).

Past-sample-distribution-dependent inference: If you expect the future person to be from a person sample similar to the past sample (distribution, ability level, but not misfit) , then base your decisions on [Table 2.5](#).

Past-sample-independent-as-possible: If you have no information about the future person, except that the latent variable will be the same as for the past sample, then base your decisions on Table 2.2.

Does the future repeat the past? Or is the future something entirely new? The truth is usually somewhere between these extremes.

Multiple-choice items

Table 2.2: shows the predicted average response on the latent variable. In this example, for item "a107", "a" (or any other incorrect option) is the predicted average response up to 3.2 logits, then "d", the correct response, becomes the average predictions. The ":" is at the transition from an average expected wrong response to an average expected "right" response, i.e., where the predicted average score on the item is 0.5, the Rasch-half-point thresholds. The "a" below "2" is positions where the expected average score on the item is 0.25. Similarly "d" would be repeated where the expected average score on the item is 0.75, according to the Rasch model.

TABLE 2.2 EXPECTED SCORE: MEAN (":" INDICATES HALF-POINT THRESHOLD) (ILLUSTRATED BY AN OBSERVED CATEGORY)

-4	-3	-2	-1	0	1	2	3	4	NUM	TOPIC	
	-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----										
a						a	:	d	55	a107 newspaper	
a						a	:	c	64	sa01 magazine	

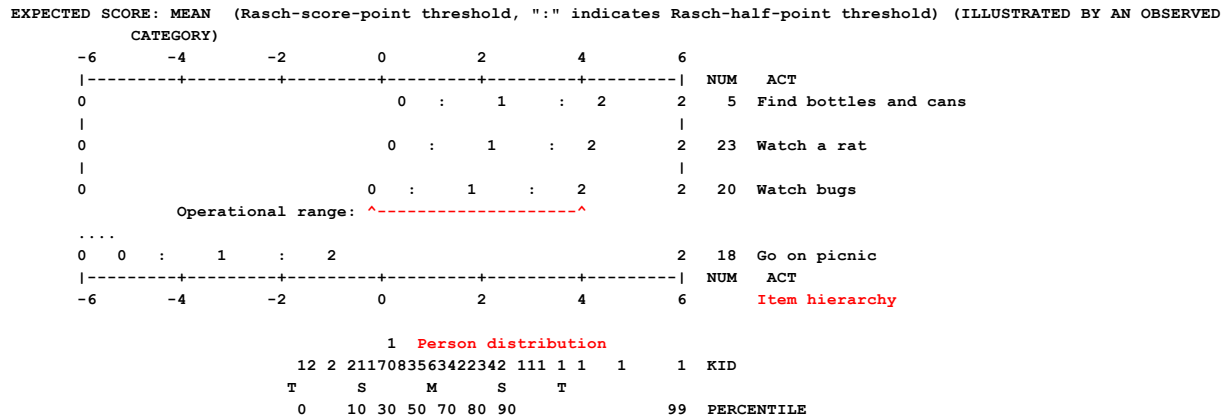
Table 2.12 is the same as Table 2, but the options are shown by their scored values, not by their codes in the data.

TABLE 2.12 EXPECTED SCORE: MEAN (":" INDICATES HALF-POINT THRESHOLD) (BY CATEGORY SCORE)

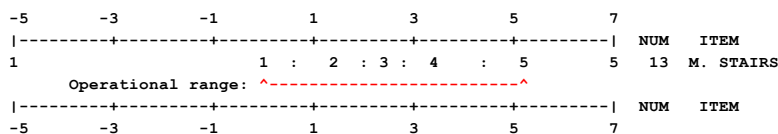
-4	-3	-2	-1	0	1	2	3	4	NUM	TOPIC	
	-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----										
0						0	:	1	55	a107 newspaper	
0						0	:	1	64	sa01 magazine	

Polytomous items

Tables 2.2 & 2.12: In the "Expected Score" Table, the standard output (or selected with [CURVES=010](#)) answers the question "what is the average rating that we expect to observe for persons of a particular measure?" This rating information is expressed in terms of expected scores (with ":" at the half-point thresholds). Extreme scores are located at expected scores .25 score points away from the extremes. This plot also indicates the operational range of each item.



In this example of a 5 category rating scale, the operation range of the top item is indicated in red.



14.3.12 Table 2.13 Median categories by category score (cumulative probabilities: Rasch-Thurstone thresholds)

(controlled by [T2SELECT=](#), [MRANGE=](#), [CATREF=](#))

[Table 2](#) for multiple-choice items.

[Table 2](#) for polytomous items.

Multiple-choice items

Table 2.13 shows the 50% cumulative probability points, the Rasch-Thurstone thresholds. The lower category ("a" and other wrong answers) has a greater than 50% probability of being observed up to 3.2 logits, when "d", the correct answer, has a higher than 50% probability.

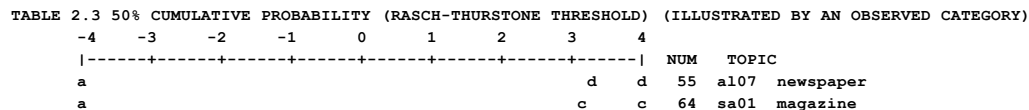
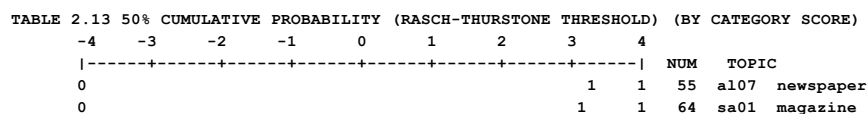


Table 2.13 is the same as [Table 2.3](#), but the options are shown by their scored values, not by their codes in the data.



Polytomous items

Tables 2.3 & 2.13: The "Cumulative Probability" Table: Rasch-Thurstone thresholds, selected with CURVES=001, answers the question "whereabouts in the category ordering is a person of a particular measure located?" This information is expressed in terms of median cumulative probabilities (the point at which the probability of scoring in or above the category is .5).

```

50% CUMULATIVE PROBABILITY: MEDIAN (equal-cumulative-probability Rasch-Thurstone thresholds) (BY CATEGORY SCORE)
-6      -4      -2      0      2      4      6
|-----+-----+-----+-----+-----+-----| NUM  ACT
0              1      2              2      5  FIND BOTTLES AND CANS
|
0              1      2              2      23 WATCH A RAT
|
0              1      2              2      20 WATCH BUGS
0              1      2              2      4  WATCH GRASS CHANGE
0              1      2              2      8  LOOK IN SIDEWALK CRACKS
|
0      1      2              2      18  GO ON PICNIC
|-----+-----+-----+-----+-----+-----| NUM  ACT
-6      -4      -2      0      2      4      6

          1
        12 2 2117083563422342 111 1 1 1 1 1 KID
      T   S   M   S   T
      0  10 30 50 70 80 90          99 PERCENTILE

```

14.3.13 Table 2.14 Andrich thresholds (structure calibrations)

(controlled by [T2SELECT=](#), [MRANGE=](#), [CATREF=](#))

[Table 2](#) for multiple-choice items.

[Table 2](#) for polytomous items.

Multiple-choice items

Table 2.14 shows the item difficulties (or more generally the Rasch-Andrich thresholds) coded by the option of the higher category. For item "a107" this is "d", the correct option.

```

TABLE 2.4 STRUCTURE MEASURES (Rasch model parameters: equal-adjacent-probability thresholds) (ILLUSTRATED BY AN OBSERVED
CATEGORY)
-4      -3      -2      -1      0      1      2      3      4
|-----+-----+-----+-----+-----+-----| NUM  TOPIC
|                                     d | 55  a107  newspaper
|                                     c | 64  sa01  magazine

```

Table 2.14 is the same as Table 4, the Rasch-Andrich thresholds, but the options are shown by their scored values, not by their codes in the data.

```

TABLE 2.14 STRUCTURE MEASURES (Rasch model parameters: equal-adjacent-probability thresholds) (BY CATEGORY SCORE)
-4      -3      -2      -1      0      1      2      3      4
|-----+-----+-----+-----+-----+-----| NUM  TOPIC
|                                     1 | 55  a107  newspaper
|                                     1 | 64  sa01  magazine

```

Polytomous items

Tables 2.4 & 2.14 show Rasch structure calibrations: Rasch-Andrich thresholds (step parameters, step measures, step difficulties, rating (or partial credit) scale calibrations). These are the relationships between adjacent categories, and correspond to the points where adjacent category probability curves cross, i.e., are equally probable of being observed according to a Rasch model.

```

STRUCTURE MEASURES (Rasch model parameters: equal-adjacent-probability Rasch-Andrich thresholds) (BY CATEGORY SCORE)
-6      -4      -2      0      2      4      6
|-----+-----+-----+-----+-----+-----| NUM  ACT
|                                     1  2 | 5  FIND BOTTLES AND CANS
|
|                                     1  2 | 23 WATCH A RAT
|      1      2 | 18  GO ON PICNIC
|-----+-----+-----+-----+-----+-----| NUM  ACT
-6      -4      -2      0      2      4      6

```

```

1
12 2 2117083563422342 111 1 1 1 1 KID
T S M S T
0 10 30 50 70 80 90 99 PERCENTILE

```

14.3.14 Table 2.15 Observed average measures of persons for scored categories (empirical averages)

(controlled by [T2SELECT=](#), [MRANGE=](#), [CATREF=](#))

[Table 2](#) for multiple-choice items.

[Table 2](#) for polytomous items.

```

OBSERVED AVERAGE MEASURES FOR KID (scored) (BY CATEGORY SCORE)
-5 -4 -3 -2 -1 0 1 2 3 4 5
|-----|-----|-----|-----|-----|
| | | | | | 0 | | | | | | | NUM TAP
| | | | | | | | | | | 1 | | 15 1-3-2-4-1-3
| | | | | | 0 | | | | | 1 | | 16 1-4-2-3-1-4
| | | | | | 0 | | | | | 1 | | 17 1-4-3-1-2-4
| | | | | | | | | | | | | |
| | | | | 0 | 1 | | | | | | | 14 1-4-2-3-4-1
| | | | | 0 | | 1 | | | | | | | 12 1-3-2-4-3
| | | | | 0 | | | 1 | | | | | | 13 1-4-3-2-4
| | | | | | 0 | | 1 | | | | | | 11 1-3-1-2-4
| | | | | | | | 1 | | | | | | | 10 2-4-3-1
| | | | | | 0 | | 1 | | | | | | | 8 1-4-2-3
| | | | | | 0 | | | 1 | | | | | | | 6 3-4-1
| | | 0 | | | 1 | | | | | | | | 9 1-3-2-4
| | 0 | | | 1 | | | | | | | | 5 2-1-4
| | 0 | | 0 | 1 | | | | | | | | 7 1-4-3-2
| | 0 | | | 1 | | | | | | | | 4 1-3-4
|-----|-----|-----|-----|-----|
-5 -4 -3 -2 -1 0 1 2 3 4 5
|-----|-----|-----|-----|-----|
1 1 2 2 2 3 2 5 4 1 2 KID
T S M S T
0 10 20 30 60 80 90 99 PERCENTILE

```

Table 2.15 is used for inference:

Past-sample-dependent inference: If you expect the future person to be from a person sample exactly like the past sample (distribution, ability level, misfit, etc.) , then base your decisions on [Table 2.7](#).

Past-sample-distribution-dependent inference: If you expect the future person to be from a person sample similar to the past sample (distribution, ability level, but not misfit) , then base your decisions on [Table 2.5](#).

Past-sample-independent-as-possible: If you have no information about the future person, except that the latent variable will be the same as for the past sample, then base your decisions on [Table 2.2](#).

Does the future repeat the past? Or is the future something entirely new? The truth is usually somewhere between these extremes.

To plot Table 2.5 in Excel:

The numbers for the Y-axis are in [IFILE=](#) (MEASURE)

The number for the X-axis are in [DISFILE=](#) (AVGE MEAS)

Multiple-choice items

Table 2.5 shows the [average measures](#) of persons choosing wrong distractors (illustrated by one of the wrong distractors, "a") and the average measures of persons choosing a correct distractor (illustrated by one of the correct distractors, "d").

TABLE 2.5 OBSERVED AVERAGE MEASURES FOR STUDENTS (scored) (ILLUSTRATED BY AN OBSERVED CATEGORY)

-4	-3	-2	-1	0	1	2	3	4	NUM	TOPIC
				a					d 55	al07 newspaper
						a	c		64	sa01 magazine

Table 2.15 is the same as Table 5, but the options are shown by their scored values, not by their codes in the data.

TABLE 2.15 OBSERVED AVERAGE MEASURES FOR STUDENTS (scored) (BY CATEGORY SCORE)

-4	-3	-2	-1	0	1	2	3	4	NUM	TOPIC
				0					1 55	al07 newspaper
						0	1		64	sa01 magazine

Polytomous items

Tables 2.5 & 2.15 plot the observed average person measures for each scored category. It reflects how this sample used these categories. The plotted values cannot fall outside the range of the sample.

The rating-scale categories, such as 0, 1 and 2, are plotted for each item (y-axis) at the average measure of the persons who responded in that category (x-axis).

OBSERVED AVERAGE MEASURES FOR KID (scored) (BY CATEGORY SCORE)									
-2	-1	0	1	2	3	4	5	NUM	ACT
			000102					5	FIND BOTTLES AND CANS
			000102					23	WATCH A RAT
		00	01		02			20	WATCH BUGS
		00	01			02		4	WATCH GRASS CHANGE
		0001			02			8	LOOK IN SIDEWALK CRACKS
OBSERVED AVERAGE MEASURES FOR KID (unscored) (BY CATEGORY SCORE)									
-2	-1	0	1	2	3	4	5	NUM	ACT
								1	KID
			12 2	2117083563422342	111 1 1	1		1	99 PERCENTILE
			T	S	M	S	T		
			0	10 30 50 70 80 90					

14.3.15 Table 2.16 Observed average measures of persons (empirical averages)

(controlled by [T2SELECT=](#), [MRANGE=](#), [CATREF=](#))

[Table 2](#) for multiple-choice items.

[Table 2](#) for polytomous items.

OBSERVED AVERAGE MEASURES FOR KID (unscored) (BY CATEGORY SCORE)												
-5	-4	-3	-2	-1	0	1	2	3	4	5	NUM	TAP
					0				1		15	1-3-2-4-1-3
					0				1		16	1-4-2-3-1-4
					0				1		17	1-4-3-1-2-4
					0		1				14	1-4-2-3-4-1
					0		1				12	1-3-2-4-3
					0			1			13	1-4-3-2-4
					0		1				11	1-3-1-2-4
					0		1				10	2-4-3-1
					0		1				8	1-4-2-3
					0		1				6	3-4-1
					0		1				9	1-3-2-4
					0		1				5	2-1-4
					0		1				7	1-4-3-2
					0		1				4	1-3-4

Table 2.17 is used for description, but can be used for inference:

Past-sample-dependent inference: If you expect the future person to be from a person sample exactly like the past sample (distribution, ability level, misfit, etc.) , then base your decisions on [Table 2.7](#).

Past-sample-distribution-dependent inference: If you expect the future person to be from a person sample similar to the past sample (distribution, ability level, but not misfit) , then base your decisions on [Table 2.5](#).

Past-sample-independent-as-possible: If you have no information about the future person, except that the latent variable will be the same as for the past sample, then base your decisions on [Table 2.2](#).

Does the future repeat the past? Or is the future something entirely new? The truth is usually somewhere between these extremes.

Multiple-choice items

Table 2.7 shows the measures that would be predicted to be observed for incorrect and correct responses if the persons responded exactly as the Rasch model predicts. "a" (an incorrect distractor) shows the average measure for persons in the sample who would be predicted to fail the item, and "d" (a correct distractor) shows the average measure for persons in the sample who would be predicted to succeed on the item. Compare this with [Tables 2.5](#) and [2.6](#) for the empirical locations, and [Table 2.2](#) for the sample-free locations.

TABLE 2.7 EXPECTED AVERAGE MEASURES FOR STUDENTS (scored) (ILLUSTRATED BY AN OBSERVED CATEGORY)

-4	-3	-2	-1	0	1	2	3	4	NUM	TOPIC
					a			d	55	a107 newspaper
						a		c	64	sa01 magazine

Table 2.17 is the same as [Table 2.7](#), but the options are shown by their scored values, not by their codes in the data.

TABLE 2.17 EXPECTED AVERAGE MEASURES FOR STUDENTS (scored) (BY CATEGORY SCORE)

-4	-3	-2	-1	0	1	2	3	4	NUM	TOPIC
					0			1	55	a107 newspaper
						0		1	64	sa01 magazine

Polytomous items

Tables 2.7 & 2.17 plot the expected average person measures for each category score. It reflects how this sample were expected to use these categories. The plotted values cannot fall outside the range of the sample. Compare this with [Tables 2.5](#) and [2.6](#) for the empirical locations, and [Table 2.2](#) for the sample-free locations.

EXPECTED AVERAGE MEASURES FOR KID (scored) (BY CATEGORY SCORE)

-6	-4	-2	0	2	4	6	NUM	ACT
			0	1	2		5	FIND BOTTLES AND CANS
			0	1	2		23	WATCH A RAT
			0	1	2		20	WATCH BUGS
			0	1	2		4	WATCH GRASS CHANGE
			0	1	2		8	LOOK IN SIDEWALK CRACKS

1	NUM	ACT
12 2 2117083563422342 111 1 1 1	1	KID
T S M S T		
0 10 30 50 70 80 90	99	PERCENTILE

14.4 Table 3.1 Summaries of persons and items

(controlled by [REALSE=](#), [UMEAN=](#), [USCALE=](#), [ISUBTOTAL=](#), [PSUBTOTAL=](#))

This table summarizes the person, item and structure information.

Table 3.1: Gives summaries for all persons and items.

[Table 3.2](#): Summary of rating categories, probability curves and category (confusion) matrix

[Table 27.3](#): Gives subtotal summaries for items, controlled by ISUBTOT=

[Table 28.3](#): Gives subtotal summaries for persons, controlled by PSUBTOT=

SUMMARY OF 34 MEASURED (NON-EXTREME) KID

	TOTAL		MEASURE	MODEL S.E.	INFIT		OUTFIT	
	SCORE	COUNT			MNSQ	ZSTD	MNSQ	ZSTD
MEAN	9.9	18.0	-.19	1.01	.99	-.17	.68	-.10
SEM	.4	.0	.34	.02	.16	.20	.22	.12
P.SD	2.1	.0	1.97	.10	.94	1.16	1.29	.68
S.SD	2.1	.0	2.00	.10	.95	1.17	1.30	.69
MAX.	14.0	18.0	3.73	1.11	4.12	2.49	6.07	2.23
MIN.	5.0	18.0	-4.32	.82	.18	-1.48	.08	-.73
REAL RMSE	1.18	TRUE SD	1.58	SEPARATION	1.34	KID	RELIABILITY	.64
MODEL RMSE	1.01	TRUE SD	1.69	SEPARATION	1.67	KID	RELIABILITY	.74
S.E. OF KID MEAN = .34								
MINIMUM EXTREME SCORE: 1 KID 2.9%								

SUMMARY OF 35 MEASURED (EXTREME AND NON-EXTREME) KID

	TOTAL		MEASURE	MODEL S.E.	INFIT		OUTFIT	
	SCORE	COUNT			MNSQ	ZSTD	MNSQ	ZSTD
MEAN	9.7	18.0	-.37	1.03				
SEM	.4	.0	.38	.03				
P.SD	2.4	.0	2.22	.17				
S.SD	2.4	.0	2.25	.18				
MAX.	14.0	18.0	3.73	1.85				
MIN.	3.0	18.0	-6.62	.82				
REAL RMSE	1.21	TRUE SD	1.86	SEPARATION	1.55	KID	RELIABILITY	.70
MODEL RMSE	1.05	TRUE SD	1.96	SEPARATION	1.87	KID	RELIABILITY	.78
S.E. OF KID MEAN = .38								

KID RAW SCORE-TO-MEASURE CORRELATION = 1.00
 CRONBACH ALPHA (KR-20) KID RAW SCORE "TEST" RELIABILITY = .75 SEM = 1.17

SUMMARY OF 14 MEASURED (NON-EXTREME) TAP

	TOTAL		MEASURE	MODEL S.E.	INFIT		OUTFIT	
	SCORE	COUNT			MNSQ	ZSTD	MNSQ	ZSTD
MEAN	16.9	35.0	.00	.71	.96	.03	.68	-.06
SEM	3.6	.0	.97	.06	.08	.20	.16	.15
P.SD	12.9	.0	3.48	.21	.28	.72	.58	.54
S.SD	13.3	.0	3.61	.22	.29	.74	.60	.56
MAX.	32.0	35.0	4.80	1.07	1.56	1.23	2.21	1.11
MIN.	1.0	35.0	-4.40	.45	.59	-1.35	.11	-.63
REAL RMSE	.77	TRUE SD	3.39	SEPARATION	4.41	TAP	RELIABILITY	.95
MODEL RMSE	.74	TRUE SD	3.40	SEPARATION	4.59	TAP	RELIABILITY	.95
S.E. OF TAP MEAN = .97								
MAXIMUM EXTREME SCORE: 3 TAP 16.7%								
MINIMUM EXTREME SCORE: 1 TAP 5.6%								

SUMMARY OF 18 MEASURED (EXTREME AND NON-EXTREME) TAP

	TOTAL		MEASURE	MODEL S.E.	INFIT		OUTFIT	
	SCORE	COUNT			MNSQ	ZSTD	MNSQ	ZSTD
MEAN	18.9	35.0	-.76	.96				
SEM	3.3	.0	1.03	.12				
P.SD	14.0	.0	4.26	.51				
S.SD	14.2	.0	4.39	.52				
MAX.	35.0	35.0	6.13	1.85				
MIN.	.0	35.0	-6.59	.45				
REAL RMSE	1.10	TRUE SD	4.12	SEPARATION	3.73	TAP	RELIABILITY	.93
MODEL RMSE	1.09	TRUE SD	4.12	SEPARATION	3.79	TAP	RELIABILITY	.93
S.E. OF TAP MEAN = 1.03								

TAP RAW SCORE-TO-MEASURE CORRELATION = -.99
 Global statistics: please see Table 44.
 UMEAN=.0000 USCALE=1.0000

EXTREME AND NON-EXTREME SCORES	All items with estimated measures
--------------------------------	-----------------------------------

NON-EXTREME SCORES ONLY	Items with non-extreme scores (omits items or persons with 0% and 100% success rates)
ITEM or PERSON COUNT	count of items or persons. "ITEM" is the name assigned with ITEM= : "PERSON" is the name assigned with PERSON=
MEAN MEASURE	average measure of items or persons.
SEM row	standard error of the mean statistic in the row above
REAL/MODEL S.E. column	standard errors of the measures (REAL = inflated for misfit). The S.E. column summarizes the S.E.s in the measurement Table. So S.E. column, S.D. row, is the S.D.s of the S.E.s in the measurement table. It is not the S.E. of the S.D. to its left in this Table.
REAL/MODEL RMSE	statistical "root-mean-square" average of the standard errors
TRUE P.SD (previously ADJ.SD)	The "true" population standard deviation is the observed population S.D. adjusted for measurement error (RMSE). This is an estimate of the measurement-error-free S.D.
REAL/MODEL SEPARATION	the separation coefficient: $G = \text{TRUE P.SD} / \text{RMSE}$ Strata = $(4 * G + 1) / 3$
REAL/MODEL RELIABILITY	the measure reproducibility = ("True" item measure variance / Observed variance) = $\text{Separation}^2 / (1 + \text{Separation}^2)$
S.E. MEAN	standard error of the mean measure of items or persons

For valid observations used in the estimation,

NON-EXTREME persons or items - summarizes persons (or items) with non-extreme scores (omits zero and perfect scores).

EXTREME AND NON-EXTREME persons or items - summarizes persons (or items) with all estimable scores (includes zero and perfect scores). Extreme scores (zero, minimum possible and perfect, maximum possible scores) have no exact measure under Rasch model conditions. Using a Bayesian technique, however, reasonable measures are reported for each extreme score, see [EXTRSC=](#). Totals including extreme scores are reported, but are necessarily less inferentially secure than those totals only for non-extreme scores. Extreme persons and extreme items (minimum possible scores and maximum possible scores) have no infit nor outfit statistics, so those statistics are omitted from "extreme and non-extreme".

RAW SCORE is the raw score (number of correct responses excluding extreme scores, [TOTALSCORE=N](#)).

TOTAL SCORE is the raw score (number of correct responses including extreme scores, [TOTALSCORE=Y](#)).

COUNT is the number of responses made.

MEASURE is the estimated measure (for persons) or calibration (for items).

REAL/MODEL: REAL is computed on the basis that misfit in the data is due to departures in the data from model specifications. This is the worst-case situation. MODEL is computed on the basis that the data fit the model, and that all misfit in the data is merely a reflection of the stochastic nature of the model. This is the best-case situation.

S.E. is the [standard error](#) of the estimate.

INFIT is an information-weighted fit statistic, which is more sensitive to unexpected behavior affecting responses to items near the person's measure level.

MNSQ is the [mean-square infit statistic](#) with expectation 1. Values substantially below 1 indicate dependency in your data; values substantially above 1 indicate noise.

ZSTD is the infit mean-square fit statistic t standardized to approximate a theoretical mean 0 and variance 1 distribution. ZSTD (standardized as a z-score) is used of a t-test result when either the t-test value has effectively infinite degrees of

freedom (i.e., approximates a unit normal value) or the Student's [t-statistic](#) distribution value has been adjusted to a unit normal value. When LOCAL=Y, then EMP is shown, indicating a local {0,1} standardization. When LOCAL=L, then LOG is shown, and the natural logarithms of the mean-squares are reported.

OUTFIT is an outlier-sensitive fit statistic, more sensitive to unexpected behavior by persons on items far from the person's measure level.

MNSQ is the [mean-square outfit statistic](#) with expectation 1. Values substantially less than 1 indicate dependency in your data; values substantially greater than 1 indicate the presence of unexpected outliers.

ZSTD is the outfit mean-square fit statistic t standardized to approximate a theoretical mean 0 and variance 1 distribution. ZSTD (standardized as a z-score) is used of a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's [t-statistic](#) distribution value has been adjusted to a unit normal value. When LOCAL=Y, then EMP is shown, indicating a local {0,1} standardization. When LOCAL=L, then LOG is shown, and the natural logarithms of the mean-squares are reported.

MEAN is the average value of the statistic.

P.SD is its [standard deviation](#) assuming that this sample of the statistic is the entire [population](#). It is not, the corrected sample S.D. = $(P.SD / \sqrt{(\text{Count of statistic}) / (\text{Count of statistic} - 1)}) 10$

P.SD = Population standard deviation (when the sample is the entire population)

S.SD = Sample standard deviation (when the sample represents the population)

MAX. is its maximum value.

MIN. is its minimum value.

MODEL RMSE is computed on the basis that the data fit the model, and that all misfit in the data is merely a reflection of the stochastic nature of the model. This is a "best case" reliability, which reports an upper limit to the reliability of measures based on this set of items for this sample. This RMSE for the person sample is equivalent to the "Test SEM (Standard Error of Measurement)" of Classical Test Theory.

REAL RMSE is computed on the basis that misfit in the data is due to departures in the data from model specifications.

This is a "worst case" reliability, which reports a lower limit to the reliability of measures based on this set of items for this sample.

RMSE is the square-root of the average error variance. It is the Root Mean Square standard Error computed over the persons or over the items. Here is how RMSE is calculated in Winsteps:

George ability measure = 2.34 logits. Standard error of the ability measure = 0.40 logits.

Mary ability measure = 3.62 logits. Standard error of the ability measure = 0.30 logits.

Error = 0.40 and 0.30 logits.

Square error = $0.40 \times 0.40 = 0.16$ and $0.30 \times 0.30 = 0.09$

Mean (average) square error = $(0.16 + 0.09) / 2 = 0.25 / 2 = 0.125$

RMSE = Root mean square error = $\sqrt{0.125} = 0.354$ logits

TRUE P.SD is the [population standard deviation](#) of the estimates (assumed to be the population) after subtracting the error variance (attributable to their standard errors of measurement) from their observed variance.

$(\text{TRUE P.SD})^2 = (\text{P.SD of MEASURE})^2 - (\text{RMSE})^2$

The TRUE P.SD is an estimate of the unobservable exact standard deviation, obtained by removing the bias caused by measurement error.

SEPARATION coefficient is the ratio of the PERSON (or ITEM) TRUE P.SD, the "true" standard deviation, to RMSE, the error standard deviation. It provides a ratio measure of [separation](#) in RMSE units, which is easier to interpret than the reliability correlation. (SEPARATION coefficient)² is the signal-to-noise ratio, the ratio of "true" variance to error variance.

RELIABILITY is a separation reliability (separation index). The PERSON (or ITEM) reliability is equivalent to KR-20, Cronbach Alpha, and the Generalizability Coefficient. See much more at [Reliability](#).

Real reliability while you are improving your results. This assumes misfit contradicts the Rasch model.

Model reliability when your results are as good as they can be. This assumes misfit is the randomness predicted by the Rasch model

S.E. OF MEAN is the standard error of the mean of the person (or item) measures for this sample.

MEDIAN is the median measure of the sample (in Tables 27, 28).

Message	Meaning for Persons or Items
MAXIMUM EXTREME SCORE	All non-missing responses are scored correct (perfect score) or in the top categories. Measures are estimated.
MINIMUM EXTREME SCORE	All non-missing responses are scored incorrect (zero score) or in the bottom categories. Measures are estimated.
LACKING RESPONSES	All responses are missing. No measures are estimated.
DELETED	Persons deleted with PDFILE= or PDELETE= . Items deleted with IDFILE= or IDELETE=
IGNORED	Entry numbers higher than highest reported entry number are deleted and not reported
CUTLO= CUTHI=	CUTLO= and CUTHI= values if these are active. They reduce the number of valid responses.

PERSON RAW SCORE-TO-MEASURE CORRELATION is the [Pearson correlation](#) between raw scores and measures, including extreme scores. When data are complete, this correlation is expected to be near 1.0 for persons.

CRONBACH ALPHA (KR-20) KID RAW SCORE "TEST" RELIABILITY is the conventional "test" reliability index. It reports an approximate test reliability based on the raw scores of this sample. It is only reported for complete data. See more at [Reliability](#). Cronbach Alpha is an estimate of the person-sample reliability (= person-score-order reproducibility). Classical Test Theory does not usually compute an estimate of the item reliability (= item-value-order reproducibility), but it could. Winsteps reports both person-sample reliability (=person-measure-order reproducibility) and item reliability (= item-measure-order-reproducibility). Cronbach Alpha is computed for both dichotomous and polytomous data. Cronbach Alpha is the same as KR-20 when the data are dichotomous. KR-20 is not defined for polytomous data. Cronbach Alpha is influenced by missing data.

High Cronbach Alpha with low Rasch reliability indicates that the raw scores reproduce reliably, but their meaning on the latent variable does not. Example: a survey in which 65% of the persons respond "Agree" to every question. The score of "Agree" is reliably predictable. The meaning of "Agree" on the latent variable is fuzzily unreliable.

SEM this is the "standard error of measurement" (the averaged S.E. of the person raw-scores) reported by Classical Test Theory = raw score S.D. * $\sqrt{1-\text{Cronbach Alpha}}$

ITEM RAW SCORE-TO-MEASURE CORRELATION is the [Pearson correlation](#) between raw scores and measures, including extreme scores. When data are complete, this correlation is expected to be near -1.0 for items. This is because higher measure implies lower probability of success and so lower item scores.

Global fit: please see [Table 44](#).

UMEAN=.000 USCALE=1.000 are the current settings of [UMEAN=](#) and [USCALE=](#).

14.5 Table 3.2+ Summary of dichotomous, rating scale or partial credit structures

(controlled by [STEPT3=](#), [STKEEP=](#), [MRANGE=](#), [MTICK=](#), [DISCRIM=](#), [CMATRIX=](#))

[Table 3.1](#): Gives summaries for all persons and items.

Table 3.2: Summary of rating categories, probability curves and category (confusion) matrix

The [average measures](#) and **category fit statistics** are how the response structure worked "for this sample" (which might have high or low performers etc.). For each observation in category k, there is a person of measure Bn and an item of measure Di. Then:

average measure = $\text{sum}(B_n - D_i) / \text{count of observations in category}$. These are **not** estimates of parameters.

The **probability curves** are how the response structure is predicted to work for any **future** sample, *provided it worked satisfactorily for this sample*.

- 1) how accurately patient measure is predicted from single observations. This is C->M.
C->M is 10% means that we can predict the estimated measure from the observed category about 1 time in 10.
 - 2) how accurately patient measure predicts single observations. This is M->C.
M->C is 50% means that we can predict the observed category from the estimated measure half the time.
- In general, the wider the categories on the latent variable (more dispersed thresholds), the higher these percents will be.

Guttman's Coefficient of Reproducibility is the count-weighted average of the M->C, i.e., $\text{Reproducibility} = \frac{\text{sum across categories (COUNT * M->C)}}{\text{sum(COUNT * 100)}}$

C->M shows what percentage of the observations in this category were produced by measures corresponding to the category. Does the category imply the measures?

Computation: For each observation in the [XFILE=](#),
 "expected response value" - round this to the nearest category number = expected average category
 if "expected average category" = "response value (after scoring and recounting)" then MC = 1, else, MC = 0
 Compute average of MC for each observed category across all the relevant data for C->M
 Compute average of MC for each expected category across all the relevant data for M->C

RMSR is the root-mean-square residual, summarizing the differences between observations in this category and their expectations (excluding observations in extreme scores).

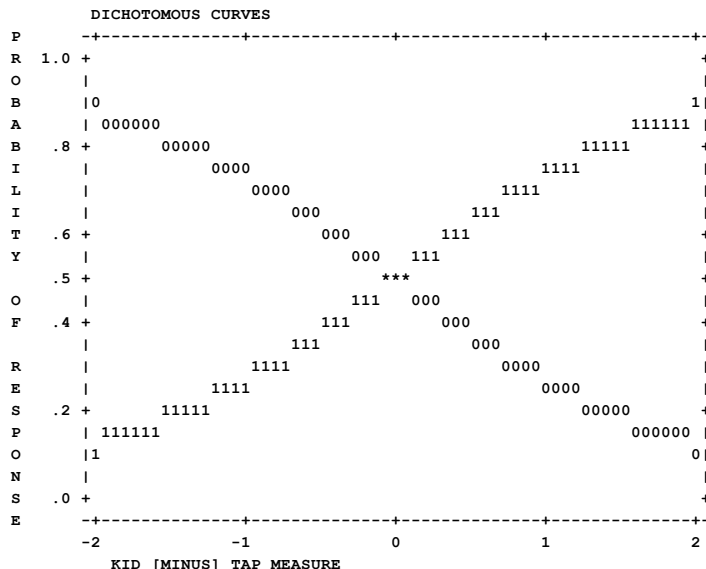
ESTIM DISCR is an estimate of the local discrimination when the model is parameterized in the form: $\log\text{-odds} = a_j (B_n - D_i - F_j)$. Its expected value is 1.0.

RESIDUAL (when shown) is the residual difference between the observed and expected counts of observations in the category. Shown as % of expected, unless observed count is zero. Then residual count is shown. Only shown if residual count is ≥ 1.0 . Indicates lack of convergence, structure anchoring, or large data set.

CATEGORY CODES are shown to the right from on [CODES=](#). The original data code is shown. If several data codes are scored to the same value, only the first data code found in CODES= is shown. Change the reported data code by changing the order of the data codes on CODES= with matching changes in [NEWSCORE=](#) and [IVALUE=](#).

CATEGORY LABELS are shown to the extreme right based on [CFILE=](#) and [CLFILE=](#). The label from CFILE= is reported, if any. If none, then the label from CLFILE= is reported, if any.

Measures corresponding to the dichotomous categories are not shown, but can be computed using the Table at "[What is a Logit?](#)" and [LOWADJ=](#) and [HIADJ=](#).



"measures of y imply ratings averaging x". This is implied by the Rasch model parameters. These are plotted in [Table 2.2](#)

"Category measures" answer the question "If there were a thousand people at the same location on the latent variable and their average rating was the category value, e.g., 2.0, then where would that location be, relative to the item?" This seems to be what people mean when they say "a performance at level 2.0". It is estimated from the Rasch expectation

To discover this ability location, we start with the Rasch model, $\log(P_{nij} / P_{ni(j-1)}) = B_n - D_i - F_j$, For known D_i , F_j and trial B_n . This produces a set of $\{P_{nij}\}$.

Compute the expected rating score: $Eni = \sum (jP_{nij})$ across the categories.

Adjust $B_n' = B_n + (\text{desired category} - Eni) / (\text{large divisor})$, until $Eni = \text{desired category}$, when B_n is the desired category measure.

Here is how to apply this formula. The purpose of this formula is to find the person ability measure corresponding to an expected score on an item.

a) Usually Winsteps has done this for us. Look at [ISFILE=](#) - the columns labeled "AT CAT", "BOT+0.25", "Top-0.25" show the person abilities, B_n , corresponding to an expected score equal to each category value of the rating scale, or a reasonable proxy for the extreme categories.

b) If you want to apply the formula yourself to find the expected value corresponding to an expected value of, say, 2.3 on an item with a 1-5 rating scale. Here's what to do.

1) Select the item, i , and the expected value, E , that you want.

2) Look at the Winsteps [EFILE=](#) output

If the expected value, E , is not there, then ...

c) set $B_n =$ the ability with the Eni closest to the E you want. Note down Eni , the expected value corresponding to B_n in the [EFILE=](#)

3) Look up the item difficulty D_i in Table 14, and the Andrich thresholds, $\{F_j\}$, in Table 3.2

4) Now start to apply the formula:

$B_n' = B_n + (\text{desired category} - Eni) / (\text{large divisor})$

desired category = your value E

B_n and Eni are from step 2)

(large divisor) = 10 - this produces a small increment along the latent variable. Its size doesn't really matter. If this iterative process does not converge, try again with a ten times larger divisor.

We now have B_n'

5) Using Excel or similar, apply the Rasch model with B_n' , D_i , $\{F_j\}$, to obtain Eni' correspond to B_n'

6) If Eni' is not close enough to your desired E , then back to step 4)

with $B_n =$ value of B_n' , and $Eni =$ value of Eni'

```

-----
|CATEGORY  STRUCTURE  | SCORE-TO-MEASURE  | 50% CUM. | COHERENCE  |ESTIM|
| LABEL    MEASURE  S.E. | AT CAT.  ---ZONE--- |PROBABLTY| M->C C->M  RMSR |DISCR|
|-----|-----|-----|-----|-----|-----|-----|
| 0        NONE      | (-3.66) -INF -2.63|          | 0%  0%  .9950| 0 Dislike
| 1        -2.54    .61 |  -.89 -2.63  .85 | -2.57 | 69% 83% .3440| 1.11| 1 Neutral
| 2         .75     .26 | ( 1.88)  .85 +INF |  .79 | 81% 72% .4222| 1.69| 2 Like
|-----|-----|-----|-----|-----|-----|
M->C = Does Measure imply Category?
C->M = Does Category imply Measure?

```

CATEGORY LABEL, the number of the category in your data set after scoring/keying.

STRUCTURE MEASURE, is the **Rasch-Andrich threshold**, the item measure add to the calibrated measure of this transition from the category below to this category. The bottom category has no prior transition, and so that the measure is shown as NONE. The Rasch-Andrich threshold is expected to increase with category value, but these can be [disordered](#). " $D_{gi} + F_{gj}$ " locations are plotted in Table 2.4, where "g" refers to the ISGROUPS= assignment. See [Rating scale conceptualization](#).

For structures with only a single polytomous item, the Partial Credit Model, **this is is the Rasch model parameter δ_{ij} in the δ_{ij} parametrization of the "Partial Credit" model.**)

Item difficulty: $D_i = \text{average}(\delta_{ij})$ for all thresholds j of item i .

Andrich threshold relative to the item difficulty: $F_{ij} = \delta_{ij} - D_i$

Andrich threshold relative to the latent variable: $STRUCTURE\ MEASURE = \delta_{ij} = D_i + F_{ij}$

When we map the location of D_i on the latent variable (or inspect the algebra of the PCM model), we discover that D_i is at the location on the latent variable where the highest and lowest categories of the item are equally probable. Conveniently, the sufficient statistic for D_i is the item's raw score.

STRUCTURE S.E. is an approximate standard error of the Rasch-Andrich threshold measure.

SCORE-TO-MEASURE

These values are plotted in Table 21, "Expected Score" ogives. They are useful for quantifying category measures. **This is implied by the Rasch model parameters.** See [Rating scale conceptualization](#).

AT CAT is the **Rasch-full-point-threshold**, the measure (on an item of 0 logit measure) corresponding to an expected score equal to the category label, which, for the rating (or partial credit) scale model, is where this category has the highest probability. See plot below.

() is printed where the matching calibration is infinite. The value shown corresponds to the measure .25 score points (or LOWADJ= and HIADJ=) away from the extreme.

--ZONE-- is the **Rasch-half-point threshold**, the range of measures from an expected score from 1/2 score-point below to the category to 1/2 score-point above it, the **Rasch-half-point thresholds**. Measures in this range (on an item of 0 measure) are expected to be observed, on average, with the category value. See plot below.

50% CUMULATIVE PROBABILITY gives the location of median probabilities, i.e. these are **Rasch-Thurstone thresholds**, similar to those estimated in the "Graded Response" or "Proportional odds" models. At these calibrations, the probability of observing the categories below equals the probability of observing the categories equal or above. The .5 or 50% cumulative probability is the point on the variable at which the category interval begins. **This is implied by the Rasch model parameters.** See [Rating scale conceptualization](#).

COHERENCE

M->C shows what percentage of the measures that were expected to produce observations in this category actually did. Do the measures imply the category?

For RSM, everything is relative to each target item in turn. So, the 69% for category 1 is a summary across the persons with abilities in the range, "zone", from -2.63 to .85 logits relative to each item difficulty, which are all the observations where $(-2.63 \leq (B_n - D_i) \leq .85)$. Of these, 69% are in category 1. Depending on the targeting of the persons and the items, there could be many eligible observations or only a few.

Guttman's Coefficient of Reproducibility is the count-weighted average of the M->C, i.e., $Reproducibility = \text{sum across categories} (COUNT * M->C) / \text{sum}(COUNT * 100)$

C->M shows what percentage of the observations in this category were produced by measures corresponding to the category. Does the category imply the measures?

RMSR is the root-mean-square residual, summarizing the differences between observations in this category and their expectations (excluding observations in extreme scores).

ESTIMDISCR (when [DISCRIM=Y](#)) is an estimate of the local discrimination when the model is parameterized in the form: $\log\text{-odds} = a_j (B_n - D_i - F_j)$

OBSERVED - EXPECTED RESIDUAL DIFFERENCE (when shown) is the residual difference between the observed and expected counts of observations in the category. This indicates that the Rasch estimates have not converged to their maximum-likelihood values. These are shown if at least one residual percent $\geq 1\%$.

residual difference % = $(\text{observed count} - \text{expected count}) * 100 / (\text{expected count})$

residual difference value = $\text{observed count} - \text{expected count}$

1. Unanchored analyses: These numbers indicate the degree to which the reported estimates have not converged. Usually performing more estimation iterations reduces the numbers.

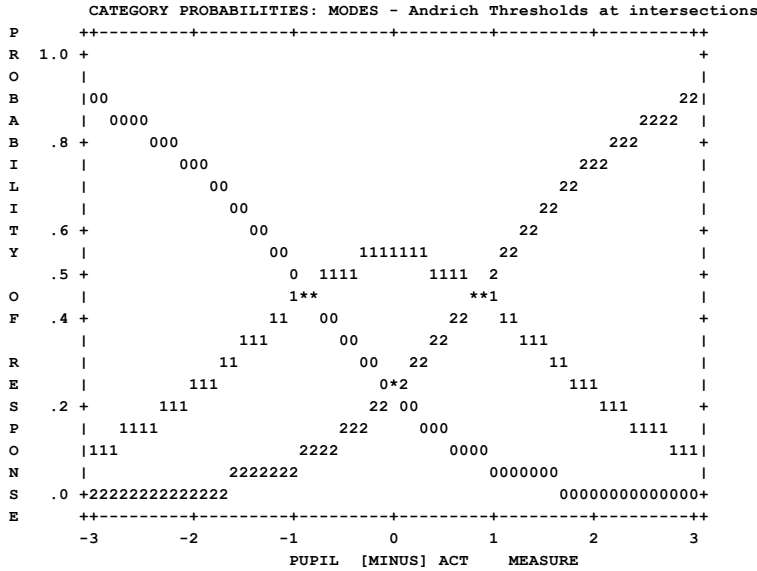
2. Anchored analyses: These numbers indicate the degree to which the anchor values do not match the current data.

For example,

(a) iteration was stopped early using Ctrl+F or the pull-down menu option.

- (b) iteration was stopped when the maximum number of iterations was reached [MJMLE=](#)
- (c) the convergence criteria [LCONV=](#) and [RCONV=](#) are not set small enough for this data set.
- (d) anchor values ([PAFILE=](#), [IAFILE=](#) and/or [SAFILE=](#)) are in force which do not allow maximum likelihood estimates to be obtained.

ITEM MEASURE ADDED TO MEASURES, is shown when the rating (or partial credit) scale applies to only one item, e.g., when ISGROUPS=0. Then all measures in these tables are adjusted by the estimated item measure.



Curves showing how probable is the observation of each category for measures relative to the item measure. Ordinarily, 0 logits on the plot corresponds to the item measure, and is the point at which the highest and lowest categories are equally likely to be observed. The plot should look like a range of hills. Categories which never emerge as peaks correspond to [disordered Rasch-Andrich thresholds](#). These contradict the usual interpretation of categories as a being sequence of most likely outcomes.

Null, Zero, Unobserved Categories

[STKEEP=YES](#) and Category 2 has no observations:

```

+-----+-----+-----+-----+-----+-----+
|CATEGORY  OBSERVED|OBSVD SAMPLE|INFIT  OUTFIT|| ANDRICH |CATEGORY|
|LABEL SCORE COUNT %|AVRGE EXPECT|  MNSQ  MNSQ||THRESHOLD| MEASURE|
+-----+-----+-----+-----+-----+-----+
| 0  0      378 20|  -.67  -.73|  .96  1.16||  NONE  |(-2.01)|
| 1  1      620 34|  -.11  -.06|  .81  .57||  -.89  | -.23 |
| 2  2         0 0|          |  .00  .00||  NULL  |  .63 |
| 3  3      852 46|  1.34  1.33|  1.00  1.64||  .89  |( 1.49)|
| 4  4         20 1|          |          ||  NULL  |  |
+-----+-----+-----+-----+-----+-----+

```

Category 2 is an incidental (sampling) zero. The category is maintained in the response structure. Category 4 has been dropped from the analysis because it is only observed in extreme scores.

- (1) Add one or more dummy person records to the data file with non-extreme scores, but with category 4 for the relevant item or rating scale. Rasch-analysis your data + dummy records, and save the Andrich thresholds: Winsteps "Output Files menu", SFILE=sf.txt
- (2) Reanalyze your data. Omit the dummy records. Include SAFILE=sf.txt to anchor the Andrich thresholds at their (1) values. This does an analysis with reasonable Andrich Thresholds for the the items, but only uses the actual data for the sufficient statistics used to estimate the Di and the theta values.

[STKEEP=NO](#) and Category 2 has no observations:

```

+-----+-----+-----+-----+-----+-----+
|CATEGORY  OBSERVED|OBSVD SAMPLE|INFIT  OUTFIT|| ANDRICH |CATEGORY|
|LABEL SCORE COUNT %|AVRGE EXPECT|  MNSQ  MNSQ||THRESHOLD| MEASURE|
+-----+-----+-----+-----+-----+-----+
| 0  0      378 20|  -.87 -1.03|  1.08  1.20||  NONE  |(-2.07)|
+-----+-----+-----+-----+-----+-----+

```

	1	1	620	34		.13	.33	.85	.69		-.86		.00	
	3	2	852	46		2.24	2.16	1.00	1.47		.86		(2.07)	

Category 2 is a structural (unobservable) zero. The category is eliminated from the response structure.

Category Matrix : Confusion Matrix : Matching Matrix

Category Matrix : Confusion Matrix : Matching Matrix					
Predicted Score-Category Frequency					
Obs Cat Freq	0	1	2	3	Total
0	12.19	10.22	2.57	.02	25.00
1	10.53	13.68	6.67	.11	31.00
2	1.93	6.57	8.63	.87	18.00
3	.33	.52	.15	.00	1.00
Total	24.98	30.99	18.03	1.01	75.00

When **CMATRIX=Yes**, a category matrix ([confusion matrix](#), matching matrix) is produced for each rating scale group defined in **ISGROUPS=**. In this example, there are four categories, 0, 1, 2, 3 observed in the data for the item. The row totals show the observed frequency of each category. According to the Rasch model, there is a probability that each observation will be in any category. These probabilities are summed for each category, and the model-predicted counts are shown in each column. When the data are strongly deterministic (Guttman-like), then the major diagonal of the matrix will dominate.

In this matrix, there is one observation of category 3 in the data (see Row Total for category 3). Since category frequencies are sufficient statistics for the maximum likelihood estimates, one observation of category 3 is predicted in these data (see Column Total for category 3). However, the Category Matrix tells us that the observation of 3 is predicted to have been a 1 (Row 3, Column 1). And an observation of category 3 would be predicted to replace an observation of 2 (Row 2, Column 3). We can compare this finding with the report in [Table 10.6](#). The observation of 3 has an expected value near 1.

TABLE 10.6
MOST UNEXPECTED RESPONSES

DATA	OBSERVED	EXPECTED	RESIDUAL	ST. RES.	MEASDIFF
3	3	.83	2.17	3.23	-2.05

Category Misfit

Usually any MNSQ (mean-square, red box in figure) less than 2.0 is OK for practical purposes. A stricter rule would be 1.5. Overfit (values less than 1.0) are almost never a problem.

CATEGORY	OBSERVED	OBSVD	SAMPLE	INFINIT	OUTFIT	STRUCTURE	CATEGORY	
LABEL	SCORE	COUNT	AVRGE	EXPECT	MNSQ	MNSQ	CALIBRATN	MEASURE
0	0	3	4	-.96	-.44	.74	.73	NONE (-3.66)
1	1	35	47	.12	.30	.70	.52	-1.64 -.89
2	2	36	49	1.60	1.38	.75	.77	1.64 (1.88)

A bigger problem than category MNSQ misfit is the disordering of the "observed averages" (blue box in Figure). These contradict the Rasch axiom that "higher measure -> higher score on the rating scale".

Also large differences between the "observed average" and the "expected average" (green box in Figure). These indicate that the misfit in the category is systematic in some way.

In principle, an "expected value" is what we would see if the data fit the Rasch model. The "observed" value is what we did see. When the "observed" and "expected" are considerably misaligned, then the validity of the data (as a basis for constructing additive measures) is threatened. However, we can usually take some simple, immediate actions to remedy these defects in the data.

Usually, a 3-stage analysis suffices:

1. Analyze the data. Identify problem areas.
2. Delete problem areas from the data ([PFILE=](#), [IDFILE=](#), [EDFILE=](#), [CUTLO=](#), etc.). Reanalyze the data. Produce item and rating-scale anchor files ([IFILE=if.txt](#), [SFILE=sf.txt](#)) which contain the item difficulties from the good data.
3. Reanalyze all the data using the item and rating-scale anchor files ([IAFILE=if.txt](#), [SAFILE=sf.txt](#)) to force the "good" item difficulties to dominate the problematic data when estimating the person abilities from all the data.

Category MnSq fit statistics

For all observations in the data:

X_{ni} is the observed value

Eni is the expected value of X_{ni}

W_{ni} is the model variance of the observation around its expectation

P_{nik} is the probability of observing X_{ni}=k

Category Outfit statistic for category k:

$[\sum ((k-Eni)^2/Wni) \text{ for all } Xni=k] / [\sum (Pnik * (k-Eni)^2/Wni) \text{ for all } Xni]$

Category Infit statistic for category k:

$[\sum ((k-Eni)^2) \text{ for all } Xni=k] / [\sum (Pnik * (k-Eni)^2) \text{ for all } Xni]$

Where does category 1 begin?

When describing a rating-scale to our audience, we may want to show the latent variable segmented into rating scale categories:

0-----01-----12-----2

There are 3 widely-used ways to do this:

1. "1" is the segment on the latent variable from where categories "0" and "1" are equally probable to where categories "1" and "2" are equally probably. These are the **Rasch-Andrich thresholds** (ANDRICH THRESHOLD) for categories 1 and 2.
2. "1" is the segment on the latent variable from where categories "0" and "1+2" are equally probable to where categories "0+1" and "2" are equally probably. These are the **Rasch-Thurstone thresholds** (50% CUM. PROBABILITY) for categories 1 and 2.
3. "1" is the segment on the latent variable from where the expected score on the item is "0.5" to where the expected score is 1.5. These are the **Rasch-half-point thresholds** (ZONE) for category 1.

Alternatively, we may want a point on the latent variable correspond to the category:

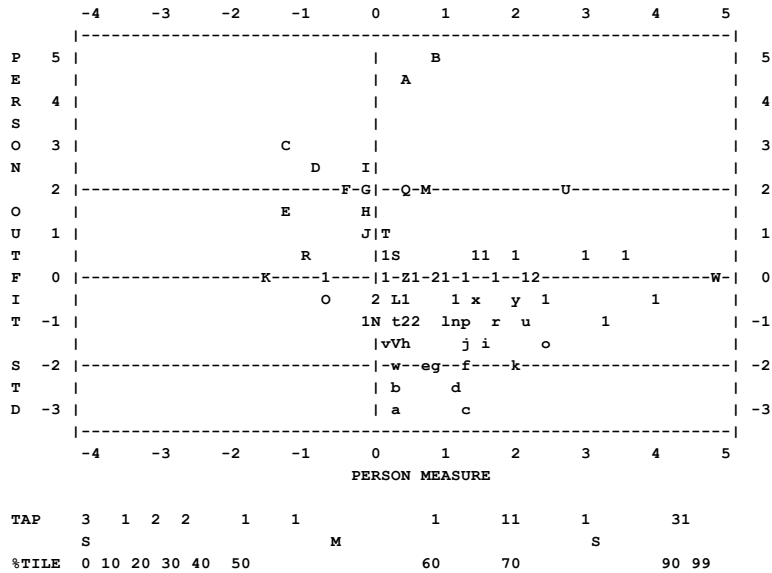
-----0-----1-----2-----

4. "1" is the point on the latent variable where the expected average score is 1.0. This is the **Rasch-Full-Point threshold** (AT CAT.) for category 1.

and should not be interpreted too rigidly. The NORMAL= variable controls the standardization method used.

Letters on the plot indicate the misfitting person or items. Numbers indicate non-extreme fit or multiple references. The letters appear on Tables 6, 17, 18, 19 for persons, and Tables 10, 13, 14, 15 for items.

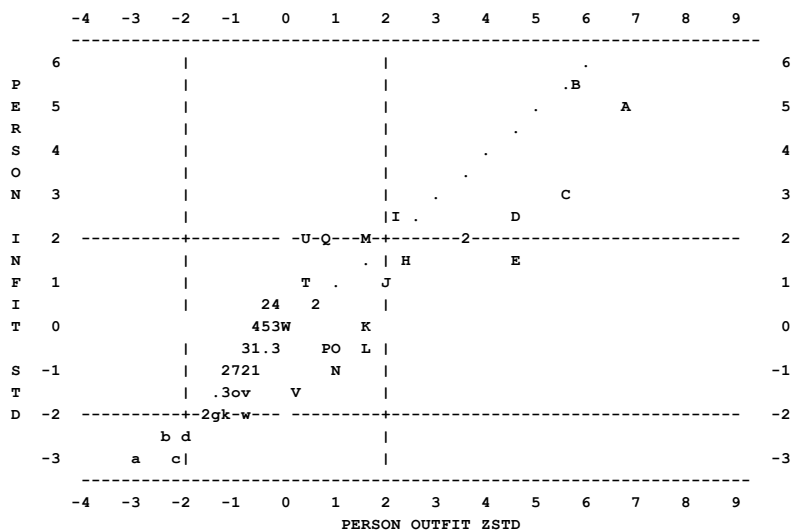
Use [MNSQ=](#) to change the y-axis to mean-squares.



The item distribution and item percentiles are shown beneath the plot.

14.8 Table 5.2 Person infit-outfit fit plot

If both [Table 4.1](#) infit and [Table 5.1](#) outfit plots are requested, then a plot of outfit against infit is also produced to assist with the identification of the different patterns of observations they diagnose. For interpretation of fit statistics, see [dichotomous](#) and [polytomous](#) fit statistics. The letters on the indicate misfitting persons in [Table 6.1](#). Numbers are the count of persons with that degree of misfit at that location on the plot.



14.9 Table 6 Person report in misfit order

14.9.1 Table 6.1 Person statistics in misfit order

(controlled by [FITP=](#), [USCALE=](#), [UIMEAN=](#), [UPMEAN=](#), [UDECIM=](#), [LOCAL=](#), [OUTFIT=](#), [PSORT=](#), [TOTAL=](#))

PERSON STATISTICS: MISFIT ORDER

For details, please see [Table 10.1](#)

Example: To eliminate misfitting persons, output Table 6.1. Copy into Notepad++ (freeware) or other software capable of rectangular copy (alt+mouse). Then paste into your Winsteps control file

[PDFILE=*](#)

pasted entry numbers - put enough blank lines here first for all the numbers

*

Another approach is to screen out very unexpected answers, rather than whole people. Do this with [CUTLO=-2](#)

14.9.2 Table 6.2 Person statistics graphically in misfit order

Please see [Table 10.2](#)

14.9.3 Table 6.4 Person most-misfitting response strings

Table 6.4 displays the unexpected responses in the up to 26 most-misfitting person-response strings.

```

MOST MISFITTING RESPONSE STRINGS
KID          OUTMNSQ |TAP
              |      111
              |47568024 <- The list column is Item 14
              |high-----
29 Anne     F    6.07 A|.0..0..1
13 Mike     M    5.28 B|.0.0..1.
 3 Walter   M    1.19 C|.....1.
24 Rick     M    .88 D|.....0.
34 Elsie    F    .71 E|.....1
23 Bert     M    .67 F|.....1
33 Martha   F    .86 I|.....1..
17 Rod      M    .97 J|..0.....
11 Pete     M    .81 M|0.....
              |-----low
              |47568111
              |      024
    
```

In this Table, the persons (rows) are sorted:

OUTFIT=Yes	Sorted by Outfit Mean-square descending
OUTFIT=No	Sorted by bigger of Outfit Mean-square and Infit mean-square descending.

The items (columns) are sorted by measure. The easiest item (highest-scoring) is on the left. The item entry-numbers are in each column (vertically). In the Figure, the first is item 18. The items shown are those with the most-unexpected responses according to Table 6.1, but excluding items with no individually misfitting observations on the reported items.

[UCOUNT=](#) sets the maximum number of items to report. [LINELENGTH=](#) controls the width of the table. "high" indicates where we expect to see high observations. "low" indicates where we expect to see low observations. Expected observations are replaced by ".".

Each row contains:

Person entry number

Person label

Person outfit mean-square

Person letter (A-Z) in the [Table 6.1](#).

The responses to the items:

0, 1, 2 and numeric values	Scored unexpected-responses
----------------------------	-----------------------------

•	Expected responses, standardized residuals < 2
(blank)	Missing data

14.9.4 Table 6.5 Person most-unexpected observations

Table 6.5 displays the person response strings with the most-unexpected responses as shown in Table 6.6.

```

MOST UNEXPECTED RESPONSES
PUPIL          MEASURE |ACT
                |1111112 122 1 22
                |89203112542669784035
                high-----
41 FATALE, NATASHA 4.77 |.....1.
17 SCHATTNER, GAIL 3.55 |.....0
71 STOLLER, DAVE .96 B|.0.10....0.....222
53 SABOL, ANDREW -1.59 L|.....1....1
                |-----low
                |11111122122619784225
                |8920311 542 6 03

```

In this Table, the persons (rows) are sorted by measure, descending.

The items (columns) are sorted by measure. The easiest item (highest-scoring) is on the left. The item entry-numbers are in each column (vertically). In the Figure, the first is item 18. The items shown are those with the most-unexpected responses as shown in Table 6.6. UCOUNT= sets the maximum number of items to report. LINELENGTH= controls the width of the table. "high" indicates where we expect to see high observations. "low" indicates where we expect to see low observations. Expected observations are replaced by "!".

Each row contains:

- Person entry number
- Person label
- Person measure
- Person letter (A-Z) in the plot in Tables 4 and 5.
- The responses to the items:

0, 1, 2 and numeric values	Scored unexpected-responses
•	Expected responses, standardized residuals < 2
(blank)	Missing data

14.9.5 Table 6.6 Person most-unexpected response list

This shows the most unexpected responses sorted by unexpectedness (standardized residual). Large standardized residuals contribute to large outfit mean-square fit statistics. UCOUNT= sets the maximum number of "most unexpected responses" to report in Tables 6.6, 10.6.

```

MOST UNEXPECTED RESPONSES
-----+-----
| DATA |OBSERVED|EXPECTED|RESIDUAL|ST. RES.|MEASDIFF| ACT  | KID  | ACT  | KID  |
|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----|
| 0 | 0 | 1.93 | -1.93 | -7.66 | 3.53 | 18 | 73 | GO ON PICNIC | SANDBERG, RYNE |
| 2 | 2 | .07 | 1.93 | 7.57 | -3.50 | 23 | 72 | WATCH A RAT | JACKSON, SOLOMON |
| 2 | 2 | .07 | 1.93 | 7.57 | -3.50 | 23 | 29 | WATCH A RAT | LANDMAN, ALAN |
| 0 | 0 | 1.93 | -1.93 | -7.41 | 3.46 | 19 | 71 | GO TO ZOO | STOLLER, DAVE |

```

- DATA is the response code in the data file
- OBSERVED is the code's value after rescoring
- EXPECTED is the predicted observation based on the person and item estimated measures
- RESIDUAL is (OBSERVED - EXPECTED), the difference between the observed and expected values
- ST. RES. is the standardized residual, the unexpectedness of the residual expressed as a unit normal deviate
- MEASDIFF is the difference between the ability and difficulty estimates. This produces the EXPECTED value.
- ACT is the item entry number
- KID is the person entry number
- ACT is the item label
- KID is the person label

14.10 Table 7.1 Person misfitting responses

(controlled by [FITP=](#), [LINELENGTH=](#), [MNSQ=](#), [OUTFIT=](#), [T7OPTIONS=](#))

These tables show the persons or items for which the t standardized outfit (or infit, if [OUTFIT=N](#)) statistic is greater than the misfit criterion ([FITP=](#)). [FITP=0](#) displays all persons. Persons are listed in descending order of misfit. The response codes are listed in their sequence order in your data file. The residuals are standardized response score residuals, which have a model led expectation of 0, and a variance of 1. Negative residuals indicate that the observed response was less correct (or, for rating (or partial credit) scales, lower down the rating scale) than expected. The printed standardized residual is truncated, not rounded, so that its actual value is at least as extreme as that shown. Standardized residuals between -1 and 1 are not printed. For exact details, see [XFILE=](#). "M" indicates a missing response.

"X" indicates that the item obtained an extreme score. "X" in Tables 7.1 and 11.1 indicates that the observation is part of an extreme score, and so does not have a standardized residual. All residuals for extreme scores are conceptually zero. Persons with "X" in Item Table 11.1 (because they have extreme scores) do not appear in Table 7.1, and *vice-versa*, because those extreme persons and items are not part of the misfit process. Basically "X" says "this observation is excluded from the misfit computation".

For Table 7, the diagnosis of misfitting persons, persons with a t standardized fit greater than [FITP=](#) are reported. Selection is based on the [OUTFIT](#) statistic, unless you set [OUTFIT=N](#) in which case the [INFIT](#) statistic is used.

TABLE OF POORLY FITTING PERSONS			(ITEMS IN ENTRY ORDER)			
NUMBER	NAME	POSITION	MEASURE	INFIT (ZSTD)	OUTFIT	MISFIT OVER 2.0
23	GEORGE		2.00	5.8	A	8.1
	RESPONSE:	1:	0 2 1 1 1	2 2 0 2 0	0 1 0 1 1	0 1 0 0 0
	Z-RESIDUAL:		X 2 3 3	2		0 1 1 0 0
	RESPONSE:	26:	1 2 0 2 1	M 0 0 1 1	1 0 1 0 0	1 0 0 0 0
	Z-RESIDUAL:		3 6 2		-2 -2	0 0 0 2 1
						4
						/ This letter on fit plots
5	MARY		2.21	5.2	B	6.5
	RESPONSE:	1:	1 2 0 0 1	2 0 0 1 0	0 2 0 0 1	1 0 0 0 0
	Z-RESIDUAL:		X 2 4	6		-2

Example: Table 7.1 with [T7OPTIONS = OERZ](#) in an analysis of [Example0.txt](#):

TABLE OF POORLY FITTING KID			(ACT IN ENTRY ORDER)			
NUMBER	NAME	POSITION	MEASURE	INFIT (MNSQ)	OUTFIT	
72	M Jackson, Solomon		-1.32	2.0	A	5.2
	OBSERVED:	1:	1 0 0	1 0	0 0	1 0 0
	EXPECTED:		0.6 0.7 0.3	0.1 0.1	0.4 0.2	0.1 0.3 1.1
	RESIDUAL:		0.4 -0.7 -0.3	0.9 -0.1	-0.4 -0.2	0.9 -0.3 -1.1
	Z-RESIDUAL:			2		2

<- [T7OPTIONS= O](#)
 <- [T7OPTIONS= E](#)
 <- [T7OPTIONS= R](#)
 <- [T7OPTIONS= Z](#)

The OBSERVED value is the scored response, before recounting due to [STKEEP=No](#), if any. The EXPECTED value is the OBSERVED value - RESIDUAL value.

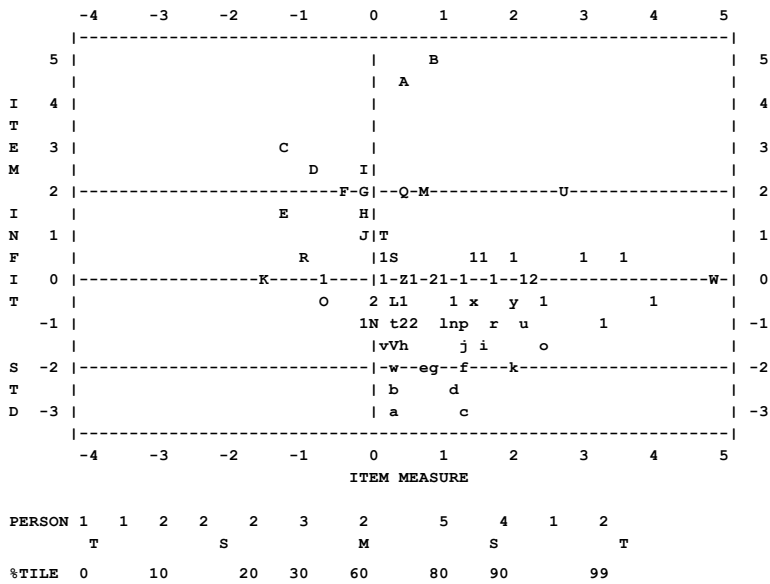
14.11 Table 8.1 Item infit fit plots

(controlled by [FRANGE=](#), [MRANGE=](#), [MTICK=](#), [LOCAL=](#), [MNSQ=](#), [OUTFIT=](#))

These tables are plots of the t standardized fit statistics, [INFIT](#) or [OUTFIT](#), against the parameter estimates. [INFIT](#) is a t standardized information-weighted mean square statistic, which is more sensitive to unexpected behavior affecting responses to items near the person's measure level. [OUTFIT](#) is a t standardized outlier-sensitive mean square fit statistic, more sensitive to unexpected behavior by persons on items far from the person's measure level. The standardization is approximate. Its success depends on the distribution of persons and items. Consequently, the vertical axis is only a guide and should not be interpreted too rigidly. The [NORMAL=](#) variable controls the standardization method used.

Letters on the plot indicate the misfitting person or items. Numbers indicate non-extreme fit or multiple references. The letters appear on Tables 6, 17, 18, 19 for persons, and Tables 10, 13, 14, 15 for items.

Use `MNSQ=` to change the y-axis to mean-squares.



The person distribution and person percentiles are shown beneath the plot.

14.12 Table 9.1 Item outfit fit plot

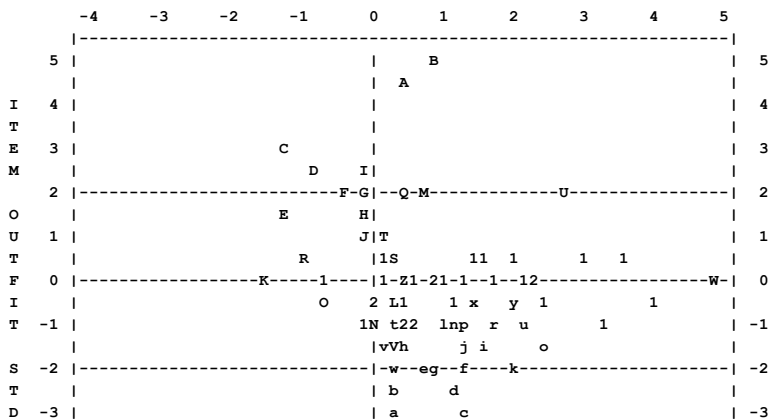
(controlled by `FRANGE=`, `MRANGE=`, `MTICK=`, `LOCAL=`, `MNSQ=`, `OUTFIT=`)

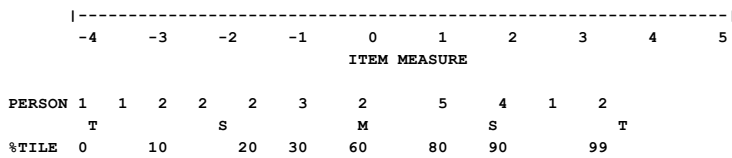
These tables are plots of the t standardized fit statistics, INFIT or OUTFIT, against the parameter estimates. INFIT is a t standardized information-weighted mean square statistic, which is more sensitive to unexpected behavior affecting responses to items near the person's measure level. OUTFIT is a t standardized outlier-sensitive mean square fit statistic, more sensitive to unexpected behavior by persons on items far from the person's measure level. The standardization is approximate. Its success depends on the distribution of persons and items. Consequently, the vertical axis is only a guide and should not be interpreted too rigidly. The `NORMAL=` variable controls the standardization method used.

Letters on the plot indicate the misfitting person or items. Numbers indicate non-extreme fit or multiple references. The letters appear on Tables 6, 17, 18, 19 for persons, and Tables 10, 13, 14, 15 for items.

Use `MNSQ=` to change the y-axis to mean-squares.

For [Table 9.2](#) Item infit-outfit fit plot.

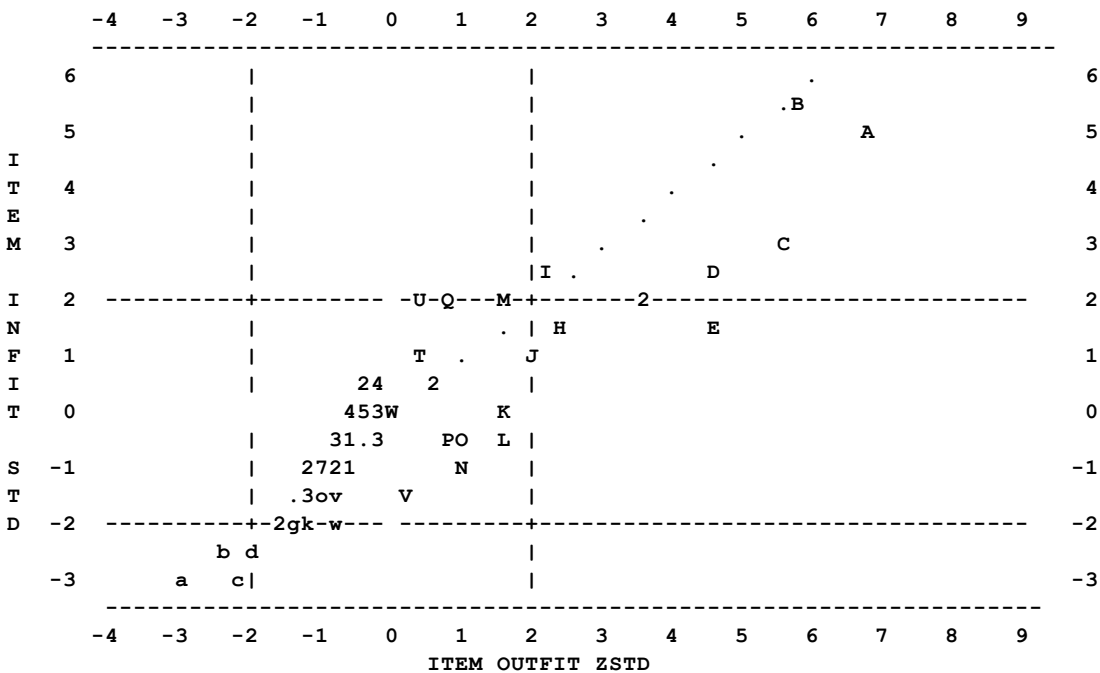




The person distribution and person percentiles are shown beneath the plot.

14.13 Table 9.2 Item infit-outfit fit plot

If both infit and outfit plots are requested, then a plot of outfit against infit is also produced to assist with the identification of the different patterns of observations they diagnose. For interpretation of fit statistics, see [dichotomous](#) and [polytomous](#) fit statistics. The letters on the indicate misfitting items in [Table 10.1](#). Numbers are the count of items with that degree of misfit at that location on the plot.



14.14 Table 10 Item report in misfit order

14.14.1 Table 10.1 Item statistics in misfit order

(controlled by [FIT=](#), [LOCAL=](#), [OUTFIT=](#), [TOTAL=](#))

ITEMS STATISTICS: MISFIT ORDER

PERSON: REAL SEP.: 1.55 REL.: .70 ... ITEM: REAL SEP.: 3.73 REL.: .93

Above the Table are shown the "real" [separation](#) coefficient and reliability (separation index) coefficients from [Table 3](#).

ENTRY	TOTAL ASYMPTOTE	TOTAL P-	TOTAL COUNT	MODEL MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	ESTIM DISCR	LOWER
1	35	35	35	-6.59A	1.85	1.00	.0	1.00	.0	A	.00	100.0	100.0	1.00	.00
	1.00	1.00	.000	.50	-.52	1-4		1							

5	31	35	-3.83	.70	1.04	.2	.52	.1	B	.55	.55	88.2	91.7	1.01	.05
1.00	.89	.248	1.00	-.87	2-1-4	2									
6	30	35	-3.38	.64	MAXIMUM MEASURE					.53	.58	100.0	100.0		
	.86	.289	1.00	-.64	3-4-1	1									
7	31	35	-3.83	.70	MINIMUM MEASURE					.40	.55	100.0	100.0		
	.89	.281	1.00	.57	1-4-3-2	2									
8	27	35			DROPPED										
					1-4-2-3	1									
10	0	35			INESTIMABLE: HIGH										
					2-4-3-1	3									
11	35	35			INESTIMABLE: LOW										
					1-3-1-2-4	4									
9		DELETED													
					1-3-2-4	2									
12		DESELECTED													
					1-3-2-4-3	1									
.....															
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----															
MEAN	18.5	35.0	-.59	.94	.96	.0	.68	-.1				89.9	90.0		
				.32											
P.SD	13.9	.0	4.21	.49	.28	.7	.58	.5				6.3	5.3		
				.27											
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----															

Column	Description
ENTRY NUMBER	the sequence number of the person, or item, in your data, and is the reference number used for deletion or anchoring.
TOTAL SCORE TOTAL COUNT	Totalscore=Yes -. This is the score when reading in the data.
NON-EXTREME SCORE NON-EXTREME COUNT	Totalscore=No - the raw score and count of response by a person on the test, or the sum of the scored responses to an item by the persons, omitting responses in extreme and inestimable scores. This is the score when estimating person abilities and item difficulties. Scored responses are transformed (re-counted) so that the lowest response is zero.
MEASURE	the estimate (or calibration) of the person ability (theta, B, beta, etc.), or the item difficulty (b, D, delta, etc.). Values are reported in logits with two decimal places, unless rescaled by USCALE= , UIMEAN= , UPMEAN= , UDECIM= . The difficulty of an item is defined to be the point on the latent variable at which its high and low categories are equally probable. SAFILE= can be used to alter this definition. If unexpected results are reported, check whether TARGET= or CUTLO= or CUTHI= or ISGROUPS= are specified.
A after MEASURE, MAXIMUM, etc.	see STATUS Table below
MODEL S.E. REAL S.E.	MODEL S.E. is the standard error of the estimate. REAL S.E. is the misfit-inflated standard error. These are commonly referred to as conditional standard errors of measurement (CSEM). Real S.E> while you are improving your results. This assumes misfit contradicts the Rasch model. Model S.E. when your results are as good as they can be. This assumes misfit is the randomness predicted by the Rasch model.
INFIT	an information-weighted statistic, which is more sensitive to unexpected behavior affecting responses to items near the person's measure level.

OUTFIT	an unweighted statistic, more sensitive to unexpected behavior by persons on items far from the person's measure level.										
MNSQ	<p>a mean-square statistic computed for all scores responses, excluding responses in extreme total scores. This is a chi-square statistic divided by its degrees of freedom. Its expectation is 1.0. Values substantially less than 1.0 indicate overfit = dependency in your data. Values values substantially greater than 1.0 indicate underfit = unmodeled noise. See dichotomous and polytomous fit statistics.</p> <table border="1" data-bbox="550 415 1463 732"> <thead> <tr> <th>Value</th> <th>Meaning</th> </tr> </thead> <tbody> <tr> <td>>2.0</td> <td>Off-variable noise is greater than useful information. Degrades measurement. Always remedy the large misfits first.</td> </tr> <tr> <td>>1.5</td> <td>Noticeable off-variable noise. Neither constructs nor degrades measurement</td> </tr> <tr> <td>0.5 - 1.5</td> <td>Productive of measurement</td> </tr> <tr> <td><0.5</td> <td>Overly predictable. Misleads us into thinking we are measuring better than we really are. (Attenuation paradox.). Misfits <1.0 are only of concern when shortening a test</td> </tr> </tbody> </table>	Value	Meaning	>2.0	Off-variable noise is greater than useful information. Degrades measurement. Always remedy the large misfits first.	>1.5	Noticeable off-variable noise. Neither constructs nor degrades measurement	0.5 - 1.5	Productive of measurement	<0.5	Overly predictable. Misleads us into thinking we are measuring better than we really are. (Attenuation paradox.). Misfits <1.0 are only of concern when shortening a test
Value	Meaning										
>2.0	Off-variable noise is greater than useful information. Degrades measurement. Always remedy the large misfits first.										
>1.5	Noticeable off-variable noise. Neither constructs nor degrades measurement										
0.5 - 1.5	Productive of measurement										
<0.5	Overly predictable. Misleads us into thinking we are measuring better than we really are. (Attenuation paradox.). Misfits <1.0 are only of concern when shortening a test										
ZSTD ZEMP LOG PROB	<p>the INFIT or OUTFIT mean-square fit statistic t standardized to approximate a theoretical "unit normal", mean 0 and variance 1, distribution. ZSTD (standardized as a z-score) is used of a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's t-statistic distribution value has been adjusted to a unit normal value. The standardization is shown on RSA, p.100-101. When LOCAL=Y, then ZEMP is shown, indicating a local {0,1} standardization. When LOCAL=LOG, then LOG is shown, and the natural logarithms of the mean-squares are reported. More exact values are shown in the Output Files. When LOCAL=PROB, the probability of the mean-square is shown.</p> <p>Ben Wright advises: "ZSTD is only useful to salvage non-significant MNSQ>1.5, when sample size is small or test length is short."</p>										
PTMEASUR-AL CORR. PTMEASUR-AL EXP.	<p>an observed point-correlation: PTBISERL-AL, PTBISERL-EX, PTMEASURE-A, PTMEASUR-EX, see Correlations. Negative reported correlations suggest that the orientation of the scoring on the item, or by the person, may be opposite to the orientation of the latent variable. This may be caused by item miskeying, reverse scoring, person special knowledge, guessing, data entry errors, or the expected randomness in the data.</p> <p>Correlations of 0.00 may be because the correlation cannot be calculated due to the structure of the data.</p> <p>In FIT ORDER Tables 6.1 and 10.1, letters A, B, ... indicating the identity of persons or items appearing on the Infit and Outfit plots, Tables 4, 5, 8, 9, precede the correlations. If the point-correlation is inestimable because items have different numbers of categories, this column does not appear, or only the FIT letters display.</p> <p>EXP. is the expected value of the point-correlation when the data fit the Rasch model with the estimated measures. See Correlations.</p>										
EXACT MATCH OBS% EXACT MATCH EXP%	<p>OBServed% is the percent of data points that are within 0.5 score points of their expected values, i.e., that match predictions.</p> <p>EXPEcted% is the percent of data points that are predicted to be within 0.5 score points of their expected values.</p>										
ESTIM DISCRIM	<p>an estimate of the 2-PL item discrimination, see DISCRIM=</p> <p>Negative discriminations are usually problematic and accompanied by negative point-biserial correlations. These indicate that the scoring on the item may be contradicting the overall latent variable. However, this is not a universal rule, so please look at the infit and</p>										

	outfit mean-squares to see whether they also indicate problems (values much greater than 1.0). Exceptions to the rule include very hard and very easy items, and situations where the person sample variance is very small.
ASYMPTOTE LOWER ASYMPTOTE UPPER	estimates of the upper and lower asymptotes for dichotomous items, see ASYMPTOTE=
P-VALUE	the observed proportion-correct (p-value) for 0/1 dichotomies, or observed average rating on the item, see PVALUE= .
RMSR	root-mean-square-residual of observations not in extreme scores, see RMSR= .
WMLE	Warm's Mean Likelihood Estimate , see WMLE=
QCML	Quasi-Conditional Maximum Likelihood Estimate , see QCML=
WEIGH	the weight assigned by IWEIGHT= or PWEIGHT= . When WEIGHT = 0.0, the item or person is estimated, but does not influence the estimate of any other item or person.
DISPLACE	<p>the displacement of the reported MEASURE from its data-derived value. This should only be shown with anchored measures. The displacement values can be seen in IFILE= and PFILE= output files.</p> <p>The displacement is an estimate of the amount to add to the MEASURE to make it conform with the data.</p> <p>+ logit displacement for a person ability indicates that the observed person score is higher (higher ability) than the expected person score based on the reported measure (anchor value).</p> <p>+ logit displacement for an item difficulty indicates that the observed item score is lower (harder item) than the expected item score based on the reported measure (anchor value).</p> <p><i>Unanchored measures:</i> If small displacements are shown, tighten the convergence criteria, CONVERGE=, LCONV=, RCONV=</p> <p><i>Anchored measures:</i> we expect half the displacements to be negative and half to be positive, and for them to be normally distributed according to the standard errors of the measures.</p>
PERSON ITEM	the name of the list of persons (data rows) or items (data columns) reported here
G	the grouping code assigned with ISGROUPS= . Table 3.2 , 3.3, etc. show the details of rating scales, etc., for each grouping code.
M	model code assigned with MODELS=
SUBSET:	the person or items are in incomparable subsets, see Subsets .

<i>Reported</i>	<i>STATUS</i> PFILE= , IFILE=	<i>Measured?</i>	<i>Reason</i>
A (after Measure)	2	Yes	Anchored (fixed) measure. The reported S.E. is that which would have been obtained if the value had been estimated. Values are reported in logits with two decimal places, unless rescaled by USCALE= , UDECIM=
(MEASURE)	1	Yes	Estimated measure
MINIMUM	0	Yes	Extreme minimum score. Measure estimated using EXTRSC= MINIMUM and MAXIMUM measures may occur because other MINIMUM and MAXIMUM measures have been dropped from the estimation. Inspect with TOTALSCORE= No.
MAXIMUM	-1	Yes	Extreme maximum score. Measure estimated using EXTRSC= MINIMUM and MAXIMUM measures may occur because other MINIMUM and MAXIMUM measures have been dropped from

			the estimation. Inspect with TOTALSCORE=No.
DROPPED	-2	No	No responses available for measurement. Perhaps due to CUTLO= , CUTHI= , CODES= , or deletion of other persons or items.
DELETED	-3	No	Deleted by user. PDELETE= , PDFILE= , IDDELETE= , IDFILE= , PSELECT= , ISELECT=
INESTIMABLE: HIGH	-4	No	Inestimable: high (all responses in the same category with ISGROUPS=0 or CUTHI=). The measure probably has a high value. For an item, this can be resolved using SAFILE= or grouping this item with a similar estimable item: SAFILE=* 23 0 0 ; item-number bottom-category 0 23 1 0 ; item-number top-category 0 *
INESTIMABLE: LOW	-5	No	Inestimable: low (all responses in the same category with ISGROUPS=0 or CUTLO=). The measure probably has a low value. For an item, this can be resolved using SAFILE= or grouping this item with a similar estimable item: SAFILE=* 23 0 0 ; item-number bottom-category 0 23 1 0 ; item-number top-category 0 *
DROPPED A (after Measure)	-6	Yes	Anchored (fixed) measure with no observed raw score
DESELECTED	-7 to -16	No	Temporarily deselected by Specification box with iSELECT= , etc. (usual STATUS - 10)
OMITTED	-17 to -26	No	Temporarily deleted by Specification box with iDELETE= , etc. (usual STATUS - 20)
REMOVED	-27 to -36	No	Temporarily deselected and deleted by Specification box with iSELECT= , etc., and iDELETE= , etc. (usual STATUS - 30)

Example: To eliminate misfitting items, output Table 10.1. Copy into Notepad++ (freeware) or other software capable of rectangular copy (alt+mouse). Then paste into your Winsteps control file

[IDFILE=*](#)

pasted entry numbers - put enough blank lines here first for all the numbers

*

Another approach is to screen out very unexpected answers, rather than whole people. Do this with [CUTLO= -2](#)

14.14.2 Table 10.2 Item statistics graphically in misfit order

Controlled by [USCALE=](#), [UMEAN=](#), [UDECIM=](#), [LOCAL=](#), [FITLOW=](#), [FITHIGH=](#), [MNSQ=](#)

Specify [CHART=YES](#) to produce Tables like this.

With [MNSQ=Yes](#):

```

PUPIL FIT GRAPH:  OUTFIT ORDER
+-----+-----+-----+-----+-----+-----+
| ENTRY | MEASURE | INFIT MEAN-SQUARE | OUTFIT MEAN-SQUARE | PUPIL |
| NUMBER | - + 0 | 0.7 1 1.3 | 2|0 0.7 1 1.3 | 2| |
+-----+-----+-----+-----+-----+-----+
| 72 | * | | | | * | A | : | : | : | * | JACKSON, SOLOMON |
| 47 | * | | | | | J | : | : | : | | VAN DAM, ANDY |
| 53 | * | | * | : | | K | : | : | : | * | SABOL, ANDREW |
| 32 | * | | * | : | | w | * | : | : | | ROSSNER, JACK |
| 21 | * | * | : | : | | a | * | : | : | | EISEN, NORM L. |
+-----+-----+-----+-----+-----+-----+

```

The fit information is shown in graphical format to aid the eye in identifying patterns and outliers. The fit bars are positioned by [FITLOW=](#) and [FITHIGH=](#). They may also be repositioned using [TFILE=](#).

With [MNSQ=No:](#), the ZSTD (*t*-standardized) statistics are reported:

ENTRY NUMBER	MEASURE	INFIT t standardized						OUTFIT t standardized						TAPS	
		-3	-2	-1	0	1	2	3	-3	-2	-1	0	1		2
18	E	:	:	:	:	:	:	:	:	:	:	:	:	:	4-1-3-4-2-1-4
15	*	:	*	:	:	:	:	*	:	:	:	:	:	:	1-3-2-4-1-3
14	*	:	:	*	:	:	:	*	:	:	:	:	:	:	1-4-2-3-4-1
12	*	:	:	*	:	:	:	*	:	:	:	:	:	:	1-3-2-4-3
13	*	:	*	:	:	:	:	*	:	:	:	:	:	:	1-4-3-2-4
11	*	:	*	:	:	:	:	*	:	:	:	:	:	:	1-3-1-2-4
10	*	:	*	:	:	:	:	*	:	:	:	:	:	:	2-4-3-1
7	*	:	*	:	:	:	:	*	:	:	:	:	:	:	1-4-3-2
4	*	:	*	:	:	:	:	*	:	:	:	:	:	:	1-3-4
1	E	:	:	:	:	:	:	:	:	:	:	:	:	:	Cube 1, then 4

"OMIT" appears in fit-ordered Tables, where items better fitting than [FITL=](#), or persons better fitting than [FITP=](#), are excluded.

14.14.3 Table 10.3 Item option & distractor frequencies in misfit order

(controlled by [Distractors=Y](#), [OSORT=](#), [CFILE=](#), [PTBIS=](#))

ITEM OPTION FREQUENCIES are output if [Distractors=Y](#). These show occurrences of each of the valid data codes in [CODES=](#), and also of [MISSCORE=](#) in the input data file. Counts of responses forming part of extreme scores are included. Only items included in the corresponding main table are listed. These statistics are also in [DISFILE=](#), which includes entries even if the code is not observed for the item. See also [Distractor Analysis](#).

[OSORT=](#) controls the ordering of options within items. The standard is the order of data codes in [CODES=](#).

ITEM CATEGORY/OPTION/DISTRACTOR FREQUENCIES: ENTRY ORDER

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	DATA %	ABILITY MEAN	P.S.D	S.E. MEAN	INFT MNSQ	OUTF MNSQ	PTMA CORR.	ITEM	
13	4	***	11	16#	.79	1.41	.45			.08	M. STAIRS	4 75% Independent
	1	1	30	52	-1.85	.99	.18	.8	1.0	-.89		1 0% Independent
	3	3	5	9	1.07	.80	.40	.6	.4	.10		3 50% Independent
	5	5	15	26	2.63	1.06	.28	1.7	1.5	.56		5 Supervision
	6	6	7	12	3.25	.81	.33	1.2	1.2	.44		6 Device
	7	7	1	2	4.63	.00		.7	.6	.23		7 Independent
	MISSING	***	1	1#	-1.99	.00				-.12		

* Average ability does not ascend with category score
Missing % includes all categories. Scored % only of scored categories

ENTRY NUMBER is the item sequence number.

The letter next to the sequence number is used on the fit plots.

DATA CODE is the response code in the data file.

MISSING means that the data code is not listed in the [CODES=](#) specification.

Codes with no observations are not listed.

SCORE VALUE is the value assigned to the data code by means of [NEWSCORE=](#), [KEY1=](#), [IVALUE=](#), etc.

*** means the data code is missing and so ignored, i.e., regarded as not administered. [MISSCORE=1](#) scores missing data as "1".

DATA COUNT is the frequency of the data code in the data file (unweighted) - this includes observations for both non-extreme and extreme persons and items. For counts weighted by [PWEIGHT=](#), see [DISFILE=](#)

DATA % is the percent of scored data codes. For dichotomies, the % are the [proportion-correct-values](#) for the options. For data with score value "****", the percent is of all data codes, indicated by "#".

ABILITY MEAN is the observed, sample-dependent, [average measure](#) of persons (relative to each item) in this analysis who responded in this category (adjusted by [PWEIGHT=](#)). This is equivalent to a "Mean Criterion Score" (MCS) expressed as a measure. It is a sample-dependent quality-control statistic for this analysis. (It is not the sample-independent value of the category, which is obtained by adding the item measure to the "score at category", in Table 3.2 or higher, for the rating (or partial credit) scale corresponding to this item.) For each observation in category k, there is a person of measure Bn and an item of measure Di. Then: average measure = sum(Bn - Di) / count of observations in category.

An "*" indicates that the average measure for a higher score value is lower than for a lower score value. This contradicts the hypothesis that "higher score value implies higher measure, and vice-versa".
 The "average ability" for missing data is the average measure of all the persons for whom there is no response to this item. This can be useful. For instance, we may expect the "missing" people to be high or low performers, or to be missing random (and so they average measure would be close to the average of the sample).
 These values are plotted in [Table 2.6](#).

ABILITY P.SD is the population standard deviation of the ABILITY values = $\sqrt{(\sum (ABILITY - (ABILITY MEAN))^2 / COUNT)}$

S.E. MEAN is the standard error of the mean (average) measure of the sample of persons from a population who responded in this category (adjusted by PWEIGHT) = $\sqrt{(\sum (ABILITY - (ABILITY MEAN))^2 / (COUNT * (COUNT - 1)))}$

INFT MNSQ is the [Infit Mean-Square](#) for observed responses in this category (weighted by [PWEIGHT=](#), and omitting responses in extreme person scores). Values greater than 1.0 indicate unmodeled noise. Values less than 1.0 indicate loss of information.

OUTF MNSQ is the [Outfit Mean-Square](#) for observed responses in this category (weighted by [PWEIGHT=](#), and omitting responses in extreme person scores). Values greater than 1.0 indicate unmodeled noise. Values less than 1.0 indicate loss of information.

PTMA CORR is the point-correlation between the data code, scored 1, or non-occurrence, scored 0, of this category or distractor and the person raw scores or measures chosen by [PTBISERIAL=](#). The computation is described in [Correlations](#). Example: for categories 0,1,2, then the correlation is between [1 for the target score (0, 1, or 2) and 0 for the other scores (1 and 2, 0 and 2, or 0 and 1)] and the person ability measures for the persons producing each score.

ITEM (here, ACT) is the name or label of the item.

Data codes and Category labels are shown to the right of the box, if CLFILE= or CFILE= is specified.

* Average ability does not ascend with category score. The average ability of the persons observed in this category is lower than the average ability of the persons in the next lower category. This contradicts the Rasch-model assumption that "higher categories <-> higher average abilities."

Missing % includes all categories. Scored % only of scored categories. The percentage for the missing category is based on all the COUNTS. The percentages for the SCOREd categories are based only on those category COUNTS.

"BETTER FITTING OMIT" appears in fit-ordered Tables, where items better fitting than [FIT=](#) are excluded.

14.14.4 Table 10.4 Item most-misfitting response strings

Table 10.4 displays the unexpected responses in the up to 26 most-misfitting item-response strings.

```

MOST MISFITTING RESPONSE STRINGS
ACT                                OUTFITMNSQ | PUPIL
|4314 452371 17667422 315231 4 4 31725
|147578953418917534874255649749502293
high-----
23 MATCH A PAT                    5.68 A|1110.0.....2.2...2...1112.2211122.
20 MATCH BUGS                     1.99 B|.....0.....2.2.....2.....11..21.
9  LEARN NEED NAMES              1.29 D|.....0.....0.....1
|-----low
|4314745237181766742243152319464531725
|1475 895341 91753487 255664 7 9 02293
    
```

In this Table, the item (rows) are sorted:

OUTFIT=Yes	Sorted by Outfit Mean-square descending
OUTFIT=No	Sorted by bigger of Outfit Mean-square and Infit mean-square descending.

The person (columns) are sorted by measure. The most able person (highest-scoring) is on the left. The person entry numbers are in each column (vertically). In the Figure, the first is person 41. The persons shown are those with the most unexpected responses. [UCOUNT=](#) sets the maximum number of persons to report. [LINELENGTH=](#) controls the width of the table. "high" indicates where we expect to see high observations. "low" indicates where we expect to see low observations. Expected observations are replaced by ".".

Each row contains:

- Item entry number
- Item label
- Item outfit mean-square
- Item letter (A-Z) in [Table 10.1](#).

The responses to the items:

0, 1, 2 and numeric values	Scored unexpected-responses
•	Expected responses, standardized residuals < 2
(blank)	Missing data

The 26 items at the top of [Table 10.1](#) are shown, excluding those with no unexpected responses on the displayed persons.

14.14.5 Table 10.5 Item most-unexpected observations

Table 10.5 displays the item response strings with the most-unexpected responses as shown in Table 10.6.

```

MOST UNEXPECTED RESPONSES
ACT          MEASURE  |KID
              |414 2377166764321 35324 14 13725
              |175753419573245754256679649520293
              high-----
18 GO ON PICNIC    -3.15 D|.....0....1...1...1.....
19 GO TO ZOO      -2.48 H|.....0.....1..1.....
20 WATCH BUGS     1.83 C|.....2...2...2.....1..21.
23 WATCH A RAT    2.18 A|1.0...2.2....2.....22.211122.
 5 FIND BOTTLES AND CANS 2.42 B|.0....2.2...2....2.212.112..11
              |-----low-
              |414723771667643214353249614513725
              |175 5341957324575 25667 49 20293
    
```

In this Table, the items (rows) are sorted by measure, ascending.

The persons (columns) are sorted by measure, descending. The most-able person (highest-scoring) is on the left. The person entry-numbers are in each column (vertically). In the Figure, the first is person 41. The persons shown are those with the most-unexpected responses in Table 10.6. [UCOUNT=](#) sets the maximum number of persons to report. [LINELENGTH=](#) controls the width of the table. "high" indicates where we expect to see high observations. "low" indicates where we expect to see low observations. Expected observations are replaced by ".".

Each row contains:

- Item entry number
- Item label
- Item measure
- Item letter (A-Z) in the plot in [Table 10.1](#).

The responses to the items:

0, 1, 2 and numeric values	Scored unexpected-responses
•	Expected responses, standardized residuals < 2
(blank)	Missing data

The 26 items with the most-unexpected individual responses are shown, excluding those with no unexpected responses on the displayed items.

14.14.6 Table 10.6 Item most-unexpected response list

This shows the most unexpected responses sorted by unexpectedness (standardized residual). Large standardized residuals contribute to large outfit mean-square fit statistics. [UCOUNT=](#) sets the maximum number of "most unexpected responses" to report in Table 10.6.

```

MOST UNEXPECTED RESPONSES
+-----+
    
```

DATA	OBSERVED	EXPECTED	RESIDUAL	ST. RES.	MEASDIFF	ACT	KID	ACT	KID
0	0	1.93	-1.93	-7.66	3.53	18	73	GO ON PICNIC	SANDBERG, RYNE
2	2	.07	1.93	7.57	-3.50	23	72	WATCH A RAT	JACKSON, SOLOMON
2	2	.07	1.93	7.57	-3.50	23	29	WATCH A RAT	LANDMAN, ALAN
0	0	1.93	-1.93	-7.41	3.46	19	71	GO TO ZOO	STOLLER, DAVE

DATA is the response code in the data file
OBSERVED is the code's value after rescoring
EXPECTED is the predicted observation based on the person and item estimated measures
RESIDUAL is (OBSERVED - EXPECTED), the difference between the observed and expected values
ST. RES. is the standardized residual, the unexpectedness of the residual expressed as a unit normal deviate
MEASDIFF is the difference between the ability and difficulty estimates. This produces the EXPECTED value.
ACT is the item entry number
KID is the person entry number
ACT is the item label
KID is the person label

14.15 Table 11.1 Item misfitting responses

(controlled by [FITI=](#), [LINELENGTH=](#), [MNSQ=](#), [OUTFIT=](#), [T7OPTIONS=](#))

These tables show the items for which the t standardized outfit (or infit, if [OUTFIT=N](#)) statistic is greater than the misfit criterion ([FITI=](#)). [FITI=0](#) displays all items. Items are listed in descending order of misfit. The response codes are listed in their sequence order in your data file. The residuals are standardized response score residuals, which have a model led expectation of 0, and a variance of 1. Negative residuals indicate that the observed response was less correct (or, for rating (or partial credit) scales, lower down the rating scale) than expected. The printed standardized residual is truncated, not rounded, so that its actual value is at least as extreme as that shown. Standardized residuals between -1 and 1 are not printed. For exact details, see [XFILE=](#). "M" indicates a missing response.

"X" indicates that the person obtained an extreme score."X" in Tables 7.1 and 11.1 indicates that the observation is part of an extreme score, and so does not have a standardized residual. All residuals for extreme scores are conceptually zero. Persons with "X" in Item Table 11.1 (because they have extreme scores) do not appear in Table 7.1, and *vice-versa*, because those extreme persons and items are not part of the misfit process. Basically "X" says "this observation is excluded from the misfit computation".

For Table 11, the diagnosis of misfitting items, items with a t standardized fit greater than [FITI=](#) are reported. Selection is based on the OUTFIT statistic, unless you set [OUTFIT=N](#) in which case the INFIT statistic is used. [T7OPTIONS=](#) selects the detail lines to display:

```

TABLE OF POORLY FITTING ITEMS      (PERSONS IN ENTRY ORDER)
NUMBER  NAME      POSITION      MEASURE  INFIT (ZSTD)  OUTFIT  MISFIT OVER 2.0

23 WATCH A RAT      2.00      5.8  A      8.1
  RESPONSE:  1:  0 2 1 1 1  2 2 0 2 0  0 1 0 1 1  0 1 0 0 0  0 1 1 0 0
  Z-RESIDUAL:      X      2 3      3      2

  RESPONSE:  26:  1 2 0 2 1  M 0 0 1 1  1 0 1 0 0  1 0 0 0 0  0 0 0 2 1
  Z-RESIDUAL:      3 6 2      -2      -2      4

                                v This is on the fit plots
5 FIND BOTTLES AND CANS      2.21      5.2  B      6.5
  RESPONSE:  1:  1 2 0 0 1  2 0 0 1 0  0 2 0 0 1  1 0 0 0 0  0 1 1 0 1
  Z-RESIDUAL:      X      2 4      6      -2

```

Example: Table 11.1 with [T7OPTIONS= OERZ](#) in an analysis of [Example0.txt](#):

```

TABLE OF POORLY FITTING ACT      (KID IN ENTRY ORDER)
NUMBER - NAME -- POSITION ----- MEASURE - INFIT (MNSQ)  OUTFIT

23 Watch a rat      2.18      2.4  A      4.1
OBSERVED:  1:  0 2 1 1 1  2 2 0 2 0  2 2 0 2 0  <- T7OPTIONS= O
EXPECTED:      0.4 X  0.5 0.3 0.1  0.2 1.2 0.5  0.2 0.3 <- T7OPTIONS= E
RESIDUAL:      -0.4  0.5 0.7 0.9  1.8 0.8 -0.5  1.8 -0.3 <- T7OPTIONS= R
Z-RESIDUAL:      2 4      4      4      <- T7OPTIONS= Z

```

The OBSERVED value is the scored response, before recounting due to STKEEP=No, if any.

The EXPECTED value is the OBSERVED value - RESIDUAL value.

14.16 Table 12 Item maps with full item labels

(controlled by [LINELENGTH=](#), [IPEXTREME=](#), [MRANGE=](#), [MTICK=](#), [MAXPAGE=](#), [NAMLMP=](#), [ISORT=](#), [T1P#=#](#))

[LINELENGTH=](#)200 or higher if the displayed labels are too short.

Table 12.2 Item map with full labels		
MEASURE	KID - MAP - ACT <more> <rare>	In Table 12.2, the full item names are shown located at their calibrated MEASURE in Table 13.1 , along with the person distribution according to Table 17.1 . You can use NAMLMP= or IMAP= to control the number of characters of each name reported. MAXPAGE= controls the length of the Table. MRANGE= controls the displayed range of measures. ISORT= controls the sort order within rows. You can adjust these values from the " Specification " pull-down menu and then select Table 12 from the Output Tables menu. Thus you can experiment without needing to rerun the analysis.
5	. + . 	Where there are more items than can be shown on one line, the extra items are printed on subsequent lines, but the latent variable "I" does not advance and is left blank.
4	. + . . T	
3	. + . T ## FIND BOTTLES AND CANS WATCH A RAT	In subtable 12.12, the items are arranged by easiness. The item hierarchy is reversed.
2	.## S+ ## LOOK IN SIDEWALK CRACKS WATCH BUGS WATCH GRASS CHANGE	
1	.### IS .#### M+ WATCH ANIMAL MOVE .## LEARN WEED NAMES MAKE A MAP TALK W/FRIENDS ABOUT PLANTS ##### LOOK AT PICTURES OF PLANTS LOOK UP STRANGE ANIMAL OR PLANT READ BOOKS ON PLANTS	T1P#=# sets the number of persons for each #
0	.##### +M FIND OUT WHAT ANIMALS EAT WATCH WHAT ANIMALS EAT # S FIND OUT WHAT FLOWERS LIVE ON READ ANIMAL STORIES WATCH BIRDS # READ BOOKS ON ANIMALS # + FIND WHERE ANIMAL LIVES WATCH BIRD MAKE NEST # IS GROW GARDEN LISTEN TO BIRD SING	Items arranged by measure: Look for the hierarchy of item names to spell out a meaningful construct from easiest (highest proportion-correct-value or highest average rating) at the bottom to hardest (lowest proportion-correct-value or lowest average rating) at the top.
-1	. T + GO TO MUSEUM GO TO ZOO T + GO ON PICNIC	Column labels in Tables 12.2, 12.5, 12.6, 12.7, 12.8
-2	<less> <freq>	PERSON= MAP ITEM=
-3	EACH "#" IS 2: EACH "." IS 1	MORE= T RARE= LESS= ↓ FREQ= frequent

Table 12.5 Item map showing expected score zones		
MEASURE	KID - MAP - ACT - Expected score zones (Rasch-half-point thresholds) <more> <rare> Neutral Like	
5	. + . 	
4	. + . . T	FIND BOTTLES AND CANS .15 WATCH A RAT .15 LOOK IN SIDEWALK CRACKS .15 WATCH BUGS .15 WATCH GRASS CHANGE .15
3	. + . T ##	WATCH ANIMAL MOVE .15 LEARN WEED NAMES .15 LOOK AT PICTURES OF PLANTS .15 MAKE A MAP .15 READ BOOKS ON PLANTS .15 TALK W/FRIENDS ABOUT PLANTS .15
2	.## S+ ## .### IS FIND BOTTLES AND CANS .05	FIND OUT WHAT ANIMALS EAT .15 LOOK UP STRANGE ANIMAL OR PLAN.15

Table 12.5 is only produced when there are polytomous items. It shows the items positioned at the lower edge of each expected score zone (the half-point thresholds). The expected score zone above "item-label .2½" extends from expected category score 2.5 to 3.5. Lower than "item-label.0½" is the zone from 0 to 0.5. These correspond to the thresholds on the [Expected Score ICC](#).

If you put the item number at the start of the item labels after

1	.#### M+	WATCH A RAT	.05	WATCH WHAT ANIMALS EAT	.15
	.##	WATCH BUGS	.05	FIND OUT WHAT FLOWERS LIVE ON	.15
		WATCH GRASS CHANGE	.05	READ ANIMAL STORIES	.15
	#####	LOOK IN SIDEWALK CRACKS	.05	WATCH BIRDS	.15
				FIND WHERE ANIMAL LIVES	.15
				READ BOOKS ON ANIMALS	.15
				WATCH BIRD MAKE NEST	.15
0	.##### +M	WATCH ANIMAL MOVE	.05	GROW GARDEN	.15
	# S	LEARN WEED NAMES	.05	LISTEN TO BIRD SING	.15
	#	LOOK AT PICTURES OF PLANTS	.05		
		MAKE A MAP	.05		
		READ BOOKS ON PLANTS	.05		
		TALK W/FRIENDS ABOUT PLANTS	.05		
-1	# +	FIND OUT WHAT ANIMALS EAT	.05	GO TO MUSEUM	.15
		LOOK UP STRANGE ANIMAL OR PLAN.	.05		
		WATCH WHAT ANIMALS EAT	.05		
	# S	FIND OUT WHAT FLOWERS LIVE ON	.05	GO TO ZOO	.15
	. T	READ ANIMAL STORIES	.05		
		WATCH BIRDS	.05		
-2	+	FIND WHERE ANIMAL LIVES	.05	GO ON PICNIC	.15
		READ BOOKS ON ANIMALS	.05		
		WATCH BIRD MAKE NEST	.05		
		GROW GARDEN	.05		
	T	LISTEN TO BIRD SING	.05		
-3	+				
		GO TO MUSEUM	.05		
		GO TO ZOO	.05		
-4	+				
		GO ON PICNIC	.05		
-5	+				
		<less> <freq> Dislike		Neutral	
		EACH "#" IS 2: EACH "." IS 1			

&END, you can show only the item numbers on this plot by using [NAMLMP=](#) or [IMAP=](#). Column headings are the category labels that match the (rescored) category numbers in [CFILE=](#) or [CLFILE=](#).

Where there are more items than can be shown on one line, the extra items are printed on subsequent lines, but the latent variable "|" does not advance and is left blank. The plotted values are in the [ISFILE=](#).

The half-point thresholds are indicated by .n½ when [ASCII=No](#) or [ASCII=Web](#), or .n5 when [ASCII=Yes](#)

Increase the length of the displayed lines with [LINELENGTH=](#)

Table 12.6 Item map showing 50% cumulative probabilities (Rasch-Thurstone thresholds)

MEASURE	KID - MAP - ACT - 50% Cumulative probabilities (Rasch-Thurstone thresholds)			
	<more> <rare>	Neutral	Like	
5	. +			
	.			
	.			
4	. +			
	.			
	. T		FIND BOTTLES AND CANS	.2
			WATCH A RAT	.2
3	. +			
	. T		LOOK IN SIDEWALK CRACKS	.2
			WATCH BUGS	.2
			WATCH GRASS CHANGE	.2
	##			
2	.## S+		WATCH ANIMAL MOVE	.2
	##		LEARN WEED NAMES	.2
			MAKE A MAP	.2
			TALK W/FRIENDS ABOUT PLANTS	.2
	.### S	FIND BOTTLES AND CANS	.1	LOOK AT PICTURES OF PLANTS
		WATCH A RAT	.1	LOOK UP STRANGE ANIMAL OR PLANT
				.2
			READ BOOKS ON PLANTS	.2
1	.#### M+		FIND OUT WHAT ANIMALS EAT	.2
			WATCH WHAT ANIMALS EAT	.2
	.##	LOOK IN SIDEWALK CRACKS	.1	FIND OUT WHAT FLOWERS LIVE ON
		WATCH BUGS	.1	READ ANIMAL STORIES
		WATCH GRASS CHANGE	.1	WATCH BIRDS
	#####		READ BOOKS ON ANIMALS	.2
0	.##### +M	WATCH ANIMAL MOVE	.1	FIND WHERE ANIMAL LIVES
			WATCH BIRD MAKE NEST	.2
	# S	LEARN WEED NAMES	.1	GROW GARDEN
		MAKE A MAP	.1	LISTEN TO BIRD SING
		TALK W/FRIENDS ABOUT PLANTS	.1	
	#	LOOK AT PICTURES OF PLANTS	.1	
		LOOK UP STRANGE ANIMAL OR PLANT	.1	
		READ BOOKS ON PLANTS	.1	
-1	# +	FIND OUT WHAT ANIMALS EAT	.1	GO TO MUSEUM
		WATCH WHAT ANIMALS EAT	.1	
	# S	FIND OUT WHAT FLOWERS LIVE ON	.1	GO TO ZOO
		READ ANIMAL STORIES	.1	
		WATCH BIRDS	.1	
	. T	READ BOOKS ON ANIMALS	.1	
-2	+	FIND WHERE ANIMAL LIVES	.1	GO ON PICNIC
		WATCH BIRD MAKE NEST	.1	

Table 12.6 is only produced when there are polytomous items.

The right-hand column shows the items positioned at median [50% cumulative probability](#) (the Rasch-Thurstone thresholds) at the lower edge of each rating probability zone. Above label.1, categories 1 and above are more probable. Below label.2, categories 1 and below are most probable. Between label.1 and label.2 is the zone which can be thought of as corresponding to a rating of 1. These correspond to the [Cumulative Probability](#) thresholds.

If you put the item number at the start of the item labels after &END, you can show only the item numbers on this plot by using [NAMLMP=](#) or [IMAP=](#). Columns are headed by the (rescored) categories in [CFILE=](#).

Where there are more items than can be shown on one line, the extra items are printed on subsequent lines, but the latent

	GROW GARDEN	.1	
	LISTEN TO BIRD SING	.1	
	T		
-3	+ GO TO MUSEUM	.1	
	GO TO ZOO	.1	
-4	+ GO ON PICNIC	.1	
-5	+ <less> <freq> Dislike		Neutral
EACH "#" IS 2: EACH "." IS 1			

variable "|" does not advance and is left blank. The plotted values are in the [ISFILE=](#).

Left-hand column shows the distribution of person ability measures along the variable.

Table 12.7 Item map with Andrich thresholds (modal categories if ordered)

MEASURE	KID - MAP - ACT - Andrich thresholds (modal categories if ordered)	Neutral	Like
5	. + . .		
4	. + . . T		
3	. + . T		FIND BOTTLES AND CANS .2 WATCH A RAT .2 LOOK IN SIDEWALK CRACKS .2 WATCH BUGS .2 WATCH GRASS CHANGE .2
2	## ## S+ ## FIND BOTTLES AND CANS .1 ### S WATCH A RAT .1		WATCH ANIMAL MOVE .2 LEARN WEED NAMES .2 LOOK AT PICTURES OF PLANTS .2 MAKE A MAP .2 READ BOOKS ON PLANTS .2 TALK W/FRIENDS ABOUT PLANTS .2
1	##### M+ WATCH BUGS .1 WATCH GRASS CHANGE .1 ## LOOK IN SIDEWALK CRACKS .1 ##### WATCH ANIMAL MOVE .1		FIND OUT WHAT ANIMALS EAT .2 LOOK UP STRANGE ANIMAL OR PLANT .2 WATCH WHAT ANIMALS EAT .2 FIND OUT WHAT FLOWERS LIVE ON .2 READ ANIMAL STORIES .2 WATCH BIRDS .2
0	##### +M LEARN WEED NAMES .1 # S LOOK AT PICTURES OF PLANTS .1 MAKE A MAP .1 READ BOOKS ON PLANTS .1 TALK W/FRIENDS ABOUT PLANTS .1 # FIND OUT WHAT ANIMALS EAT .1 LOOK UP STRANGE ANIMAL OR PLANT .1 WATCH WHAT ANIMALS EAT .1		FIND WHERE ANIMAL LIVES .2 READ BOOKS ON ANIMALS .2 WATCH BIRD MAKE NEST .2 GROW GARDEN .2
-1	# + # S FIND OUT WHAT FLOWERS LIVE ON .1 READ ANIMAL STORIES .1 WATCH BIRDS .1 . T FIND WHERE ANIMAL LIVES .1 READ BOOKS ON ANIMALS .1 WATCH BIRD MAKE NEST .1		GO TO MUSEUM .2 GO TO ZOO .2
-2	+ GROW GARDEN .1 LISTEN TO BIRD SING .1 T		GO ON PICNIC .2
-3	+ GO TO MUSEUM .1 GO TO ZOO .1 		
-4	+ GO ON PICNIC .1 		
-5	+ <less> <freq> Dislike		Neutral
EACH "#" IS 2: EACH "." IS 1			

Table 1.7 is only produced when there are polytomous items.

The right-hand column shows the items positioned at their Andrich thresholds (points of equal probability of adjacent categories) at the lower edge of each rating probability zone. When the thresholds are ordered, above label.2, the most probable category is 2. Below label.1, the most probable category is 0. Between label.1 and label.2 is the zone which can be thought of as corresponding to a rating of 1.

If you put the item number at the start of the item labels after &END, you can show only the item numbers on this plot by using [NAMLMP=](#) or [IMAP=](#). Columns are headed by the (rescored) categories in [CFILE=](#).

Where there are more items than can be shown on one line, the extra items are printed on subsequent lines, but the latent variable "|" does not advance and is left blank. The plotted values are in the [ISFILE=](#).

Left-hand column shows the distribution of person ability measures along the variable.

Table 12.8 Item map of measures for category scores (maximum probability of observing a category)

MEASURE	KID - MAP - ACT - Measures for category scores (maximum probability of observing a category)	Dislike	Neutral	Like
7	. + . .			
6	. +			

Table 12.8 is only produced when there are polytomous items.

The right-hand column shows the items

<pre> 3 X X + X T XX XXXX S 2 XXX + XX XX XXXX S XXX 1 XXXXXX + GO TO ZOO.2 WATCH GRASS CHAN.3 WATCH A RAT.2 WATCH A RAT.3 FIND BOTTLES AND.2 WATCH BUGS.3 XXXXX M FIND BOTTLES AND.3 WATCH A RAT.4 XXX LOOK IN SIDEWALK.2 MAKE A MAP.3 WATCH BUGS.2 LOOK IN SIDEWALK.3 XXXXXXXX GO ON PICNIC.2 WATCH WHAT ANIMA.3 WATCH GRASS CHAN.2 WATCH BIRDS.3 LEARN WEED NAMES.2 READ ANIMAL STOR.3 WATCH ANIMAL MOV.2 0 XXXXXXXX +M MAKE A MAP.2 READ BOOKS ON AN.3 LISTEN TO BIRD S.3 GO ON PICNIC.3 FIND WHERE ANIMA.3 GROW GARDEN.3 WATCH BIRD MAKE.3 X TALK W/FRIENDS A.2 LOOK AT PICTURES.2 X S FIND OUT WHAT AN.2 GO TO MUSEUM.3 LOOK UP STRANGE.2 READ BOOKS ON PL.2 READ ANIMAL STOR.2 FIND OUT WHAT FL.2 WATCH BIRD MAKE.2 READ BOOKS ON AN.2 WATCH WHAT ANIMA.2 XX GROW GARDEN.2 GO TO ZOO.3 FIND WHERE ANIMA.2 WATCH BIRDS.2 -1 XX + XX S LISTEN TO BIRD S.2 X T -2 + </pre>	<p>If you put the item number at the start of the item labels after &END, you can show only the item numbers on this plot by using NAMLMP= or IMAP=. Columns are headed by the (rescored) categories in CFILE=.</p> <p>Where there are more items than can be shown on one line, the extra items are printed on subsequent lines, but the latent variable "]" does not advance and is left blank. The plotted values are in the ISFILE=.</p> <p>Left-hand column shows the distribution of person ability measures along the variable.</p>
--	--

Table 12.12 Item map with full labels (reversed measures)

```

MEASURE      KID - MAP - ACT
              <more>||<freq>
5            .  ++
              .  ||

```

In Table 12.12, the items are arranged by easiness. The item hierarchy is [reversed](#).

You can use [NAMLMP=](#) or [IMAP=](#) to control the number of characters of each name reported. [MAXPAGE=](#) controls the length of the Table. [MRANGE=](#) controls the displayed range of

```

4      .  ||
      .  ++
      .  ||
      .  T||
3      .  ++ GO ON PICNIC
      .  ||T
      ##  || GO TO ZOO
2      .## S++ GO TO MUSEUM
      ##  ||
      .### ||S GROW GARDEN
      LISTEN TO BIRD SING
1      .#### M++ FIND WHERE ANIMAL LIVES
      WATCH BIRD MAKE NEST
      .##  || READ BOOKS ON ANIMALS
      ##### || FIND OUT WHAT FLOWERS LIVE ON
      READ ANIMAL STORIES
      WATCH BIRDS
0      .##### ++M FIND OUT WHAT ANIMALS EAT
      WATCH WHAT ANIMALS EAT
      # S|| LOOK AT PICTURES OF PLANTS
      LOOK UP STRANGE ANIMAL OR PLANT
      READ BOOKS ON PLANTS
      #  || LEARN WEED NAMES
      MAKE A MAP
      TALK W/FRIENDS ABOUT PLANTS
-1     #  ++ WATCH ANIMAL MOVE
      #  ||S
      . T|| LOOK IN SIDEWALK CRACKS
      WATCH BUGS
      WATCH GRASS CHANGE
-2     ++
      || FIND BOTTLES AND CANS
      WATCH A RAT
      ||T
-3     ++
      <less>||<rare>
EACH "#" IS 2: EACH "." IS 1

```

measures. [ISORT=](#) controls the sort order within rows. You can adjust these values from the "[Specification](#)" pull-down menu and then select Table 12 from the [Output Tables](#) menu. Thus you can experiment without needing to rerun the analysis.

Where there are more items than can be shown on one line, the extra items are printed on subsequent lines, but the latent variable "|" does not advance and is left blank.

[T1P#](#) sets the number of persons for each #

Items arranged by easiness. The item difficulty measures are reversed relative to the mean person measure, "M". Look for the hierarchy of item names to spell out a meaningful construct from easiest at the top to hardest at the bottom.

The double line || indicates the two sides have opposite orientations.

Each "X" is one person.

Column Labels

```

PERSON=  MAP  ITEM=
MORE=    T    FREQ=
LESS=    |    RARE=

```

MEASURE	Location of person or item on the unidimensional latent variable, in logits or USCALE= units
<more>	higher person "ability" = higher raw score with complete data
<less>	lower person "ability" = lower raw score with complete data
<rare>	higher item "difficulty" = lower item "easiness" = lower score on the item with complete data: PVALUE=
<frequ> <frequent>	lower item "difficulty" = higher item "easiness" = higher score on the item with complete data: PVALUE=
M	Mean of person or item distribution
S	One standard deviation from the person or item mean
T	Two standard deviations from the person or item mean
X	one person or item
#	several persons or items, e.g., EACH "#" = 4. The value for # can be changed with T1I# for items, T1P# for persons.
.	between 1 and (# - 1) persons or items, e.g., if # = 4, then "." = 1 to 3
	Variable advances. Lines without " " have the same measure as the nearest " " above.
	The double line indicates the two sides have opposite orientations: person ability and item easiness. The item difficulties are reversed . This is useful if the items and persons are being compared to the responses.

<i>On plot:</i>	<i>Description:</i>
MEASURE	Location on the unidimensional latent variable, in logits or USCALE= units
<more>	higher person "ability" = higher raw score with complete data
<less>	lower person "ability" = lower raw score with complete data
<rare>	higher item "difficulty" = lower score on the item with complete data: PVALUE=
<frequent>	lower item "difficulty" = higher score on the item with complete data: PVALUE=

M	Mean of person or item distribution
S	One standard deviation from the person or item mean
T	Two standard deviations from the person or item mean
X	one person or item
#	several persons or items, e.g., "#" = 4. The value for # can be changed with I1#=# , I1P#=# .
.	between 1 and #-1 persons or items, e.g., "." = 1 to 3
or +	Variable advances. Lines without " " or "+" have the same measure as the nearest " " or "+" above.

14.17 Table 13 Item report in measure order

14.17.1 Table 13.1 Item statistics in measure order

(controlled by [USCALE=#](#), [UMEAN=#](#), [UDECI=#](#), [LOCAL=#](#), [TOTAL=#](#), [DISCRIMINATION=#](#), [ASYMPTOTE=#](#), [PVALUE=#](#))

ITEM STATISTICS: MEASURE ORDER

For details, please see [Table 10.1](#)

14.17.2 Table 13.2 Item statistics graphically in measure order

Please see [Table 10.2](#)

14.17.3 Table 13.3 Item option & distractor frequencies in measure order

(controlled by [Distractors=Y](#), [OSORT=#](#), [CFILE=#](#), [PTBIS=#](#))

ITEM OPTION FREQUENCIES are output if [Distractors=Y](#). These show occurrences of each of the valid data codes in [CODES=#](#), and also of [MISSCORE=#](#) in the input data file. Counts of responses forming part of extreme scores are included. Only items included in the corresponding main table are listed. These statistics are also in [DISFILE=#](#), which includes entries even if the code is not observed for the item. See also [Distractor Analysis](#).

[OSORT=#](#) controls the ordering of options within items. The standard is the order of data codes in [CODES=#](#).

```

ITEM CATEGORY/OPTION/DISTRACTOR FREQUENCIES:  ENTRY ORDER
-----
|ENTRY  DATA  SCORE |   DATA   |   ABILITY   S.E.  INFT  OUTF  PTMA |
|NUMBER CODE  VALUE |  COUNT  % |   MEAN  P.SD  MEAN  MNSQ  MNSQ  CORR. | ITEM
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|  13  4      *** |  11 16# |   .79   1.41 .45          .08 |M. STAIRS
|      1      1 |  30 52 | -1.85   .99 .18   .8  1.0  -.89 |
|      3      3 |   5  9 |   1.07   .80 .40   .6  .4  .10 |
|      5      5 |  15 26 |   2.63  1.06 .28  1.7  1.5  .56 |
|      6      6 |   7 12 |   3.25   .81 .33  1.2  1.2  .44 |
|      7      7 |   1  2 |   4.63   .00          .7  .6  .23 |
|      MISSING *** |   1 1# | -1.99   .00          -1.12 |
|-----|-----|-----|-----|-----|-----|-----|-----|
* Average ability does not ascend with category score
# Missing % includes all categories. Scored % only of scored categories

```

ENTRY NUMBER is the item sequence number.

The letter next to the sequence number is used on the fit plots.

DATA CODE is the response code in the data file.

MISSING means that the data code is not listed in the [CODES=#](#) specification.

Codes with no observations are not listed.

SCORE VALUE is the value assigned to the data code by means of [NEWSCORE=#](#), [KEY1=#](#), [IVALUE=#](#), etc.

*** means the data code is missing and so ignored, i.e., regarded as not administered. [MISSCORE=1](#) scores missing data as "1".

DATA COUNT is the frequency of the data code in the data file (unweighted) - this includes observations for both non-extreme and extreme persons and items. For counts weighted by [PWEIGHT=](#), see [DISFILE=](#)

DATA % is the percent of scored data codes. For dichotomies, the % are the [proportion-correct-values](#) for the options. For data with score value "****", the percent is of all data codes, indicated by "#".

ABILITY MEAN is the observed, sample-dependent, [average measure](#) of persons (relative to each item) in this analysis who responded in this category (adjusted by [PWEIGHT=](#)). This is equivalent to a "Mean Criterion Score" (MCS) expressed as a measure. It is a sample-dependent quality-control statistic for this analysis. (It is not the sample-independent value of the category, which is obtained by adding the item measure to the "score at category", in Table 3.2 or higher, for the rating (or partial credit) scale corresponding to this item.) For each observation in category k, there is a person of measure B_n and an item of measure D_i . Then: average measure = $\text{sum}(B_n - D_i) / \text{count of observations in category}$.

An "*" indicates that the average measure for a higher score value is lower than for a lower score value. This contradicts the hypothesis that "higher score value implies higher measure, and vice-versa".

The "average ability" for missing data is the average measure of all the persons for whom there is no response to this item. This can be useful. For instance, we may expect the "missing" people to be high or low performers, or to be missing random (and so they average measure would be close to the average of the sample).

These values are plotted in [Table 2.6](#).

ABILITY P.SD is the population standard deviation of the ABILITY values = $\sqrt{(\sum (ABILITY - (ABILITY MEAN))^2 / COUNT)}$

S.E. MEAN is the standard error of the mean (average) measure of the sample of persons from a population who responded in this category (adjusted by [PWEIGHT=](#)) = $\sqrt{(\sum (ABILITY - (ABILITY MEAN))^2 / (COUNT * (COUNT - 1)))}$

INFT MNSQ is the [Infit Mean-Square](#) for observed responses in this category (weighted by [PWEIGHT=](#), and omitting responses in extreme person scores). Values greater than 1.0 indicate unmodeled noise. Values less than 1.0 indicate loss of information.

OUTF MNSQ is the [Outfit Mean-Square](#) for observed responses in this category (weighted by [PWEIGHT=](#), and omitting responses in extreme person scores). Values greater than 1.0 indicate unmodeled noise. Values less than 1.0 indicate loss of information.

PTMA CORR is the point-correlation between the data code, scored 1, or non-occurrence, scored 0, of this category or distractor and the person raw scores or measures chosen by [PTBISERIAL=](#). The computation is described in [Correlations](#). Example: for categories 0,1,2, then the correlation is between [1 for the target score (0, 1, or 2) and 0 for the other scores (1 and 2, 0 and 2, or 0 and 1)] and the person ability measures for the persons producing each score.

ITEM (here, ACT) is the name or label of the item.

Data codes and Category labels are shown to the right of the box, if [CLFILE=](#) or [CFILE=](#) is specified.

* Average ability does not ascend with category score. The average ability of the persons observed in this category is lower than the average ability of the persons in the next lower category. This contradicts the Rasch-model assumption that "higher categories <-> higher average abilities."

Missing % includes all categories. Scored % only of scored categories. The percentage for the missing category is based on all the COUNTS. The percentages for the SCOREd categories are based only on those category COUNTS.

""BETTER FITTING OMIT" appears in fit-ordered Tables, where items better fitting than [FITI=](#) are excluded.

14.18 Table 14 Item report in entry order

14.18.1 Table 14.1 Item statistics in entry order

(controlled by [USCALE=](#), [UMEAN=](#), [UDECIM=](#), [LOCAL=](#), [TOTAL=](#), [DISCRIMINATION=](#), [ASYMPTOTE=](#), [PVALUE=](#))

ITEM STATISTICS: ENTRY ORDER

For details, please see [Table 10.1](#)

14.18.2 Table 14.2 Item statistics graphically in entry order

Please see [Table 10.2](#)

14.18.3 Table 14.3 Item option & distractor frequencies in entry order

(controlled by [Distractors=Y](#), [OSORT=](#), [CFILE=](#), [PTBIS=](#))

ITEM OPTION FREQUENCIES are output if Distractors=Y. These show occurrences of each of the valid data codes in CODES=, and also of MISSSCORE= in the input data file. Counts of responses forming part of extreme scores are included. Only items included in the corresponding main table are listed. These statistics are also in [DISFILE=](#), which includes entries even if the code is not observed for the item. See also [Distractor Analysis](#).

[OSORT=](#) controls the ordering of options within items. The standard is the order of data codes in CODES=.

ITEM CATEGORY/OPTION/DISTRACTOR FREQUENCIES: ENTRY ORDER

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	ABILITY MEAN	P.S.D	S.E. MEAN	INFT MNSQ	OUTF MNSQ	PTMA CORR.	ITEM	
13	4	***	11	16#	.79	1.41	.45			.08	M. STAIRS	4 75% Independent
	1	1	30	52	-1.85	.99	.18	.8	1.0	-.89		1 0% Independent
	3	3	5	9	1.07	.80	.40	.6	.4	.10		3 50% Independent
	5	5	15	26	2.63	1.06	.28	1.7	1.5	.56		5 Supervision
	6	6	7	12	3.25	.81	.33	1.2	1.2	.44		6 Device
	7	7	1	2	4.63	.00		.7	.6	.23		7 Independent
	MISSING	***	1	1#	-1.99	.00				-.12		

* Average ability does not ascend with category score
 # Missing % includes all categories. Scored % only of scored categories

ENTRY NUMBER is the item sequence number.

The letter next to the sequence number is used on the fit plots.

DATA CODE is the response code in the data file.

MISSING means that the data code is not listed in the CODES= specification.

Codes with no observations are not listed.

SCORE VALUE is the value assigned to the data code by means of NEWScore=, KEY1=, IVALUEA=, etc.

*** means the data code is missing and so ignored, i.e., regarded as not administered. [MISSSCORE=1](#) scores missing data as "1".

DATA COUNT is the frequency of the data code in the data file (unweighted) - this includes observations for both non-extreme and extreme persons and items. For counts weighted by [PWEIGHT=](#), see [DISFILE=](#)

DATA % is the percent of scored data codes. For dichotomies, the % are the [proportion-correct-values](#) for the options. For data with score value "****", the percent is of all data codes, indicated by "#".

ABILITY MEAN is the observed, sample-dependent, [average measure](#) of persons (relative to each item) in this analysis who responded in this category (adjusted by [PWEIGHT=](#)). This is equivalent to a "Mean Criterion Score" (MCS) expressed as a measure. It is a sample-dependent quality-control statistic for this analysis. (It is not the sample-independent value of the category, which is obtained by adding the item measure to the "score at category", in Table 3.2 or higher, for the rating (or partial credit) scale corresponding to this item.) For each observation in category k, there is a person of measure Bn and an item of measure Di. Then: average measure = sum(Bn - Di) / count of observations in category.

An "*" indicates that the average measure for a higher score value is lower than for a lower score value. This contradicts the hypothesis that "higher score value implies higher measure, and vice-versa".

The "average ability" for missing data is the average measure of all the persons for whom there is no response to this item. This can be useful. For instance, we may expect the "missing" people to be high or low performers, or to be missing random (and so they average measure would be close to the average of the sample).

These values are plotted in [Table 2.6](#).

ABILITY P.SD is the population standard deviation of the ABILITY values = $\sqrt{(\sum (ABILITY - (ABILITY MEAN))^2/COUNT)}$

S.E. MEAN is the standard error of the mean (average) measure of the sample of persons from a population who responded in this category (adjusted by PWEIGHT=) = $\sqrt{(\sum (ABILITY - (ABILITY MEAN))^2/(COUNT*(COUNT-1)))}$

INFT MNSQ is the [Infit Mean-Square](#) for observed responses in this category (weighted by [PWEIGHT=](#), and omitting responses in extreme person scores). Values greater than 1.0 indicate unmodeled noise. Values less than 1.0 indicate loss of information.

OUTF MNSQ is the [Outfit Mean-Square](#) for observed responses in this category (weighted by [PWEIGHT=](#), and omitting responses in extreme person scores). Values greater than 1.0 indicate unmodeled noise. Values less than 1.0 indicate loss of information.

PTMA CORR is the point-correlation between the data code, scored 1, or non-occurrence, scored 0, of this category or distractor and the person raw scores or measures chosen by [PTBISERIAL=](#). The computation is described in [Correlations](#). Example: for categories 0,1,2, then the correlation is between [1 for the target score (0, 1, or 2) and 0 for the other scores (1 and 2, 0 and 2, or 0 and 1)] and the person ability measures for the persons producing each score.

ITEM (here, ACT) is the name or label of the item.

Data codes and Category labels are shown to the right of the box, if [CLFILE=](#) or [CFILE=](#) is specified.

* Average ability does not ascend with category score. The average ability of the persons observed in this category is lower than the average ability of the persons in the next lower category. This contradicts the Rasch-model assumption that "higher categories <-> higher average abilities."

Missing % includes all categories. Scored % only of scored categories. The percentage for the missing category is based on all the COUNTS. The percentages for the SCOREd categories are based only on those category COUNTS.

""BETTER FITTING OMIT" appears in fit-ordered Tables, where items better fitting than [FITE=](#) are excluded.

14.19 Table 15 Item report in label order

14.19.1 Table 15.1 Item statistics in label order

(controlled by [USCALE=](#), [UMEAN=](#), [UDECEM=](#), [LOCAL=](#), [ISORT=](#), [TOTAL=](#), [DISCRIMINATION=](#), [ASYMPTOTE=](#), [PVALUE=](#))

[ISORT=](#) \$S1W22 ; sets the item sort order for Table 15.1

ITEM STATISTICS: ALPHA ORDER ON COLUMN: \$S1W22

For details, please see [Table 10.1](#)

14.19.2 Table 15.2 Item statistics graphically in label order

Please see [Table 10.2](#)

14.19.3 Table 15.3 Item option & distractor frequencies in label order

(controlled by [Distractors=Y](#), [OSORT=](#), [CFILE=](#), [PTBIS=](#))

ITEM OPTION FREQUENCIES are output if [Distractors=Y](#). These show occurrences of each of the valid data codes in [CODES=](#), and also of [MISSCORE=](#) in the input data file. Counts of responses forming part of extreme scores are included. Only items included in the corresponding main table are listed. These statistics are also in [DISFILE=](#), which includes entries even if the code is not observed for the item. See also [Distractor Analysis](#).

[OSORT=](#) controls the ordering of options within items. The standard is the order of data codes in [CODES=](#).

ITEM CATEGORY/OPTION/DISTRACTOR FREQUENCIES: ENTRY ORDER												
ENTRY	DATA	SCORE	DATA		ABILITY		S.E.	INFT	OUTF	PTMA		
NUMBER	CODE	VALUE	COUNT	%	MEAN	P.SD	MEAN	MNSQ	MNSQ	CORR.	ITEM	
13	4	***	11	16#	.79	1.41	.45			.08	M. STAIRS	4 75% Independent
	1	1	30	52	-1.85	.99	.18	.8	1.0	-.89		1 0% Independent
	3	3	5	9	1.07	.80	.40	.6	.4	.10		3 50% Independent
	5	5	15	26	2.63	1.06	.28	1.7	1.5	.56		5 Supervision
	6	6	7	12	3.25	.81	.33	1.2	1.2	.44		6 Device
	7	7	1	2	4.63	.00		.7	.6	.23		7 Independent
	MISSING	***	1	1#	-1.99	.00				-.12		

* Average ability does not ascend with category score
 # Missing % includes all categories. Scored % only of scored categories

ENTRY NUMBER is the item sequence number.
 The letter next to the sequence number is used on the fit plots.

DATA CODE is the response code in the data file.
 MISSING means that the data code is not listed in the [CODES=](#) specification.
 Codes with no observations are not listed.

SCORE VALUE is the value assigned to the data code by means of [NEWSCORE=](#), [KEY1=](#), [IVALUEA=](#), etc.
 *** means the data code is missing and so ignored, i.e., regarded as not administered. [MISSCORE=1](#) scores missing data as "1".

DATA COUNT is the frequency of the data code in the data file (unweighted) - this includes observations for both non-extreme and extreme persons and items. For counts weighted by [PWEIGHT=](#), see [DISFILE=](#)

DATA % is the percent of scored data codes. For dichotomies, the % are the [proportion-correct-values](#) for the options.
 For data with score value "****", the percent is of all data codes, indicated by "#".

ABILITY MEAN is the observed, sample-dependent, [average measure](#) of persons (relative to each item) in this analysis who responded in this category (adjusted by [PWEIGHT=](#)). This is equivalent to a "Mean Criterion Score" (MCS) expressed as a measure. It is a sample-dependent quality-control statistic for this analysis. (It is not the sample-independent value of the category, which is obtained by adding the item measure to the "score at category", in Table 3.2 or higher, for the rating (or partial credit) scale corresponding to this item.) For each observation in category k, there is a person of measure Bn and an item of measure Di. Then: average measure = sum(Bn - Di) / count of observations in category.
 An "*" indicates that the average measure for a higher score value is lower than for a lower score value. This contradicts the hypothesis that "higher score value implies higher measure, and vice-versa".
 The "average ability" for missing data is the average measure of all the persons for whom there is no response to this item. This can be useful. For instance, we may expect the "missing" people to be high or low performers, or to be missing random (and so they average measure would be close to the average of the sample).
 These values are plotted in [Table 2.6](#).

ABILITY P.SD is the population standard deviation of the ABILITY values = $\sqrt{(\sum (ABILITY - (ABILITY MEAN))^2/COUNT)}$

S.E. MEAN is the standard error of the mean (average) measure of the sample of persons from a population who responded in this category (adjusted by [PWEIGHT=](#)) = $\sqrt{(\sum (ABILITY - (ABILITY MEAN))^2/(COUNT*(COUNT-1))}$

INFT MNSQ is the [Infit Mean-Square](#) for observed responses in this category (weighted by [PWEIGHT=](#), and omitting responses in extreme person scores). Values greater than 1.0 indicate unmodeled noise. Values less than 1.0 indicate loss of information.

OUTF MNSQ is the [Outfit Mean-Square](#) for observed responses in this category (weighted by [PWEIGHT=](#), and omitting responses in extreme person scores). Values greater than 1.0 indicate unmodeled noise. Values less than 1.0 indicate loss of information.

PTMA CORR is the point-correlation between the data code, scored 1, or non-occurrence, scored 0, of this category or

distractor and the person raw scores or measures chosen by [PTBISERIAL=](#). The computation is described in [Correlations](#). Example: for categories 0,1,2, then the correlation is between [1 for the target score (0 , 1, or 2) and 0 for the other scores (1 and 2, 0 and 2, or 0 and 1)] and the person ability measures for the persons producing each score.

ITEM (here, ACT) is the name or label of the item.

Data codes and Category labels are shown to the right of the box, if CLFILE= or CFILE= is specified.

* Average ability does not ascend with category score. The average ability of the persons observed in this category is lower than the average ability of the persons in the next lower category. This contradicts the Rasch-model assumption that "higher categories <-> higher average abilities."

Missing % includes all categories. Scored % only of scored categories. The percentage for the missing category is based on all the COUNTS. The percentages for the SCOREd categories are based only on those category COUNTS.

""BETTER FITTING OMIT" appears in fit-ordered Tables, where items better fitting than [FIT=](#) are excluded.

14.20 Table 16 Person maps with full person labels

(controlled by [IPEXTREME=](#), [MRANGE=](#), [MTICK=](#), [MAXPAGE=](#), [NAMLMP=](#), [PSORT=](#), [T1#=](#), [LINELENGTH=](#))

Table 16.3 Person map with full labels		
MEASURE ACT - MAP - KID <rare> <more>		
5	+ FATALE, NATASHA	ROSSNER, LAWRENCE F.
	PASTER, RUTH	
4	+	
	SCHATTNER, GAIL	
	T	
	DOEPPNER, TOM	
3	+ MCLOUGHLIN, BILLY	
	T WRIGHT, BENJAMIN	
X	BUFF, MARGE BABY	CHAZELLE, BERNIE
	CLAPP, LB	SEILER, KAREN
X	S ERNST, RICHARD MAX	SQURREL, ROCKY J.
2	+ CLAPP, CHARLIE	EASTWOOD, CLINT
	VROOM, JEFF	
XX	BADENOV, BORIS	KENT, CLARK
X	MOOSE, BULLWINKLE	ROSSNER, BESS
	CORLEONE, VITO	MAN, SPIDER
	PATRIARCA, RAY	
S	HSIEH, PAUL FRED	SQUILLY, MICHAEL
X	CLAPP, DOCENT	HOGAN, KATHLEEN
	ROSSNER, TOBY G.	
1	+M BECKER, SELWIN	BLOFELD, VILLAIN
	FONTANILLA, HAMES	LAMBERT, MD., ROSS W.
	SABILE, JACK	TOZER, AMY ELIZABETH
XX	ALLEN, PETER	BABBOO, BABOO
	CORLEONE, MICHAEL	PINHEAD, ZIPPY
XXX	BEISER, ED	HSIEH, DANIEL SEB
	MALAPROP, MRS.	SANDBERG, RYNE
X	CIANCI, BUDDY	EISEN, NORM L.
XX	AIREHEAD, JOHN	ANGUIANO, ROB
0	M+ AMIRAULT, ZIPPY	DRISKEL, MICHAEL
XXX	S	
X	BAUDET, GERARD	
X	BOND, JAMES	ROSSNER, REBECCA A.
-1	X + LIEBERMAN, DANIEL	NORDGREN, JAN SWEDE
X	S JACKSON, SOLOMON	LANDMAN, ALAN
X	T	
	SABOL, ANDREW	

Right-hand column shows the full person labels located by the person ability measures along the variable. Abbreviated person labels are shown in [Table 1.3](#). The persons often have a normal distribution.

Left-hand column locates the item difficulty measures along the variable. The item labels are abbreviated to fit on one page. [Table 12](#) has the full item labels. Items arranged by measure: look for the hierarchy of item names to spell out a meaningful construct from easiest at the top to hardest at the bottom.

For dichotomous items, look for an even spread of items along the variable (the y-axis) with no gaps, indicating poorly defined or tested regions of the variable. Good tests usually have the items targeted (lined up with) the persons.

For polytomous items, [Table 2.2](#) shows the operational range of the item with its rating scale.

In Table 16.3, the full person names are shown with an item distribution. You can use [NAMLMP=](#) or [PMAP=](#) to control the number of characters of each name reported. [MAXPAGE=](#) controls the length of the Table. [MRANGE=](#) controls the displayed range of measures. [PSORT=](#) controls the sort order within rows.

Persons arranged by measure: Look for the hierarchy of person names to spell out a meaningful distribution from highest scoring at the top to lowest scoring at the bottom.

Column Labels

<pre> -2 X + <freq> <less> </pre>	<pre> ITEM= MAP PERSON= RARE= T MORE= ↓ LESS= </pre>
<p>Table 16.13 Person map with full labels and reversed item distribution</p> <pre> MEASURE TAP - MAP - KID <freq> <more> 5 XXX ++ X 4 XX ++T XX S Rick M Susan F ++ Frank M 3 X ++S Bert M Betty F Elsie F Thomas M 2 X ++ Britton F Carol F David M Joe M Zula F 1 M++M Anne F Audrey F Barbara F Dorothy F James M Kim F Linda F Mike M Ron M Tracie F Walter M William M -1 X ++ Janet F Lisa F Rod M -2 XX ++S Brenda F Pete M -3 ++ Adam M Richard M X S Blaise M Martha F -4 ++T Don M -5 XXXX ++ Helen F <rare> <less> </pre>	<p>Right-hand column is the person details, see Table 16.3. Abbreviated person labels are shown in Table 1.13.</p> <p>Left-hand column locates the item distribution by their easiness measures along the latent variable. These are the difficulty measures reversed around the mean person difficulty "M".</p> <p style="text-align: center;">Column Labels</p> <pre> ITEM= MAP PERSON= FREQ= T MORE= ↓ LESS= </pre>

MEASURE	Location of person or item on the unidimensional latent variable, in logits or USCALE= units
<more>	higher person "ability" = higher raw score with complete data
<less>	lower person "ability" = lower raw score with complete data
<rare>	higher item "difficulty" = lower item "easiness" = lower score on the item with complete data: PVALUE=
<frequ> <frequent>	lower item "difficulty" = higher item "easiness" = higher score on the item with complete data: PVALUE=
M	Mean of person or item distribution
S	One standard deviation from the person or item mean
T	Two standard deviations from the person or item mean
X	one person or item
#	several persons or items, e.g., EACH "#" = 4. The value for # can be changed with T1#=# for items, T1P#=# for persons.
.	between 1 and (# - 1) persons or items, e.g., if # = 4, then "." = 1 to 3
	Variable advances. Lines without " " have the same measure as the nearest " " above.
	The double line indicates the two sides have opposite orientations: person ability and item easiness. The item difficulties are reversed . This is useful if the items and persons are being compared to the responses.

<i>On plot:</i>	<i>Description:</i>
MEASURE	Location on the unidimensional latent variable, in logits or USCALE= units
<more>	higher person "ability" = higher raw score with complete data
<less>	lower person "ability" = lower raw score with complete data
<rare>	higher item "difficulty" = lower score on the item with complete data: PVALUE=
<frequent>	lower item "difficulty" = higher score on the item with complete data: PVALUE=

M	Mean of person or item distribution
S	One standard deviation from the person or item mean
T	Two standard deviations from the person or item mean
X	one person or item
#	several persons or items, e.g., "#" = 4. The value for # can be changed with I1# , I1P# .
.	between 1 and #-1 persons or items, e.g., "." = 1 to 3
or +	Variable advances. Lines without " " or "+" have the same measure as the nearest " " or "+" above.

14.21 Table 17.1 Person statistics in measure order

(controlled by [USCALE](#), [UMEAN](#), [UDECIM](#), [LOCAL](#), [TOTAL](#))

PERSON STATISTICS: MEASURE ORDER

For details, please see [Table 10.1](#)

14.22 Table 17.2 Person statistics graphically in measure order

Please see [Table 10.2](#)

14.23 Table 18.1 Person statistics in entry order

(controlled by [USCALE](#), [UMEAN](#), [UDECIM](#), [LOCAL](#), [TOTAL](#))

PERSON STATISTICS: ENTRY ORDER

For details, please see [Table 10.1](#)

14.24 Table 18.2 Person statistics graphically in entry order

Please see [Table 10.2](#)

14.25 Table 19.1 Person statistics in label order

(controlled by [USCALE](#), [UMEAN](#), [UDECIM](#), [LOCAL](#), [PSORT](#), [TOTAL](#))

[PSORT](#) @TIMEPOINT ; sets the person sort order for Table 19.1

PERSON STATISTICS: ALPHA ORDER ON COLUMN: @TIMEPOINT

For details, please see [Table 10.1](#)

14.26 Table 19.2 Person statistics graphically in label order

Please see [Table 10.2](#)

14.27 Table 20.1 Complete score-to-measure table on test of all items

[Table 20.2](#) Person score and measure distribution

[Table 20.3](#) Complete score-to-calibration table for tests based on whole sample

TABLE OF MEASURES ON TEST OF 13 ITEM

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
13	-6.34E	1.82	40	-1.35	.28	67	1.33	.35
14	-5.15	.99	41	-1.27	.28	68	1.46	.36
...								
36	-1.65	.28	63	.84	.35	90	6.04	1.03
37	-1.58	.28	64	.96	.35	91	7.29E	1.84
38	-1.50	.28	65	1.08	.35			
39	-1.43	.28	66	1.21	.35			

```

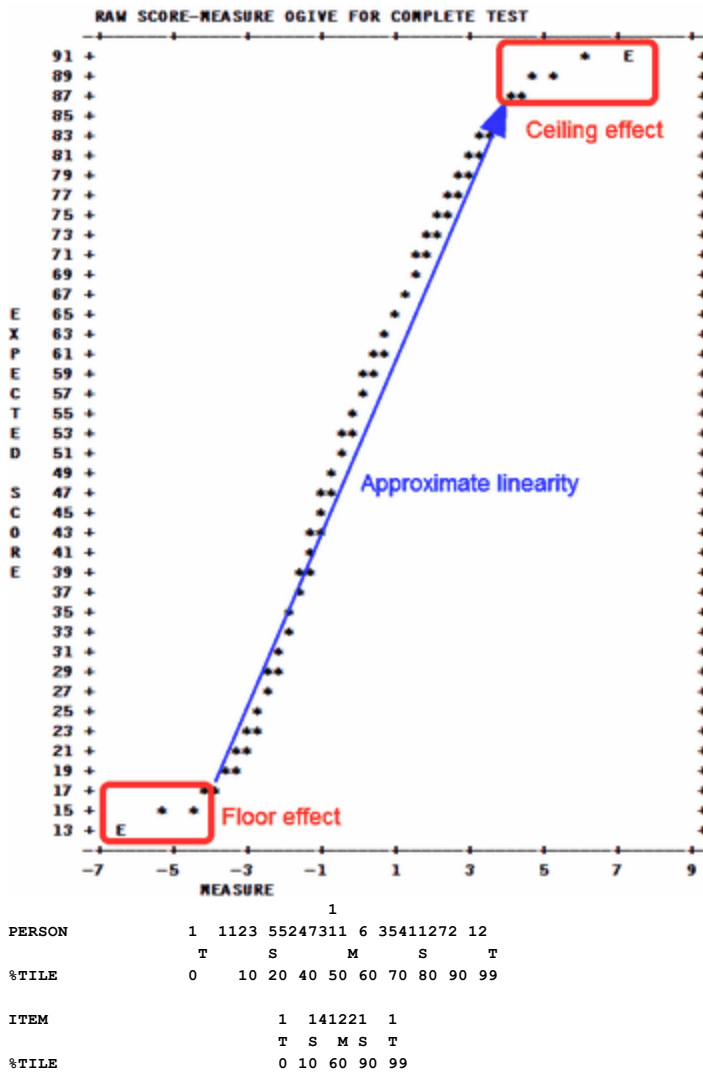
CURRENT VALUES, UMEAN=.0000 USCALE=1.0000
TO SET MEASURE RANGE AS 0-100, UMEAN=46.5329 USCALE=7.3389
TO SET MEASURE RANGE TO MATCH RAW SCORE RANGE, UMEAN=49.2956 USCALE=5.7243
Predicting Score from Measure: Score = Measure * 8.2680 + 38.9511
Predicting Measure from Score: Measure = Score * .1158 + -4.5099
Maximum statistically different levels of performance (strata) = 3.8
Wright's Sample-independent Person (Test) Reliability based on maximum strata = .94
    
```

In Table 20.1	Explanation
<p>TABLE OF MEASURES ON TEST OF 13 ITEM</p> <p>TOTALSCORE=Yes includes extreme items TOTALSCORE=No excludes extreme items (if any)</p>	<p>Raw score-to-Measure Table for all raw scores on the complete set of all active non-extreme calibrated items. This can also be output with SCOREFILE=.</p> <p>Table 20.1 shows the score-to-measure table for persons with complete data. For persons with missing data, this table does not apply.</p> <p>Score-to-measure tables for subtests or persons with missing data can be produced by using ISELECT= or IDDELETE= from the Specification menu before requesting Table 20.</p> <p>If there are subsets, then the reported Table 20 is one of an infinite number of possible Table 20s.</p> <p>If you want the score file for person measures including the extreme (zero, perfect) items, then</p> <ol style="list-style-type: none"> 1. Run a standard analysis. 2. Output: IFILE=if.txt SFILE=sf.txt Eliminate unwanted items from IFILE= and remove any; 3. Run the analysis again, with Extra specifications: IAFILE=if.txt SAFILE=sf.txt 4. The person measures will have altered somewhat to adjust for the imputed difficulties of the extreme items. 5. Output Table 20 and SCOREFILE=
SCORE	<p>Raw score on the complete set of non-extreme items. If IWEIGHT= produces decimal scores, then scores with one decimal place are shown. Some SCOREs may not be observable. Some observable scores may not be shown, but may be better approximated from TCCFILE=</p>

MEASURE	<p>Estimated measure for the SCORE raw score on all the items, adjusted by USCALE= and UIMEAN=, if active. The MEASURE may differ slightly from Table 17. Tighten convergence to make them agree exactly.</p> <p>The SCOREFILE= and Table 20 person ability estimates are estimated on the basis that the current item difficulty estimates are the "true" estimates. These are the person estimates if you anchored (fixed) the items at their reported estimates. The convergence criterion used are (LCONV=) *.01 and (RCONV=)*.01 - these are considerably tighter than for the main analysis. So Table 20 is a more precise estimate of the person measures based on the final set of item difficulties.</p> <p>PFILE=, Table 17 and the other Person Measure Tables show the person abilities that are the maximum likelihood estimates at the current stage of estimation. To make these two sets of estimates coincide, please tighten the convergence criteria in your Winsteps control file: CONVERGE=L LCONV=*.001 ; or tighter</p> <p>If STBIAS=YES is used, then score table measures are dependent on the data array, even if items are anchored.</p>
-6.34E	<p>"E" (Extreme, Extrapolated) is a warning that the accompanying value is not estimated in the usual way, but is an approximation based on arbitrary decisions. For extreme scores, the Rasch estimated would be infinite. So the arbitrary decision has been made to make the extreme score more central by the EXTREMESCORE= amount (or its default value), and to report the measure corresponding to that score. The Rasch standard error of an extreme score is also infinite, so the reported standard error of an "E" value is the standard error of the finite reported measure.</p>
S.E.	<p>The model-based standard error of the MEASURE, not adjusted for misfit.</p>
CURRENT VALUES, UMEAN=.0000 USCALE=1.0000	<p>This shows the current user-scaling. Changing the user-scaling in the Specification menu dialog box or the Help menu Scaling calculator changes Table 20 immediately. Use this to experiment with different values of USCALE= and UIMEAN=.</p>
TO SET MEASURE RANGE AS 0-100, UMEAN=46.5329 USCALE=7.3389	<p>Suggested UIMEAN= and USCALE= values for transformation of the MEASUREs into user-friendly values in the range 0-100 resembling percentages.</p>
TO SET MEASURE RANGE TO MATCH RAW SCORE RANGE, UMEAN=49.2956 USCALE=5.7243	<p>Suggested UIMEAN= and USCALE= values for transformation of the MEASUREs into user-friendly values into the range or raw scores so that the MEASUREs resemble raw scores.</p>
Predicting Score from Measure: Score = Measure * 8.2680 + 38.9511	<p>A linear approximation of the score-to-measure ogive useful for predicting raw scores on this set of items from measures.</p>
Predicting Measure from Score: Measure = Score * .1158 + -4.5099	<p>A linear approximation of the score-to-measure ogive useful for predicting measures from raw scores on this set of items.</p>

Maximum statistically different levels of performance (strata) = 3.8
 Wright's Sample-independent Person (Test) Reliability based on maximum strata = 0.94

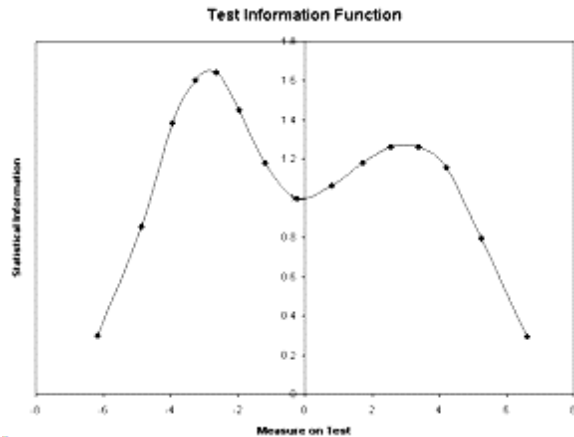
Across the raw-score range in Table 20.1, statistically significant different raw scores are $1.96 * (M_1 - M_2) / \sqrt{(SE_1^2 + SE_2^2)}$ apart where M_1 and M_2 are the two measures. SE_1 and SE_2 are their two standard errors. See www.rasch.org/rmt/rmt144k.htm. The number of different levels (strata) resembles a [separation](#) index. Wright's matching "Test" Reliability is shown. These are equivalent to the Separation and Reliability of a person sample with a uniform distribution having the range of the MEASUREs in Table 20.



A graph of the score to measure conversion is also reported. "*" indicates the conversion. This is the Test Characteristic Curve, TCC. Table 20.1 gives the score-to-measure conversion for a complete test, when going from the y-axis (score) to the x-axis (measure). When going from the x-axis (measure) to the y-axis (score), it predicts what score on the complete test is expected to be observed for any particular measure on the x-axis. For CAT tests and the like, no one takes the complete test, so going from the y-axis to the x-axis does not apply. But going from the x-axis to the y-axis predicts what the raw score on the complete bank would have been, i.e., the expected total score, if the whole bank had been administered.

Measure distributions for the persons and items are shown below the score-to-measure table. M is the mean, S is one P.SD (population standard deviation) from the mean. T is two P.SDs from the mean. %ILE is the percentile, the percentage below the measure. Percentiles have the range 0-99.

Test Information



The statistical information is $(USCALE/S.E.)^2$. These are plotted in the [test information function](#) in the [Graph](#) window. You can also plot the values reported in [TCCFILE=](#)

Score-to-measure Table 20 is to be produced from known item and rating scale structure difficulties

In your Winsteps control file:

[I/AFILE=](#)if.txt ; usually values from IFILE=if.txt of another analysis ; the item anchor file containing the known item difficulties

[S/AFILE=](#)sf.txt ; usually values from SFILE=sf.txt of another analysis ; the structure/step anchor file (only for polytomies)

CONVERGE=L ; only logit change is used for convergence

LCONV=0.0001 ; logit change too small to appear on any report.

STBIAS=NO ; no estimation bias correction with anchor values

TFILE=*

20 ; the score table

*

The data file comprises two dummy data records, so that every item has a non extreme score, e.g.,

For dichotomies:

CODES = 01

Record 1: 10101010101

Record 2: 01010101010

For a rating scale from 1 to 5:

CODES = 12345

Record 1: 15151515151

Record 2: 51515151515

14.28 Table 20.2 Person score and measure distribution

[Table 20.1](#) Complete score-to-measure table on test of all items

[Table 20.2](#) Person score and measure distribution

[Table 20.3](#) Complete score-to-calibration table for tests based on whole sample

TABLE OF SAMPLE NORMS (500/100) AND FREQUENCIES CORRESPONDING TO COMPLETE TEST

SCORE	MEASURE	S.E.	NORMED	S.E.	FREQUENCY	%	CUM.FREQ.	%	PERCENTILE
0	-6.17E	1.83	147	107	0	.0	0	.0	0
1	-4.86	1.08	225	63	0	.0	0	.0	0

	2	-3.94	.85	278	50	1	2.9	1	2.9	1	
	3	-3.27	.79	318	46	2	5.9	3	8.8	6	
	4	-2.64	.78	355	46	2	5.9	5	14.7	12	
	5	-1.97	.83	394	49	2	5.9	7	20.6	18	
	6	-1.19	.92	440	54	3	8.8	10	29.4	25	
	7	-.23	1.00	496	59	12	35.3	22	64.7	47	
	8	.80	.97	557	57	5	14.7	27	79.4	72	
	9	1.72	.92	610	54	4	11.8	31	91.2	85	
	10	2.55	.89	660	52	1	2.9	32	94.1	93	
	11	3.37	.89	707	52	2	5.9	34	100.0	97	
	12	4.21	.93	756	54	0	.0	34	100.0	100	
	13	5.23	1.12	817	66	0	.0	34	100.0	100	
	14	6.60E	1.84	897	108	0	.0	34	100.0	100	

-----+
 THE NORMED SCALE IS EQUIVALENT TO UIMEAN= 516.7919 USCALE= 45.0404

The columns in the Table of Sample Norms and Frequencies are:

Measures on the Complete Test:	
SCORE	raw score on a complete test containing all calibrated items: TOTALSCORE= Yes includes extreme items. TOTALSCORE= No excludes extreme items (if any)
MEASURE	measure corresponding to score
	If a person did not take all items or items are weighted, then that person is stratified with the measure on the complete test nearest the person's estimated measure (as reported in Table 18), regardless of that person's observed score.
S.E.	standard error of the measure (model)
	The statistical information is (USCALE/S.E.) ²
Statistics for this sample:	
NORMED	measures linearly locally-rescaled so that the mean person measure for this sample is 500 and the population standard deviation is 100. Equivalent to UPMEAN= 500, USCALE= 100/(Person P.SD)
S.E.	standard error of the normed measure (for a score based on all the items)
FREQUENCY	count of sample with measures at or near (for missing data) the complete test measure
%	percentage of sample included in FREQUENCY
CUM.FREQ.	count of sample with measures near or below the test measure, the cumulative frequency.
%	percentage of sample include in CUM. FREQ.
PERCENTILE	The percentile is the cumulative frequency percent for the score below + half the frequency percent for the current score, half-rounded, and constrained to the range 1-99 for non-zero frequencies.
NORMED SCALE	This shows the UIMEAN= and USCALE= values in order to report the person MEASURES at their NORMED (500/100) values.

Logit measures support direct probabilistic inferences about relative performances between persons and absolute performances relative to items. Normed measures support descriptions about the location of subjects within a sample (and maybe a population). Report the measures which are most relevant to your audience.

This Table is easy to paste into Excel. Use Excel's "data", "text to columns" feature to put the scores and measures into columns.

Example: I want to stratify my sample into low, medium, high ability groups.

The [separation index](#) is based on the statistical fiction that your data accord exactly with a normal distribution and that the average measurement error (RMSE) precisely summarizes the precision of your data. In practice, these assumptions are only met approximately.

If your data are "complete" (everyone responds to every item), then a convenient places to start is Table 20.2. Starting at the lowest score, look down the scores until you find the score that best characterizes (as a first guess) your "low group". Then mentally multiply its S.E. by 3, and add it to the measure for the low group. This will take you to the measure for the middle group, which will be approximately statistically significantly different ($p \leq .05$) from the low group. Do the same again for the

middle group, and it will take you to the high group. Same again may take you to an even higher group, or up into the outliers at the top of the test. The cut-points will be half-way between the group centers that you have identified.

Do the same process from the top score downwards for another version of the stratification.

Then synthesize the two stratifications by adjusting the group by moving their center scores further apart (not closer together).

The Percentile Computation

14.29 Table 20.3 Complete score-to-calibration table for tests based on whole sample

[Table 20.1](#) Complete score-to-measure table on test of all items

[Table 20.2](#) Person score and measure distribution

Table 20.3 Complete score-to-calibration table for tests based on whole sample

This Table, which must be selected explicitly with [TFILE=](#), or as a subtable with "[Request Subtable](#)" on the "[Output Tables](#)" menu, shows an estimated item calibration for all possible rates of success by persons, i.e., the item measure corresponding to every observable proportion-correct-value or item marginal raw score for the entire sample.

Output Tables

Request Subtables

20.3

OK

or enter into your control file:

TFILE=*

20.3

*

TABLE OF ITEM MEASURES ON COMPLETE NON-EXTREME SAMPLE FOR TAP LIKE 4 1-3-4								
SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
0	6.17E	1.87	12	.81	.45	24	-1.58	.48
1	4.84	1.08	13	.61	.44	25	-1.82	.50
2	3.98	.82	14	.41	.44	26	-2.08	.52
3	3.40	.70	15	.22	.44	27	-2.36	.54
4	2.96	.64	16	.03	.43	28	-2.67	.57
5	2.58	.59	17	-.16	.43	29	-3.01	.60
6	2.26	.55	18	-.34	.43	30	-3.40	.64
7	1.97	.53	19	-.53	.44	31	-3.85	.71
8	1.71	.50	20	-.73	.44	32	-4.43	.82
9	1.46	.49	21	-.93	.45	33	-5.29	1.08
10	1.23	.47	22	-1.14	.46	34	-6.61E	1.87
11	1.01	.46	23	-1.35	.47			

SCORE raw score on this item of a complete sample containing all calibrated persons. Table 20.3 shows the score-to-measure table for items with complete data. For items with missing data, this table does not apply. [TOTALSCORE=Yes](#) includes extreme persons

TOTALSCORE=No excludes extreme persons (if any).

MEASURE measure corresponding to score (user-scaled with [UJMEAN=](#), [USCALE=](#)).

S.E. standard error of the measure.

The statistical information in an item estimate for the entire sample is $(USCALE/S.E.)^2$

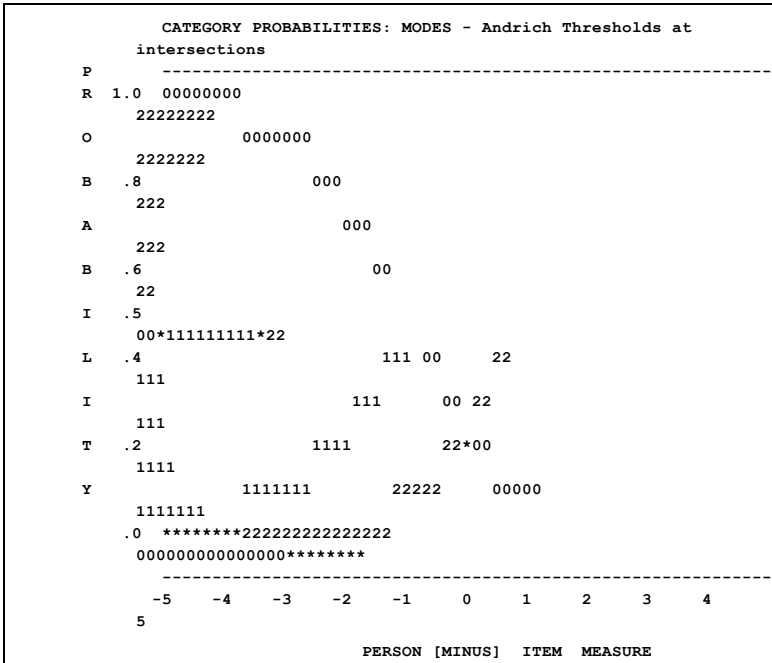
The real difficulty of an item is its difficulty relative to ...

the average difficulty of all the items	This is the Winsteps default UIMEAN= 0
the average ability of all the persons	Specify: UPMEAN= 0
a person who scores 50% on all the items in the test	a) Run a standard analysis b) Output Table 20 c) In Table 20.1 , look for the measure corresponding to 50% success on the test d) Specify UIMEAN = - (corresponding to 50% success on the test) e) rerun the analysis f) look at Table 20.1 , 50% success on the test should be "0".
an item on which 50% of all the persons succeed	a) run a standard analysis b) Menu: Output Tables c) Click on first entry "Request Subtables" 20.3 Click on OK c) In Table 20.3 , look for the measure corresponding to 50% success by the sample d) Specify UIMEAN = - (corresponding to 50% success by the sample) e) rerun the analysis f) look again at Table 20.3 , 50% success by the sample should be "0".

14.30 Table 21 Category Probability curves and Expected score ogive

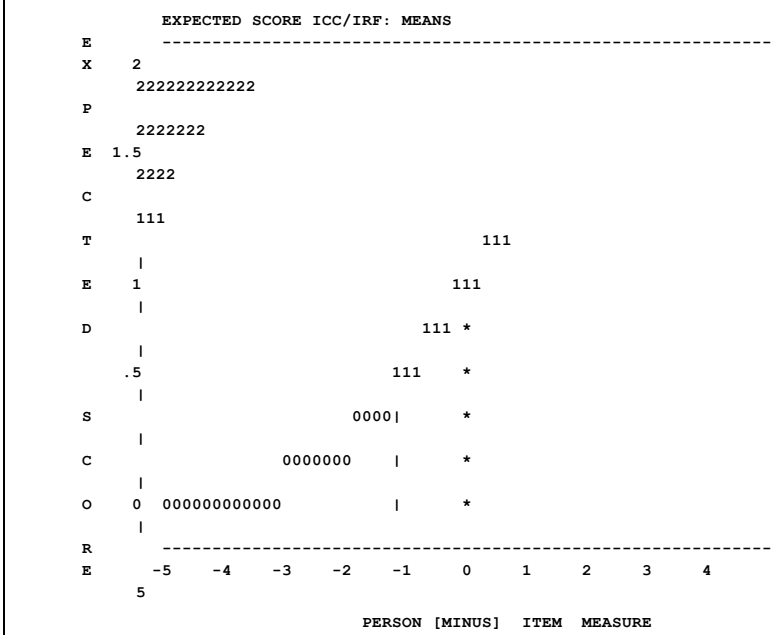
(controlled by [MRANGE=](#), [CURVES=](#)) - also called Item Response Curves, IRCs

<p>21.1 Dichotomies: Category probabilities</p> <pre> DICHOTOMOUS CURVES ----- P R 1.0 0000000000000000 1111111111111111 O 111111 000000 B .8 111 000 A 11 00 B .6 11 00 I *** L .4 00 11 I 00 11 T .2 000 111 Y 000 111111 000000 .0 1111111111111111 0000000000000000 ----- -6 -4 -2 0 2 4 6 PERSON [MINUS] ITEM MEASURE </pre>	<p>The probability of each response is shown across the measurement continuum. The measure to be used for determining the probability of any particular response is the difference between the measure of the person and the calibration of the item. For dichotomies, only one curve is shown plotting the probability of scoring a "1" (correct), and also of scoring a "0" (incorrect) for any measure relative to item measure. For 'S' and 'F' models these curves are approximations.</p>
<p>21.1 Polytomies: Individual category probabilities</p>	<p>See Rating scale conceptualization. When there are more than two categories, the probability of each category is shown. The points</p>



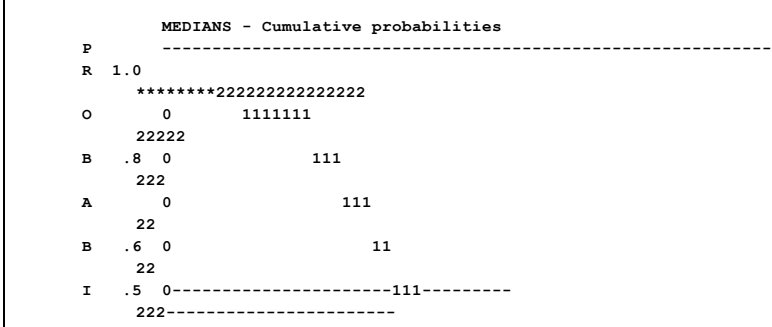
of intersection of adjacent categories are the Rasch-Andrich thresholds (structure calibrations).

21.2 Polytomies: Expected average ratings

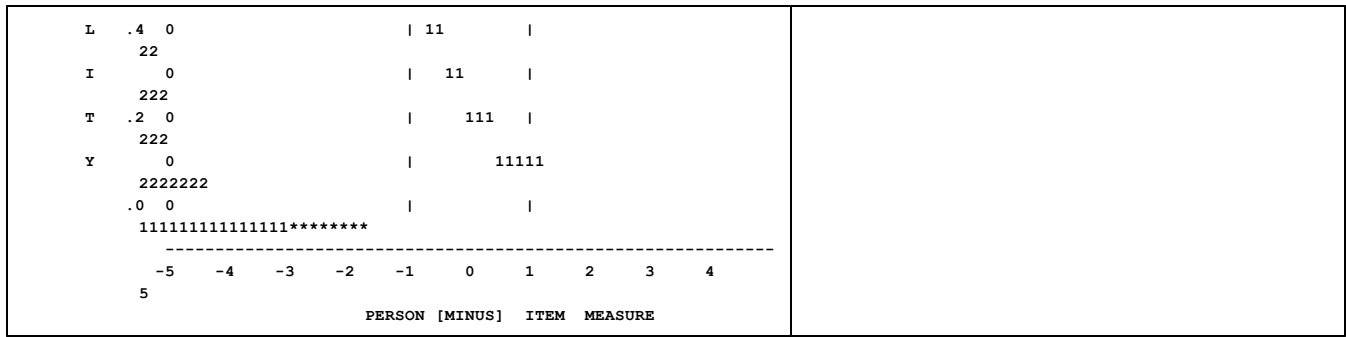


The Expected score ogive is also called the Model Item Characteristic Curve (ICC) and Item Response Function (IRF). For response structures with three or more categories, two further graphs can be drawn. The second graph depicts the expected score ogive. The vertical "*" characters correspond to integer expected scores, and the "|" characters correspond to half-point expected scores, the Rasch-half-point thresholds. The intervals between the Rasch-half-point thresholds can be thought of as the intervals corresponding to the observed categories. For the purposes of inference, measures in the zone on the x-axis between '|' and '|' correspond, on average, to the rating given on the 'y' axis, '1'. Similarly ratings on the y-axis can be thought of as corresponding to measures in the matching zone on the x-axis. The degree to which the data support this is given by the COHERENCE statistics in Table 3.2. Empirical item characteristic curves are shown in [Table 29](#) and from the [Graphs menu](#).

21.3 Polytomies: Cumulative category probabilities



The third graph is of the zone curves which indicate the probability of an item score at or below the stated category for any particular difference between person measure and item calibration. The area to the left of the "0" ogive corresponds to "0". The right-most area corresponds to the highest category. The P=0.5 intercepts are the median cumulative probabilities. "|" indicate the Rasch-Thurstone thresholds.



If ISGROUPS= defines more than one item group, then one set of these curves is shown for each item group: 21.1, 21.2, 21.3, then 21.11, 21.12, 21.13, then 21.21, 21.22, 21.23, etc. Each item group is identified by an example item from the group.

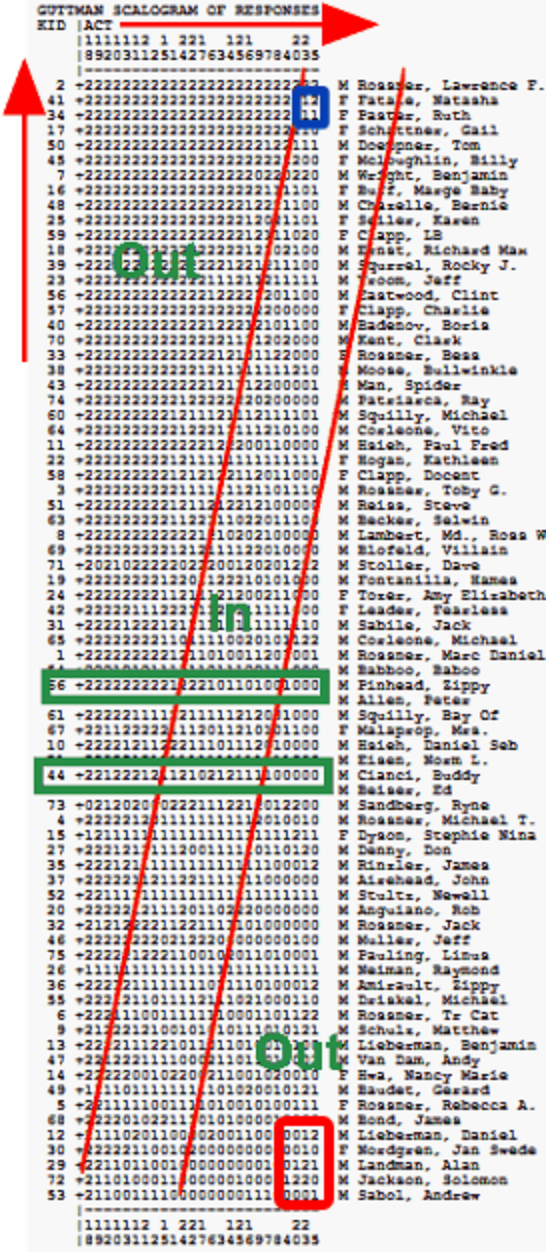
To produce these using other software, see [GRFILE=](#) or [Graphs Window](#), Copy Data.

14.31 Table 22.1, 22.2, 22.3 Sorted response matrix (Guttman scalogram)

(controlled by [LINELENGTH=](#)).

<p>Table 22.1 Sorted observed data matrix - Guttman scalogram</p> <p>GUTTMAN SCALOGRAM OF RESPONSES: KID ACT 1111112 1 221 121 22 8920311251427634569784035 ----- 2 +222222222222222222222222222222 M ROSSNER, LAWRENCE F. 41 +222222222222222222222222222212 F FATALE, NATASHA 34 +222222222222222222222222222211 F PASTER, RUTH 17 +222222222222222222222222222210 F SCHATTNER, GAIL 50 +2222222222222222222222122111 M DOEPPNER, TOM 45 +222222222222222222222222221200 F MCLOUGHLIN, BILLY 7 +2222222222222222222220220220 M WRIGHT, BENJAMIN 16 +2222222222222222222222111101 F BUFF, MARGE BABY 48 +2222222222222222222212221100 M CHAZELLE, BERNIE 25 +2222222222222222222212021101 F SEILER, KAREN 59 +2222222222222222222212111020 F CLAPP, LB 18 +2222222222222222222212102100 M ERNST, RICHARD MAX</p>	<p>The observations are printed in order of person and item measures as a Guttman Scalogram, with most able persons listed first, the easiest items printed on the left. This scalogram shows the extent to which a Guttman pattern is approximated. See also Guttman Coefficient of Reproducibility.</p>
<p>Table 22.2 Guttman scalogram of zoned responses</p> <p>GUTTMAN SCALOGRAM OF ZONED RES KID ACT 1111112 1 221 121 22 8920311251427634569784035 ----- 2 +222222222222222222222222222222 M ROSSNER, LAWRENCE F. 41 +2222222222222222222222222222A2 F FATALE, NATASHA 34 +2222222222222222222222222222AA F PASTER, RUTH 17 +2222222222222222222222222222A@ F SCHATTNER, GAIL 50 +2222222222222222222222A2A11 M DOEPPNER, TOM 45 +22222222222222222222222222AB@@ F MCLOUGHLIN, BILLY 7 +2222222222222222222222B@BB@ M WRIGHT, BENJAMIN 16 +2222222222222222222222A111@1 F BUFF, MARGE BABY 48 +222222222222222222A22B11@@ M CHAZELLE, BERNIE 25 +222222222222222222A2@B11@1 F SEILER, KAREN 59 +222222222222222222A2A11@B@ F CLAPP, LB 18 +222222222222222222A21@B1@@ M ERNST, RICHARD MAX</p>	<p>The scalogram is that of Table 22.1, but with each observation marked as to whether it conforms with its expectation or not. Observations within 0.5 rating points of their expectation are deemed to be in their expected categories, and are reported with their category values, e.g., '1', '2', etc. These ratings support the overall inferential relationship between observations and measures. Observations more than 0.5 rating points away from their expectations, i.e., in a "wrong" category, are marked with a letter equivalent: 'A' = '0', 'B' = '1', 'C' = '2', etc. These contradict observation-to-measure inferences. The proportion of in- and out-of-category observations are reported by the COHERENCE statistics in Table 3.2.</p>
<p>Table 22.3 Guttman scalogram of original response codes</p> <p>GUTTMAN SCALOGRAM OF ORIGINAL RESPONSES: KID ACT</p>	<p>The observations are printed in order of person and item measures, with most able persons listed first, the easiest items printed on the left. This scalogram shows the original codes in the data file.</p>

1111112	1	221	121	22	
8920311251427634569784035					
2	+222222222222222222222222	M ROSSNER, LAWRENCE F.			
41	+222222222222222222222222	F FATALE, NATASHA			
34	+222222222222222222222222	F PASTER, RUTH			
17	+222222222222222222222222	F SCHATTNER, GAIL			
50	+222222222222222222222222	M DOEPPNER, TOM			
45	+222222222222222222222222	F MCMLOUGHLIN, BILLY			
7	+22222222222222222222020220	M WRIGHT, BENJAMIN			
16	+222222222222222222222222	F BUFF, MARGE BABY			
48	+222222222222222222222222	M CHAZELLE, BERNIE			
25	+222222222222222222222222	F SEILER, KAREN			
59	+222222222222222222222222	F CLAPP, LB			
18	+222222222222222222222222	M ERNST, RICHARD MAX			



Interpreting a Scalogram. Example of Table 22.1 for [Example0.txt](#). A scalogram orders the persons from high measure to low measure as rows, and the items from low measure (easy) to high measure (hard) as columns. The green boxes show the probabilistic advance from responses in high categories on the easy items to responses in low categories on the hard items.

Top left corner: where the “more able” (more liking) children respond to the “easier” (to like) items. So we expect to see responses of “Like” (2). We do! Unexpected responses in this area influence the Outfit statistics more strongly.

Top right corner: (blue box) where the “most liking” children and the “hardest to like items” meet - you can see some ratings of 1.

Bottom right corner: where “less able” (less liking) children respond to the “harder” (to like) items. So we expect to see responses of “Dislike” (0). But do we?? Something has gone wrong! There are 1’s and 2’s where we expected all 0’s. Unexpected responses in this area influence the Outfit statistics more strongly.

Diagonal transition zone: Between the red diagonals lies the transition zone where we expect 1’s. In this zone Infit is more sensitive to unexpected patterns of responses. More categories in the rating scale means a wider transition zone. Then the transition zone can be wider than the observed responses.

Example of Table 22.2. Five categories: 0,1,2,3,4. Categories in the wrong score bands are shown as @ (=0), A (=1), B (=2), C (=3), D (=4). The 5 score bands have different colors here. Band 0 is yellow. Band 1 is green. Band 2 is blue. Band 3 is white. Band 4 is mauve.

+44444	444	44	C44	C	4	4	D	D	DDD	D	B	B	3DB	2D	A@	@	2A	C1
+44444	444	C44	4	4	44	DDD	3DD3	BD	DAB3	DA	A	C	2	A	2			
+C4C4C	CCC	4C	CCC	3	3	3	3	33D	3	3	D	22C	2C	22	2B	CB	@1	
+44444	444	44	444	3	3	3	B	B	DBB	D	3	A	22D	A2	2@	AB	BC	@1
+44444	444	DDD	D	D	BD	D3B	DBDD	B3B	D2AC	C@	A	@	@	1@	@			
+44444	444	ADD	D	DBDD	DAD	DD3DD	ADD	D@2A	AA	@	@	@	B@1	1				
+44C44	444	3DD	3	33D3	3AB	DB33B	D3A	2A2A	2C	1	1	C	111	1				
+4C444	3	3D	3	3D	3	D3	D3	3DA	3C	CC@	@	A	AA2B1	11	@	@	11	
+4CD3	D	DB	D	D3	3	33	3B	3@3	3C	2AC	2	A2	2C1@B	@1	B	1	1@	
+CB3	3	3	333	3	333B	BBB	3BB32	2C2	CC2C	2B	B	B	1	BB1	1			
+4D3	3	3	333	D	AD33	DB3	3CCCC	A22	2A22	1B	1	B	1	1@1	0			
+43D	D	3	333	3	3BAB	33A	DCACA	C22	AC22	1B	B	1	B	1C1	A			
+444D3	3D3	BD	3DD	3	3	B	D	DD2	2	C	A	AB@	@@	1@	@1	@1	00	
+4C4D3	3DD	DB	DD3	3	B	B	B	3	CCC	A	2	A	C11	11	1@	11	@@	AA
+444DD	DDD	DB	3DD	D	D	A	B	C	CCD	A	2	@	AC@	B@	@@	@@	@@	00
+444DD	DDB	DD	D33	D	B	D	3	2	@AD	@	@	D	B1@	@1	11	C@	@@	00
+4DD3	B	DB	D	3A	D	BB	3B	CDA	CD	A22	2	D@	1CC@1	@@	1	@	@@	
+4D3D	3	3D	3	B3	3	3B	C2	2CA	A2	2C@	2	A2	1C1B1	B1	@	1	00	
+CD33	D	3D	B	3D	B	3B	2C	2AA	22	222	D	AA	1111B	11	@	1	AA	
+4DDD	D	DD	3	BD	A	BD	AC	2CC	DA	CAC	C	@@	111@@	@@	@	1	00	
+4DD3	3	3B	3	DA	3	@D	2A	D2C	CC	2	C	2	@B	11@B1	@1	B	@	00
+C333	3	B3	3	B3	B	CC	2C	2CC	A@	22C	2	D@	@C1@1	1@	1	@	00	
+4DDD	D	3D	D	3A	D	AA	CD	CDA	AA	CAA	2	1@	1@@B@	1@	@	0	00	
+CDD	3	3	B33	B	33	CC	C2@	AC	222	@C@1	@@	B	B	1	@@	0	00	0
+444DD	3DD	BD	DD	@	C	D	A	2D@		B	B	1	@@	@1	@@	@@	@@	00
+443D	3BB	BB	3	B	C	2	2	2	A@A	1	1	CB1	1@	C@	@@	AA	00	
+CABB	3	BB	B	B3	A	CA	2A	A2C	2A	2BB	B	1B	11C11	11	B	A	AA	
+4C43D	3B3	33	B@3	2	2	@	C	@	2@@	C	1	D	@@C	B@	@C	CO	00	A@
+D333	3	B3	3	B3	C	22	CC	AA@	2A	2BB	B	@B	1@@1B	@A	A	O	00	
+DAB	B	A	DD3	A	C	22	@CC	2CDA2	A@D	@BC1	11	@	@	B	00A	0		
+3AA	3	3	333	B	CCA2	AC2	22@A2	CAA	11BC	1B	@	1	C	00A	A			
+3BB3	B	3B	B	3	2	C	2@	@@2	A	@@1	B	BC	111CB	AB	A	A	BO	
+B33	3	D	3BA	3	CC22	A2C	2A2AA	22A	BC1@	@@	B	@	A	0A0	A			
+ADBD	D	3D	A	3D	C	AC	2@	D@A	22	1@1	B	@@	111@@	AO	O	A	0B	
+DD3	3	B	332	C	A	DA	AD2	@@2@C	BB1	@C@B	1@	@	O	O	000	0		
+DB3B	2	C2	2	22	@	D1	@D	@CC	@@	1B@	@	C1	1@@00	@@	O	O	00	

Example of Table 22.3. Here is the Scalogram for [Example 5](#), a computer-adaptive, multiple-choice test. The original responses are shown.

GUTTMAN SCALOGRAM OF ORIGINAL RESPONSES:

STUDENT	TOPIC	1	11	3	11	1	3212232212132425654421434145625	36555366356465464633654														
			640215038189677748390992315641640517264579268221076889430372139553458																			
3 +							d	b	c	db	d	c	b	cd	S	STA						
6 +								c		b	c	d	d	c	cd	S	CHU					
27 +							c	c	c		c	cbb	a	dd	c	b	dd	S	ERI			
4 +								c	b	a	d	cb	c	dd	d	b	dc	S	CHE			
9 +							d	c	a	a	cbc	c	ba	b	dca	c	db	dc	bbdbb	c	S	MEL
25 +							c	b	d	c	bc	a	a	d	a	c	dba	d	cbccbd	c	S	DAN

```

24 +      a      d      ac      b      cd      c      ba      a      AP SAR
12 +      c      b      c      c      bb      b      ba      cb      cc      c      A ALL
28 +      c      a      bcc      b      c      acc      d      b      bc      b      c      bdcdb      d      IH GRE
21 + b      c      b      a      c      b      a      bc      bb      ab      a      db      bbd      cc      b      b      AP MIL
11 + a      d      c      b      cc      cbad      d      ca      c      d      ca      ba      b      ac      bac      bc      d      c      cc      S CHR
17 +      d      dc      c      acac      cac      b      c      a      a      b      c      c      bc      b      ac      ac      IH FRA
7 +      d      c      d      ccaaac      d      bcd      ca      b      cb      c      b      a      a      a      ab      b      bb      ac      IH JAN
1 +      c      c      ab      ca      dc      b      a      a      c      a      c      IM CAT
20 +      bd      b      c      b      cad      cd      a      bac      b      a      b      c      b      IL JAM
22 + d      a      d      c      b      d      cdabcc      bbc      cbdb      a      a      dbac      c      dc      d      a      IM BAR
30 +      d      a      a      cab      cdd      d      d      da      c      d      c      c      db      IL HOL
14 + b      d      c      db      caa      caa      da      c      dcd      b      c      a      cd      d      c      c      d      b      IL TAK
23 +      d      cab      dcaa      cadbbba      a      c      a      db      c      b      a      c      c      ab      c      IL MEI
26 +      d      a      cc      b      ac      cdd      b      a      d      db      a      d      b      d      d      cb      c      IL MIT
15 + d      d      a      a      caba      dcbbdc      c      d      b      b      a      c      bb      NH MAR
13 + d      d      d      c      bac      dc      bd      d      dcd      d      a      d      d      c      b      b      NH KEL
19 + b      dbd      bdcabd      baaacbd      c      dc      ab      a      dd      d      c      b      c      NH RIC
8 + bbdadb      d      abaa      acbcbccacbcadab      c      d      cdab      a      d      a      a      c      c      cb      IL ROB
16 + b      da      b      bdcbbcccbadaabba      ccd      c      ac      bd      a      a      b      NM HEL
18 + b      a      a      d      bca      bc      a      cccc      a      db      c      a      a      d      d      NH JOE
10 + b      badbcbcccadacbababca      cba      a      a      acd      a      d      a      b      NM JON
19 + bbdad      dacbdb      cabdbc      c      ad      d      ab      b      c      a      b      NM THE
2 + bbdad      a      d      aaaa      c      c      a      cbd      ab      a      d      c      NM KAT
5 +      aad      a      ab      a      b      a      NL SMI
|-----|
| 11 3 11 1 3212232212132425654421434145625 36555366356465464633654
|640215038189677748390992315641640517264579268221076889430372139553458

```

Example: To display Scalogram response strings on the person Tables [6](#), [17](#), [18](#), [19](#).

1. Perform your standard Winsteps analysis.
2. Output Table 22, with a LINELENGTH= big enough for all observations to be on one line.
3. Save Table 22 to your Desktop (or wherever).
4. Open the Table 22 file in software that can do a rectangular copy (NotePad++, TextPad, Word, etc.)
5. Open your Winsteps data file using the same rectangular-copy software.
6. Rectangular copy the Scalogram immediately adjacent to the person labels.
7. Adjust Winsteps control [NAMELENGTH=](#) etc for the new data file format.
8. Save the control and data file(s)
9. Perform your revised Winsteps analysis.
10. Table 6 should now display the Scalogram as part of the person label.

14.32 Table 23 Item multidimensionality

14.32.1 Table 23.0 Variance components scree plot for items

The dimensionality analysis of [Table 23](#) stratifies the items into three clusters for each Contrast (Principal Component). The items are listed in [Table 23.2](#). For each Contrast, each person is measured on each cluster of items. These measures are then correlated for each pair of clusters. If the correlations approach 1.0, then empirically the clusters of items are measuring approximately the same thing. Since measurement error (low test reliability) lessens the correlations, the disattenuated correlations are also reported.

According to Mundfrom et al. (2005), the sample size needs to be at least 6 x (number of items) for stable results: Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum Sample Size Recommendations for Conducting Factor Analyses. *International Journal of Testing*, 5(2), 159-168.

[Table 23.0](#) Variance components scree plot for items

[Table 23.1, 23.11](#) Principal components plots of item loadings

[Table 23.2, 23.12](#) Item Principal components analysis/contrast of residuals

[Table 23.3, 23.13](#) Item contrast by persons

[Table 23.4, 23.14](#) Item contrast loadings sorted by measure

[Table 23.5, 23.15](#) Item contrast loadings sorted by entry number

[Table 23.6, 23.16](#) Person measures for item clusters in contrast. [Cluster Measure Plot](#) for Table 23.6.

[Table 23.99](#) Largest residual correlations for items

[Youtube video](#) explaining Table 23

Extreme scores produce estimated measures with high standard errors. Correlations are reported excluding and also including (+Extreme) person estimates for extreme scores.

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = ACT information units				
	Eigenvalue	Observed	Expected	
Total raw variance in observations =	50.9521	100.0%	100.0%	
Raw variance explained by measures =	25.9521	50.9%	50.7%	
Raw variance explained by persons =	10.3167	20.2%	20.2%	
Raw Variance explained by items =	15.6354	30.7%	30.6%	
Raw unexplained variance (total) =	25.0000	49.1%	100.0%	49.3%
Unexplned variance in 1st contrast =	4.6287	9.1%	18.5%	
Unexplned variance in 2nd contrast =	2.9434	5.8%	11.8%	
Unexplned variance in 3rd contrast =	2.2957	4.5%	9.2%	
Unexplned variance in 4th contrast =	1.7322	3.4%	6.9%	
Unexplned variance in 5th contrast =	1.6327	3.2%	6.5%	

STANDARDIZED RESIDUAL VARIANCE SCREE PLOT

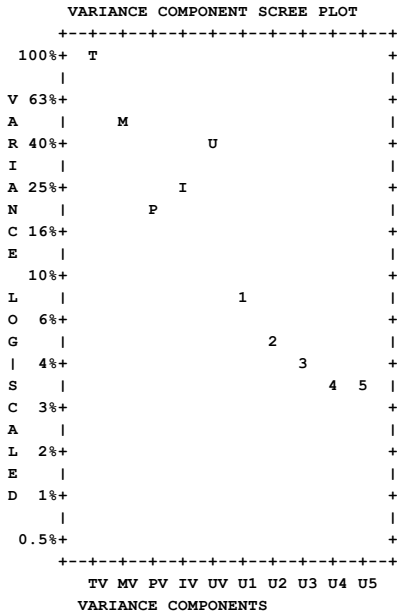


Table 23.0 shows a variance decomposition of the observations for the items. This is not produced for [PRCOMP=O](#).

If your Table says "**Total variance in observations**", instead of "Total raw variance in observations", then please update to the current version of Winsteps, or produce this Table with [PRCOMP=R](#).

Extreme items and persons (minimum possible and maximum possible raw scores) are omitted from this computation because their correlations are 0.

Simulation studies, and the empirical results of Winsteps users, indicated that the previous computation of "variance explained" was over-optimistic in explaining variance. So a more conservative algorithm was implemented. Technically, the previous computation of "variance explained" used standardized residuals (by default). These are generally considered to have better statistical properties than the raw residuals. But the raw residuals (PRCOMP=R) were found to provide more realistic explanations of variance, so the current Winsteps computation uses raw residuals for "variance explained" in the top half of the variance table.

The "Unexplained variance" is controlled by [PRCOMP=](#), which defaults to standardized residuals (PRCOMP=S). Set PRCOMP=R to express the entire table in terms of raw residuals.

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = ITEM information units				
	Eigenvalue	Observed percentage of total variance	Observed percentage of unexplained variance	Expected percentage of total variance
Total raw variance in observations =	50.9	100.0%	← Expected values if these data fit the Rasch model perfectly →	100.0%

Raw variance explained by measures =	25.9	50.9%	← If these match reasonably, then the measures explain the expected amount of variance in the data →	46.5%
Raw variance explained by persons =	10.3	20.2%		18.5%
Raw Variance explained by items =	15.6	30.7%		28.0%
Raw unexplained variance (total) =	25.0 = count of items (or persons)	49.1%	100.0%	53.5%
Unexplned variance in 1st contrast =	4.6	9.1%	18.5%	← Use simulations to estimate the Rasch-model-expected values SIFILE=
Unexplned variance in 2nd contrast =	2.9	5.8%	11.8%	
Unexplned variance in 3rd contrast =	2.3	4.5%	9.2%	
Unexplned variance in 4th contrast =	1.7	3.4%	6.9%	
Unexplned variance in 5th contrast =	1.6	3.2%	6.5%	

Table of STANDARDIZED RESIDUAL variance: the standardized residuals form the basis of the "unexplained variance" computation, set by [PRCOMP=](#)

in Eigenvalue units: [variance components](#) are rescaled so that the total unexplained variance has its expected summed eigenvalue.

ITEM information units: the eigenvalue units are rescaled to match the number of items, so these values are equivalent to "strength in item units".

Observed: variance components for the observed data

Expected: variance components expected for these data if they exactly fit the Rasch model, i.e., the variance that would be explained if the data accorded with the Rasch definition of unidimensionality.

If **Observed** and **Expected** differ noticeably, then there is a problem in the estimation. This is not a symptom of multidimensionality. Check that iteration was not stopped manually (ctrl+f) or by [MUJCON=MJMLE=](#), also that [RCONV=](#) and [LCONV=](#) are reasonably small. However for very large or very sparse data or very long rating scales, convergence can be difficult to obtain. Then big differences between observed and expected in Table 23 have the same meaning as unexpected displacements in the Measure tables. There is a problem with the estimation.

Total raw variance in observations: total raw-score variance in the observations

Raw variance explained by measures: raw-score variance in the observations explained by the Rasch item difficulties, person abilities and polytomous scale structures.

Raw variance explained by persons: raw-score variance in the observations explained by the Rasch person abilities (and apportioned polytomous scale structures) - this is equivalent to Jelle Goeman & Nivja de Jong's Summability Index (2013)

Raw variance explained by items: raw-score variance in the observations explained by the Rasch item difficulties (and apportioned polytomous scale structures)

Raw unexplained variance (total): raw-score variance in the observations not explained by the Rasch measures

Unexplned variance in 1st, 2nd, ... contrast: variance that is not explained by the Rasch measures is decomposed into Principal Component Analysis, PCA, components = Contrasts. The size of the first, second, ... [contrast](#) (component) in the PCA decomposition of standardized residuals (or as set by [PRCOMP=](#)), i.e., variance that is not explained by the Rasch measures, but that is explained by the contrast. At most, 5 contrasts are reported. If less than 5 are reported, then the other contrasts have negative eigenvalues, usually due to the data overfitting the Rasch model. You may have cleaned the data too much, see [When to stop removing items and persons in Rasch analysis?](#)

The important lines in this Table are "contrasts". If the first contrast is much larger than the size of an Eigenvalue expected by chance, usually less than 2 - www.rasch.org/rmt/rmt191h.htm - please inspect your [Table 23.3](#) to see the contrasting content of the items which is producing this large off-dimensional component in your data, or [Table 24.3](#) to see the contrasting persons. The threat to Rasch measurement is not the ratio of unexplained (by the model) to explained (by the model), or the amount of explained or unexplained. The threat is that there is another non-Rasch explanation for the "unexplained". This is what the "contrasts" are reporting.

How Variance Decomposition is done ...

1. A central person ability and a central item difficulty are estimated. When the central ability is substituted for the estimated person abilities for each observation, the expected total score on the instrument across all persons equals the observed total score. Similarly, when the central ability is substituted for the estimated item difficulties for each observation, the expected total score on the instrument across all items equals the observed total score.
2. For each observation, a central value is predicted from the central person ability and the central item difficulty and the estimated rating scale (if any).

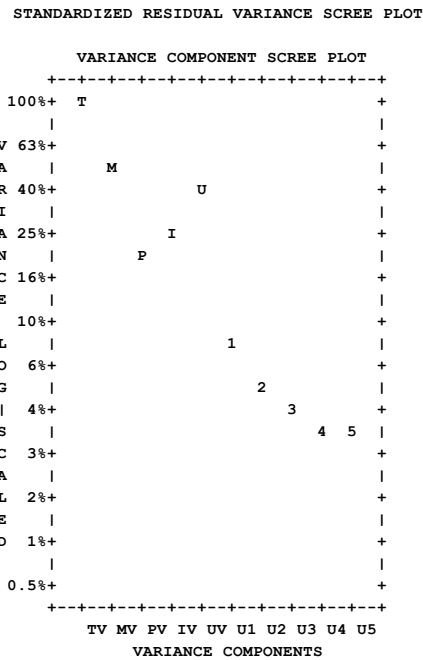
In the "Observed" columns:

3. "Total raw variance in observations =" the sum-of-squares of the observations around their central values.
4. "Raw unexplained variance (total)=" is the sum-of-squares of the difference between the observations and their Rasch predictions, the raw residuals.
5. "Raw variance explained by measures=" is the difference between the "Total raw variance" and the "Raw unexplained variance".
6. "Raw variance explained by persons=" is the fraction of the "Raw variance explained by measures=" attributable to the person measure variance (and apportioned rating scale structures).
7. "Raw variance explained by items=" is the fraction of the "Raw variance explained by measures=" attributable to the item measure variance (and apportioned rating scale structures).
8. The reported variance explained by the items and the persons is normalized to equal the variance explained by all the measures. This apportions the variance explained by the rating scale structures.
9. The observation residuals, as transformed by [PRCOMP=](#), are summarized as an inter-person correlation matrix, with as many columns as there are non-extreme persons. This correlation matrix is subjected to Principle Components Analysis, PCA.
10. In PCA, each diagonal element (correlation of the person with itself) is set at 1.0. Thus the eigenvalue of each person is 1.0, and the total of the eigenvalues of the matrix is the number of persons. This is the sum of the variance modeled to exist in the correlation matrix, i.e., the total of the unexplained variance in the observations.
11. For convenience the size of the "Raw unexplained variance (total)" is rescaled to equal the total of the eigenvalues. This permits direct comparison of all the variance terms.
12. The correlation matrix reflects the Rasch-predicted randomness in the data and also any departures in the data from Rasch criteria, such as those due to multidimensionality in the persons.
13. PCA reports components. If the data accord with the Rasch model, then each person is locally independent and the inter-person correlations are statistically zero. The PCA analysis would report each person as its own component. Simulation studies indicate that even Rasch-conforming data produce eigenvalues with values up to 2.0, i.e., with the strength of two persons.
14. Multidimensionality affects the pattern of the residuals. The residual pattern should be random, so the "contrast" eigenvalue pattern should approximately match the eigenvalue pattern from simulated data. When there is multidimensionality the residuals align along the dimensions, causing the early contrast eigenvalues to be higher than those from random (simulated) data. So multidimensionality inflates the early PCA contrasts above the values expected from random data, and correspondingly must lower the later ones, because the eigenvalue total is fixed.
15. "Unexplned variance in 1st contrast =" reports the size of the first PCA component. This is termed a "contrast" because the substantive differences between persons that load positively and negatively on the first component are crucial. It may reflect a systematic second dimension in the persons.
16. "Unexplned variance in 2nd contrast =" . Consecutively smaller contrasts are reported (up to 5 contrasts). These may also contain systematic multi-dimensionality in the persons.

In the "Expected" columns:

17. "Raw variance explained by measures=" is the sum-of-squares of the Rasch-predicted observations (based on the item difficulties, person abilities, and rating scale structures) around their central values.
18. "Raw variance explained by persons=" is the fraction of the "Raw variance explained by measures=" attributable to the person measure variance (and apportioned rating scale structures).
19. "Raw variance explained by items=" is the fraction of the "Raw variance explained by measures=" attributable to the item measure variance (and apportioned rating scale structures).
20. The reported variance explained by the items and the persons is normalized to equal the variance explained by all the measures. This apportions the variance explained by the rating scale structures.
21. "Raw unexplained variance (total)=" is the summed Rasch-model variances of the observations around their expectations, the unexplained residual variance predicted by the Rasch model.

22. "Total raw variance in observations =" is the sum of the Rasch-model "Raw variance explained by measures=" and the "Raw unexplained variance (total)="
23. The "Model" and the "Empirical" values for the "Total raw variance in observations =" are both rescaled to be 100%.
24. Use the `SIFILE=` option in order to simulate data. From these data predicted model values for the contrast sizes can be obtained.



Scree plot of the variance component percentage sizes, logarithmically scaled:

On plot	On x-axis	Meaning
T	TV	total variance in the observations, always 100%
M	MV	variance explained by the Rasch measures
P	PV	variance explained by the person abilities
I	IV	variance explained by the item difficulties
U	UV	unexplained variance
1	U1	first contrast (component) in the residuals
2	U2	second contrast (component) in the residuals, etc.

For the observations (`PRCOMP=Obs`), a standard Principal Components Analysis (without rotation, and with orthogonal axes) is performed based on the scored observations.

Table of OBSERVATION variance (in Eigenvalue units)

			-- Observed --
Raw unexplained variance (total)	=	13.0	100.0%
Unexplnd variance in 1st contrast	=	9.9	76.1%
Unexplnd variance in 2nd contrast	=	.9	7.2%
Unexplnd variance in 3rd contrast	=	.6	4.3%
Unexplnd variance in 4th contrast	=	.3	2.6%
Unexplnd variance in 5th contrast	=	.3	2.0%

Here "contrast" means "component" or "factor".

Approximate relationships between the KID measures

PCA	ACT	Pearson	Disattenuated	Pearson+Extr	Disattenuated+Extr
Contrast	Clusters	Correlation	Correlation	Correlation	Correlation
1	1 - 3	0.1404	0.2175	0.1951	0.2923
1	1 - 2	0.2950	0.4675	0.3589	0.5485

1	2 - 3	0.8065	1.0000	0.8123	1.0000
2	1 - 3	0.5319	0.7042	0.5640	0.7464
2	1 - 2	0.6546	0.9232	0.6782	0.9434
2	2 - 3	0.6271	0.8799	0.6672	0.9217
3	1 - 3	0.2320	0.4686	0.2765	0.5197
3	1 - 2	0.8083	1.0000	0.8080	1.0000
3	2 - 3	0.4327	0.7342	0.5060	0.7971
4	1 - 3	0.6955	1.0000	0.7088	1.0000
4	1 - 2	0.7467	1.0000	0.7496	1.0000
4	2 - 3	0.7731	1.0000	0.7601	1.0000
5	1 - 3	0.5854	0.9622	0.6242	1.0000
5	1 - 2	0.6667	1.0000	0.6986	1.0000
5	2 - 3	0.7665	1.0000	0.7913	1.0000

For explanation, see [Table 23.1](#)

Example 1:

We are trying to explain the data by the estimated Rasch measures: the person abilities and the item difficulties. The Rasch model also predicts random statistically-unexplained variance in the data. This unexplained variance should not be explained by any systematic effects.

Table of RAW RESIDUAL variance (in Eigenvalue units)

		Empirical		Modeled
Total raw variance in observations	=	19.8	100.0%	100.0%
is composed of				
Raw variance explained by measures	=	7.8	39.3%	39.1%
Raw unexplained variance (total)	=	12.0	60.7% 100.0%	60.9%

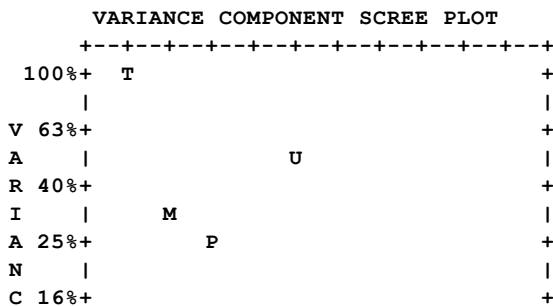
Nothing is wrong so far. The measures are central, so that most of the variance in the data is unexplained. The Rasch model predicts this unexplained variance will be random.

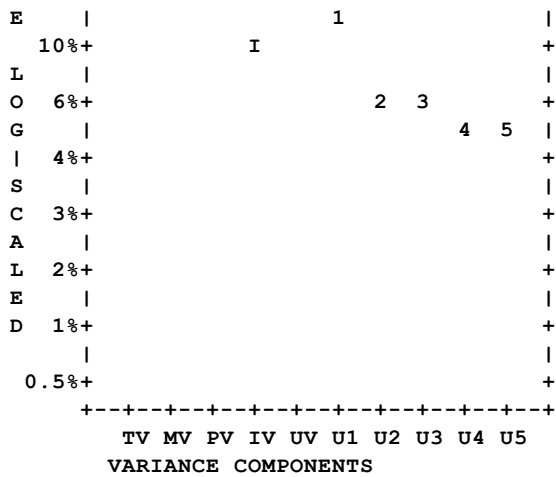
Raw variance explained by measures	=	7.8	39.3%	39.1%
is composed of				
Raw variance explained by persons	=	5.7	28.9%	28.8%
Raw Variance explained by items	=	2.1	10.4%	10.3%

Nothing is wrong so far. The person measures explain much more variance in the data than the item difficulties. This is probably because the person measure S.D. is bigger than the item difficulty S.D. in [Table 3.1](#).

Raw unexplained variance (total)	=	12.0	60.7%	100.0%	60.9%
is composed of					
Unexplned variance in 1st contrast	=	2.6	13.1%	21.5%	
Unexplned variance in 2nd contrast	=	1.4	6.9%	11.4%	
Unexplned variance in 3rd contrast	=	1.3	6.4%	10.6%	
Unexplned variance in 4th contrast	=	1.1	5.6%	9.3%	
Unexplned variance in 5th contrast	=	1.1	5.4%	8.9%	
(and about 7 more)					

Now we have multidimensionality problems. According to Rasch model simulations, it is unlikely that the 1st contrast in the "unexplained variance" (residual variance) will have a size larger than 2.0. Here it is 2.6. Also the variance explained by the 1st contrast is 13.1%. this is larger than the variance explained by the item difficulties 10.4%. A secondary dimension in the data appears to explain more variance than is explained by the Rasch item difficulties. Below is the scree plot showing the relative sizes of the variance components (logarithmically scaled).





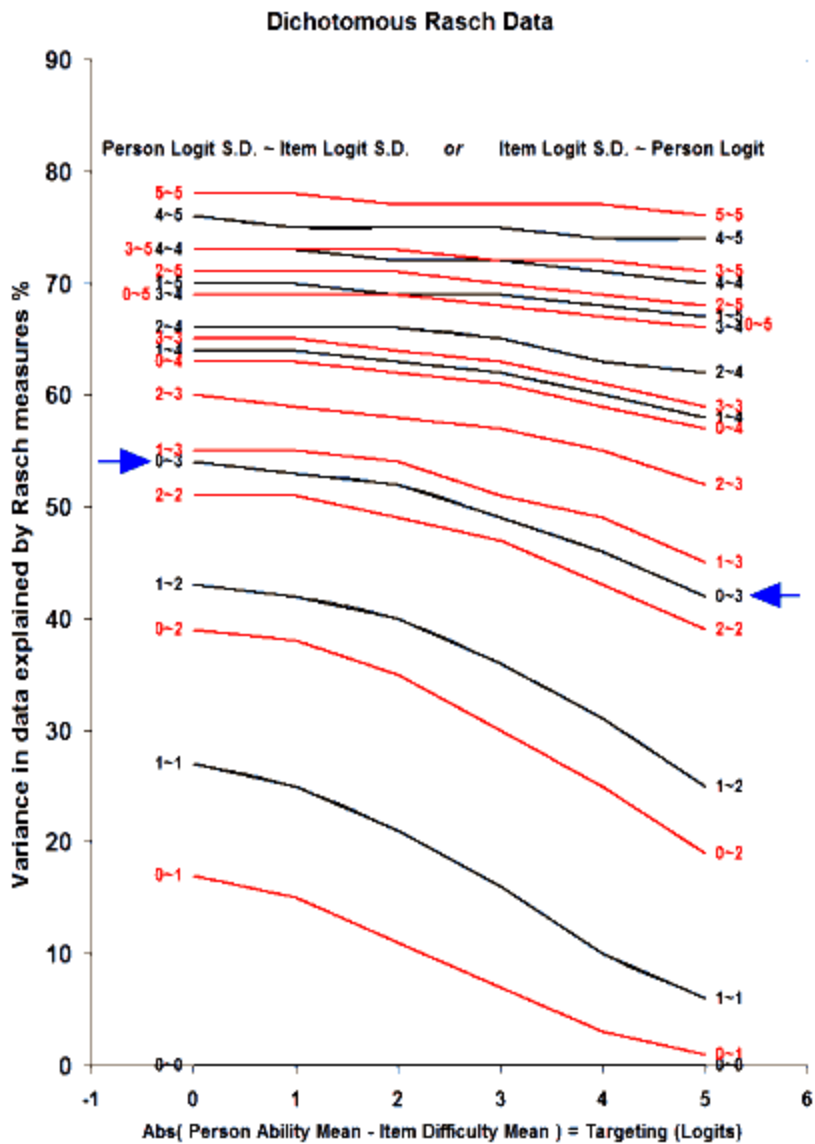
Example 2:

Question: My Rasch dimension only explains 45.5% of the variance in the data and there is no clear secondary dimension. How can I increase the "variance explained"?

Reply: If there is "no clear secondary dimension" and no excessive amount of misfitting items or persons, then your data are under statistical control and your "variance explained" is as good as you can get without changing the sample or the instrument.

Predicted Explained-Variance

A Rasch model predicts that there will be a random aspect to the data. This is well understood. But what does sometimes surprise us is how large the random fraction is. The Figure shows the proportion of "variance explained" predicted to exist in dichotomous data under various conditions.



The x-axis is the absolute difference between the mean of the person and item distributions, from 0 logits to 5 logits. The y-axis is the percent of variance in the data explained by the Rasch measures. Each plotted line corresponds to one combination of standard deviations. The lesser of the person S.D. and the item S.D. is first, 0 to 5 logits, followed by "~". Then the greater of the person S.D. and the item S.D. Thus, the arrows indicate the line labeled "0-3". This corresponds to a person S.D. of 0 logits and an item S.D. of 3 logits, or a person S.D. of 0 logits and an item S.D. of 3 logits. The Figure indicates that, with these measure distributions about 50% of the variance in the data is explained by the Rasch measures. When the person and item S.D.s, are around 1 logit, then only 25% of the variance in the data is explained by the Rasch measures, but when the S.D.s are around 4 logits, then 75% of the variance is explained. Even with very wide person and item distributions with S.D.s of 5 logits only 80% of the variance in the data is explained.

In general, to increase the variance explained, there must be a wide range of person measures and/or of item difficulties. We can obtain this in three ways:

1. Increase the person S.D.: Include in the sample more persons with measures less central than those we currently have (or omit from the sample persons with measures in the center of the person distribution)
2. Increase the item S.D.: Include in the test more items with measures less central than those we currently have (or omit from the test items with measures in the center of the item distribution)
3. Make the data more deterministic (Guttman-like) so that the estimated Rasch measures have a wider logit range:

- a.) Remove "special causes" (to use quality-control terminology) by trimming observations with extreme standardized residuals.
- b.) Reduce "common causes" by making the items more discriminating, e.g., by giving more precise definitions to rating scale categories, increasing the number of well-defined categories, making the items more similar, etc.

However, for a well-constructed instrument administered in a careful way to an appropriate sample, you may already be doing as well as is practical.

For comparison, here are some percents for other instruments:		
exam12.txt	78.7%	(FIM sample chosen to exhibit a wide range of measures)
exam1.txt	71.1%	(Knox Cube Test)
example0.txt	50.8%	(Liking for Science)
interest.txt	37.5%	(NSF survey data - 3 category rating scale)
agree.txt	30.0%	(NSF survey data - 4 category rating scale)
exam5.txt	29.5%	(CAT test) - as CAT tests improve, this % will decrease!
coin-toss	0.0%	

In Winsteps Table 23.0, the "Model" column gives the "Variance Explained" value that you could expect to see if your data had perfect fit to the Rasch model with the current degree of randomness. The "Model" value is usually very close to the empirical value. This is because some parts of your data underfit the model (too little variance explained) and some parts overfit (too much variance explained).

Relationship to Bigsteps and earlier versions of Winsteps:

My apologies for the difficulties caused by the change in computation to "Variance Explained".

The earlier computation was based on the best statistical theory available to us at the time. Further developments in statistical theory, combined with practical experience, indicated that the previous computation was too generous in assigning "variance explained" to the Rasch measures. The current computation is more accurate.

Set PRCOMP=R (raw residuals) in Bigsteps and earlier versions of Winsteps, and you will obtain approximately the same explained/unexplained variance proportions as the current version of Winsteps.

Research in the last couple of years has demonstrated that PRCOMP=R gives a more realistic estimate of the variance explained than PRCOMP=S (standardized residuals). PRCOMP=S overestimates the explained variance as a proportion of the total variance.

But for decomposing the unexplained variance into "contrasts", PRCOMP=S is better. So this mixed setting is now the default for Winsteps.

The Eigenvalue reported for the 1st contrast has not changed. If this is much larger than the size of an Eigenvalue expected by chance, usually less than 2 - www.rasch.org/rmt/rmt191h.htm - Please inspect your Table 23.3 to see the contrasting content of the items which is producing this large off-dimensional component in your data.

14.32.2 Table 23.1, 23.11 Principal components plots of item loadings

Please do not interpret this as a usual factor analysis. These plots show contrasts between opposing factors, identified as "A,B,..." and "a,b,..." , not loadings on one factor. For more discussion, see [dimensionality and contrasts](#).

Quick summary:

(a) the X-axis is the measurement axis. So we are not concerned about quadrants, we are concerned about vertical differences. The Table 23 plots show contrasts between types of items: those at the top vs. those at the bottom. The [Table 24](#) plots show contrasts between types of persons: those at the top vs. those at the bottom.

(b) "How much" is important. See the Variance Table explained in [Table 23.0](#). Important differences have eigenvalues greater than 2.0.

(c) If the difference is important, it suggests that we divide the test into pieces, clustering the items in the top half of the plot and the items in the bottom half. Winsteps estimates a measure for each person on each cluster of items. The correlations of the measures are reported. Small disattenuated correlations indicate that the clusters of items measure different sub-dimensions.

(d) Perform separate analyses for the target clusters of items. Cross-plot and correlate the person measures. We will then see for whom the differences are important. Usually, for a carefully designed instrument, it is such a small segment that we decide it is not worth thinking of the test as measuring two dimensions. [Tables 23.4](#) also helps us think about this.

1. Put a code into the item label to indicate the item subset to which the item belongs.

2a. Use [ISELECT=](#) for each subset code, and produce a person measure ([PFILE=](#)). [Cross-plot](#) the person measures.

or

2b. Do a Differential Person Functioning ([DPF=](#)) analysis based on the subset code. [Table 31.1](#) will give you an inter-item-subset t-test for each person.

[Table 23.0](#) Variance components scree plot for items

[Table 23.1, 23.11](#) Principal components plots of item loadings

[Table 23.2, 23.12](#) Item Principal components analysis/contrast of residuals

[Table 23.3, 23.13](#) Item contrast by persons

[Table 23.4, 23.14](#) Item contrast loadings sorted by measure

[Table 23.5, 23.15](#) Item contrast loadings sorted by entry number

[Table 23.6, 23.16](#) Person measures for item clusters in contrast. [Cluster Measure Plot](#) for Table 23.6.

[Table 23.99](#) Largest residual correlations for items

[Youtube video](#) explaining Table 23

These plots show the [contrasts](#) by plotting the unstandardized "raw" loading on each component against the item calibration (or person measure). The contrast shows items (or persons) with different residual patterns. A random pattern with few high loadings is expected.

The horizontal axis is the Rasch dimension. This has been extracted from the data prior to the analysis of residuals.

Letters "A,B,C,..." and "a,b,c,..." identify items (persons) with the most opposed loadings on the first contrast in the residuals. On subsequent contrasts, the items retain their first contrast identifying letters. When there are 9 items (persons) or less, the item number is displayed.

The items are clustered into 3 clusters on the right-side of the plot. This is because interpreting the PCA Components ("Contrasts") usually requires us to compare the items at the top of the plot against the items at the bottom, often ignoring the middle items. The purpose of the [plot](#) associated with [Table 23.6](#) is to help us see whether the 3 clusters of items are truly measuring different things. If they are measuring the same thing statistically, then usually no action is needed. If the clusters of items are measuring different things, then the analyst must decide what to do.

In the residuals, each item (person) is modeled to contribute one unit of randomness. Thus, there are as many residual variance units as there are items (or persons). For comparison, the amount of person (item) variance explained by the item (person) measures is approximated as units of that same size.

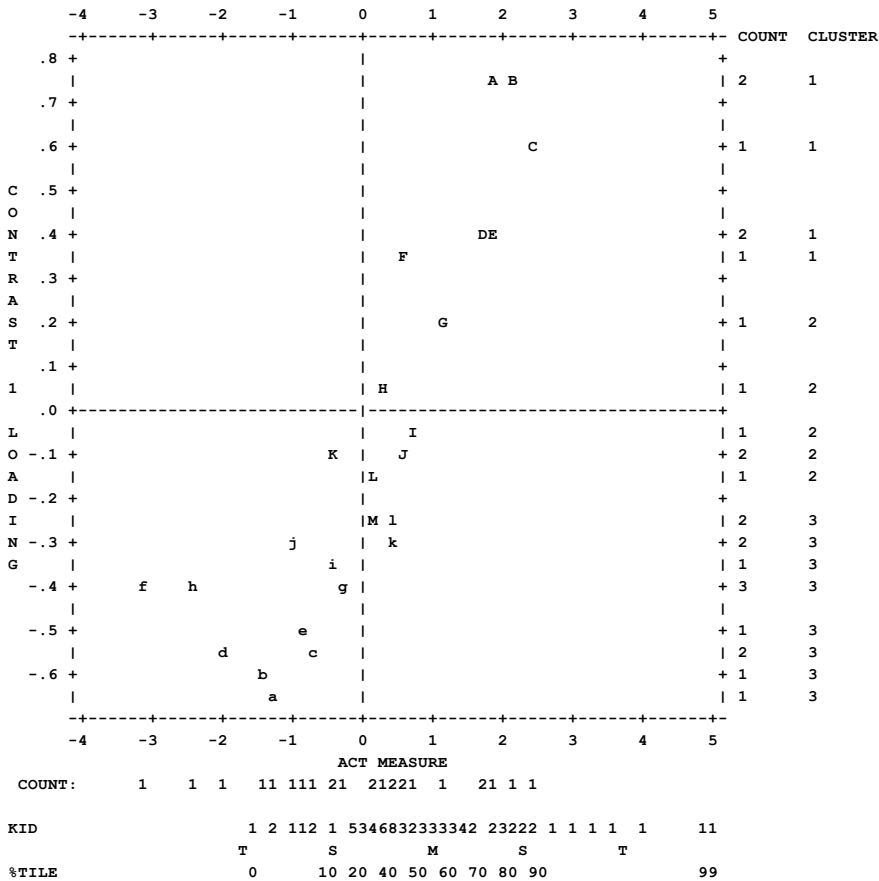
In the Figure below from [Example0.txt](#), the first contrast in the standardized residuals separates the items into 3 clusters. To identify the items, see [Tables 23.3, 24.3](#). In this example, the dimension is noticeable, with strength of around 5 out of 25 items. This is in the residual variance, i.e., in the part of the observations unexplained by the measurement model. But, hopefully, most of the variance in the observations has been explained by the model. The part of that explained variance attributable to the Persons is shown in variance units locally-rescaled to accord with the residual variances. In this example, the variance explained by the person measures is equivalent to 10 items. Consequently, the secondary dimension in the items is noticeable. The disattenuated correlation between person measures on item in Cluster 1 and person measures on items in Cluster 3 is less than 0.3. The secondary dimension underlying the 1st Contrast is biasing the person measures.

For items:

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = ACT information units

	Eigenvalue	Observed	Expected
Total raw variance in observations	= 50.9521	100.0%	100.0%
Raw variance explained by measures	= 25.9521	50.9%	50.7%
Raw variance explained by persons	= 10.3167	20.2%	20.2%
Raw Variance explained by items	= 15.6354	30.7%	30.6%
Raw unexplained variance (total)	= 25.0000	49.1%	100.0%
Unexplained variance in 1st contrast	= 4.6287	9.1%	18.5%

STANDARDIZED RESIDUAL CONTRAST 1 PLOT



Example: A teacher survey had hundreds of items contributed by numerous stakeholders (with special agendas), but almost no hypothesized constructs. In other words, it was a mess! PCA of residuals was used to discover subsets of items that cooperated. A useful strategy is to use WEIGHT=. Weight a core subset of items on a construct of interest "1", and weight all the other items "0". Then do a Winsteps analysis. All the items will be calibrated, but only the weighted items will contribute to the person measures. Then do a PCA of residuals. All the items will participate. The core items will cluster. Unweighted items that cluster with the core can be inspected, and added to the core if suitable. Then the process is repeated. When all the items for one construct have been identified, those items can be deleted from the dataset. The process begins again with the next core subset of items.

Correlation Table of person Measures on each of the Clusters:

Approximate relationships between the KID measures					
PCA Contrast	ACT Clusters	Pearson Correlation	Disattenuated Correlation	Pearson+Extr Correlation	Disattenuated+Extr Correlation
1	1 - 3	0.1404	0.2175	0.1951	0.2923
1	1 - 2	0.2950	0.4675	0.3589	0.5485
1	2 - 3	0.8065	1.0000	0.8123	1.0000

The items are anchored (fixed) at their difficulties from the main analysis (as reported in Table 14.1). Then the items are segmented into subtests according to their cluster numbers, and each person is measured on each subtest. The person

measures for each cluster of items are correlated with their measures from the other clusters of items, and reported here as the "Pearson Correlation". Each person measure for each cluster of items has a standard error. These error variances are removed to produce the "[Disattenuated Correlation](#)". If the disattenuated correlation approaches 1.0, then the person measures from the two clusters of items are statistically the same. We cannot reject the hypothesis that the two clusters of items are measuring the same thing. If the disattenuated correlation is out of range or undefined, it is reported as (1.00) or (-1.00) matching the sign of the observed correlation.

For the Pearson Correlation and the Disattenuated (Pearson) Correlation, persons with extreme scores on a cluster are omitted. For the Pearson+Extr(eme) Correlation and the Disattenuated (Pearson)+Extr(eme) Correlation, persons with extreme scores on a cluster are included.

These correlations are approximate. For more accurate correlations, please perform separate analyses of each cluster of items, and then use the [Scatterplot](#) function to investigate the relationships between the person measures estimated from the different clusters.

Roughly speaking, we look at the disattenuated correlations (otherwise measurement error clouds everything):

Correlations below 0.57 indicate that person measures on the two item clusters have half as much variance in common as they have independently. (*Cut-off for probably different latent variables?*)

Correlations above 0.71 indicate that person measures on the two item clusters have more than half their variance in common, so they are more dependent (= more the same thing) than independent (= different things).

Correlations above 0.82, twice as dependent as independent. (*Cut-off for probably the same latent variable?*)

Correlations above 0.87, three times as dependent as independent (*definitely the same thing*).

Example 1: From Table 23.3 of the analysis of a large empirical dataset:

Approximate relationships between the person measures

Contrast 1	Pearson	Disattenuated
Item Clusters	Correlation	Correlation
1 - 2	.5151	.5762

Actual relationships of person measures estimated from separate analyses of the two clusters of items:

Contrast 1	Pearson	Disattenuated
Item Clusters	Correlation	Correlation
1 - 2	.5083	.6352

14.32.3 Table 23.2, 23.12 Item Principal components analysis/contrast of residuals

Please do not interpret this as a usual factor analysis. These plots show contrasts between opposing factors, not loadings on one factor. For more discussion, see [dimensionality and contrasts](#).

This Table decomposes the matrix of item correlations based on residuals to identify possible other contrasts (dimensions) that may be affecting response patterns. Specify [PRCOMP=S](#) or =R or =L to obtain this Table.

[Table 23.0](#) Variance components scree plot for items

[Table 23.1, 23.11](#) Principal components plots of item loadings

[Table 23.2, 23.12](#) Item Principal components analysis/contrast of residuals

[Table 23.3, 23.13](#) Item contrast by persons

[Table 23.4, 23.14](#) Item contrast loadings sorted by measure

[Table 23.5, 23.15](#) Item contrast loadings sorted by entry number

[Table 23.6, 23.16](#) Person measures for item clusters in contrast. [Cluster Measure Plot](#) for Table 23.6.

[Table 23.99](#) Largest residual correlations for items


[Youtube video](#) explaining Table 23

Prior to this first contrast, the Rasch dimension has been extracted from the data. Residuals are those parts of the observations not explained by the Rasch dimension. According to Rasch specifications, these should be random and show no structure. The contrasts show conflicting local patterns in inter-item correlations based on residuals or their transformations. Letters "E", "b", etc. relate items to their unstandardized "raw" loadings on the first contrast. In this Table, "bugs", "rat" and "cans" contrast with "Grow garden". Since "bugs", "rat" and "can" misfit conspicuously, they load on a second dimension in the data.

CON-	TRAST	LOADING	MEASURE	INFINIT	OUTFIT	ENTRY	NUMBER	ACT	G
			MNSQ	MNSQ	MNSQ				R
1	1	.76	1.81	1.32	1.84	A	20	Watch bugs	2
1	1	.75	2.16	2.40	4.19	B	23	Watch a rat	2
1	1	.62	2.44	2.31	3.57	C	5	Find bottles and cans	1
1	1	.40	1.68	1.11	1.21	D	8	Look in sidewalk cracks	1
1	1	.40	1.77	.90	.92	E	4	Watch grass change	1
1	1	.38	.60	.95	.94	F	16	Make a map	2
1	1	.19	1.11	.99	1.02	G	7	Watch animal move	1
1	2	.06	.30	.83	.77	H	6	Look up strange animal or plan	1
1	2	-.64	-1.33	1.24	.95	a	13	Grow garden	1
1	2	-.61	-1.52	.80	.58	b	10	Listen to bird sing	1
1	2	-.56	-.73	.95	.74	c	2	Read books on animals	1
1	2	-.53	-2.09	.71	.51	d	12	Go to museum	1
1	2	-.49	-.83	.83	.63	e	21	Watch bird make nest	2
1	2	-.41	-3.10	1.50	1.23	f	18	Go on picnic	2
1	2	-.39	-.29	.88	.77	g	24	Find out what flowers live on	2
1	2	-.38	-2.44	1.08	1.10	h	19	Go to zoo	2
1	2	-.36	-.46	.76	.62	i	15	Read animal stories	2
1	2	-.33	-.99	.64	.50	j	11	Find where animal lives	1
1	2	-.30	.41	.59	.55	k	3	Read books on plants	1
1	2	-.25	.42	.81	.74	l	14	Look at pictures of plants	2
1	2	-.24	.17	.64	.57	m	17	Watch what animals eat	2
1	2	-.15	.09	.81	.72	l	22	Find out what animals eat	2
1	2	-.11	-.42	.56	.50	K	1	Watch birds	1
1	2	-.08	.70	1.20	1.19	J	9	Learn weed names	1
1	2	-.07	.53	.78	.72	I	25	Talk w/friends about plants	2

Contrast these

with these

 cluster number

The loading is that on the first PCA contrast. It is unstandardized. In the factor analysis literature, values of ± 4 or more extreme are considered substantive. To standardize the loading, divide the loadings by their root-mean-square. The measures and mean-square statistics are the same as those reported in [Table 10.1](#) etc. The letters under "ENTRY NUMBER" refer to the plots in [Table 23.2](#)

The "cluster number" indicates a statistical clustering of the loadings (useful for splitting these items into unidimensional substests). The clusters are obtained by doing a cluster-analysis of the loadings. The Fisher-linearized loadings are assigned to three (or less) clusters based on the centroids of the clusters.

To copy numbers out of this Table, use WORD to copy a [rectangle of text](#) or copy-and-paste into [Excel, then "text to columns"](#).

14.32.4 Table 23.3, 23.13 Item contrast by persons

Please do not interpret this as a usual factor analysis. These plots show contrasts between opposing factors, not loadings on one factor. For more discussion, see [dimensionality and contrasts](#).

The effect on the persons of the contrast between the oppositely loading items ([Table 23.1](#)) at the top and bottom of the contrast plot is shown here.

[Table 23.0](#) Variance components scree plot for items

[Table 23.1, 23.11](#) Principal components plots of item loadings

[Table 23.2, 23.12](#) Item Principal components analysis/contrast of residuals

[Table 23.3, 23.13](#) Item contrast by persons

[Table 23.4, 23.14](#) Item contrast loadings sorted by measure

[Table 23.5, 23.15](#) Item contrast loadings sorted by entry number

[Table 23.6, 23.16](#) Person measures for item clusters in contrast. [Cluster Measure Plot](#) for Table 23.6.

[Table 23.99](#) Largest residual correlations for items

[Youtube video](#) explaining Table 23

Responses by up to 9 persons (or on up to 9 items) to the items (or by the persons) with extreme positive loadings and the items (or persons) with extreme negative loadings are identified. These responses are seen to be higher, as expected, or lower than the model predicts. Counts of these are obtained for each person (or item). The persons (or items) showing the biggest impact of this contrast are listed first. Items and persons showing the most contrast are chosen for this Table based on the "Liking for Science" data.

Table 23.3 shows the impact of the first contrast in Table 23.1 on persons:

ACT contrast 1 CONTRASTING RESPONSES BY PUPILS

FAVORS TOP										
TOP 7 ACTS					BOTTOM 7 ACTS					
HIGH	EXP.	LOW	HIGH	EXP.	LOW	HIGH	EXP.	LOW		
5	2	0	0	3	4	15	DYSON, STEPHIE NINA			Persons scoring higher than expected on the top 7 items
5	2	0	0	3	4	26	NEIMAN, RAYMOND			
FAVORS BOTTOM										
TOP 7 ACTS					BOTTOM 7 ACTS					
HIGH	EXP.	LOW	HIGH	EXP.	LOW		items			
0	3	4	0	7	0	8	LAMBERT, MD., ROSS W.			Persons scoring higher than expected on the bottom 7 items
0	3	4	0	7	0	11	HSIEH, PAUL FRED			
0	5	2	3	3	1	32	ROSSNER, JACK			
0	3	4	0	7	0	51	REISS, STEVE			

FAVORS TOP	These persons score higher on the items at the top of the Contrast 1 plot in Table 23.1 than the items at the bottom of the contrast plot
FAVORS BOTTOM	These persons score higher on the items at the bottom of the Contrast 1 plot in Table 23.1 than the items at the top of the contrast plot
TOP 7 ACTS	Responses for the top 7 items (or top half of the items if less than 7) on the Contrast plot
BOTTOM 7 ACTS	Responses for the bottom 7 items (or top half of the items if less than 7) on the Contrast plot
HIGH	Person response is higher than expected on this many items
EXP.	Person response is expected (within 0.5 score points) on this many items
LOW	Person response is lower than expected on this many items
15 DYSON, STEPHIE NINA	Person entry number and person label. Persons with greatest differential performance between TOP and BOTTOM are shown with strongest difference listed first.
15 DYSON, STEPHIE NINA 8 LAMBERT, MD., ROSS W.	The effect of the 1st Contrast is more strongly seen in the performance of these two persons. Is there something about them that interacts with the set of items?

14.32.5 Table 23.4, 23.14 Item contrast loadings sorted by measure

CONTRAST 1 FROM PRINCIPAL COMPONENT ANALYSIS
STANDARDIZED RESIDUAL LOADINGS FOR TAP (SORTED BY MEASURE)

CON CL	LOADING	MEASURE	IN FIT	OUT FIT	ENTRY	NUMBER	TAP
1	3	-.31	4.80	.74	.11	d	15 1-3-2-4-1-3

1 1	.58	4.80	.74	.11	B	16	1-4-2-3-1-4	
1 1	.58	4.80	.74	.11	C	17	1-4-3-1-2-4	
1 3	-.52	3.37	1.56	1.49	b	14	1-4-2-3-4-1	
1 3	-.62	2.24	1.16	1.06	a	12	1-3-2-4-3	
1 2	.28	1.95	.70	.38	F	13	1-4-3-2-4	
1 3	-.41	.79	1.07	.79	c	11	1-3-1-2-4	
1 2	-.12	-1.57	1.06	.83	f	10	2-4-3-1	
1 1	.45	-2.35	.59	.43	E	8	1-4-2-3	
1 1	.52	-3.38	1.17	.96	D	6	3-4-1	
1 2	-.06	-3.38	.62	.21	g	9	1-3-2-4	
1 2	-.02	-3.83	1.04	.52	G	5	2-1-4	
1 1	.73	-3.83	1.33	2.21	A	7	1-4-3-2	
1 2	-.16	-4.40	.90	.35	e	4	1-3-4	

See [Table 23.2](#)

[Table 23.0](#) Variance components scree plot for items

[Table 23.1, 23.11](#) Principal components plots of item loadings

[Table 23.2, 23.12](#) Item Principal components analysis/contrast of residuals

[Table 23.3, 23.13](#) Item contrast by persons

[Table 23.4, 23.14](#) Item contrast loadings sorted by measure

[Table 23.5, 23.15](#) Item contrast loadings sorted by entry number

[Table 23.6, 23.16](#) Person measures for item clusters in contrast. [Cluster Measure Plot](#) for Table 23.6.

[Table 23.99](#) Largest residual correlations for items

[Youtube video](#) explaining Table 23

14.32.6 Table 23.5, 23.15 Item contrast loadings sorted by entry number

CONTRAST 1 FROM PRINCIPAL COMPONENT ANALYSIS
STANDARDIZED RESIDUAL LOADINGS FOR TAP (SORTED BY ENTRY)

CON CL		INFIT OUTFIT			ENTRY			
TRA US	LOADING	MEASURE	MNSQ	MNSQ	NUMBER TAP			
1 2	-.16	-4.40	.90	.35	e	4	1-3-4	
1 2	-.02	-3.83	1.04	.52	G	5	2-1-4	
1 1	.52	-3.38	1.17	.96	D	6	3-4-1	
1 1	.73	-3.83	1.33	2.21	A	7	1-4-3-2	
1 1	.45	-2.35	.59	.43	E	8	1-4-2-3	
1 2	-.06	-3.38	.62	.21	g	9	1-3-2-4	
1 2	-.12	-1.57	1.06	.83	f	10	2-4-3-1	
1 3	-.41	.79	1.07	.79	c	11	1-3-1-2-4	
1 3	-.62	2.24	1.16	1.06	a	12	1-3-2-4-3	
1 2	.28	1.95	.70	.38	F	13	1-4-3-2-4	
1 3	-.52	3.37	1.56	1.49	b	14	1-4-2-3-4-1	
1 3	-.31	4.80	.74	.11	d	15	1-3-2-4-1-3	
1 1	.58	4.80	.74	.11	B	16	1-4-2-3-1-4	
1 1	.58	4.80	.74	.11	C	17	1-4-3-1-2-4	

See [Table 23.2](#)

[Table 23.0](#) Variance components scree plot for items

[Table 23.1, 23.11](#) Principal components plots of item loadings

[Table 23.2, 23.12](#) Item Principal components analysis/contrast of residuals

[Table 23.3, 23.13](#) Item contrast by persons

[Table 23.4, 23.14](#) Item contrast loadings sorted by measure

[Table 23.5, 23.15](#) Item contrast loadings sorted by entry number

[Table 23.6, 23.16](#) Person measures for item clusters in contrast. [Cluster Measure Plot](#) for Table 23.6.

[Table 23.99](#) Largest residual correlations for items

[Youtube video](#) explaining Table 23

14.32.7 Table 23.6, 23.16 Person measures for item clusters in contrast

CONTRAST 1 CLUSTER MEASURES									
		CLUSTER 1		CLUSTER 2		CLUSTER 3			
KID	STATUS	MEASURE	S.E.	MEASURE	S.E.	MEASURE	S.E.	KID LABEL	
1	1	-2.44	1.28	-3.37	1.11	-.48	1.98	Richard M	1111111000
2	1	.90	3.21	.37	1.84	-.50	2.00	Tracie F	1111111111

	3	1	.90	3.21	-1.99	1.28	1.22	1.34	Walter	M	1111111110	
	4	1	-3.95	1.27	-3.37	1.11	-.48	1.98	Blaise	M	1111001010	
	5	1	.90	3.21	.37	1.84	-.50	2.00	Ron	M	1111111111	
	6	1	.90	3.21	.37	1.84	-.50	2.00	William	M	1111111111	
	7	1	.90	3.21	2.83	2.11	4.39	1.34	Susan	F	1111111111	
	8	1	.90	3.21	.37	1.84	-.50	2.00	Linda	F	1111111111	
	9	1	.90	3.21	.37	1.84	-.50	2.00	Kim	F	1111111111	
	10	1	.90	3.21	.37	1.84	1.22	1.34	Carol	F	1111111111	
	11	1	.90	3.21	-3.37	1.11	-.50	2.00	Pete	M	1110111110	
	12	1	-3.95	1.27	.36	1.83	-.48	1.98	Brenda	F	1111101011	
	13	1	-3.95	1.27	.35	1.83	2.80	1.22	Mike	M	1111100111	
	14	1	.90	3.21	.37	1.84	1.22	1.34	Zula	F	1111111111	
	15	1	.90	3.21	2.83	2.11	2.80	1.22	Frank	M	1111111111	
	16	1	.90	3.21	-1.99	1.28	1.22	1.34	Dorothy	F	1111111110	
	17	1	.90	3.21	-1.99	1.28	-.50	2.00	Rod	M	1111011111	
	18	1	.90	3.21	2.83	2.11	-.50	2.00	Britton	F	1111111111	
	19	1	.90	3.21	-1.99	1.28	-.50	2.00	Janet	F	1111111110	
	20	1	.90	3.21	2.83	2.11	-.50	2.00	David	M	1111111111	
	21	1	.90	3.21	2.83	2.11	1.22	1.34	Thomas	M	1111111111	
	22	1	.90	3.21	.37	1.84	2.80	1.22	Betty	F	1111111111	
	23	1	.90	3.21	2.83	2.11	1.22	1.34	Bert	M	1111111111	
	24	1	6.51	1.96	2.83	2.11	1.22	1.34	Rick	M	1111111111	
	25	1	-3.95	1.27	-4.66	1.22	-.48	1.98	Don	M	1110110000	
	26	1	.90	3.21	.37	1.84	-.50	2.00	Barbara	F	1111111111	
	27	1	-2.44	1.28	-3.37	1.11	-.48	1.98	Adam	M	1111111000	
	28	1	.90	3.21	-1.99	1.28	1.22	1.34	Audrey	F	1111111110	
	29	1	-3.95	1.27	.35	1.83	2.80	1.22	Anne	F	1111110011	
	30	1	.90	3.21	-1.99	1.28	-.50	2.00	Lisa	F	1111111110	
	31	1	.90	3.21	.37	1.84	-.50	2.00	James	M	1111111111	
	32	1	.90	3.21	.37	1.84	1.22	1.34	Joe	M	1111111111	
	33	1	-3.95	1.27	-3.37	1.11	-.48	1.98	Martha	F	1111001001	
	34	1	.90	3.21	.37	1.84	2.80	1.22	Elsie	F	1111111111	
	35	-1	-5.50	1.92	-6.16	1.91	-.50	1.99	Helen	F	1110000000	

Reliability: 0.20 0.43 0.00

Approximate relationships between the KID measures

PCA	ACT	Pearson	Disattenuated	Pearson+Extr	Disattenuated+Extr
Contrast	Clusters	Correlation	Correlation	Correlation	Correlation
1	1 - 3	0.1404	0.2175	0.1951	0.2923
1	1 - 2	0.2950	0.4675	0.3589	0.5485
1	2 - 3	0.8065	1.0000	0.8123	1.0000

This subtable shows the estimated person measures and standard errors for each cluster of items in the PCA contrast. The cluster correlations are computed from these measures and reported in [Table 23.0](#). Scatterplots of the person measures in Table 23.6 for each item cluster in [Table 23.1](#) are shown in the [Cluster Measure Plot](#).

For Smith's (2002) "dimensionality" method:

1. Analyze all the data. Table 23.1 identifies Item cluster 1 and Item cluster 3
2. Table 23.6 has the person measures and S.E.s.
3. Conduct *t*-tests.

Smith EV Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas.* 2002; 3(2):205-31.

[Table 23.0](#) Variance components scree plot for items

[Table 23.1, 23.11](#) Principal components plots of item loadings

[Table 23.2, 23.12](#) Item Principal components analysis/contrast of residuals

[Table 23.3, 23.13](#) Item contrast by persons

[Table 23.4, 23.14](#) Item contrast loadings sorted by measure

[Table 23.5, 23.15](#) Item contrast loadings sorted by entry number

[Table 23.6, 23.16](#) Person measures for item clusters in contrast. [Cluster Measure Plot](#) for Table 23.6.

[Table 23.99](#) Largest residual correlations for items

[Youtube video](#) explaining Table 23

14.32.8 Table 23.99 Largest residual correlations for items

These Tables show items (Table 23.99) that may be locally dependent. Specify [PRCOMP=R](#) (for score residuals, Yen Q3) or [PRCOMP=S](#) or [Y](#) (for standardized residuals) or [PRCOMP=L](#) (for logit residuals) to obtain this Table. Residuals are those parts of the data not explained by the Rasch model. High correlation of residuals for two items (or persons) indicates

that they may not be locally independent, either because they duplicate some feature of each other or because they both incorporate some other shared dimension.

[Table 23.0](#) Variance components scree plot for items

[Table 23.1, 23.11](#) Principal components plots of item loadings

[Table 23.2, 23.12](#) Item Principal components analysis/contrast of residuals

[Table 23.3, 23.13](#) Item contrast by persons

[Table 23.4, 23.14](#) Item contrast loadings sorted by measure

[Table 23.5, 23.15](#) Item contrast loadings sorted by entry number

[Table 23.6, 23.16](#) Person measures for item clusters in contrast. [Cluster Measure Plot](#) for Table 23.6.

[Table 23.99](#) Largest residual correlations for items

[Youtube video](#) explaining Table 23

Missing data are deleted pairwise if both of a pair are missing or PRCOMP=O (for observations), otherwise missing data are replaced by their Rasch expected residuals of 0.

LARGEST STANDARDIZED RESIDUAL CORRELATIONS
USED TO IDENTIFY DEPENDENT TAP

CORRELATION	ENTRY		ENTRY	
	NUMBER	TAP	NUMBER	TAP
1.00	16	1-4-2-3-1-4	17	1-4-3-1-2-4
.52	6	3-4-1	7	1-4-3-2
.45	7	1-4-3-2	8	1-4-2-3
.28	8	1-4-2-3	9	1-3-2-4
-.67	8	1-4-2-3	14	1-4-2-3-4-1
-.58	7	1-4-3-2	14	1-4-2-3-4-1
-.49	6	3-4-1	12	1-3-2-4-3
-.48	7	1-4-3-2	11	1-3-1-2-4
-.43	15	1-3-2-4-1-3	16	1-4-2-3-1-4
-.43	15	1-3-2-4-1-3	17	1-4-3-1-2-4
-.38	7	1-4-3-2	12	1-3-2-4-3
-.37	12	1-3-2-4-3	16	1-4-2-3-1-4
-.37	12	1-3-2-4-3	17	1-4-3-1-2-4
-.32	5	2-1-4	10	2-4-3-1
-.31	6	3-4-1	10	2-4-3-1
-.29	12	1-3-2-4-3	13	1-4-3-2-4
-.29	11	1-3-1-2-4	13	1-4-3-2-4
-.28	6	3-4-1	11	1-3-1-2-4
-.28	4	1-3-4	8	1-4-2-3
-.26	4	1-3-4	5	2-1-4

Note: Redundant correlations of 1.0 are not printed. If A has a correlation of 1.0 with B, and also with C, assume that B and C also have a correlation of 1.0. After eliminating redundant correlations, the largest correlations are shown in the Table. To see all correlations, output [ICORFILE=](#) or [PCORFILE=](#)

In this Table, high positive residual correlations may indicate local item dependency (LID) between pairs of items or persons. When raw score residual correlations are computed, [PRCOMP=R](#), it corresponds to Wendy Yen's Q3 statistic. It is used to detect dependency between pairs of items or persons. Wendy Yen suggests a small positive adjustment to the correlation of size $1/(L-1)$ where L is the test length.

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 8, 125-145.

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, 30, 187-213.

Local dependence would be a large positive correlation. Highly locally dependent items (Corr. > +.7), such as items "Q." and "R." share more than half their "random" variance, suggesting that only one of the two items is needed for measurement. But, in classical test theory terms, these items may have the highest point-biserial correlations and so be the "best" items.

A large negative correlation indicates the opposite of local dependence, as usually conceptualized. If you look at the item fit tables, item "J." or "R." is likely to have large misfit.

Remember that "common variance = correlation²", so items 10 and 11 only share .40*.40 = 16% of the variance in their residuals in common. 84% of each of their residual variances differ. In this Table we are usually only interested in correlations that approach 1.0 or -1.0, because that may indicate that the pairs of items are duplicative or are dominated by a shared factor.

Suggestion: simulate Rasch-fitting data like yours using the Winsteps [SIFILE=](#) option. Analyze these data with Winsteps. Compare your correlation range with that of the simulated data.

An influential paper says "[Readers] expect to see that this issue has been dealt with, if only to report that no response dependency was found." However, even simulated Rasch-fitting data will report some accidental response dependency, so this statement is too extreme. Let's amend it. to say "if only to report that no **consequential** response dependency was found."

Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum. 2007 Dec 15;57(8):1358-62.

Locally-Dependent Items

In practical terms, a correlation of $r=0.40$ is low dependency. The two items only have $0.4*0.4=0.16$ of their variance in common. Correlations need to be around 0.7 before we are really concerned about dependency.

If you want to create one super-item out of two dependent items, then use Excel (or similar) to add the scored responses on the two item together. Include the super-item in the data file instead of the dependent items. Change `CODES=` and use `ISGROUPS=` to model the additional super-item. You will notice a very small reduction in the variances of the measures.

Procedure:

1. Winsteps analyze the original data
 2. Output the scored responses: "Output file", "RFILE=", responses.xls
 3. In Excel, open responses.xls
 4. Sum the scored responses of dependent items into a new item
 5. Delete the original dependent items
 6. Save responses.xls
 7. In Winsteps, "Excel/RSSST" menu, "Excel", Import responses.xls
 8. Create Winsteps file of the new set of items (you may need edit ISGROUPS=)
 9. Analyze the new Winsteps control file
-

Item Calibration without Local Item Dependency (LID)

If you need item difficulties from your dataset, but without the effect of LID, here is a procedure:

1. Analyze all the dataset with Winsteps with `PRCOMP=R` to produce the raw residuals.
Output `IFILE=iforig.txt` as a reference for the original data.
2. Output [ICORFILE=](#) for the raw residuals to Excel in list format
3. Sort the list by correlation
4. Look at the top and bottom of the list. How many inter-item residual correlations are $>.02$ (or your correlation cut-off value)? (In my empirical dataset about .1%)
Save these item pairings to avoid administering these pairs of item together in CAT or test forms.
5. Make an [IDFILE=LID.txt](#) list of one item from each item pair with inter-item residual correlation $>.02$ (or your correlation cut-off value)
6. Reanalyze all the data with Winsteps and `IDFILE=LID.txt`
7. Output [PFILE=pf.txt](#) - these are the person measures matching only the LID-free items
8. Reanalyze all the data with Winsteps without `IDFILE=`. Include `PAFILE=pf.txt`
This forces all the item difficulties and rating-scale structures to conform with the LID-free person measures.
Anchoring the person measures prevents LID from impacting the item difficulties.

9. Output [IFILE=if.txt](#), [SFILE=sf.txt](#) - these are the item difficulty and rating-scale threshold values for the item bank or whatever.

10. Scatterplot if.txt against iforig.txt. How much is the impact of LID on iforig.txt? Was all this work worth it?

14.33 Table 24 Person multidimensionality

14.33.1 Table 24.0 Variance components scree plot for persons

Table 24.0 shows a variance decomposition of the observations for the persons. This is not produced for [PRCOMP=0](#).

[Table 24.0](#) Variance components scree plot for persons

[Table 24.1, 24.11](#) Principal components plots of person loadings

[Table 24.2, 24.12](#) Person Principal components analysis/contrast of residuals

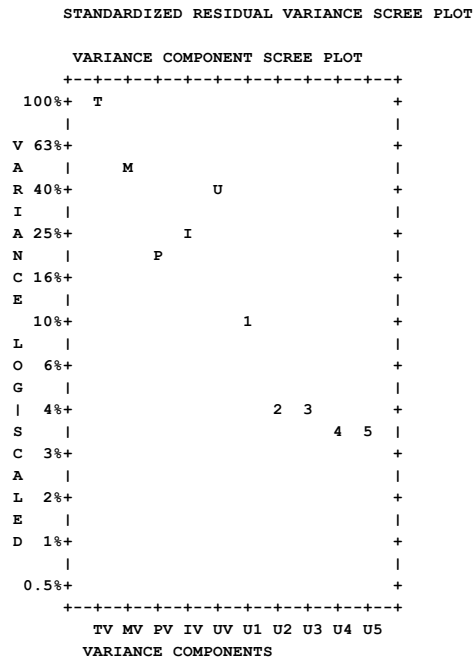
[Table 24.3, 24.13](#) Person contrast by items

[Table 24.4, 24.14](#) Person contrast loadings sorted by measure

[Table 24.5, 24.15](#) Person contrast loadings sorted by entry number

[Table 24.99](#) Largest residual correlations for persons

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = KID information units			
		Eigenvalue	Observed Expected
Total raw variance in observations	=	150.8183	100.0% 100.0%
Raw variance explained by measures	=	76.8183	50.9% 50.7%
Raw variance explained by persons	=	30.5375	20.2% 20.2%
Raw Variance explained by items	=	46.2809	30.7% 30.6%
Raw unexplained variance (total)	=	74.0000	49.1% 100.0% 49.3%
Unexplned variance in 1st contrast	=	18.2333	12.1% 24.6%
Unexplned variance in 2nd contrast	=	6.9841	4.6% 9.4%
Unexplned variance in 3rd contrast	=	6.1491	4.1% 8.3%
Unexplned variance in 4th contrast	=	4.9831	3.3% 6.7%
Unexplned variance in 5th contrast	=	4.8045	3.2% 6.5%



For details, please see [Table 23.0](#)

Example:

In this example, 50.9% of the variance in the data is explained by the measures. If the data fit the model perfectly, 50.7% would be explained. These percentages are close, indicating that the estimation of Rasch measures has been successful.

The variance in the data explained by the item difficulties, 30.7%, is larger than the variance explained by the person abilities, 20.2%. For dichotomies, this usually reflects the person and item measure standard deviations in [Table 3.1](#).

The unexplained variance in the data is 49.1%. This includes the Rasch-predicted randomness and any departures in the data from Rasch criteria, such as those due to multidimensionality in the items.

The first contrast has an eigenvalue of 18.2, corresponding to a strength of over 18 persons, and considerably larger than the largest eigenvalue expected by chance. It corresponds to 12.1% of the variance in the data.

14.33.2 Table 24.1, 24.11 Principal components plots of person loadings

Please do not interpret this as a usual factor analysis. These plots show contrasts between opposing factors, identified as "A,B,.. " and "a,b,..", not loadings on one factor. For more discussion, see [dimensionality and contrasts](#).

Quick summary:

(a) the X-axis is the measurement axis. So we are not concerned about quadrants, we are concerned about vertical differences. The Table 24 plots show contrasts between types of persons: those at the top vs. those at the bottom.

(b) "How much" is important. See the Variance Table explained in [Table 24.0](#). Important differences have eigenvalues greater than 2.0.

(c) If the difference is important, it suggests that we divide the dataset into two pieces: the persons in the top half of the plot and the persons in the bottom half. Perform two separate analyses and cross-plot and correlate the item calibrations. We will then see for which items the differences are important. Usually, for a carefully designed instrument, it is such a small segment that we decide it is not worth thinking of the test as measuring two dimensions. [Table 24.4](#) also helps us think about this.

1. Put a code into the person label to indicate the subset to which the item belongs.

2a. Use [PSELECT=](#) for each subset code, and produce a person measure ([IFILE=](#)). [Cross-plot](#) the item calibrations.
or

2b. Do a Differential Item Functioning ([DIF=](#)) analysis based on the subset code. [Table 30.1](#) will give you an inter-person-subset t-test for each item.

Alternatively, [transpose](#) the dataset, and investigate persons as though they are items.

[Table 24.0](#) Variance components scree plot for persons

[Table 24.1, 24.11](#) Principal components plots of person loadings

[Table 24.2, 24.12](#) Person Principal components analysis/contrast of residuals

[Table 24.3, 24.13](#) Person contrast by items

[Table 24.4, 24.14](#) Person contrast loadings sorted by measure

[Table 24.5, 24.15](#) Person contrast loadings sorted by entry number

[Table 24.99](#) Largest residual correlations for persons

These plots show the [contrasts](#) by plotting the unstandardized "raw" loading on each component against the person measure. The contrast shows persons with different residual patterns. A random pattern with few high loadings is expected.

The horizontal axis is the Rasch dimension. This has been extracted from the data prior to the analysis of residuals.

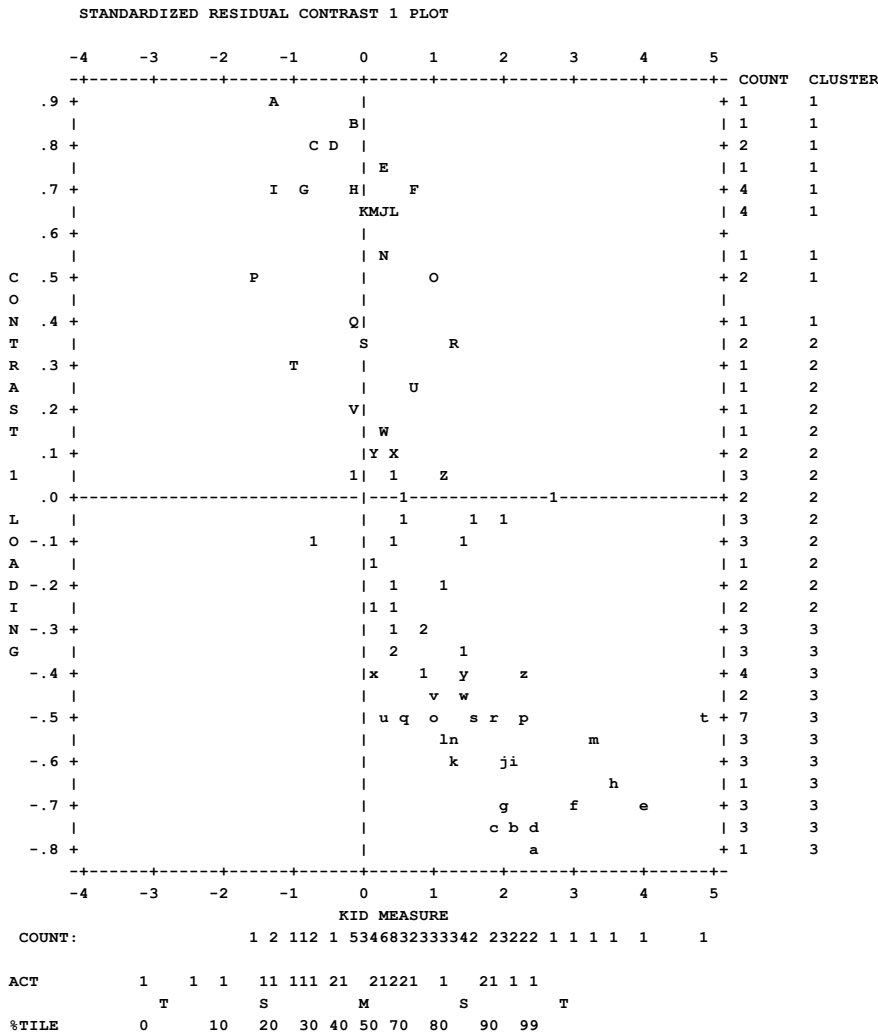
Letters "A,B,C,..." and "a,b,c,..." identify persons with the most opposed loadings on the first contrast in the residuals. On subsequent contrasts, the items retain their first contrast identifying letters. When there are 9 persons or less, the person number is displayed.

In the residuals, each person is modeled to contribute one unit of randomness. Thus, there are as many residual variance units as there are persons. For comparison, the amount of variance explained by the measures is approximated as units of that same size.

In the Figure below , the first contrast in the standardized residuals separates the persons into 3 clusters. To identify the persons, see [Table 24.3](#). In this example, the dimension is noticeable, with strength of around 18 out of 74 persons. This is in the residual variance, i.e., in the part of the observations unexplained by the measurement model. But, hopefully, most of the variance in the observations has been explained by the model. The part of that explained variance attributable to the persons is shown in variance units locally-rescaled to accord with the residual variances. In this example, the variance explained by the person measures is equivalent to 30 persons. Consequently, the secondary dimension (or whatever) in the persons is noticeable.

For persons:

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = KID information units			
		Eigenvalue	Observed Expected
Total raw variance in observations	=	150.8183	100.0% 100.0%
Raw variance explained by measures	=	76.8183	50.9% 50.7%
Raw variance explained by persons	=	30.5375	20.2% 20.2%
Raw Variance explained by items	=	46.2809	30.7% 30.6%
Raw unexplained variance (total)	=	74.0000	49.1% 100.0% 49.3%
Unexplned variance in 1st contrast	=	18.2333	12.1% 24.6%



The plot shows a contrast in the residuals for PERSONS. Each letter is a person up to a maximum of 52 persons, A-Z and a-z. For persons 53-74, "1" means that there is one person at that location on the plot. "2" means that there are two persons, etc.

14.33.3 Table 24.2, 24.12 Person Principal components analysis/contrast of residuals

Please do not interpret this as a usual factor analysis. These plots show contrasts between opposing factors, not loadings on one factor. For more discussion, see [dimensionality and contrasts](#).

This Table decomposes the matrix of person correlations based on residuals to identify possible other contrasts (dimensions) that may be affecting response patterns. Specify [PRCOMP=S](#) or =R or =L to obtain this Table.

[Table 24.0](#) Variance components scree plot for persons

[Table 24.1, 24.11](#) Principal components plots of person loadings

[Table 24.2, 24.12](#) Person Principal components analysis/contrast of residuals

[Table 24.3, 24.13](#) Person contrast by items

[Table 24.4, 24.14](#) Person contrast loadings sorted by measure

[Table 24.5, 24.15](#) Person contrast loadings sorted by entry number

[Table 24.99](#) Largest residual correlations for persons

Prior to this first contrast, the Rasch dimension has been extracted from the data. Residuals are those parts of the observations not explained by the Rasch dimension. According to Rasch specifications, these should be random and show no structure. The contrasts show conflicting local patterns in inter-item (or inter-person) correlations based on residuals or their transformations. Letters "E", "b", etc. relate persons to their unstandardized "raw" loadings on the first contrast.

In this Table 24.3, Landman and Rossner contrast with Chazelle and Ernst. Since Landman and Rossner misfit, they load on a second dimension in the data.

CON-	TRAST	LOADING	MEASURE	INFIT	OUTFIT	ENTRY	NUMBER	KID
			MNSQ	MNSQ				
1	1	.90	-1.32	1.40	3.94	A	29	M Landman, Alan
1	1	.83	-.08	1.62	2.39	B	6	M Rossner, Tr Cat
1	1	.80	-.67	.88	1.33	C	5	F Rossner, Rebecca A.
1	1	.79	-.43	1.53	2.48	D	49	M Baudet, Gerard
1	1	.76	.26	.56	.74	E	52	M Stultz, Newell
1	1	.72	.73	1.65	1.59	F	65	M Corleone, Michael
1	1	.71	-.92	1.87	3.51	G	12	M Lieberman, Daniel
1	1	.69	-.08	1.41	1.83	H	9	M Schulz, Matthew
1	2	-.78	2.49	.86	.49	a	48	M Chazelle, Bernie
1	2	-.75	2.11	.93	.58	b	18	M Ernst, Richard Max
1	2	-.74	1.79	.68	.50	c	40	M Badenov, Boris
1	2	-.73	2.49	.52	.31	d	16	F Buff, Marge Baby
1	2	-.72	4.06	.61	.20	e	34	F Paster, Ruth
1	2	-.71	2.95	1.17	.50	f	45	F Mccloughlin, Billy
1	2	-.70	1.95	1.10	.70	g	57	F Clapp, Charlie
1	2	-.66	3.59	1.17	.39	h	17	F Schattner, Gail

Contrast these

with these

Cluster number

The loading is that on the first PCA contrast. It is unstandardized. In the factor analysis literature, values of ± 0.4 or more extreme are considered substantive. To standardize the loading, divide the loadings by their root-mean-square.

The measures and mean-square statistics are the same as those reported in [Table 6](#) etc.

The letters under "ENTRY NUMBER" refer to the plots in [Table 24.2](#).

The "cluster number" indicates a statistical clustering of the loadings (useful for splitting these persons into two or more unidimensional sub-samples). The clusters are obtained by doing a cluster-analysis of the loadings.

To copy numbers out of this Table, use WORD to copy a [rectangle of text](#) or copy-and-paste into [Excel](#), then ["text to columns"](#).

14.33.4 Table 24.3, 24.13 Person contrast by items

Please do not interpret this as a usual factor analysis. These plots show contrasts between opposing factors, not loadings on one factor. For more discussion, see [dimensionality and contrasts](#).

The effect on the items of the contrast between the oppositely loading persons (Table 24.1) at the top and bottom of the contrast plot is shown here.

- [Table 24.0](#) Variance components scree plot for persons
- [Table 24.1, 24.11](#) Principal components plots of person loadings
- [Table 24.2, 24.12](#) Person Principal components analysis/contrast of residuals
- [Table 24.3, 24.13](#) Person contrast by items
- [Table 24.4, 24.14](#) Person contrast loadings sorted by measure
- [Table 24.5, 24.15](#) Person contrast loadings sorted by entry number
- [Table 24.99](#) Largest residual correlations for persons

Responses by up to 9 persons (or on up to 9 items) to the items (or by the persons) with extreme positive loadings and the items (or persons) with extreme negative loadings are identified. These responses are seen to be higher, as expected, or lower than the model predicts. Counts of these are obtained for each person (or item). The persons (or items) showing the biggest impact of this contrast are listed first. Items and persons showing the most contrast are chosen for this Table based on the "Liking for Science" data.

Table 24.3 shows the impact of the first [contrast](#) in the person residuals on items:

```

PUPIL contrast 1 CONTRASTING RESPONSES BY ACTS
+-----+
|          FAVORS TOP          |
| TOP 9 PUPILS | BOTTOM 9 PUPILS|
| HIGH EXP. LOW | HIGH EXP. LOW |          Items biased towards top 9 persons
+-----+
|   9   0   0 |   0   0   9 | 23 WATCH A RAT
|   8   1   0 |   0   1   8 |  5 FIND BOTTLES AND CANS
|   7   2   0 |   1   7   1 | 20 WATCH BUGS
+-----+
|          FAVORS BOTTOM          |
| TOP 9 PUPILS | BOTTOM 9 PUPILS|
| HIGH EXP. LOW | HIGH EXP. LOW |          Items biased towards bottom 9 persons
+-----+
|   0   4   5 |   0   9   0 | 13 GROW GARDEN
|   0   5   4 |   0   9   0 |  2 READ BOOKS ON ANIMALS
|   0   5   4 |   0   9   0 | 15 READ ANIMAL STORIES
|   0   5   4 |   0   9   0 | 24 FIND OUT WHAT FLOWERS LIVE ON
+-----+

```

FAVORS TOP	These items score higher on the persons at the top of the Contrast 1 plot in Table 24.1 than the persons at the bottom of the contrast plot
FAVORS BOTTOM	These items score higher on the persons at the bottom of the Contrast 1 plot in Table 24.1 than the persons at the top of the contrast plot
TOP 7 PUPILS	Responses for the top 7 persons (or top half of the persons if less than 7) on the Contrast plot
BOTTOM 7 PUPILS	Responses for the bottom 7 persons (or top half of the persons if less than 7) on the Contrast plot
HIGH	item response is higher than expected on this many persons
EXP.	item response is expected (within 0.5 score points) on this many persons
LOW	item response is lower than expected on this many persons
23 WATCH A RAT	item entry number and item label.

	items with greatest differential performance between TOP and BOTTOM are shown with strongest difference listed first.
23 WATCH A RAT 13 GROW GARDEN	The effect of the 1st Contrast is more strongly seen in the performance of these two items. Is there something about them that interacts with the set of persons?

14.33.5 Table 24.4, 24.14 Person contrast loadings sorted by measure

CONTRAST 1 FROM PRINCIPAL COMPONENT ANALYSIS
STANDARDIZED RESIDUAL LOADINGS FOR KID (SORTED BY MEASURE)

```

-----
|CON CL|      |          INFIT OUTFIT| ENTRY |
|TRA US|LOADING|MEASURE  MNSQ MNSQ |NUMBER KID |
|-----+-----+-----+-----+-----+-----+-----+-----|
| 1 2 | -.12 | 3.73 .95 .39 |h 7 Susan F |
| 1 2 | .13 | 3.73 1.94 .88 |O 24 Rick M |
| 1 3 | -.22 | 2.85 .36 .15 |f 15 Frank M |
| 1 2 | .00 | 1.94 .52 .18 |p 21 Thomas M |
| 1 3 | -.22 | 1.94 .65 .22 |g 22 Betty F |
| 1 2 | .46 | 1.94 1.60 .67 |J 23 Bert M |
| 1 2 | .32 | 1.94 1.74 .71 |L 34 Elsie F |
| 1 2 | -.07 | .92 .39 .13 |l 10 Carol F |
| 1 2 | -.07 | .92 .39 .13 |m 14 Zula F |
| 1 2 | .46 | .92 .99 .32 |H 18 Britton F |
| 1 2 | .46 | .92 .99 .32 |I 20 David M |

```

See [Table 24.2](#)

- [Table 24.0](#) Variance components scree plot for persons
- [Table 24.1, 24.11](#) Principal components plots of person loadings
- [Table 24.2, 24.12](#) Person Principal components analysis/contrast of residuals
- [Table 24.3, 24.13](#) Person contrast by items
- [Table 24.4, 24.14](#) Person contrast loadings sorted by measure
- [Table 24.5, 24.15](#) Person contrast loadings sorted by entry number
- [Table 24.99](#) Largest residual correlations for persons

14.33.6 Table 24.5, 24.15 Person contrast loadings sorted by entry number

CONTRAST 1 FROM PRINCIPAL COMPONENT ANALYSIS
STANDARDIZED RESIDUAL LOADINGS FOR KID (SORTED BY ENTRY)

```

-----
|CON CL|      |          INFIT OUTFIT| ENTRY |
|TRA US|LOADING|MEASURE  MNSQ MNSQ |NUMBER KID |
|-----+-----+-----+-----+-----+-----+-----+-----|
| 1 2 | -.11 | -2.94 .61 .29 |i 1 Richard M |
| 1 1 | .95 | -.26 .18 .08 |A 2 Tracie F |
| 1 3 | -.37 | -.26 1.94 1.19 |c 3 Walter M |
| 1 2 | .00 | -3.61 .78 .36 |P 4 Blaise M |
| 1 1 | .95 | -.26 .18 .08 |B 5 Ron M |
| 1 1 | .95 | -.26 .18 .08 |C 6 William M |
| 1 2 | -.12 | 3.73 .95 .39 |h 7 Susan F |
| 1 1 | .95 | -.26 .18 .08 |D 8 Linda F |
| 1 1 | .95 | -.26 .18 .08 |E 9 Kim F |
| 1 2 | -.07 | .92 .39 .13 |l 10 Carol F |
| 1 2 | .00 | -2.23 1.02 .81 |Q 11 Pete M |
| 1 2 | .29 | -2.23 1.10 .51 |M 12 Brenda F |
| 1 3 | -.37 | -.26 4.12 5.28 |d 13 Mike M |
| 1 2 | -.07 | .92 .39 .13 |m 14 Zula F |

```

See [Table 24.2](#)

- [Table 24.0](#) Variance components scree plot for persons
- [Table 24.1, 24.11](#) Principal components plots of person loadings
- [Table 24.2, 24.12](#) Person Principal components analysis/contrast of residuals
- [Table 24.3, 24.13](#) Person contrast by items
- [Table 24.4, 24.14](#) Person contrast loadings sorted by measure

[Table 24.5.24.15](#) Person contrast loadings sorted by entry number

[Table 24.99](#) Largest residual correlations for persons

14.33.7 Table 24.99 Largest residual correlations for persons

Table 24.99 shows persons that may be locally dependent. Specify PRCOMP=R (for score residuals) or PRCOMP=S or Y (for standardized residuals) or PRCOMP=L (for logit residuals) to obtain this Table. Residuals are those parts of the data not explained by the Rasch model. High correlation of residuals for two items (or persons) indicates that they may not be locally independent, either because they duplicate some feature of each other or because they both incorporate some other shared dimension.

Missing data are deleted pairwise if both of a pair are missing or PRCOMP=O (for observations), otherwise missing data are replaced by their Rasch expected residuals of 0.

```

LARGEST STANDARDIZED RESIDUAL CORRELATIONS
USED TO IDENTIFY DEPENDENT KID
-----
|CORREL-| ENTRY      | ENTRY      |
| ATION|NUMBER KID  |NUMBER KID  |
|-----+-----+-----|
| 1.00 | 1 Richard M | 27 Adam    M |
| 1.00 | 2 Tracie F  | 5 Ron      M |
| 1.00 | 2 Tracie F  | 6 William  M |
| 1.00 | 2 Tracie F  | 8 Linda    F |
| 1.00 | 2 Tracie F  | 9 Kim      F |
| 1.00 | 2 Tracie F  | 26 Barbara F |
| 1.00 | 2 Tracie F  | 31 James   M |
| 1.00 | 10 Carol F  | 14 Zula    F |
| 1.00 | 10 Carol F  | 32 Joe     M |
| 1.00 | 16 Dorothy F| 28 Audrey  F |
| 1.00 | 18 Britton F| 20 David   M |
| 1.00 | 19 Janet F  | 30 Lisa    F |
|-----+-----+-----|
| -.82 | 4 Blaise M  | 25 Don     M |
| -.75 | 21 Thomas M| 34 Elsie   F |
| -.75 | 19 Janet F  | 33 Martha  F |
| -.75 | 30 Lisa F   | 33 Martha  F |
| -.70 | 22 Betty F  | 23 Bert    M |
| -.67 | 18 Britton F| 22 Betty   F |
| -.67 | 20 David M  | 22 Betty   F |
| -.65 | 16 Dorothy F| 33 Martha  F |
-----

```

For more about this Table, see [Table 23.99](#)

14.34 Table 25 Item report in displacement order

14.34.1 Table 25.1 Item statistics in displacement order

(controlled by [USCALE=](#), [UMEAN=](#), [UDECI=](#), [LOCAL=](#), [TOTAL=](#), [DISCRIMINATION=](#), [ASYMPTOTE=](#), [PVALUE=](#))

ITEM STATISTICS: DISPLACEMENT ORDER

For details, please see [Table 10.1](#)

14.34.2 Table 25.2 Item statistics graphically

Please see [Table 10.2](#)

14.34.3 Table 25.3 Item option & distractor frequencies

(controlled by [Distractors=Y](#), [OSORT=](#), [CFILE=](#), [PTBIS=](#))

ITEM OPTION FREQUENCIES are output if Distractors=Y. These show occurrences of each of the valid data codes in [CODES=](#), and also of [MISSCORE=](#) in the input data file. Counts of responses forming part of extreme scores are included. Only items included in the corresponding main table are listed. These statistics are also in [DISFILE=](#), which includes entries even if the code is not observed for the item. See also [Distractor Analysis](#).

[OSORT=](#) controls the ordering of options within items. The standard is the order of data codes in [CODES=](#).

ITEM CATEGORY/OPTION/DISTRACTOR FREQUENCIES: ENTRY ORDER

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	DATA %	ABILITY MEAN	P. SD	S.E. MEAN	INFT MNSQ	OUTF MNSQ	PTMA CORR.	ITEM	
13	4	***	11	16#	.79	1.41	.45			.08	M. STAIRS	4 75% Independent
1	1		30	52	-1.85	.99	.18	.8	1.0	-.89		1 0% Independent
	3		5	9	1.07	.80	.40	.6	.4	.10		3 50% Independent
	5		15	26	2.63	1.06	.28	1.7	1.5	.56		5 Supervision
	6		7	12	3.25	.81	.33	1.2	1.2	.44		6 Device
	7		1	2	4.63	.00		.7	.6	.23		7 Independent
	MISSING	***	1	1#	-1.99	.00				-.12		

* Average ability does not ascend with category score
Missing % includes all categories. Scored % only of scored categories

ENTRY NUMBER is the item sequence number.
The letter next to the sequence number is used on the fit plots.

DATA CODE is the response code in the data file.
MISSING means that the data code is not listed in the [CODES=](#) specification.
Codes with no observations are not listed.

SCORE VALUE is the value assigned to the data code by means of [NEWSCORE=](#), [KEY1=](#), [IVALUEA=](#), etc.
*** means the data code is missing and so ignored, i.e., regarded as not administered. [MISSCORE=1](#) scores missing data as "1".

DATA COUNT is the frequency of the data code in the data file (unweighted) - this includes observations for both non-extreme and extreme persons and items. For counts weighted by [PWEIGHT=](#), see [DISFILE=](#)

DATA % is the percent of scored data codes. For dichotomies, the % are the [proportion-correct-values](#) for the options.
For data with score value "****", the percent is of all data codes, indicated by "#".

ABILITY MEAN is the observed, sample-dependent, [average measure](#) of persons (relative to each item) in this analysis who responded in this category (adjusted by [PWEIGHT=](#)). This is equivalent to a "Mean Criterion Score" (MCS) expressed as a measure. It is a sample-dependent quality-control statistic for this analysis. (It is not the sample-independent value of the category, which is obtained by adding the item measure to the "score at category", in Table 3.2 or higher, for the rating (or partial credit) scale corresponding to this item.) For each observation in category k, there is a person of measure Bn and an item of measure Di. Then: average measure = sum(Bn - Di) / count of observations in category.
An "*" indicates that the average measure for a higher score value is lower than for a lower score value. This contradicts the hypothesis that "higher score value implies higher measure, and vice-versa".
The "average ability" for missing data is the average measure of all the persons for whom there is no response to this item. This can be useful. For instance, we may expect the "missing" people to be high or low performers, or to be missing random (and so they average measure would be close to the average of the sample).
These values are plotted in [Table 2.6](#).

ABILITY P.SD is the population standard deviation of the ABILITY values = $\sqrt{(\sum (ABILITY - (ABILITY MEAN))^2/COUNT)}$

S.E. MEAN is the standard error of the mean (average) measure of the sample of persons from a population who responded in this category (adjusted by [PWEIGHT=](#)) = $\sqrt{(\sum (ABILITY - (ABILITY MEAN))^2/(COUNT*(COUNT-1))}$

INFT MNSQ is the [Infit Mean-Square](#) for observed responses in this category (weighted by [PWEIGHT=](#), and omitting responses in extreme person scores). Values greater than 1.0 indicate unmodeled noise. Values less than 1.0 indicate loss of information.

OUTF MNSQ is the [Outfit Mean-Square](#) for observed responses in this category (weighted by [PWEIGHT=](#), and omitting responses in extreme person scores). Values greater than 1.0 indicate unmodeled noise. Values less than 1.0 indicate loss of information.

PTMA CORR is the point-correlation between the data code, scored 1, or non-occurrence, scored 0, of this category or distractor and the person raw scores or measures chosen by [PTBISERIAL=](#). The computation is described in

Correlations. Example: for categories 0,1,2, then the correlation is between [1 for the target score (0 , 1, or 2) and 0 for the other scores (1 and 2, 0 and 2, or 0 and 1)] and the person ability measures for the persons producing each score.

ITEM (here, ACT) is the name or label of the item.

Data codes and Category labels are shown to the right of the box, if CLFILE= or CFILE= is specified.

* Average ability does not ascend with category score. The average ability of the persons observed in this category is lower than the average ability of the persons in the next lower category. This contradicts the Rasch-model assumption that "higher categories <-> higher average abilities."

Missing % includes all categories. Scored % only of scored categories. The percentage for the missing category is based on all the COUNTS. The percentages for the SCOREd categories are based only on those category COUNTS.

""BETTER FITTING OMIT" appears in fit-ordered Tables, where items better fitting than FIT are excluded.

14.35 Table 26 Item report in correlation order

14.35.1 Table 26.1 Item statistics in correlation order

(controlled by USCALE=, UMEAN=, UDECIM=, LOCAL=, ISORT=, TOTAL=, DISCRIMINATION=, ASYMPTOTE=, PVALUE=, PTBIS=)

Negative point-biserial correlations or point-measure correlations, PTBIS=, indicate that the responses to the item contradict the latent variable defined by the consensus of the items. The items may need to be omitted or rescored in the opposite direction. This may also be an indication that a multiple-choice item has the wrong option keyed as correct. Always investigate negative point correlations before the Rasch fit statistics.

ITEM STATISTICS: CORRELATION ORDER

For details, please see [Table 10.1](#)

14.35.2 Table 26.2 Item statistics graphically

Please see [Table 10.2](#)

14.35.3 Table 26.3 Item option & distractor frequencies

(controlled by Distractors=Y, OSORT=, CFILE=, PTBIS=)

ITEM OPTION FREQUENCIES are output if Distractors=Y. These show occurrences of each of the valid data codes in CODES=, and also of MISSSCORE= in the input data file. Counts of responses forming part of extreme scores are included. Only items included in the corresponding main table are listed. These statistics are also in DISFILE=, which includes entries even if the code is not observed for the item. See also [Distractor Analysis](#).

OSORT= controls the ordering of options within items. The standard is the order of data codes in CODES=.

ITEM CATEGORY/OPTION/DISTRACTOR FREQUENCIES: ENTRY ORDER												
ENTRY	DATA	SCORE	DATA		ABILITY		S.E.	INFT	OUTF	PTMA		
NUMBER	CODE	VALUE	COUNT	%	MEAN	P.SD	MEAN	MNSQ	MNSQ	CORR.	ITEM	
13	4	***	11	16#	.79	1.41	.45			.08	M. STAIRS	4 75% Independent
	1	1	30	52	-1.85	.99	.18	.8	1.0	-.89		1 0% Independent
	3	3	5	9	1.07	.80	.40	.6	.4	.10		3 50% Independent
	5	5	15	26	2.63	1.06	.28	1.7	1.5	.56		5 Supervision
	6	6	7	12	3.25	.81	.33	1.2	1.2	.44		6 Device
	7	7	1	2	4.63	.00		.7	.6	.23		7 Independent
	MISSING	***	1	1#	-1.99	.00				-.12		

* Average ability does not ascend with category score

Missing % includes all categories. Scored % only of scored categories

ENTRY NUMBER is the item sequence number.

The letter next to the sequence number is used on the fit plots.

DATA CODE is the response code in the data file.

MISSING means that the data code is not listed in the CODES= specification.

Codes with no observations are not listed.

SCORE VALUE is the value assigned to the data code by means of NEWScore=, KEY1=, IVALUEA=, etc.

*** means the data code is missing and so ignored, i.e., regarded as not administered. [MISSCORE=1](#) scores missing data as "1".

DATA COUNT is the frequency of the data code in the data file (unweighted) - this includes observations for both non-extreme and extreme persons and items. For counts weighted by [PWEIGHT=](#), see [DISFILE=](#)

DATA % is the percent of scored data codes. For dichotomies, the % are the [proportion-correct-values](#) for the options. For data with score value "****", the percent is of all data codes, indicated by "#".

ABILITY MEAN is the observed, sample-dependent, [average measure](#) of persons (relative to each item) in this analysis who responded in this category (adjusted by [PWEIGHT=](#)). This is equivalent to a "Mean Criterion Score" (MCS) expressed as a measure. It is a sample-dependent quality-control statistic for this analysis. (It is not the sample-independent value of the category, which is obtained by adding the item measure to the "score at category", in Table 3.2 or higher, for the rating (or partial credit) scale corresponding to this item.) For each observation in category k, there is a person of measure B_n and an item of measure D_i. Then: average measure = $\sum (B_n - D_i) / \text{count of observations in category}$.

An "*" indicates that the average measure for a higher score value is lower than for a lower score value. This contradicts the hypothesis that "higher score value implies higher measure, and vice-versa".

The "average ability" for missing data is the average measure of all the persons for whom there is no response to this item. This can be useful. For instance, we may expect the "missing" people to be high or low performers, or to be missing random (and so they average measure would be close to the average of the sample).

These values are plotted in [Table 2.6](#).

ABILITY P.SD is the population standard deviation of the ABILITY values = $\sqrt{(\sum (ABILITY - (ABILITY MEAN))^2 / COUNT)}$

S.E. MEAN is the standard error of the mean (average) measure of the sample of persons from a population who responded in this category (adjusted by [PWEIGHT=](#)) = $\sqrt{(\sum (ABILITY - (ABILITY MEAN))^2 / (COUNT * (COUNT - 1)))}$

INFT MNSQ is the [Infit Mean-Square](#) for observed responses in this category (weighted by [PWEIGHT=](#), and omitting responses in extreme person scores). Values greater than 1.0 indicate unmodeled noise. Values less than 1.0 indicate loss of information.

OUTF MNSQ is the [Outfit Mean-Square](#) for observed responses in this category (weighted by [PWEIGHT=](#), and omitting responses in extreme person scores). Values greater than 1.0 indicate unmodeled noise. Values less than 1.0 indicate loss of information.

PTMA CORR is the point-correlation between the data code, scored 1, or non-occurrence, scored 0, of this category or distractor and the person raw scores or measures chosen by [PTBISERIAL=](#). The computation is described in [Correlations](#). Example: for categories 0,1,2, then the correlation is between [1 for the target score (0, 1, or 2) and 0 for the other scores (1 and 2, 0 and 2, or 0 and 1)] and the person ability measures for the persons producing each score.

ITEM (here, ACT) is the name or label of the item.

Data codes and Category labels are shown to the right of the box, if [CLFILE=](#) or [CFILE=](#) is specified.

* Average ability does not ascend with category score. The average ability of the persons observed in this category is lower than the average ability of the persons in the next lower category. This contradicts the Rasch-model assumption that "higher categories <-> higher average abilities."

Missing % includes all categories. Scored % only of scored categories. The percentage for the missing category is based on all the COUNTS. The percentages for the SCOREd categories are based only on those category COUNTs.

""BETTER FITTING OMIT" appears in fit-ordered Tables, where items better fitting than [FITI=](#) are excluded.

14.36 Table 27 Item subtotals

14.36.1 Table 27.1 Item subtotal summaries on one line

(controlled by [ISUBTOT=](#), [UDECIMALS=](#), [REALSE=](#))

These summarize the measures from the main analysis for all items selected by [ISUBTOT=](#) (Table 27), including extreme scores.

Table

[27.2](#) Measure sub-totals bar charts, controlled by [ISUBTOT=](#)

[27.3](#) Measure sub-totals summary statistics, controlled by [ISUBTOT=](#)

Subtotal specification is: ISUBTOTAL=\$S1W1

ALL ITEM SCORES ARE NON-EXTREME

ACT COUNT	MEAN MEASURE	S.E. MEAN	P.SD	S.SD	MEDIAN	MODEL SEPARATION	MODEL RELIABILITY	RMSE	TRUE SD	MEAN OUTFIT	CODE
25	5.00	2.87	14.05	14.34	6.59	5.86	.97	2.36	13.85	1.08	*
4	8.08	7.35	12.73	14.69	3.86	5.89	.97	2.13	12.55	1.41	F
4	-17.42	3.90	6.75	7.79	-17.63	1.62	.72	3.55	5.74	.94	G
5	8.23	5.12	10.24	11.44	9.17	4.81	.96	2.08	10.02	.89	L
1	10.98	-	.00	-	10.98	.00	.00	1.90	.00	.95	M
3	2.42	3.44	4.86	5.95	.18	2.14	.82	2.06	4.40	.63	R
1	10.26	-	.00	-	10.26	.00	.00	1.90	.00	.74	T
7	13.24	4.51	11.04	11.92	16.02	5.29	.97	2.05	10.84	1.37	W

SUBTOTAL RELIABILITY: inestimable

UMEAN=0 USCALE=1

Subtotal specification is: ISUBTOTAL=\$S1W1	identifies the columns in the item label to be used for classifying the item by \$S1W1 or whatever, using the column selection rules .
EXTREME AND NON-EXTREME KID SCORES ALL SCORES ARE NON-EXTREME NON-EXTREME SCORES ONLY	The items included in this summary table. Items with non-extreme scores (omits items with 0% and 100% success rates)
ITEM COUNT	count of items. "ITEM" is the name assigned with ITEM=
MEAN MEASURE	average measure of items
S.E. MEAN	standard error of the average measure of items
P.SD	population standard deviation of the item measures.
S.SD	sample standard deviation of the item measures.
MEDIAN	the measure of the middle item
REAL/MODEL SEPARATION	the separation coefficient: the "true" adjusted standard deviation / root-mean-square measurement error of the items (REALSE= inflated for misfit).
REAL/MODEL RELIABILITY	the item measure reproducibility = ("True" item measure variance / Observed variance) = Separation ² / (1 + Separation ²)
RMSE	Statistical average of the standard errors of the measures
TRUE SD	Observed population S.D. adjusted for measurement error
MEAN OUTFIT	Average outfit mean-square for the group. Expectation near 1.0
ITEM CODE	the classification code in the item label. The first line, "*", is the total for all items. The remaining codes are those in the item columns specified by \$S1W1 or whatever, using the column selection rules .

SUBTOTAL RELIABILITY	the reliability (reproducibility) of the means of the subtotals = true variance / observed variance = (observed variance - error variance) / observed variance. Observed variance = variance of MEAN MEASURES Error variance = mean-square of the S.E. MEAN <i>inestimable</i> = some subtotal counts are too small to estimate Reliability
UMEAN=0 USCALE=1	Current user-scaling

```

-----
|      ITEM      MEAN DIFFERENCE      Welch-2sided |
| CODE CODE MEASURE  S.E.    t    d.f.  Prob. |
|-----|
|  0   1      -9.06   .57 -15.95   10   .000 |
|  0   2      -9.72   .87 -11.14   10   .000 |
|  0   4      -6.29   .94  -6.71   11   .000 |
|  1   2       -6.66   .66  -1.00    2   .423 |
|  1   4       2.77   .75   3.71    3   .034 |
|  2   4       3.43   1.00   3.44    3   .041 |
|-----|

```

ITEM CODE	the classification code in the item label for subtotal "1"
CODE	the classification code in the item label for subtotal "2"
MEAN DIFFERENCE	difference between the mean measures of the two CODE subtotals, "1" and "2"
MEASURE	size of the difference between "1" and "2"
S.E.	standard error of the difference = sqrt ((S.E. Mean "1") ² + (S.E. Mean "2") ²)
t	Student's t = MEASURE / S.E.
Welch2-sided	2-sided t-test using Welch's adaptation of Student's t-test.
d.f.	Welch's degrees of freedom
Prob.	two-sided probability of Student's t. See t-statistics .

One-way ANOVA of subtotal means and variances

This reports a one-way analysis of variance for the subtotal means. Are they the same (statistically) as the overall mean?

```

-----
| ANOVA -      KID      |
| Source Sum-of-Squares  d.f.  Mean-Squares  F-test  Prob>F |
|-----|
| @TOPIC          1.70    1.00          1.70    1.89    .1761 |
| Error          26.91   30.00           .90 |
| Total          28.61   31.00           .92 |
|-----|
| Fixed-Effects Chi-square: 1.7026 with 1 d.f., prob. .1919 |
|-----|

```

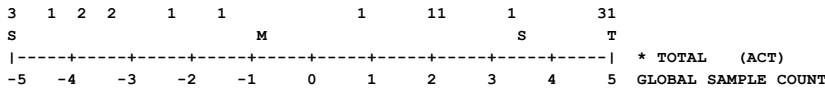
Source	the variance component.
@TYPE (the specified ISUBTOTAL= classification)	the variation of the subtotal mean measures around the grand mean.
Error	Error is the part of the total variation of the measures around their grand mean not explained by the @TYPE
Total	total variation of the measures around their grand mean
Sum-of-Squares	the variation around the relevant mean
d.f.	the degrees of freedom corresponding to the variation (= number of measures - 1)
Mean-Squares	Sum-of-Squares divided by d.f.
F-test	@TYPE Mean-Square / Error Mean-Square
Prob>F	the right-tail probability of the F-test value with (@TYPE, Error) d.f. A probability less than .05 indicates statistically significant differences between the means.

Fixed-Effects Chi-Square (of Homogeneity)	a test of the hypothesis that all the subtotal means are the same, except for sampling error
d.f.	degrees of freedom of chi-square = number of sub-totals - 1
prob.	probability of observing this value of the chi-square or larger if the hypothesis is true. A probability less than .05 indicates statistically significant differences between the means.
inestimable	some item counts are too small and/or some variances are zero.

14.36.2 Table 27.2 Item subtotal measure bar charts

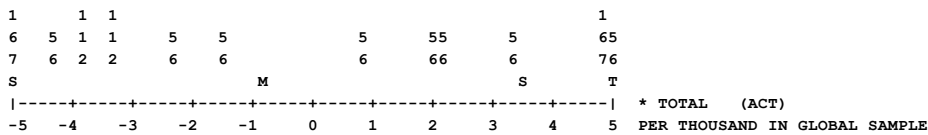
These show the distributions of the measures from the main analysis for each sub-sample. Summaries are shown in Tables [27.1](#) (Items, [ISUBTOT=](#)).

Here is the measure distribution of the total sample:



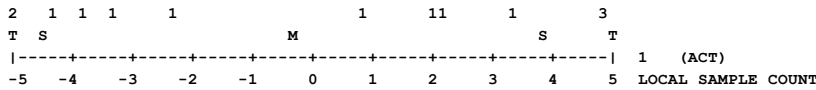
M= Mean, S = one [standard deviation](#) from mean, T = two standard deviations from mean.

Here is the measure distribution of the total sample standardized to a sample size of 1000:

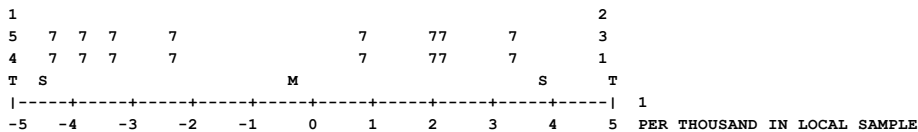


Read the counts vertically so that the first left-hand count is 167.

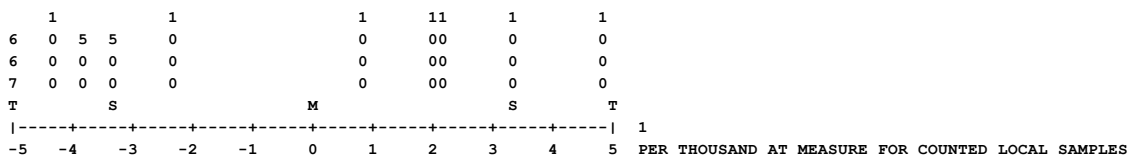
Here is the measure distribution of one sub-sample, specified as \$S9W1="F", using the column [selection rules](#).



Here is the measure distribution of the sub-sample standardized to a sample size of 1000:



Here is the measure distribution of the sub-sample standardized so that the total at any measure is 1,000, but in proportion to the observed counts in each sub-sample:



Here is the measure distribution of the sub-sample standardized so that the total any measure is 1,000, but based on the proportion observed when the sizes of all sub-samples are adjusted to be equal:



T	S	M	S	T	
-5	-4	-3	-2	-1	0
1	2	3	4	5	1
PER THOUSAND AT MEASURE FOR EQUAL LOCAL SAMPLES					

14.36.3 Table 27.3+ Item subtotal detailed summary statistics

(controlled by [ISUBTOT=](#), [UDECIMALS=](#))

These summarize the measures from the main analysis for all items selected by [ISUBTOT=](#). [Table 27.1](#) shows one-line summary statistics. Bar charts are shown in [Table 27.2](#). Detailed summary statistics in [Table 27.3](#), [27.4](#), ...

```

TOTAL FOR ALL 14 NON-EXTREME TAP
-----
|          TOTAL          MODEL          INFIT          OUTFIT          |
|          SCORE          COUNT          MEASURE          S.E.          MNSQ          ZSTD          MNSQ          ZSTD          |
|-----|-----|-----|-----|-----|-----|-----|-----|
| MEAN          16.9          35.0          .00          .71          .96          .0          .68          -.1          |
| P.SD          12.9          .0          3.48          .21          .28          .7          .58          .5          |
| S.SD          13.3          .0          3.61          .22          .29          .7          .60          .6          |
| MAX.          32.0          35.0          4.80          1.07          1.56          1.2          2.21          1.1          |
| MIN.          1.0          35.0          -4.40          .45          .59          -1.3          .11          -.6          |
|-----|-----|-----|-----|-----|-----|-----|-----|
| REAL RMSE          .77          TRUE SD          3.39          SEPARATION          4.41          TAP          RELIABILITY          .95          |
| MODEL RMSE          .74          TRUE SD          3.40          SEPARATION          4.59          TAP          RELIABILITY          .95          |
| S.E. OF TAP MEAN = .97          |
| MEDIAN = -.39          |
|-----|-----|-----|-----|-----|-----|-----|

MAXIMUM EXTREME SCORE: 3 TAP 16.7%
MINIMUM EXTREME SCORE: 1 TAP 5.6%

TOTAL FOR ALL 18 EXTREME AND NON-EXTREME TAP
-----
|          TOTAL          MODEL          INFIT          OUTFIT          |
|          SCORE          COUNT          MEASURE          S.E.          MNSQ          ZSTD          MNSQ          ZSTD          |
|-----|-----|-----|-----|-----|-----|-----|-----|
| MEAN          18.9          35.0          -.76          .96          .96          .0          .52          .0          |
| P.SD          14.0          .0          4.26          .51          .96          .0          .52          .0          |
| S.SD          14.4          .0          4.39          .52          .96          .0          .52          .0          |
| MAX.          35.0          35.0          6.13          1.85          .96          .0          .52          .0          |
| MIN.          .0          35.0          -6.59          .45          .96          .0          .52          .0          |
|-----|-----|-----|-----|-----|-----|-----|-----|
| REAL RMSE          1.10          TRUE SD          4.12          SEPARATION          3.73          TAP          RELIABILITY          .93          |
| MODEL RMSE          1.09          TRUE SD          4.12          SEPARATION          3.79          TAP          RELIABILITY          .93          |
| S.E. OF TAP MEAN = 1.03          |
| MEDIAN = -1.96          |
|-----|-----|-----|-----|-----|-----|-----|

```

For details, please see [Table 3.1](#)

14.37 Table 28 Person subtotals

14.37.1 Table 28.1 Person subtotal summaries on one line

(controlled by [PSUBTOT=](#), [UDECIMALS=](#), [REALSE=](#))

These summarize the measures from the main analysis for persons selected by [PSUBTOT=](#) (Table 28), including extreme scores. [PSUBTOTAL=](#) is useful for quantifying the impact of a test on different types of test-takers.

Table

[28.2](#) Measure sub-totals bar charts, controlled by [PSUBTOT=](#)

[28.3](#) Measure sub-totals summary statistics, controlled by [PSUBTOT=](#)

Subtotal specification is: [PSUBTOTAL=@GENDER](#)

Subtotals

```

EXTREME AND NON-EXTREME KID SCORES
-----
|          KID          MEAN          S.E.          MODEL          MODEL          |
|          COUNT          MEASURE          MEAN          P.SD          S.SD          MEDIAN          SEPARATION          RELIABILITY          CODE          |
|-----|-----|-----|-----|-----|-----|-----|-----|
|          35          -.37          .38          2.22          2.25          -.26          1.87          .78          *          |
|          18          -.68          .47          1.93          1.98          -.26          1.61          .72          F          |
|          17          -.05          .61          2.45          2.53          -.26          2.04          .81          M          |
|-----|-----|-----|-----|-----|-----|-----|-----|

```

SUBTOTAL RELIABILITY: .00
 UMEAN=0 USCALE=1

Subtotal specification is: PSUBTOTAL=@GENDER	identifies the columns in the Person label to be used for classifying the Person by @GENDER or whatever, using the column selection rules .
EXTREME AND NON-EXTREME SCORES	All persons with estimated measures
NON-EXTREME SCORES ONLY	Persons with non-extreme scores (omits Persons with 0% and 100% success rates)
PERSON COUNT	count of Persons. "PERSON" is the name assigned with PERSON=
MEAN MEASURE	average measure of Persons
S.E. MEAN	standard error of the average measure of Persons
P.SD	population standard deviation of the Persons.
S.SD	sample standard deviation of the Persons.
MEDIAN	the measure of the middle Person
REAL/MODEL SEPARATION	the separation coefficient: the "true" adjusted standard deviation / root-mean-square measurement error of the Persons (REAL = inflated for misfit).
REAL/MODEL RELIABILITY	the Person measure reproducibility = ("True" Person measure variance / Observed variance) = Separation ² / (1 + Separation ²)
RMSE	Statistical average of the standard errors of the measures
TRUE SD	Observed population S.D. adjusted for measurement error
MEAN OUTFIT	Average outfit mean-square for the group. Expectation near 1.0
CODE	the classification code in the Person label. The first line, "", is the total for all Persons. The remaining codes are those in the Person columns specified by @GENDER or whatever, using the column selection rules . In this example, "F" is the code for "Female" in the data file. "M" for "Male". It is seen that the two distributions are almost identical.
SUBTOTAL RELIABILITY	the reliability (reproducibility) of the means of the subtotals = true variance / observed variance = (observed variance - error variance) / observed variance. Observed variance = variance of MEAN MEASURES Error variance = mean-square of the S.E. MEAN

Independent-samples t-test of pairs of subtotal means

```

-----
| PERSON   MEAN DIFFERENCE      Welch |
| CODE CODE MEASURE   S.E.   t   d.f. Prob. |
|-----|
| F   M       -.62   .77   -.81  33   .424 |
|-----|

```

PERSON CODE	the classification code in the Person label for subtotal "1"
CODE	the classification code in the Person label for subtotal "2"
MEAN DIFFERENCE	difference between the mean measures of the two CODE subtotals, "1" and "2"
MEASURE	size of the difference between "1" and "2"
S.E.	standard error of the difference = sqrt ((S.E. Mean "1") ² + (S.E. Mean "2") ²)
t	Student's t = MEASURE / S.E.
d.f.	Welch's degrees of freedom
Prob.	two-sided probability of Student's t. See t-statistics .

One-way ANOVA of subtotal means and variances

This reports a one-way analysis of variance for the subtotal means. Are they the same (statistically) as the overall mean?

```

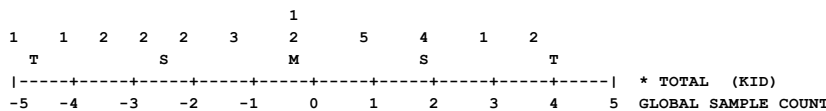
-----
| ANOVA -      KID                               |
| Source  Sum-of-Squares  d.f.  Mean-Squares  F-test  Prob>F |
|-----|-----|-----|-----|-----|
| @GENDER          3.41    1.00      3.41    .67    .5743 |
| Error           169.12   33.00      5.12           |
| Total           172.53   34.00      5.07           |
|-----|-----|-----|-----|
| Fixed-Effects Chi-square: .6565 with 1 d.f., prob. .4178 |
|-----
    
```

Source	the variance component.
@GENDER (the specified PSUBTOTAL= classification)	the variation of the subtotal mean measures around the grand mean.
Error	Error is the part of the total variation of the measures around their grand mean not explained by the @GENDER
Total	total variation of the measures around their grand mean
Sum-of-Squares	the variation around the relevant mean
d.f.	the degrees of freedom corresponding to the variation (= number of measures - 1)
Mean-Squares	Sum-of-Squares divided by d.f.
F-test	@GENDER Mean-Square / Error Mean-Square
Prob>F	the right-tail probability of the F-test value with (@GENDER, Error) d.f. A probability less than .05 indicates statistically significant differences between the means.
Fixed-Effects Chi-Square (of Homogeneity)	a test of the hypothesis that all the subtotal means are the same, except for sampling error
d.f.	degrees of freedom of chi-square = number of sub-totals - 1
prob.	probability of observing this value of the chi-square or larger if the hypothesis is true. A probability less than .05 indicates statistically significant differences between the means.
inestimable	some person counts are too small and/or some variances are zero.

14.37.2 Table 28.2 Person subtotal measure bar charts

These show the distributions of the measures from the main analysis for each sub-sample. Summaries are shown in Tables 28.1 (Items, [PSUBTOT=](#)).

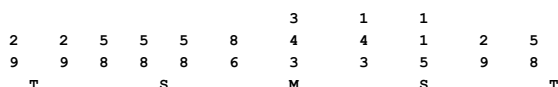
Here is the measure distribution of the total sample:



Read the counts vertically so that the center count is 12 observed near -0.4.

M= Mean, S = one [standard deviation](#) from mean, T = two standard deviations from mean.

Here is the measure distribution of the total sample standardized to a sample size of 1000:



```

|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----| * TOTAL (KID)
-5 -4 -3 -2 -1 0 1 2 3 4 5 PER THOUSAND IN GLOBAL SAMPLE

```

Here is the measure distribution of one sub-sample, specified as \$S9W1="F", using the column [selection rules](#).

```

1      1      1  2  7  3  2      1
T      S      M      S      T
|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----| F (KID)
-5 -4 -3 -2 -1 0 1 2 3 4 5 LOCAL SAMPLE COUNT

```

Here is the measure distribution of the sub-sample standardized to a sample size of 1000:

```

          1  3  1  1
5      5  5  6  8  6  1  5
6      6  6  2  9  7  2  6
T      S      M      S      T
|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----| F (KID)
-5 -4 -3 -2 -1 0 1 2 3 4 5 PER THOUSAND IN LOCAL SAMPLE

```

Here is the measure distribution of the sub-sample standardized so that the total at any measure is 1,000, but in proportion to the observed counts in each sub-sample:

```

1
0      5  5  6  5  6  5  5
0      0  0  6  8  0  0  0
0      0  0  7  4  0  0  0
T      S      M      S      T
|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----| F (KID)
-5 -4 -3 -2 -1 0 1 2 3 4 5 PER THOUSAND AT MEASURE FOR COUNTED LOCAL SAMPLES

```

Here is the measure distribution of the sub-sample standardized so that the total any measure is 1,000, but based on the proportion observed when the sizes of all sub-samples are adjusted to be equal:

```

1
0      4  4  6  5  5  4  4
0      8  8  5  7  8  8  8
0      6  6  4  0  7  6  6
T      S      M      S      T
|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----| F (KID)
-5 -4 -3 -2 -1 0 1 2 3 4 5 PER THOUSAND AT MEASURE FOR EQUAL LOCAL SAMPLES

```

14.37.3 Table 28.3+ Person subtotal detailed summary statistics

(controlled by [PSUBTOT=](#), [UDECIMALS=](#))

These summarize the measures from the main analysis for all persons selected by [PSUBTOT=](#). [Table 28.1](#) shows one-line summary statistics. Bar charts are shown in [Table 28.2](#). Detailed summary statistics in [Table 28.3](#).

TOTAL FOR ALL 34 NON-EXTREME KID

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	9.9	18.0	-.19	1.01	.99	-.2	.68	-.1
P.SD	2.1	.0	1.97	.10	.94	1.2	1.29	.7
S.SD	2.1	.0	2.00	.10	.95	1.2	1.30	.7
MAX.	14.0	18.0	3.73	1.11	4.12	2.5	6.07	2.2
MIN.	5.0	18.0	-4.32	.82	.18	-1.5	.08	-.7
REAL RMSE	1.18	TRUE SD	1.58	SEPARATION	1.34	KID	RELIABILITY	.64
MODEL RMSE	1.01	TRUE SD	1.69	SEPARATION	1.67	KID	RELIABILITY	.74
S.E. OF KID MEAN = .34								
MEDIAN = -.26								

MINIMUM EXTREME SCORE: 1 KID 2.9%

TOTAL FOR ALL 35 EXTREME AND NON-EXTREME KID

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	9.7	18.0	-.37	1.03				
P.SD	2.4	.0	2.22	.17				
S.SD	2.4	.0	2.25	.18				
MAX.	14.0	18.0	3.73	1.85				
MIN.	3.0	18.0	-6.62	.82				

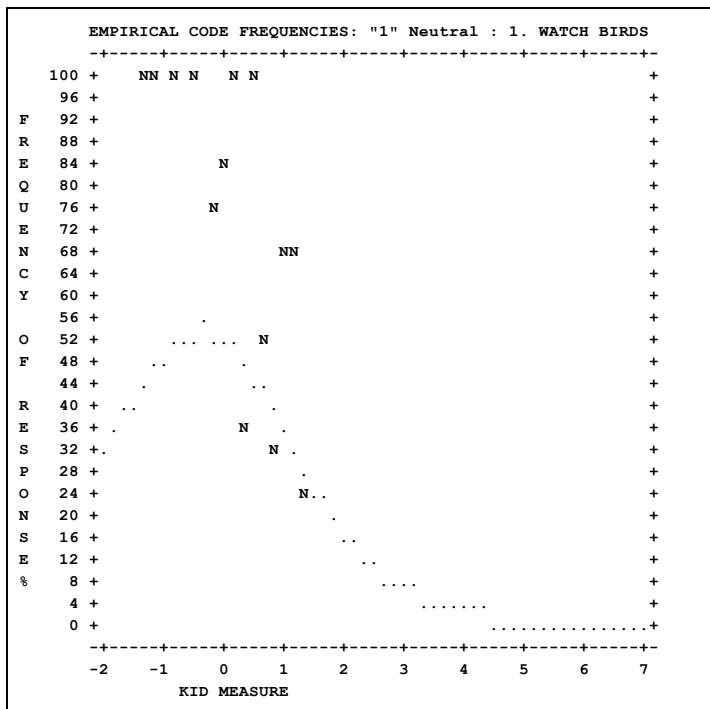
REAL RMSE	1.21	TRUE SD	1.86	SEPARATION	1.55	KID	RELIABILITY	.70
MODEL RMSE	1.05	TRUE SD	1.96	SEPARATION	1.87	KID	RELIABILITY	.78
S.E. OF KID MEAN = .38								
MEDIAN = -.26								

For details, please see Table 3.1

14.38 Table 29 Empirical ICCs and option frequencies

A subtable of plots is shown for each item. The subtables are numbered by Item Entry number. These plots only appear if there is something to show.

<p>EMPIRICAL & MODEL ICCs : 1. WATCH BIRDS</p>	<p>EMPIRICAL & MODEL ICCs</p> <p>This Table displays, as dots, the model expected score ogive, also called the predicted Item Characteristic Curve (ICC) and the predicted Item Response Function (IRF). The model ICC is shown by '.'. this is the expected average score on the item for persons at each measure relative to the item.</p> <p>The empirical observed average score per measure, the empirical ICC, is shown by 'x's.</p> <p>Observed average scores on the items are shown by 'x'. When '.' and 'x' coincide, '*' is shown.</p> <p>You can use this plot to compute the empirical item discrimination.</p> <p>See also Graphs menu: Expected and Empirical ICCs</p>
<p>EMPIRICAL CODE FREQUENCIES: "C" : 2. 2</p>	<p>EMPIRICAL CODE FREQUENCIES: multiple-choice items</p> <p>This shows the relative frequency of each item response code for each measure level.</p> <p>For a multiple-choice item, with C as an incorrect distractor, this could like the plot below. The dots are the Rasch ogive for an incorrect answer.</p> <p>- See also Graphs menu: Empirical category category curves</p>



EMPIRICAL CODE FREQUENCIES: polytomies
 For a polytomous item, the percentage frequencies of responses, and the model ordinal probability, of each category are shown at each measure level. The dots are the model curve. "N" is the empirical frequency for category N = Neutral.

14.39 Table 30 DIF Differential Item Functioning

14.39.1 Table 30.1 Differential item functioning DIF pairwise

Table 30 supports the investigation of item bias, [Differential Item Functioning](#) (DIF), i.e., interactions between individual items and types of persons.

Table

30.1 is best for pairwise comparisons, e.g., Females vs. Males.

[30.2](#) DIF report (measure list: person class within item)

[30.3](#) DIF report (measure list: item within person class)

[30.4](#) DIF report (item-by-person class chi-squares)

[30.5](#) Within-class fit report (person class within item)

[30.6](#) Within-class fit report (item within person class)

[30.7](#) Item measure profiles for classes of persons

Excel DIF [Plots](#)

Excel DIF [Scatterplots](#)

You need to choose a baseline item difficulty for your DIF comparisons.

In Table 30.1, we usually choose one group (the majority group) to be the baseline, and DIF is computed pairwise relative to that group. Both groups have statistical uncertainty.

In Table 30.2, we have many roughly equally-sized groups, such as age groups, and we take the average of all the groups (the item difficulty from the main analysis) as the baseline. Then DIF is relative to this baseline which is regarded as a known value. Only the focus group has statistical uncertainty.

The rules for DIF reporting are the same for Tables 30.1 and 30.2, but the underlying computations are somewhat different.

In [Table 30.1](#) - the hypothesis is "this item has the same difficulty for two groups"

In [Table 30.2, 30.3](#) - the hypothesis is "this item has the same difficulty as its average difficulty for all groups"

In [Table 30.4](#) - the hypothesis is "this item has no overall DIF across all groups"

Table 30.1 reports a probability and a size for DIF statistics. Usually we want:

1. probability so small that it is unlikely that the DIF effect is merely a random accident

2. size so large that the DIF effect has a substantive impact on scores/measures on the test

A general thought: Significance tests, such as DIF tests, are always of doubtful value in a Rasch context, because differences can be statistically significant, but far too small to have any impact on the meaning, or practical use, of the measures. So we need both statistical significance and substantive difference before we take action regarding bias, etc.

Table 30.1 is a pairwise DIF (bias) analysis: this is testing "item difficulty for Group A vs. item difficulty for Group B". Table 30.1 makes sense if there are only two groups, or there is one majority reference group.

Tables [30.2](#) and [30.3](#) are a global DIF (bias) analysis: this is testing "item difficulty for Group A vs. item difficulty for all groups combined." Tables 30.2 and 30.3 make sense when there are many small groups, e.g., age-groups in 5 year increments from 0 to 100.

DIF results are considerably influenced by sample size, so if you have only two person-groups, go to Table 30.1. If you have lots of person-groups go to Table 30.2

Specify [DIF=](#) for person classifying indicators in person labels. Item bias and DIF are the same thing. The widespread use of "item bias" dates to the 1960's, "DIF" to the 1980's. The reported DIF is corrected to test impact, i.e., differential average performance on the whole test. Use ability stratification to look for [non-uniform DIF](#) using the [selection rules](#). Tables 30.1 and 30.2 present the same information from different perspectives.

From the Output Tables menu, the [DIF/DPF dialog](#) is displayed.

[Table 31](#) supports person bias, Differential Person Functioning (DPF), i.e., interactions between individual persons and classifications of items.

[Table 33](#) reports bias or interactions between classifications of items and classifications of persons.

In these analyses, persons with extreme scores are excluded, because they do not exhibit differential ability across items. For background discussion, see [DIF and DPF considerations](#).

Example output:

You want to examine item bias (DIF) between Females and Males. You need a column in your Winsteps person label that has two (or more) demographic codes, say "F" for female and "M" for male (or "0" and "1" if you like dummy variables) in column 9.

Table 30.1 is best for pairwise comparisons, e.g., Females vs. Males.

DIF class specification is: DIF=@GENDER

KID	Obs-Exp	DIF	DIF	KID	Obs-Exp	DIF	DIF	DIF	JOINT	Rasch-Welch	Mantel-Haenszel	Size	Active	TAP				
CLASS	Average	MEASURE	S.E.	CLASS	Average	MEASURE	S.E.	CONTRAST	S.E.	t	d.f.	Prob.	Chi-squ	Prob.	CUMLOR	Slices	Number	Name
F	.00	-6.59E	.00	M	.00	-6.59E	.00	.00	.00	.00	0	1.000					1	1-4
F	.04	-5.24	1.90	M	-.04	-3.87	.90	-1.37	2.10	-.65	28	.5194	.0000	1.0000		7	4	1-3-4
F	.01	-1.67	.68	M	-.01	-1.48	.70	-.19	.97	-.19	31	.8468	.1316	.7167	-.06	7	10	2-4-3-
M	.00	-6.59E	.00	F	.00	-6.59E	.00	.00	.00	.00	30	1.000					1	1-4

Width of Mantel-Haenszel slice: MHSlice = .010 logits

The most important numbers in Table 30.1: The DIF CONTRAST is the difference in difficulty of the item between the two groups. This should be at least 0.5 logits for DIF to be noticeable. "Prob." shows the probability of observing this amount of contrast by chance, when there is no systematic item bias effect. For statistically significance DIF on an item, Prob. ≤ .05.

DIF class specification defines the columns used to identify DIF classifications, using [DIF=](#) and the [selection rules](#). For summary statistics on each class, use [Table 28](#).

Reading across the Table 30.1 columns:

PERSON CLASS identifies the CLASS of persons. PERSON is specified with [PERSON=](#), e.g., the first here is CLASS is "A".

Obs-Exp Average is the average difference between the observed and expected responses for the Class on the Item. When this is positive, the Class has higher ability than expected or the item is easier than expected.

DIF estimates with the [iterative-logit \(Rasch-Welch\)](#) method:

DIF MEASURE is the difficulty of this item for this class, with all else held constant, e.g., -.40 is the local difficulty for Class A of Item 1. The more difficult, the higher the DIF measure. The measures are conveniently listed in the Excel file for the DIF [plots](#), or copy them from the [Table into Excel](#).

For the raw scores corresponding to these measures, see [Table 30.2](#)

-.52> reports that this measure corresponds to an extreme maximum person-class score. [EXTRSCORE=](#) controls extreme score estimate.

1.97< reports that this measure corresponds to an extreme minimum person-class score. [EXTRSCORE=](#) controls extreme score estimate.

-6.91E reports that this measure corresponds to an item with an extreme score, which cannot exhibit DIF

DIF MEASURE is the same doing a full analysis of the data, outputting PFILE=pf.txt and SFILE=sf.txt, then doing another analysis with PAFILE=pf.txt and SAFILE=sf.txt and PSELECT=@DIF=code

DIF S.E. is the standard error of the DIF MEASURE. A value of ".00" indicates that DIF cannot be observed in these data.

PERSON CLASS identifies the CLASS of persons, e.g., the second CLASS is "D".

DIF MEASURE is the difficulty of this item for this class, with all else held constant, e.g., -.52 is the local difficulty for Class D of Item 1. > means "extreme maximum score".

DIF S.E. is the standard error of the second DIF MEASURE

DIF CONTRAST is the "effect size" in logits (or [USCALE=](#) units), the difference between the two DIF MEASURE, i.e., size of the DIF across the two classifications of persons, e.g., -.40 -- .52 = .11 (usually in logits). A positive DIF contrast indicates that the item is more difficult for the first, left-hand-listed CLASS.

If you want a sample-based effect size, then
effect size = DIF CONTRAST / (person sample measure S.D.)

JOINT S.E. is the standard error of the DIF CONTRAST = $\sqrt{\text{first DIF S.E.}^2 + \text{second DIF S.E.}^2}$, e.g., 2.50 = $\sqrt{.11^2 + 2.49^2}$

Welch t gives the DIF significance as a Welch's (Student's) [t-statistic](#) » DIF CONTRAST / JOINT S.E. The t-test is a two-sided test for the difference between two means (i.e., the estimates) based on the standard error of the means (i.e., the standard error of the estimates). The null hypothesis is that the two estimates are the same, except for measurement error.

d.f. is the joint degrees of freedom, computed according to [Welch-Satterthwaite](#). When the d.f. are large, the t statistic can be interpreted as a unit-normal deviate, i.e., z-score.

INF means "the degrees of freedom are so large they can be treated as infinite", i.e., the reported t-value is a unit normal deviate.

Prob. is the two-sided probability of Student's *t*. See [t-statistics](#).

[MantelHanzel](#) reports Mantel-Haenszel (1959) DIF test for dichotomies or Mantel (1963) for polytomies using [MHSLICE=](#). Statistics are reported when computable.

Chi-squ. is the Mantel-Haenszel for dichotomies or Mantel for polytomies chi-square with 1 degree of freedom.

Prob. is the probability of observing these data (or worse) when there is no DIF based on a chi-square value with 1 d.f.

Size CUMLOR (cumulative log-odds ratio in logits) is an estimate of the DIF (scaled by [USCALE=](#)). When the size is not estimable, +. and -. indicate direction.

Active Slices is a count of the estimable stratified cross-tabulations used to compute MH. MH is sensitive to score frequencies. If you have missing data, or only small or zero counts for some raw scores, the MH statistic can go wild or not be estimable. Please try different values of [MHSLICE=](#) (thin and thick slicing) to see how robust the MH estimates are.

ITEM Number is the item entry number. ITEM is specified by [ITEM=](#)

Name is the item label.

Below "----", each line in the Table is repeated with the CLASSes reversed.

ETS DIF Category	with DIF Contrast and DIF Statistical Significance	
C = moderate to large	 DIF ≥ 0.64 logits	prob(DIF ≤ 0.43 logits) ≤ .05 (2-sided) <i>approximately: DIF > 0.43 logits + 2 * DIF S.E.</i>

B = slight to moderate	 DIF ≥ 0.43 logits	prob(DIF = 0 logits) ≤ .05 (2-sided) <i>approximately: DIF > 2 * DIF S.E</i>
A = negligible	-	-
C-, B- = DIF against focal group; C+, B+ = DIF against reference group		
ETS (Educational Testing Service) use Delta δ units. 1 logit = 2.35 Delta δ units. 1 Delta δ unit = 0.426 logits.		
Zwick, R., Thayer, D.T., Lewis, C. (1999) An Empirical Bayes Approach to Mantel-Haenszel DIF Analysis. . Journal of Educational Measurement, 36, 1, 1-28 More explanation at www.ets.org/Media/Research/pdf/RR-12-08.pdf pp. 3,4		

For meta-analysis, the DIF Effect Size = DIF Contrast / S.D. of the "control" CLASS (or the pooled CLASSES). The S.D. for each CLASS is shown in [Table 28](#).

Example: The estimated item difficulty for Females, the DIF MEASURE, is 2.85 logits, and for males the DIF MEASURE is 1.24 logits. So the DIF CONTRAST, the apparent bias of the item against Females, is 1.61 logits. An alternative interpretation is that the Females are 1.61 logits less able on the item than the males.

```

Item 13: +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
          -1         0         1.24      +2         2.85      DIF measure
          +-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
                                     +-----+-----+-----+-----+-----+-----+
                                     = 1.61 DIF contrast

```

14.39.2 Table 30.2, 30.3 Differential item functioning DIF list

Table 30 supports the investigation of item bias, [Differential Item Functioning](#) (DIF), i.e., interactions between individual items and types of persons. Specify **DIF=** for person classifying indicators in person labels. Item bias and DIF are the same thing. The item measures by person class are plotted in the [DIF Plot](#).

In [Table 30.1](#) - the hypothesis is "this item has the same difficulty for two groups"

In [Table 30.2, 30.3](#) - the hypothesis is "this item has the same difficulty as its average difficulty for all groups"

In [Table 30.4](#) - the hypothesis is "this item has no overall DIF across all groups"

Example output:

You want to examine item bias (DIF) between Females and Males in [Exam1.txt](#). You need a column in your Winsteps person label that has two (or more) demographic codes, say "F" for female and "M" for male (or "0" and "1" if you like dummy variables) in column 9.

[Table 30.1](#) is best for pairwise comparisons, e.g., Females vs. Males. Use [Table 30.1](#) if you have two classes.

Table 30.2 or Table 30.3 are best for multiple comparisons, e.g., regions against the national average. Table 30.2 sorts by item then person class then item. Table 30.3 sorts by person class then item.

```

-----
| KID      OBSERVATIONS      BASELINE      DIF      DIF      DIF      DIF      DIF
| TAP      |
| CLASS    COUNT  SCORE AVERAGE EXPECT MEASURE  SCORE MEASURE  SIZE  S.E.  t      d.f.  Prob.
| Number  Name  |
-----
| F        18      18      1.00      1.00      -6.59      .00      -6.59      .00      .00      .00      1 1.000
| 1 1-4   |
| M        17      17      1.00      1.00      -6.59      .00      -6.59      .00      .00      .00      1 1.000
| 1 1-4   |
| F        18      16      .89      .92      -4.40      -.03      -3.93      .48      .89      .54      16 .5998
| 4 1-3-4 |
| M        17      16      .94      .91      -4.40      .03      -5.15>  -.75  1.91  -.39  14 .7020
| 4 1-3-4 |

```

F	18	15	.83	.88	-3.83	-.05	-3.22	.61	.79	.77	16	.4504
5 2-1-4												
M	17	16	.94	.89	-3.83	.05	-5.14>	-1.30	1.90	-.69	14	.5036
5 2-1-4												

This displays a list of the local difficulty/ability estimates underlying the paired DIF analysis. These can be plotted directly from the [Plots](#) menu.

DIF class specification identifies the columns containing DIF classifications, with [DIF=](#) set to [@GENDER](#) using the [selection rules](#).

The DIF effects are shown ordered by CLASS within item (column of the data matrix).

KID CLASS identifies the CLASS of persons. KID is specified with [PERSON=](#), e.g., the first CLASS is "F"

OBSERVATIONS are what are seen in the data

COUNT is the number of observations of the classification used for DIF estimation, e.g., 18 non-extreme F persons responded to TAP item 1.

AVERAGE is the average observation on the classification, e.g., 0.89 is the p-value, proportion-correct-value, of item 4 for F persons.

COUNT * AVERAGE = total score of person class on the item

BASELINE is the prediction without DIF

EXPECT is the expected value of the average observation when there is no DIF, e.g., 0.92 is the expected proportion-correct-value for F without DIF.

MEASURE is the what the overall measure would be without DIF, e.g., -4.40 is the overall item difficulty of item 4 as reported in [Table 14](#).

DIF: Differential Item Functioning

DIF SCORE is the difference between the observed and the expected average observations, e.g., $0.92 - 0.89 = -0.03$

DIF MEASURE is the item difficulty for this class, e.g., item 4 has a local difficulty of -3.93 for CLASS F.

The average of DIF measures across CLASS for an item is not the BASELINE MEASURE because score-to-measure conversion is non-linear. ">" (maximum score), "<" (minimum score) indicate measures corresponding to extreme scores.

DIF SIZE is the difference between the DIF MEASURE for this class and the BASELINE DIFFICULTY, i.e., $-3.93 - -4.40 = .48$. Item 4 is .48 logits more difficult for class F than expected.

DIF S.E. is the approximate standard error of the difference, e.g., 0.89 logits

DIF t is an approximate Student's [t-statistic](#) test, estimated as DIF SIZE divided by the DIF S.E.

d.f. t has approximately (COUNT-2) degrees of freedom.

Prob. is the two-sided probability of Student's *t*. See [t-statistics](#).

These numbers are plotted in the [DIF plot](#). Here item 4 is shown. The y-axis is the "DIF Measure".

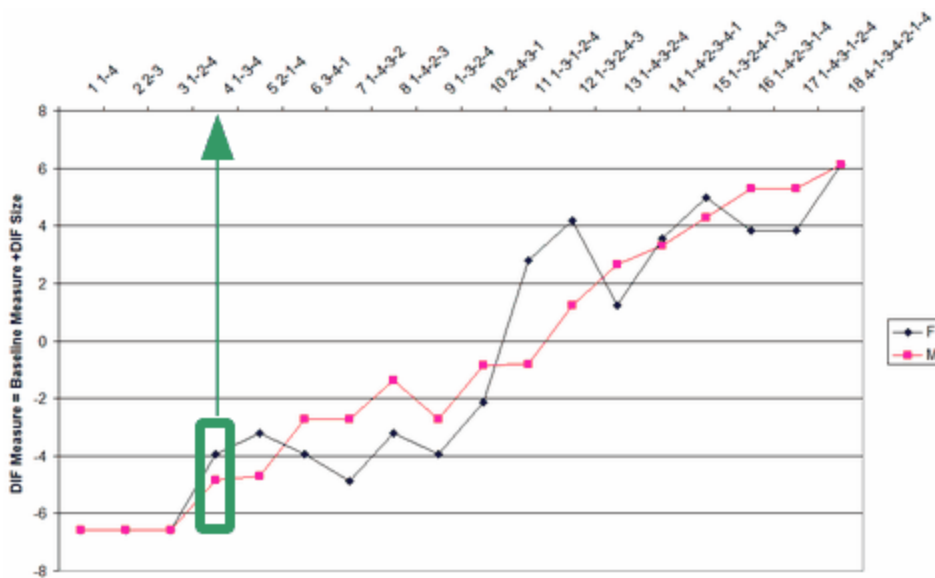


Table 31.1 Differential person

functioning DPF pairwise

Example 1: Where do I extract appropriate difficulties for my classes for both items that exhibit DIF and those that don't?

The DIF-sensitive difficulties are shown as "DIF Measure" in Table 30.1. They are more conveniently listed in Table 30.2. The "DIF Size" in Table 30.2 or Table 30.3 shows the size of the DIF relative to the overall measure in the IFILE=.

To apply the DIF measures as item difficulties, you would need to produce a list of item difficulties for each group, then analyze that group (e.g., with PSELECT=) using the specified list of item difficulties as an anchor file (IAFILE=).

My approach would be to copy Table 30.3 into an Excel spreadsheet, then use "Data", "Text to Columns" to put each Table column into a separate Excel column. The anchor file would have the item number in the first column, and either the overall "baseline measure" or the group "DIF measure" in the second column. Then copy and paste these two columns into a .txt anchor file.

Example 2: What is the impact of DIF on person measures?

Please look at Winsteps Table 30.2 or 30.3. The average effect of the DIF on person measures of the person group is (DIF Size for the item for the person group)/(total number of items). DIF has raised the person measures if the observed average of the scored responses is greater than expected, and vice-versa.

14.39.3 Table 30.4 Differential item functioning DIF fit summary

In Table 30.1 - the hypothesis is "this item has the same difficulty for two groups"

In Table 30.2, 30.3 - the hypothesis is "this item has the same difficulty as its average difficulty for all groups"

In Table 30.4 - the hypothesis is "this item has no overall DIF across all groups"

Table 30.4 summarizes the CLASS/GROUP Differential Item Functioning statistics for each item shown in Table 30.2. Table 30.2 shows a t-statistic for each CLASS/GROUP. These are summarized as chi-square statistics for each item, indicating whether the observed DIF within each item is due to chance alone. The null hypothesis is that the DIF is statistically zero across the classes. Please do not use anchoring (IAFILE=, PAFILE=, SAFILE=) when investigating this Table.

DIF class/group specification is: DIF=@GENDER

KID	SUMMARY DIF			BETWEEN-CLASS		TAP	
	CLASSES	CHI-SQUARE	D.F.	UNWTD	MNSQ		
					t=ZSTD	Number Name	
1	2	.8772	1	.3490	.9200	.4132	6 3-4-1
1	2	2.2427	1	.1342	2.5792	1.2592	7 1-4-3-2
1	2	1.9207	1	.1658	1.9963	1.0211	13 1-4-3-2-4

KID CLASSES/GROUPS is the count of person CLASSES or GROUPS for the item: 2 = Male and Female.

SUMMARY DIF CHI-SQUARE is the sum of the *t*-statistic values from [Table 30.2](#), squared and normalized. See Example 1 below.

D.F. is the degrees of freedom, the count of CLASSES or GROUPS contributing to the chi-square less 1.

PROB. is the probability of the chi-square. Values less than .05 indicate statistical significance.

BETWEEN-CLASS are [Between-Group Fit Statistics](#), testing the hypothesis: "The dispersion of the group measures accords with Rasch model expectations."

UNWTD MNSQ is the unweighted mean-square (chi-square divided by its degrees of freedom). It is the size of the misfit (expectation = 1.0, overfit <1.0, underfit >1.0).

t=ZSTD is the significance of the MEAN-SQUARE standardized as a unit-normal deviate (*t*-statistic with infinite degrees of freedom).

TAP is the item. Number is the item entry number. Name is the item label.

Item-Trait Chi-Square

The SUMMARY DIF CHI-SQUARE values in Table 30.4 are equivalent to the item-trait chi-square statistics reported by RUMM with [DIF=MA3](#) (or however many strata are chosen in RUMM). In RUMM, the trait CLASSES are obtained by ordering the persons by measure, omitting extreme scores, and dividing the ordered list as equally as possible into equal size classes, including all persons with the same measure in the same class.

The Root Mean Square Error of Approximation (RMSEA)

RMSEA is a transformation of the Item-Trait Chi-Square and so of the SUMMARY DIF CHI-SQUARE with [DIF=MA3](#). Its computation is

$$RMSEA = \sqrt{\max\left(\left[\frac{\chi^2}{df} - 1\right] / (N - 1), 0\right)}$$

where

χ^2 = SUMMARY DIF CHI-SQUARE

d.f. = D.F.

N is the person sample size

Example 1. SUMMARY DIF CHI-SQUARE.

Let's do an analysis of Exam 1.txt with Udecimals=4. Output Table 30 with [DIF=@Gender](#)

In Table 30.2, the values for item 13:

CLASS F: COUNT= 18, DIF SIZE= -.7268, DIF S.E.= .6764

CLASS M: COUNT= 16, DIF SIZE= -.7075, DIF S.E.= .7497

So the *t*-statistics are approximately:

CLASS F: *t* = -.7268/.6764 with (18-1) d.f. -> -1.0745 with 17 d.f.

CLASS M: *t* = -.7075/.7497 with (16-1) d.f. -> .9437 with 15 d.f.

Squared and normalized. The normalizing transformation is Peizer and Pratt (1968) in the *Handbook of the Normal Distribution, Second Edition* (Patel and Read), p. 214.

CLASS F: *t***2 = 1.1546 -> normalized(*t***2) = 1.0843

CLASS M: *t***2 = .8905 -> normalized(*t***2) = .8363

Chi-square = 1.0843 + .8363 = 1.9207 with (2-1) = 1 d.f.

14.39.4 Table 30.5, 30.6 Within-class fit report

Table 30.5 Within-class fit report (person class within item).

```
DIF class specification is: DIF=@GENDER
```

KID	OBSERVATIONS	BASELINE			W-INFIT		W-OUTFIT		TAP	
		COUNT	AVERAGE	EXPECT MEASURE	MNSQ	ZSTD	MNSQ	ZSTD	Number	Name
F	18	.89	.85	-3.38	.46	-1.2	.19	-.4	9	1-3-2-4
M	16	.88	.92	-3.38	.88	.0	.23	-.3	9	1-3-2-4
F	18	.72	.65	-1.57	.80	-.6	.76	.0	10	2-4-3-1
M	16	.69	.77	-1.57	1.47	1.0	.91	.4	10	2-4-3-1

Table 30.6 Within-class fit report (item within person class)

```
DIF class specification is: DIF=@GENDER
```

KID	OBSERVATIONS	BASELINE	W-INFIT	W-OUTFIT	TAP
-----	--------------	----------	---------	----------	-----

CLASS	COUNT	AVERAGE	EXPECT	MEASURE	MNSQ	ZSTD	MNSQ	ZSTD	Number	Name
F	18	.89	.85	-3.38	.46	-1.2	.19	-.4	9	1-3-2-4
F	18	.72	.65	-1.57	.80	-.6	.76	.0	10	2-4-3-1
M	16	.88	.92	-3.38	.88	.0	.23	-.3	9	1-3-2-4
M	16	.69	.77	-1.57	1.47	1.0	.91	.4	10	2-4-3-1

For a general description, see [Table 30.2](#)

The DIF effects are shown ordered by CLASS within item (column of the data matrix).

KID CLASS identifies the CLASS of persons. KID is specified with [PERSON=](#), e.g., the first CLASS is "F"

OBSERVATIONS are what are seen in the data

COUNT is the number of observations of the classification used for DIF estimation, e.g., 18 F persons responded to TAP item 1.

AVERAGE is the average observation on the classification, e.g., 0.89 is the proportion-correct-value of item 4 for F persons.

COUNT * AVERAGE = total score of person class on the item

BASELINE is the prediction without DIF

EXPECT is the expected value of the average observation when there is no DIF, e.g., 0.92 is the expected proportion-correct-value for F without DIF.

MEASURE is the what the overall measure would be without DIF, e.g., -4.40 is the overall item difficulty of item 4 as reported in [Table 14](#).

W-INFIT is the [Within-Group infit statistic](#) for the person CLASS on the item.

W-OUTFIT is the [Within-Group outfit statistic](#) for the person CLASS on the item.

MNSQ is the mean-square statistic (chi-square divided by its degrees of freedom).

ZSTD is the mean-square statistic expressed as a z-score (unit normal deviate).

14.39.5 Table 30.7 Item measure profiles for classes of persons

Table 30.7 shows the item difficulty and summary scores for each person classification selected by [DIF=](#)

TABLE 30.7 LIKING FOR SCIENCE (Wright & Masters ZOU211WS.TXT Apr 30 2017 13: 3
INPUT: 75 KID 25 ACT REPORTED: 75 KID 25 ACT 3 CATS WINSTEPS 3.93.0

DIF class specification is: DIF=@GENDER

* ALL KID ON ACT				F KID CLASS				M KID CLASS				ACT
COUNT	T.SCORE	MEASURE	S.E.	COUNT	T.SCORE	MEASURE	S.E.	COUNT	T.SCORE	MEASURE	S.E.	ENTRY LABEL
75	109	-.40	.21	18	29	-.48	.49	57	80	-.40	.23	1 Watch birds
75	116	-.71	.23	18	29	-.48	.49	57	87	-.76	.24	2 Read books on animals

same with [BOXSHOW=](#)No, useful for pasting into Excel:

* ALL KID ON ACT				F KID CLASS				M KID CLASS				ACT
COUNT	T.SCORE	MEASURE	S.E.	COUNT	T.SCORE	MEASURE	S.E.	COUNT	T.SCORE	MEASURE	S.E.	ENTRY LABEL
75	109	-.40	.21	18	29	-.48	.49	57	80	-.40	.23	1 Watch birds
75	116	-.71	.23	18	29	-.48	.49	57	87	-.76	.24	2 Read books on animals

Column	Meaning
COUNT	Count of non-missing responses to this item by this person classification group.
T.SCORE	Total score on this item by this person classification group.
MEASURE	Item difficulty for this person group = overall item difficulty + item DIF for the person group
S.E.	S.E. of the MEASURE

14.40 Table 31 DPF Differential Person Functioning

Enter topic text here.

14.40.1 Table 31.1 Differential person functioning DPF pairwise

Table 31 supports person bias, Differential Person Functioning (DPF), i.e., interactions between individual persons and classifications of items. This is useful for estimating sub-test, domain and strand measures for individuals in the context of an overall measure.

Tables:

[31.2](#) DPF report (measure list: item class within person)

[31.3](#) DPF report (measure list: person within item class)

[31.4](#) DPF report (person by item-class chi-squares)

[31.5](#) Within-class fit report (item class within person)

[31.6](#) Within-class fit report person class within item)

[31.7](#) Person measure profiles for classes of items

DPF [plots](#)

DPF [Scatterplots](#)

Table 31.1 reports a probability and a size for DPF statistics. Usually we want:

1. probability so small that it is unlikely that the DPF effect is merely a random accident
2. size so large that the DPF effect has a substantive impact on scores/measures on the test

Specify [DPF=](#) for classifying indicators in item labels. Use difficulty stratification to look for non-uniform DPF using the [selection rules](#).

From the Output Tables menu, the [DPF dialog](#) is displayed.

[Table 30](#) supports the investigation of item bias, Differential Item Functioning (DIF), i.e., interactions between individual items and types of persons.

[Table 33](#) reports bias or interactions between classifications of items and classifications of persons.

In these analyses, persons and items with extreme scores are excluded, because they do not exhibit differential ability across items. For background discussion, see [DIF and DPF concepts](#).

Example output:

Table 31.1

```
DPF class specification is: DPF=$S1W1
-----
| TAP  Obs-Exp  DPF  DPF  TAP  Obs-Exp  DPF  DPF  DPF  JOINT  Rasch-Welch  KID  |
| CLASS Average MEASURE S.E. CLASS Average MEASURE S.E. CONTRAST S.E. t d.f. Prob. Number Name |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1      -.05  -3.52  1.05  2      .04  -2.80  1.65  -.73  1.95  -.37  2  .7459  1 Adam  M1|
| 1      -.05  -3.52  1.05  3      .39  -2.78> 2.07  -.74  2.32  -.32  0  .0000  1 Adam  M1|
| 1      -.05  -3.52  1.05  4      .00  -2.94E .00  -.58  .00  .00  0  1.000  1 Adam  M1|
```

DPF Specification defines the columns used to identify Differential Person Function classifications, using the [selection rules](#).

TAP CLASS is the item class

Obs-Exp Average is the average difference between the observed and expected responses for the Class by the person.

When this is positive, the Class is easier than expected or the person has higher ability than expected.

DPF MEASURE is the ability of the person for this item class, with all else held constant. This is output in the Excel file for the DPF [plots](#).

DPF MEASURE is the same doing a full analysis of the data, outputting IFILE=if.txt and SFILE=sf.txt, then doing another analysis with IAFILE=if.txt and SAFILE=sf.txt and ISELECT=@DPF=code. ">" and "<" indicate that the scores for the group are extreme.

DPF S.E. is the standard error of the measure

DPF CONTRAST is the difference in the person ability measures, i.e., size of the DPF, for the two classifications of items. JOINT S.E. is the standard error of the DPF CONTRAST

DPF estimates with the the [iterative-logit \(Rasch-Welch\)](#) method:

t gives the DPF significance as a Student's [t-statistic](#) test. The t-test is a two-sided test for the difference between two means (i.e., the estimates) based on the standard error of the means (i.e., the standard error of the estimates). The null hypothesis is that the two estimates are the same, except for measurement error.

d.f. is the joint degrees of freedom. This is shown as the sum of the counts (see [Table 31.2](#)) of two classifications - 2 for the two measure estimates, but this estimate of d.f. is somewhat high, so interpret the t-test conservatively. When the d.f. are large, the t statistic can be interpreted as a unit-normal deviate, i.e., z-score.

Prob. is the two-sided probability of Student's t. See [t-statistics](#).

-5.24> reports that this measure corresponds to an extreme maximum score. [EXTRSCORE=](#) controls extreme score estimate.

5.30< reports that this measure corresponds to an extreme minimum score. [EXTRSCORE=](#) controls extreme score estimate.

14.40.2 Table 31.2, 31.3 Differential person functioning DPF list

Table 31.2 sorts by class then item. Table 31.3 sorts by item then class. The person measures by item class are plotted in the [DPF Plot](#).

Table 31 supports the investigation of item bias, [Differential Person Functioning](#) (DPF), i.e., interactions between individual persons types of items. Specify [DPF=](#) for item classifying indicators in item labels. Person bias and DPF are the same thing.

Example output:

You want to examine person bias (DPF) between starting-blocks in [Exam1.txt](#). You need a column in your Winsteps item label that has two (or more) item type codes.

[Table 31.1](#) is best for pairwise comparisons, e.g., Positive vs. Negative items. Use [Table 31.1](#) if you have two classes.

Table 31.2 or Table 31.3 are best for multiple comparisons, e.g., regions against the national average.

Table 31.2 sorts by person then item class.

```

-----
| TAP      OBSERVATIONS      BASELINE      DPF      DPF      DPF      DPF      DPF
| KID      |
| CLASS    COUNT  SCORE AVERAGE EXPECT MEASURE  SCORE MEASURE  SIZE  S.E.  t    d.f.  Prob.
| Number   Name    |
-----
| 1        13        4    .31    .35   -2.94   -.04   -3.54   -.60  1.05  -.57   9   .5857
| 1 Adam   M|
| 2        3        2    .67    .64   -2.94   .03   -2.70   .24  1.65  .15   1   .9083
| 1 Adam   M|
| 3        1        1    1.00   .61   -2.94   .39   -2.53> .41  2.18  .19   1   .8819
| 1 Adam   M|
| 4        1        0    .00    .00   -2.94   .00   -2.94   .00  .00   .00   1   1.000
| 1 Adam   M|
| 1        13        6    .46    .48   -.26   -.02   -.77   -.51  1.35  -.38   9   .7161
| 2 Anne   F|

```

Table 31.3 sorts by item class then person.

```

-----
| TAP      OBSERVATIONS      BASELINE      DPF      DPF      DPF      DPF      DPF
| KID      |

```

CLASS	COUNT	SCORE	AVERAGE	EXPECT	MEASURE	SCORE	MEASURE	SIZE	S.E.	t	d.f.	Prob.
Number	Name											
1	13	4	.31	.35	-2.94	-.04	-3.54	-.60	1.05	-.57	9	.5857
	1 Adam M											
1	13	6	.46	.48	-.26	-.02	-.77	-.51	1.35	-.38	9	.7161
	2 Anne F											
1	13	7	.54	.55	.92	-.01	.81	-.11	1.13	-.10	9	.9261
	3 Audrey F											
1	13	4	.31	.30	-3.61	.01	-3.54	.07	1.05	.07	9	.9483
	4 Barbara F											

This displays a list of the local difficulty/ability estimates underlying the paired DPF analysis. These can be plotted directly from the [Plots](#) menu.

DPF class specification identifies the columns containing DPF classifications, with `DPF=` set to `$$S1W2` using the [selection rules](#).

The DPF effects are shown ordered by CLASS within person (row of the data matrix).

TAP CLASS identifies the CLASS of items. KID is specified with `ITEM=`, e.g., the first CLASS is "1-"

OBSERVATIONS are what are seen in the data

COUNT is the number of observations of the classification used for DPF estimation, e.g., 11 responses were made by person 1 to non-extreme items in item class "1-".

AVERAGE is the average observation on the classification, e.g., 0.18 is the average score class "1-" items by person 1.

COUNT * AVERAGE = total score of person on the item class.

BASELINE is the prediction without DPF

EXPECT is the expected value of the average observation when there is no DPF, e.g., 0.92 is the expected average for person 1 on item class "1-" without DPF.

MEASURE is the what the overall ABILITY measure would be without DPF, e.g., -2.94 is the overall person ability of person 1 as reported in [Table 18](#).

DPF: Differential Person Functioning

DPF SCORE is the difference between the observed and the expected average observations, e.g., $0.92 - 0.89 = -0.03$

DPF MEASURE is the person ability for this item class, e.g., person 1 has a local ability of -3.52 for item CLASS "1-".

The average of DPF measures across CLASS for an item is not the BASELINE MEASURE because score-to-measure conversion is non-linear. ">" (maximum score), "<" (minimum score) indicate measures corresponding to extreme scores.

DPF SIZE is the difference between the DPF MEASURE for this class and the BASELINE measure ability, i.e., $-3.93 - -4.40 = .48$. Item 4 is .48 logits more difficult for class F than expected.

DPF S.E. is the approximate standard error of the difference, e.g., 0.89 logits

DIF t is an approximate Student's [t-statistic](#) test, estimated as DIF SIZE divided by the DIF S.E.

d.f. t has approximately (COUNT-2) degrees of freedom.

Prob. is the two-sided probability of Student's t. See [t-statistics](#).

Person performance profiles by subtest

- Put a subtest code in each item label
@subtest=1W1 ; position of subtest code in item label
- Run Table 31 Differential Person Functioning for @subtest.
- Output the DPF Excel Plots
- The [Excel Worksheet](#) lists each person with subtest measures

14.40.3 Table 31.4 Differential person functioning DPF fit summary

Table 31.4 summarizes the CLASS [Differential Person Functioning](#) statistics for each person shown in [Table 31.2](#). Table 31.2 shows a t-statistic for each CLASS. These are summarized as chi-square statistics for each person, indicating whether the observed DPF within each person is due to chance alone. The null hypothesis is that the DPF is statistically zero across the classes. Please do not use anchoring (`IAFILE=`, `PAFILE=`, `SAFILE=`) when investigating this Table.

```
DPF class/group specification is: DPF=$S1W1
```

TAP	SUMMARY DPF			BETWEEN-CLASS		KID	
CLASSES	CHI-SQUARE	D.F.	PROB.	MEAN-SQUARE	t=ZSTD	Number	Name
1	2	.2914	1	.5893	.2291	-.3520	1 Adam M

TAP CLASSES/GROUPS is the count of item CLASSES or GROUPS for the person.
SUMMARY DIF CHI-SQUARE is the sum of the squared normalized t-statistic values from [Table 31.2](#).
D.F. is the degrees of freedom, the count of CLASSES or GROUPS contributing to the chi-square less 1.
PROB. is the probability of the chi-square. Values less than .05 indicate statistical significance.

BETWEEN-CLASS/GROUP are [Between-Group Fit Statistics](#), testing the hypothesis: "The dispersion of the group measures accords with Rasch model expectations."

MEAN-SQUARE is chi-square divided by its degrees of freedom. It is the size of the misfit (expectation = 1.0, overfit <1.0, underfit >1.0).

t=ZSTD is the significance of the MEAN-SQUARE standardized as a unit-normal deviate (t-statistic with infinite degrees of freedom).

KID is the person. Number is the person entry number. Name is the person label.

14.40.4 Table 31.5, 31.6 Within-class fit report

Table 31.5 Within-class fit report (item class within person).

Table 31.6 Within-class fit report (person class within item)

```
DIF class specification is: DIF=@GENDER
```

TAP	OBSERVATIONS		BASELINE		W-INFIT		W-OUTFIT		KID	
CLASS	COUNT	AVERAGE	EXPECT	MEASURE	MNSQ	ZSTD	MNSQ	ZSTD	Number	Name
1	11	.18	.23	-2.94	.71	-.5	.25	-.3	1 Adam	M

For a general description, see [Table 31.2](#)

The DPF effects are shown ordered by CLASS within person (row of the data matrix).

TAP CLASS identifies the CLASS of items. KID is specified with [ITEM=](#), e.g., the first CLASS is "1-"

OBSERVATIONS are what are seen in the data

COUNT is the number of observations of the classification used for DPF estimation, e.g., 11 "1-" items responses were made by person 1.

AVERAGE is the average observation on the classification, e.g., 0.18 is the average score class "1-" items by person 1.

COUNT * AVERAGE = total score of person on the item class.

BASELINE is the prediction without DPF

EXPECT is the expected value of the average observation when there is no DPF, e.g., 0.92 is the expected average for person 1 on item class "1-" without DPF.

MEASURE is the what the overall ABILITY measure would be without DPF, e.g., -2.94 is the overall person ability of person 1 as reported in [Table 18](#).

W-INFIT is the [Within-Group infit statistic](#) for the item CLASS on the person.

W-OUTFIT is the [Within-Group outfit statistic](#) for the item CLASS on the person.

MNSQ is the mean-square statistic (chi-square divided by its degrees of freedom).

ZSTD is the mean-square statistic expressed as a z-score (unit normal deviate).

14.40.5 Table 31.7 Person measure profiles for classes of items

.Table 31.7 shows the person ability and summary scores for each item classification selected by [DPF=](#)

```
TABLE 31.7 An MCQ Test: administration was Compu ZOU749WS.TXT Apr 30 2017 13:11
INPUT: 30 STUDENT 69 TOPIC REPORTED: 30 STUDENT 48 TOPIC 2 CATS WINSTEPS 3.93.0
```

```
DPF class specification is: DPF=1
```

* ALL TOPIC ON STUDENT				i TOPIC CLASS				n TOPIC CLASS				STUDENT	
COUNT	T.SCORE	MEASURE	S.E.	COUNT	T.SCORE	MEASURE	S.E.	COUNT	T.SCORE	MEASURE	S.E.	ENTRY	LABEL
21	9	-1.92	.00	5	0	-3.39	1.64	16	9	-1.50	.59	2 NM	KAT
9	2	-3.09	.00	3	1	-2.18	1.40	6	1	-3.63	1.19	5 NL	SMI

same with `BOXSHOW=No`, useful for pasting into Excel:

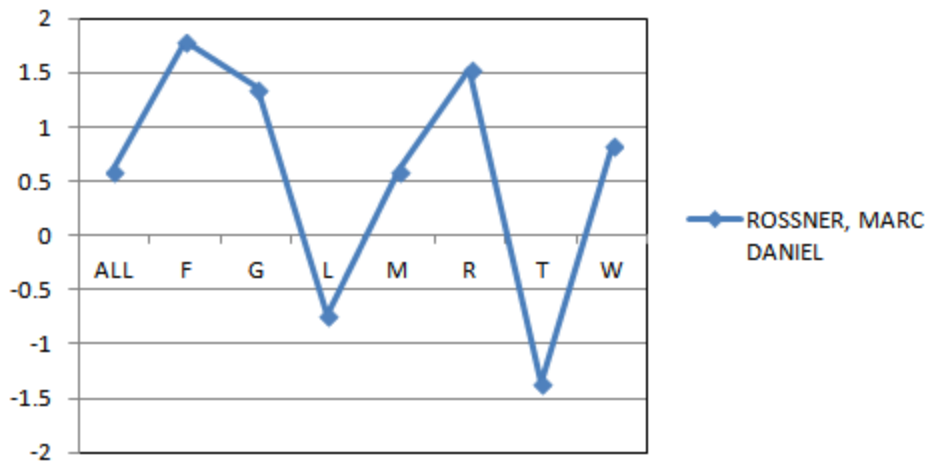
* ALL TOPIC ON STUDENT				i TOPIC CLASS				n TOPIC CLASS				STUDENT	
COUNT	T.SCORE	MEASURE	S.E.	COUNT	T.SCORE	MEASURE	S.E.	COUNT	T.SCORE	MEASURE	S.E.	ENTRY	LABEL
21	9	-1.92	.00	5	0	-3.39	1.64	16	9	-1.50	.59	2 NM	KAT
9	2	-3.09	.00	3	1	-2.18	1.40	6	1	-3.63	1.19	5 NL	SMI

Excel procedure:

1. Output Table 31.7
2. Paste into a new Excel worksheet
3. In Excel, "Data", "Text to Columns", "Fixed width"
4. "Insert", "Line Chart with Markers"
5. "Select Data", the Excel spreadsheet row. "Category labels", the top row of the Excel spreadsheet.
6. X-axis, text labels.

Here is the plot for Example0.txt across items with different first letters:

ROSSNER, MARC DANIEL



This can be automated with an Excel macro:

Sub Profile()

firstsubtest = "A" ' in Excel spreadsheet, delete unwanted columns. First subtest column

lastsubtest = "H" ' total score + 7 subtests

firstperson = 2 ' first person in row 2 of the worksheet

lastperson = 5 ' last person in row 76 of the spreadsheet

personname = "I" ' person name in column 9 = I

For thisperson = lastperson To firstperson Step -1

thispersonrow = Trim(thisperson)

Set objchart = Charts.Add

objchart.ChartType = xlLineMarkers

objchart.SeriesCollection.NewSeries

objchart.SeriesCollection(1).Values = "=" & Sheet1!\$" + firstsubtest + "\$" + thispersonrow + ":" + lastsubtest + "\$" + thispersonrow

objchart.SeriesCollection(1).Name = "=" & Sheet1!\$" + personname + "\$" + thispersonrow

objchart.SeriesCollection(1).XValues = "=" & Sheet1!\$" + firstsubtest + "\$1:\$" + lastsubtest + "\$1"

Next thisperson

' and to print the charts

For Each ch In ActiveWorkbook.Charts

ch.PrintOut

Next ch

End Sub

14.41 Table 32 Control specifications

This gives the current setting of every Winsteps control variable. This is the default setting for unchanged variables. This can be accessed as Table 32 from a control file, or from the [Output Files](#) pull-down menu "Control variable list=". It is written to a temporary text file, but can be "saved as" to a permanent file.

```

; Values of Control Specifications
; CONTROL FILE = C:\data\kct.txt - WINSTEPS 3.90.0
; OUTPUT REPORT FILE = C:\data\ZOU713ws.txt
; DATE AND TIME = Feb 11 23:29 2004
ALPHANUM =
ASCII = Y
BATCH = N
CATREF = 0
CFILE =
CHART = N
CLFILE =
CODES = 01
CONVERGE = E
....

```

14.42 Table 33 DGF Differential Group Functioning

Enter topic text here.

14.42.1 Table 33.1, 33.2 Differential group functioning DGF pairwise

This Table identifies Differential Group Functioning DGF interactions between classification-groups of persons (identified by [DIF=](#), referencing the person labels) and classification-groups of items (identified by [DPF=](#), referencing the item labels) using the column [selection rules](#). Differential average classification-group performance (DGF) is powerful when looking for [latent classes](#) among the persons. Differential bundle functioning (DBF) is powerful when looking for local dependence among the items. For more details, see [Table 30](#) (DIF) and [Table 31](#) (DPF). A graphing technique can be used to display [DIF item characteristic curves](#) for non-uniform DIF.

Table

- 33.1 DGF report (paired person classes on each item class)
- 33.2 DGF report (paired item classes on each person class)
- [33.3](#) DGF report (list of person classes within item class)
- [33.4](#) DGF report (list of item classes within person class)
- [33.7](#) DGF Item group-Person group profiles
- [33.8](#) DGF Item group-Person group profiles

We put a code in the person label of every person indicating whether the person belongs to the Control Group, "C", or the Treatment Group, "T". The column for this code is the Differential Item Functioning [DIF=](#) column. Then, if we want a report for every item, it is a DIF analysis, Winsteps [Table 30](#).

If we want to group items, then we put a code in the item label of every item which indicates the item group to which the item belongs. The column for this code is the Differential Person Functioning, [DPF=](#) column. Then, if we want a report for every person, it is a DPF analysis, Winsteps [Table 31](#).

If we want to do "person group" with "item group" . Then this is DGF (Differential Group Functioning). Person [DIF=](#) column and Item [DPF=](#) column, Winsteps Table 33.

DGF analysis: the log-odds model for an individual person in the group and item in the group is:

$$\log(P_{ni1}/P_{ni0}) = B_{gn} - D_{hi} - M_{gh}$$

where

B_{gn} is the overall estimate from the main analysis of the ability of person n (who is in group DIF g)

D_{hi} is the overall estimate from the main analysis of the difficulty of item i (which is in DPF group h)

M_{gh} is the interaction (bias, DGF) for person group g on item group h. This is estimated from all persons in group g combined with all items in group h.

The [DGF dialog](#) displays when Table 33 is called from the [Output Tables](#) or [Plots](#) menus.

Table 33.1

DGF CLASS-LEVEL BIAS/INTERACTIONS FOR DIF=@GENDER AND DPF=\$\$S1W1													
PERSON CLASS	DGF SCORE	DGF SIZE	DGF S.E.	PERSON CLASS	DGF SCORE	DGF SIZE	DGF S.E.	DGF CONTRAST	JOINT S.E.	Rasch-Welch t	d.f.	Prob.	ITEM CLASS
F	.00	.06	.28	M	.01	-.16	.27	.22	.39	.57	452	.5708	1
F	.01	-.18	.50	M	.00	.00	.61	-.18	.79	-.23	101	.8208	2
F	.04	-.50	.83	M	-.04	.76	1.04	-1.26	1.33	-.95	32	.3484	3
F	-.01	.00<	3.00	M	-.01	.00<	2.40	.00	3.84	.00	32	1.0000	4
M	.01	-.16	.27	F	.00	.06	.28	-.22	.39	-.57	452	.5708	1
M	.00	.00	.61	F	.01	-.18	.50	.18	.79	.23	101	.8208	2
M	-.04	.76	1.04	F	-.04	-.50	.83	1.26	1.33	.95	32	.3484	3
M	-.01	.00<	2.40	F	-.01	.00<	3.00	.00	3.84	.00	32	1.0000	4

The most important numbers in Table 33.1: The DGF CONTRAST is the difference in difficulty of the item between the two groups. This should be at least 0.5 logits for DGF to be noticeable. "Prob." shows the probability of observing this amount of contrast by chance, when there is no systematic item bias effect. For statistical significance DGF on an item, Prob. ≤ .05.

Table 33.2

DGF CLASS-LEVEL BIAS/INTERACTIONS FOR DIF=@GENDER AND DPF=\$\$S1W1													
ITEM CLASS	DGF SCORE	DGF SIZE	DGF S.E.	ITEM CLASS	DGF SCORE	DGF SIZE	DGF S.E.	DGF CONTRAST	JOINT S.E.	Rasch-Welch t	d.f.	Prob.	PERSON CLASS
1	.00	.06	.28	2	.01	-.18	.50	.24	.57	.42	119	.6774	F
1	.00	.06	.28	3	.04	-.50	.83	.56	.87	.64	30	.5261	F
1	.01	-.16	.27	2	.00	.00	.61	-.16	.67	-.24	100	.8097	M
1	.01	-.16	.27	3	-.04	.76	1.04	-.92	1.07	-.86	25	.3973	M
2	.01	-.18	.50	1	.00	.06	.28	-.24	.57	-.42	119	.6774	F
2	.01	-.18	.50	3	.04	-.50	.83	.32	.97	.33	39	.7420	F

This Table contrasts, for each item class, the size and significance of the Differential Item Functioning for pairs of person classifications.

DGF class specification defines the columns used to identify DGF classifications, using [DIF=](#) and [DPF=](#), see the [selection rules](#).

Reading across the Table 33.1 columns:

PERSON CLASS identifies the CLASS of persons specified with [DIF=](#), e.g., the first here is CLASS is "F".

DGF estimates with the [iterative-logit](#) (Rasch-Welch) method:

DGF SCORE is the average response score-point difference between the observed and the expected scores for this PERSON CLASS on this ITEM CLASS. Higher scores mean locally higher ability or locally lower difficulty relative to each groups performance overall.

DGF SIZE is the differential difficulty of this item (scaled by [USCALE=](#)) for this class, with all else held constant, e.g., .07 is the relative difficulty for Kid Class F on Item Class1. The more difficult, the higher the DGF measure.

-.52> reports that this measure corresponds to an extreme maximum person-class score. [EXTRSCORE=](#) controls extreme score estimate.

1.97< reports that this measure corresponds to an extreme minimum person-class score. [EXTRSCORE=](#) controls extreme score estimate.

-6.91E reports that this measure corresponds to an item with an extreme score, which cannot exhibit DIF

DGF S.E. is the standard error of the DGF SIZE (scaled by [USCALE=](#)).

PERSON CLASS identifies the CLASS of persons, e.g., the second CLASS is "M".

DGF SCORE is the average response score-point difference between the observed and the expected scores for this PERSON CLASS on this ITEM CLASS.

DGF SIZE is the differential difficulty of this item for this class, with all else held constant, e.g., -.15 is the relative difficulty for Kid Class M on Item Class1. The more difficult, the higher the DGF measure.

DGF S.E. is the standard error of the second DGF SIZE.

DGF CONTRAST is the difference between the two DGF SIZE, i.e., size of the DGF across the two classifications of persons, e.g., .07 - .15 = .23 (usually in logits). A positive DGF contrast indicates that the item is more difficult for the first, left-hand-listed CLASS. See details in Table 33.3.

JOINT S.E. is the standard error of the DGF CONTRAST = $\sqrt{\text{first DIF S.E.}^2 + \text{second DIF S.E.}^2}$, e.g., .38 = $\sqrt{.27^2 + .27^2}$ t gives the DGF significance as a Student's t-statistic = DGF CONTRAST / JOINT S.E. The t-test is a two-sided test for the difference between two means (i.e., the estimates) based on the standard error of the means (i.e., the standard error of the estimates). The null hypothesis is that the two estimates are the same, except for measurement error.

d.f. is the joint degrees of freedom. This is shown as the sum of the sizes of two classifications (see Table 33.3 less 2 for the two measure estimates, but this estimate of d.f. is somewhat high, so interpret the t-test conservatively, e.g., d.f. = (426 A + 1 D - 2) = 425. When the d.f. are large, the t statistic can be interpreted as a unit-normal deviate, i.e., z-score.

INF means "the degrees of freedom are so large they can be treated as infinite", i.e., the reported t-value is a unit normal deviate.

Prob. is the two-sided probability of Student's t. See [t-statistics](#).

ITEM CLASS is the item classification specified by [DPF=](#). Here the first ITEM CLASS is "1"

Each line in the Table is repeated with the PERSON CLASSes in reversed order.

14.42.2 Table 33.3, 33.4 Differential group functioning DGF list

Table 33 supports the investigation of item bias, [Differential Group Functioning](#) (DGF), i.e., interactions between classes of items and types of persons. Specify [DIF=](#) for person classifying indicators in person labels, and [DPF=](#) for item classifying indicators in the item labels.

Example output:

You want to examine item bias (DIF) between Females and Males in [Exam1.txt](#). You need a column in your Winsteps person label that has two (or more) demographic codes, say "F" for female and "M" for male (or "0" and "1" if you like dummy variables) in column 9.

[Table 33.1](#) is best for pairwise comparisons, e.g., Females vs. Males. Use [Table 33.1](#) if you have two classes of persons, and [Table 33.2](#) if you have two classes of items.

Table 33.3 or Table 33.4 are best for multiple comparisons, e.g., regions against the national average. Table 33.3 sorts by item class then person class. Table 33.4 sorts by person class then item class.

Table 33.3

```

DGF CLASS-LEVEL BIAS/INTERACTIONS FOR DIF=@GENDER AND DPF=$$SIW1
-----
| PERSON      OBSERVATIONS      BASELINE      DGF      DGF      DGF      DGF      ITEM |
| CLASS      COUNT AVERAGE EXPECT      SCORE     SIZE  S.E.   t   Prob.  CLASS |
-----
| F          234      .46      .46          .00      .06   .28   -.22  .8294  1 |
| F           54      .85      .84          .01     -.18   .50   .36  .7224  2 |
| F           18      .89      .85          .04     -.50   .83   .60  .5541  3 |
| F           18      .00      .01         -.01     .00<  3.00  .00  1.000  4 |
| M          221      .51      .50          .01     -.16   .27   .59  .5552  1 |
| M           51      .86      .86          .00      .00   .61   .00  1.000  2 |
| M           17      .82      .87         -.04     .76  1.04  -.74  .4734  3 |
| M           17      .00      .01         -.01     .00<  2.40  .00  1.000  4 |
-----

```

Table 33.4

```

-----
| ITEM      OBSERVATIONS      BASELINE      DGF      DGF      DGF      DGF      PERSON |
| CLASS      COUNT AVERAGE EXPECT      SCORE     SIZE  S.E.   t   Prob.  CLASS |
-----
| 1          234      .46      .46          .00      .06   .28   -.22  .8294  F |
| 1          221      .51      .50          .01     -.16   .27   .59  .5552  M |
| 2           54      .85      .84          .01     -.18   .50   .36  .7224  F |
| 2           51      .86      .86          .00      .00   .61   .00  1.000  M |
| 3           18      .89      .85          .04     -.50   .83   .60  .5541  F |
| 3           17      .82      .87         -.04     .76  1.04  -.74  .4734  M |
| 4           18      .00      .01         -.01     .00<  3.00  .00  1.000  F |
| 4           17      .00      .01         -.01     .00<  2.40  .00  1.000  M |
-----

```

This displays a list of the local difficulty/ability estimates underlying the paired DGF analysis. These can be plotted directly from the [Plots](#) menu.

DGF class specification identifies the person-label columns containing DIF classifications, with [DIF=](#) set to [@GENDER](#) using the [selection rules](#). The item-label columns for item classes are specified by [DPF=](#).

Table 33.3. The DGF effects are shown ordered by Person CLASS within item class.

Table 33.4. The DGF effects are shown ordered by Person CLASS within Item CLASS.

KID CLASS identifies the CLASS of persons. KID is specified with [PERSON=](#), e.g., the first CLASS is "F"

OBSERVATIONS are what are seen in the data

COUNT is the number of observations of the classification used for DIF estimation, e.g., 18 F persons responded to TAP item 1.

AVERAGE is the average observation on the classification, e.g., 0.89 is the proportion-correct-value of item 4 for F persons.

COUNT * AVERAGE = total score of person class on the item

BASELINE is the prediction without DGF

EXPECT is the expected value of the average observation when there is no DIF, e.g., 0.92 is the expected proportion-correct-value for F without DGF.

DGF: Differential Group Functioning

DGF SCORE is the difference between the observed and the expected average observations, e.g., 0.92 - 0.89 = -0.03

DGF SIZE is the relative difficulty for this class, e.g., person CLASS F has a relative difficulty of .07 for item CLASS 1-.

">" (maximum score), "<" (minimum score) indicate measures corresponding to extreme scores.

DGF S.E. is the approximate standard error of the difference, e.g., 0.89 logits

DGF t is an approximate Student's [t-statistic](#) test, estimated as DGF SIZE divided by the DGF S.E. with a little less than (COUNT-2) degrees of freedom.

Prob. is the two-sided probability of Student's *t*. See [t-statistics](#).

ITEM CLASS identifies the CLASS of items.

14.42.3 Table 33.7, 33.8 Item group-Person group profiles

The measurement profile for each person group across the item groups, and *vice-versa*, are shown. The profiles are similar to [Table 31.7](#).

TABLE 33.7 KNOX CUBE TEST ZOU182WS.TXT Jun 4 2019 16:37
 KID class/group specification is: DIF=@GENDER TAP class/group specification is: DPF=\$S1W1

1 TAP CLASS/GROUP				2 TAP CLASS/GROUP				3 TAP CLASS/GROUP				4 TAP CLASS/GROUP				KID
1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	KID
COUNT	T.SCORE	SIZE	S.E.	COUNT	T.SCORE	SIZE	S.E.	COUNT	T.SCORE	SIZE	S.E.	COUNT	T.SCORE	SIZE	S.E.	CLASS/GROUP
234	108	.07	.27	54	46	-.18	.51	18	16	-.55	.90	18	0	-.96	1.98	F
221	113	-.15	.27	51	44	-.03	.60	17	14	.68	.90	17	0	-.49	1.90	M

TABLE 33.8 KNOX CUBE TEST ZOU182WS.TXT Jun 4 2019 16:37
 KID class/group specification is: DIF=@GENDER TAP class/group specification is: DPF=\$S1W1

F KID CLASS/GROUP				M KID CLASS/GROUP				TAP
F	F	F	F	M	M	M	M	TAP
COUNT	T.SCORE	SIZE	S.E.	COUNT	T.SCORE	SIZE	S.E.	CLASS/GROUP
234	108	.07	.27	221	113	-.15	.27	1
54	46	-.18	.51	51	44	-.03	.60	2
18	16	-.55	.90	17	14	.68	.90	3
18	0	-.96	1.98	17	0	-.49	1.90	4

14.43 Table 34.1 Columnar statistical comparison and scatterplot

Entry numbers	Differenc	Measure	Comparison
18 -13	-8 -7		7 NUM TAP
*		.	1 1-4
*	*	.	2 2-3
*	*	.	3 1-2-4
*	*	.	4 1-3-4
*	*	.	5 2-1-4
*	*	.	6 3-4-1

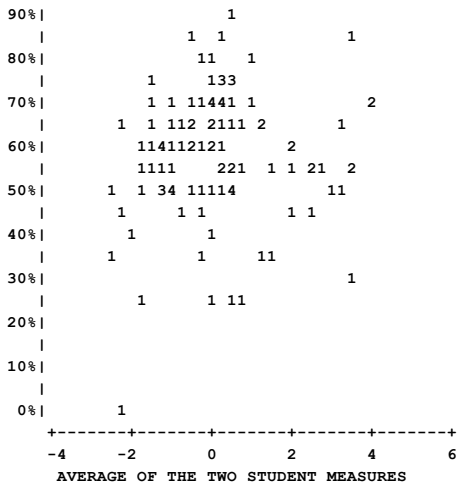


Table 35.3, 35.13, 35.23, ...

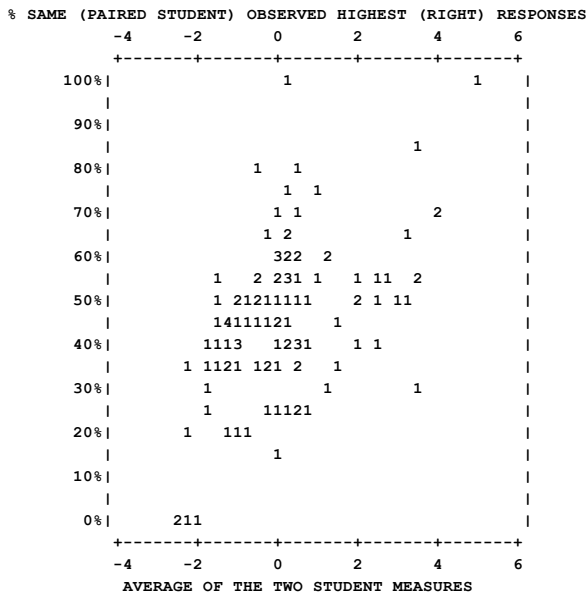


Table 35.4, 35.14, 35.24, ...

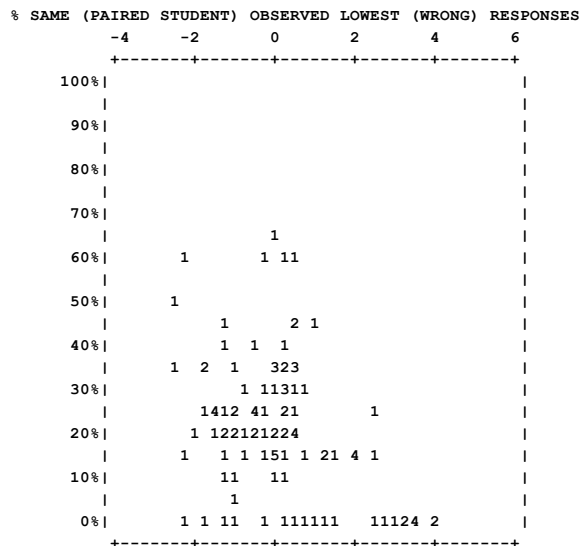


Table 35.3, 35.13, 35.23, ...

y-axis: "% Same (paired student) observed highest (right) responses" = the count of items for which both persons have the same highest scored response divided by the count of items for which one or both persons have a highest scored response %.

Cheating detection: shared knowledge

Table 35.4, 35.14, 35.24, ...

y-axis: "% Same (paired student) observed lowest (wrong) responses" = the count of items for which both persons have the same lowest scored response divided by the count of items for which one or both persons have a lowest scored response %.

Cheating detection: shared ignorance - a better indication of cheating than shared knowledge.

		-4	-2	0	2	4	6
		AVERAGE OF THE TWO STUDENT MEASURES					
Table 35.5, 35.15, 35.25, ...							
		% SAME (PAIRED STUDENT) MISSING RESPONSES					
		-4	-2	0	2	4	6
		+-----+-----+-----+-----+-----+					
100%							
90%					1	2	1
80%			1	1			
70%			1	11	11	1	1
60%		11	3331	421	11	1	21
50%			125	1222	1	11	
40%		111	21231	55	1	1	1
30%			1	1	3421	1	
20%				1	11	1	1
10%					1	2	
0%					11		
		+-----+-----+-----+-----+-----+					
		-4	-2	0	2	4	6
		AVERAGE OF THE TWO STUDENT MEASURES					
Table 35.5, 35.15, 35.25, ...							
		<i>y-axis: "% Same (paired student) missing responses" = the count of items for which both persons have a missing response divided by the count of items for which one or both persons have a missing response %.</i>					
		<i>CAT testing: to detect test overlap, which we want to minimize in order avoid item over-exposure and "tracking" through the item bank.</i>					

14.45 Table 36 PKMAP person diagnostic maps

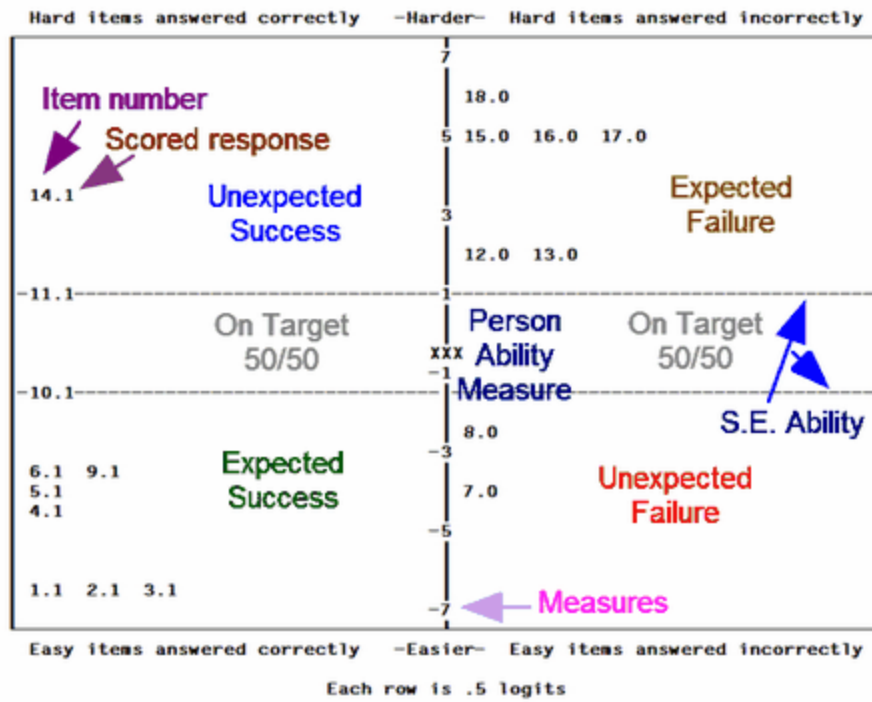
PKMAP diagnostic maps are a means of displaying the responses of a person to the items. They are described in [KIDMAP: Person-by-Item Interaction Mapping](#). The [Table Heading](#) is only shown at the start of Table 36, not for each subtable.

For dichotomies, text output:

```
PKMAP=*
MAXPAGE=40
ASCII=No
*
```

Name: Pete M
 Ref. Number: 25
 Test: KNOX CUBE TEST

Measure: -.26 S.E. 1.11 Score: 10

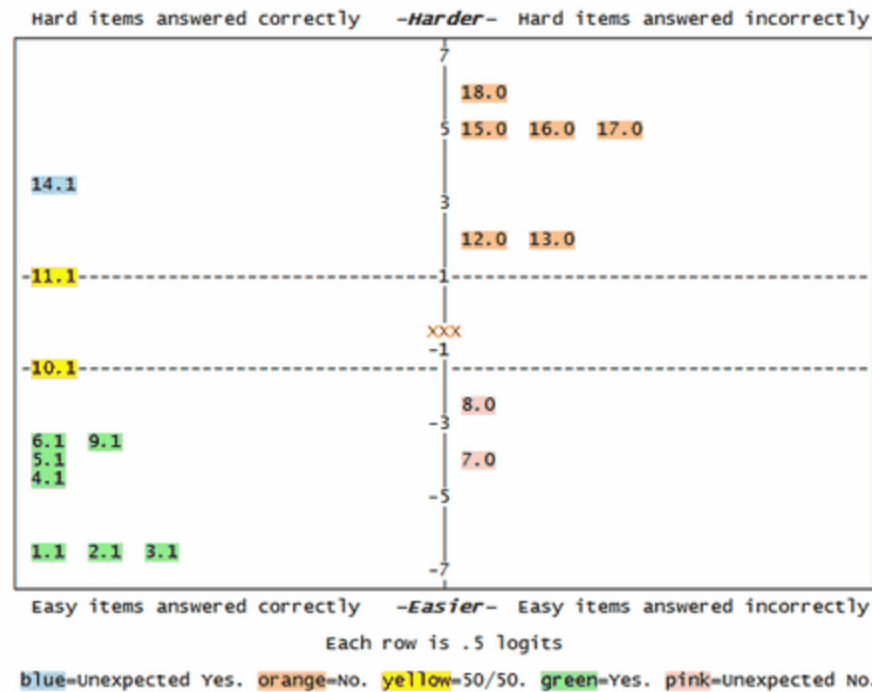


For dichotomies, webpage output, with automatic page-breaks when printed:

[PKMAP=*](#)
[MAXPAGE=40](#)
[ASCII=Webpage](#)
 *

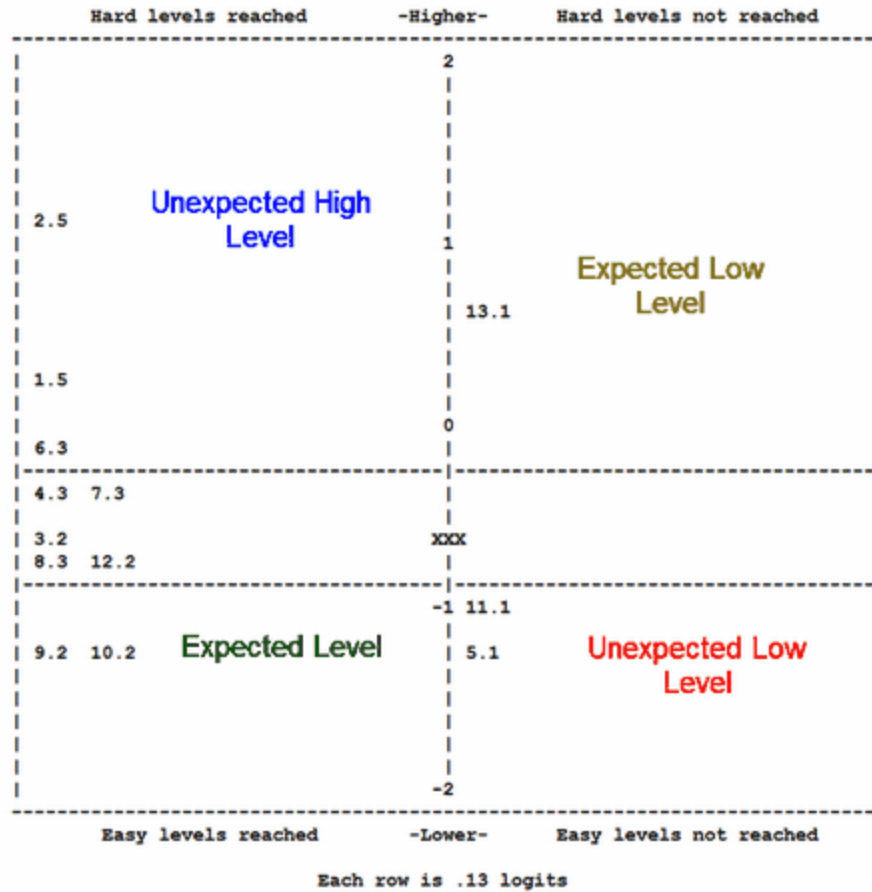
Name: Pete M
 Ref. Number: 25
 Test: KNOX CUBE TEST

Measure: -.26 S.E. 1.11 Score: 10



For polytomies:

Name: 21101 A
 Ref. Number: 1 Measure: -.59 S.E. .29 Score: 33
 Test: GENERIC ARTHRITIS FIM CONTROL FILE

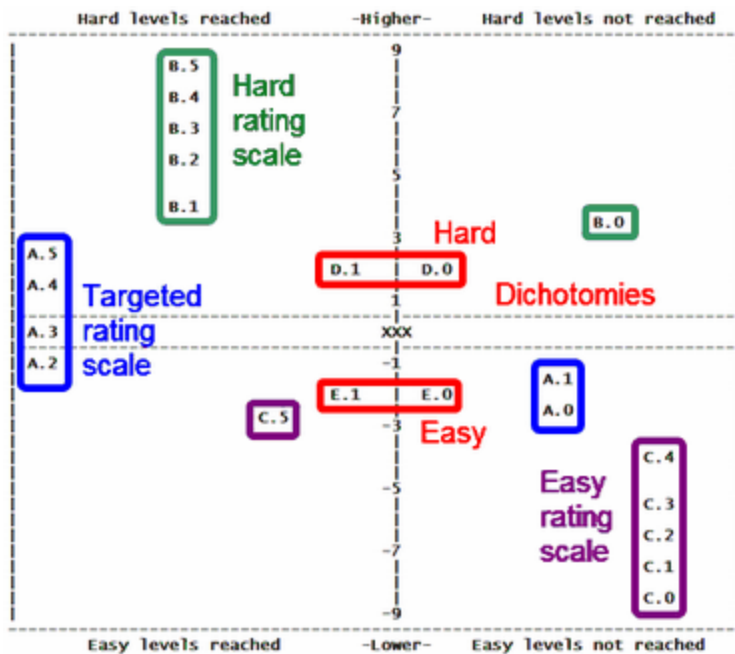


Continuation lines:

When there are too many item numbers to fit on one line, they are continued on the next line, indicated by ":"

89 134 155 296 303 357 409 465	82 93 122 139 141 166 183 293 294
	358 392
104 161 363 454 459 689	88 106 120 127 154 170 290 334 356

What observation goes where:



This Figure (using the Lucida Console font) shows the location in a PKMAP of every possible response to 6-category (0-5) rating-scale items A, B, C and two-category (0-1) dichotomous items, D, E, by a person of 0 logits ability.

For details of the PKMAP control variables, please see [PKMAP=](#)

PKMAP Size

The internal height of PKMAP grid box is 61 or [MAXPAG=](#) value - 8.

The internal width of the PKMAP grid box is [LINELENGTH=](#) value - 6.

PKMAP page breaks

Each diagnostic map automatically starts on a new page when Table 36 is output as a webpage (ASCII=W) or as a Word document (ASCII=D).

Adjusting PKMAP wording

1. With the current PKMAP settings, "Output Tables" menu. Table 36. To list the settings at the start of Table 36, include in your control file:

```
PKMAP=*
1D = Yes
*
```

2. Table 36 processing may be interrupted with Ctrl+F.

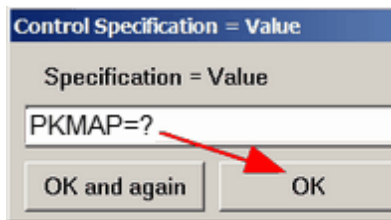
Processing Table 36

```
>=====<
```

3. Copy the PKMAP instructions from Table 36 into a new text file.

```
1= Name: #NAME#
2= Ref. Number: #ENTRY#
3= Test: #TITLE#
4=Hard items answered incorrectly
5=Hard items answered correctly
6=XXX
7=Easy items answered incorrectly
8=Easy items answered correctly
Measure: #MEASURE# S.E. #ERROR# Score: #SCORE#
```

4. Edit the PKMAP instructions in the new file
5. Save the new file
6. Specification box: PKMAP=?



7. A file dialog box displays. Navigate to the new file. Then click OK.

8. "Output Tables" menu. Table 36. Again

Questions:

1. Can I select cases, e.g., all cases with infit and outfit > 1.33? Very important to diagnose potential atypical cases.

Reply: Include in your PKMAP= file:
1F = 1.33

You wrote:

2. Can I select a particular person, e.g., #123 is interesting because of misfit on infit and outfit?

Reply:
Winsteps specification menu box: PDELETE=+123
Output Table 36
Winsteps specification menu box: PDELETE=

You wrote: Select on other criteria, e.g., age, gender?
Windows specification menu box:
PSELECT=?????M

3. Can I display the item labels instead of just the item entry numbers on the map?

Reply: Include in your PKMAP= file:
1R = #LABEL#
or, if you only want part of the label:
1R =#\$S3E5\$# ; for columns 3,4,5 in the item label

4. It does not appear that all cases get PKMAPs, do they? How are the persons selected?

Reply:
Usually everyone has a PKMAP. Your control file or PKMAP= file contains a PSELECT= or PDELETE= instruction. This removes some people.

14.46 Tables 37-41 Person Keyforms

Enter topic text here.

14.46.1 Table 37 Person KeyForms - Measure order

Table 37 Person KeyForms - Measure order. Table 37 was Table 17.3-
Table 38 Person KeyForms - Entry order. Table 38 was Table 18.3-
Table 39 Person KeyForms - Alphabetical order. Table 39 was Table 19.3-.
Table 40 Person KeyForms - Misfit order. Table 40 was Table 7.2-.
[Table 41](#) Person KeyForms Misfit order, only unexpected responses
Items are always in difficulty order, descending.
See also [Plots](#) menu, [Keyform plot](#).

This set of subtable displays a version of [Table 2.2](#) with the responses filled in for each person. These can be useful for diagnosis and individual reporting. The responses are the original values in the data file, as far as possible.

KEY: .1.=OBSERVED, 1=EXPECTED, (1)=OBSERVED, BUT VERY UNEXPECTED.

NUMBER	NAME	MEASURE	INFIT (MNSQ)	OUTFIT	S.E.	NUM	ACT
72	JACKSON, SOLOMON	-1.32	2.0	A	5.2		

MEASURE	ACT	NUM	ACT
.0.		5	FIND BOTTLES AND CANS
0	(2)	23	WATCH A RAT
0	(2)	20	WATCH BUGS
0	(1)	4	WATCH GRASS CHANGE
0	(1)	8	LOOK IN SIDEWALK CRACKS
.0.		7	WATCH ANIMAL MOVE
.0.		9	LEARN WEED NAMES
.0.		16	MAKE A MAP
0	.1.	25	TALK W/FRIENDS ABOUT PLANTS
.0.		3	READ BOOKS ON PLANTS
.0.		14	LOOK AT PICTURES OF PLANTS
.0.		6	LOOK UP STRANGE ANIMAL OR PLANT
.0.		17	WATCH WHAT ANIMALS EAT
.0.		22	FIND OUT WHAT ANIMALS EAT
.0.	1	24	FIND OUT WHAT FLOWERS LIVE ON
.1.		1	WATCH BIRDS
.1.		15	READ ANIMAL STORIES
.0.	1	2	READ BOOKS ON ANIMALS
.0.	1	21	WATCH BIRD MAKE NEST
.0.	1	11	FIND WHERE ANIMAL LIVES
.1.		13	GROW GARDEN
.0.	1	10	LISTEN TO BIRD SING
.1.		12	GO TO MUSEUM
.1.		19	GO TO ZOO
.2.		18	GO ON PICNIC

The vertical line of numbers corresponds to the person measure, and indicates the expected (average) responses. Responses marked .0. or .1. or .2. are observed and expected. Responses shown merely as 0 or 1 are expected, but not observed. Responses (1) and (2) are observed and statistically significantly unexpected, $|t| \geq 2$, $p < 5\%$. Table 41 shows only the unexpected responses.

Response code with no dots or parentheses: this only happens in the ability column and indicates the response we expected a person of that ability to make. If there are several similar responses, e.g., wrong answers to an MCQ item, then one response is chosen.

Response code with .dots. This is the response the person actually made, but is not a statistically unexpected response.

Response code with () This is the response the person actually made, but is a statistically unexpected response, $p < .05$ double-sided.

Table 41 Person KeyForms Misfit order, only unexpected responses

This table displays a version of [Table 2.2](#) with only the unexpected responses filled in for each person. These can be useful for diagnosis and individual reporting. The responses are the original values in the data file, as far as possible. Persons without unexpected responses are omitted.

KEY: .1.=OBSERVED, 1=EXPECTED, (1)=OBSERVED, BUT VERY UNEXPECTED.

NUMBER	NAME	MEASURE	INFIT (MNSQ)	OUTFIT	S.E.	NUM	TAP
29	Anne F	-.26	4.1	A	6.1		

MEASURE	TAP	NUM	TAP
0	(1)	14	1-4-2-3-4-1
0	.1.	11	1-3-1-2-4
(0)	1	8	1-4-2-3
0)	1	7	1-4-3-2

The vertical line of numbers corresponds to the person measure, and indicates the expected (average) responses. Responses (1) and (2) are observed and statistically significantly unexpected, $|t| \geq 2$, $p < 5\%$. Responses shown merely as 0 or 1 are expected, but not observed. The full list of responses is in Table 40.

14.46.2 Table 38 Person KeyForms - Entry order

See [Table 37](#)

14.46.3 Table 39 Person KeyForms - Alphabetical order

See [Table 37](#)

14.46.4 Table 40 Person KeyForms - Misfit order

See [Table 37](#)

14.46.5 Table 41 Person KeyForms - Misfit order, only unexpected responses

See [Table 37](#)

14.47 Table 42.1 Person statistics in displacement order

(controlled by [USCALE=](#), [UMEAN=](#), [UDECIM=](#), [LOCAL=](#), [TOTAL=](#))

PERSON STATISTICS: DISPLACEMENT ORDER

For details, please see [Table 10.1](#)

14.48 Table 42.2 Person statistics graphically in displacement order

Please see [Table 10.2](#)

14.49 Table 43.1 Person statistics in correlation order

(controlled by [USCALE=](#), [UMEAN=](#), [UDECIM=](#), [LOCAL=](#), [TOTAL=](#))

Table 43 is not output when person-point correlations are inestimable because items have different numbers of categories.

PERSON STATISTICS: CORRELATION ORDER

For details, please see [Table 10.1](#)

14.50 Table 43.2 Person statistics graphically in correlation order_2

Please see [Table 10.2](#). Table 43 is not output when person-point correlations are inestimable because items have different numbers of categories.

14.51 Table 44.1 Global statistics

```
Global Statistics:
Active KID: 36, weighted: 35.50, non-extreme: 35
Active TAP: 18, weighted: 17.50, non-extreme: 14
Active datapoints: 643 = 99.2% of Active+Missing datapoints, weighted: 616.2500
Missing datapoints: 5 = .8% of Active+Missing datapoints
```

Non-extreme datapoints: 487 weighted: 462.7500, $\ln(487) = 6.1883$
 Standardized residuals $N(0,1)$: mean: .00 P.SD: .86 count: 462.75
 Log-Likelihood Degrees of freedom (d.f.) by simulation = 219 +- 3
 Log-Likelihood chi-square: 217.5709 with approximately 219 d.f., probability = .5146
 Estimated Parameters = Non-extreme KID + Non-extreme TAP - 1 + sum(Thresholds - 1) = 48
 Akaike Information Criterion, AIC = (2 * parameters) + chi-square = 313.5709
 Schwarz Bayesian Information Criterion, BIC = (parameters * $\ln(\text{non-extreme datapoints})$)
 + chi-square = 512.1559
 Global Weighted Root-Mean-Square Residual: .2308 with expected value: .2341 count:
 616.25
 Capped Weighted Binomial Deviance: .0805 with expected value: .0859 count: 616.25
 dichotomies

Global statistics:	Statistics based on the currently-selected data, omitting permanently and temporarily deleted, deselected and dropped items and persons
Active person, weighted, non-extreme	Persons currently active in this analysis. Weighted count only if active persons are weighted. Non-extreme persons do not have extreme scores (0%, 100% scores)
Active item, weighted, non-extreme	Items currently active in this analysis. Weighted count only if active items are weighted.
Active datapoints, weighted, non-extreme	Observations/responses currently active in this analysis (includes extreme scores). Weighted count only if active items or persons are weighted. Non-extreme items do not have extreme scores (0%, 100% scores)
Missing datapoints, % of Active+Missing datapoints	Observations/responses not active in this analysis, usually because they do not have scored values. If any potentially active data are missing, then percent that are active
Non-extreme datapoints, weighted, non-extreme	Observations/responses currently active in this analysis (excludes extreme scores). Weighted count if any active persons or items are weighted $\ln()$ = natural logarithm of count of datapoints
Standardized residuals $N(0,1)$: mean: P.SD: count:	Standardized Residuals are modeled to have a unit normal distribution. Large departures from mean of 0.0 and standard deviation of 1.0 indicate that the data do not conform to the basic Rasch model specification that randomness in the data be normally distributed do that standardized residuals to be close to mean 0.0, P.SD 1.0. The count is of relevant observations.
Log-Likelihood degrees of freedom (d.f.) by simulation	The degrees of freedom are obtained by performing 200 simulations of Rasch-conforming data matching the active datapoints. The estimated d.f. will vary slightly with each output of Table 44, as indicated by +/- . Each output of Table 44 includes a fresh computation of the d.f. To keep the reported d.f. constant, specify SISEED= a value 2 or greater. There is an alternative computation of d.f. in Global fit statistics .
Log-likelihood chi-square with approximately d.f. +-, probability =	The chi-square value asymptotically = $-2 * \log\text{-likelihood}$ of the active datapoints. It is based on the currently-reported estimates which may depart noticeably from the "true" maximum likelihood estimates for these datapoints. The probability that these data fit the Rasch model globally. Despite good global fit, there can be considerable local misfit in Tables 6 and 10 .
Estimated Parameters	An estimate of the number of parameters (ignoring anchoring) = Non-extreme persons + Non-extreme items - 1 + sum(Thresholds - 1)
Akaike Information Criterion, AIC	$AIC = 2*k - 2*\ln(L)$ where k = parameters, and $-2*\ln(L)$ is the log-likelihood chi-square. Used for model comparison: higher values indicate worse adjusted fit
Schwarz Bayesian Information Criterion, BIC	$BIC = \ln(n)*k - 2*\ln(L)$ where n is the number of Non-extreme datapoints. Used for model comparison: higher values indicate worse adjusted fit
Global (Weighted) Root-Mean-Square Residual (RMSR) with expected value:	This is $\sqrt{(\sum(X-E)^2)}$ where the sum is across X, each of the observations, and E, the expectation of each observation according to the Rasch model. Weighting is applied to the data if IWEIGHT= or PWEIGHT= are specified. The expected value

	of the RMSR according to the Rasch model. RMSR values smaller than the expected value indicate better fit (or overfit) to the Rasch model.
Capped (Weighted) Binomial Deviance (CBD) = ... with expected value ... for ... dichotomies	This is the average of $-[X \cdot \text{LOG}_{10}(E) + (1-X) \cdot \text{LOG}_{10}(1-E)]$ for all dichotomous observations where $X=0,1$ is the observation and E is its Rasch-model expectation. E is limited to the range 0.01 to 0.99. Weighting is applied to the data if IWEIGHT= or PWEIGHT= are specified. Glickman, Mark E. "Parameter estimation in large dynamic paired comparison experiments." <i>Journal of the Royal Statistical Society: Series C (Applied Statistics)</i> 48.3 (1999): 377-394. The expected value of the CBD according to the Rasch model. CBD values smaller than the expected value indicate better fit (or overfit) to the Rasch model.

Example: Rating Scale Model (RSM) and Partial Credit Model (PCM) of the same dataset. When the models are nested (as they are with RSM and PCM), then we have:

RSM chi-square and RSM d.f.

PCM chi-square (which should be smaller) and PCM d.f. (which will be smaller)

Then the model choice could be based on: (RSM LL chi-square - PCM LL chi-square) with (RSM-PCM) d.f., However, global fit statistics obtained by analyzing your data with conditional Rasch estimation, [CMLE](#), or log-linear models (e.g., in SPSS) will be more exact than those produced by Winsteps. For the more information about the choice between RSM and PCM, see www.rasch.org/rmt/rmt143k.htm.

if global fit statistics are the decisive evidence for choice of analytical model, then Rasch analysis may not be appropriate. In the statistical philosophy underlying Rasch measurement, the decisive evidence for choice of model is "which set of measures is more meaningful and useful" (a practical decision), not "which set of measures fit the model better" (a statistical decision).

14.52 Table 45 Person measures after each item

Table 45 shows the measures and fit statistics for the persons after the administration of each item, starting with the first item. This is useful for tracking how the person measures change as the items are administered. The item difficulties are based on all the items and persons. If items are deleted or selected in the "Specification" menu box, then the person measures are estimated only on the active items. The final measures in Table/Plot 45 will differ from the reported measures in [Table 18](#).

CUMULATIVE KID MEASURES AFTER EACH ACT

```

-----
| KID      | 1 -1.49 Watch birds          | 2 -1.49 Read books on animals
| KID      |                               |
| NUMBER | SCORE COUNT OBS MEASURE M.S.E. IMSQ OMSQ | SCORE COUNT OBS MEASURE M.S.E. IMSQ OMSQ
| LABEL   |                               |
|-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 1 | 1 1 1 -1.47 4.50 .00 .00 | 3 2 2 2.17 1.41 1.00 1.00
| M Rossner, Marc Daniel |
| 2 | 2 1 2 3.01 2.18 .00 .00 | 4 2 2 3.90 1.98 .00 .00
| M Rossner, Lawrence F. |
| 3 | 2 1 2 3.01 2.18 .00 .00 | 4 2 2 3.90 1.98 .00 .00
| M Rossner, Toby G. |
| 4 | 1 1 1 -1.47 4.50 .00 .00 | 1 2 0 -5.14 1.41 1.00 1.00
| M Rossner, Michael T. |
-----

```

The displayed fields are controlled by T45OPTIONS= and the Table 45 dialog box.

Field	Meaning	Table 45	Table 45 Plot
-------	---------	----------	---------------

KID NUMBER	Person entry number in data file	First column	in first column of Excel Worksheet
1	Item entry number in data file	Top row	x-axis of plot and second row of worksheet
-1.49	Item difficulty	Top row	
Watch birds	Item label (name)	Top row	
SCORE	total score by person so far	T45OPTIONS=1??????	y-axis choice
COUNT	total count of valid observations so far	T45OPTIONS=?1?????	y-axis choice
OBS	scored observation. Missing is usually blank or "."	T45OPTIONS=??1????	y-axis choice
MEASURE	estimated person measure so far	T45OPTIONS=???1???	y-axis choice
M.S.E. R.S.E.	"Model" standard error of the measure assuming data fit the Rasch model. "Real" standard error of the measure assuming data misfit the Rasch model when REALSE=Yes	T45OPTIONS=????1??	y-axis choice
IMSQ	Infit mean-square fit statistic	T45OPTIONS=?????1?	y-axis choice
OMSQ	Outfit mean-square fit statistic	T45OPTIONS=???????1	y-axis choice
KID LABEL	Person label (name)	"Person Label field" column selection in dialog box.	"Person Legend field" column selection in dialog box. Last column of worksheet.
In dialog box:			
Start at first observation	Are values reported in cells with missing (not administered) observations?	values for items before the first valid observation are displayed as missing "." or not plotted	
Skip missing observations		values for items after the first valid observation and before the last valid observation are displayed as missing "." or not plotted	
End at last observation		values for items after the last valid observation are displayed as missing "." or not plotted	
Omit extreme items			Extreme, unanchored items are omitted from the plot. They do not change the measures, S.E.s or Mean-Squares.
(Sort order)	Sequence of persons displayed.	Person entry order	Person Legends are ordered by last displayed value which is in "Sort" column of worksheet

Please select output for Table 45

Display Table 45 person measures after each item
 T45OPTIONS= for Table 45:

- Score so far
- Count so far
- Observation
- Measure so far
- S.E. so far
- Infit Mean-Square so far
- Outfit Mean-Square so far

Person Label field:

Display plot of person measures after each item
 Table 45 plot choices:

- Score so far
- Count so far
- Observation
- Measure so far
- S.E. so far
- Infit Mean-Square so far
- Outfit Mean-Square so far

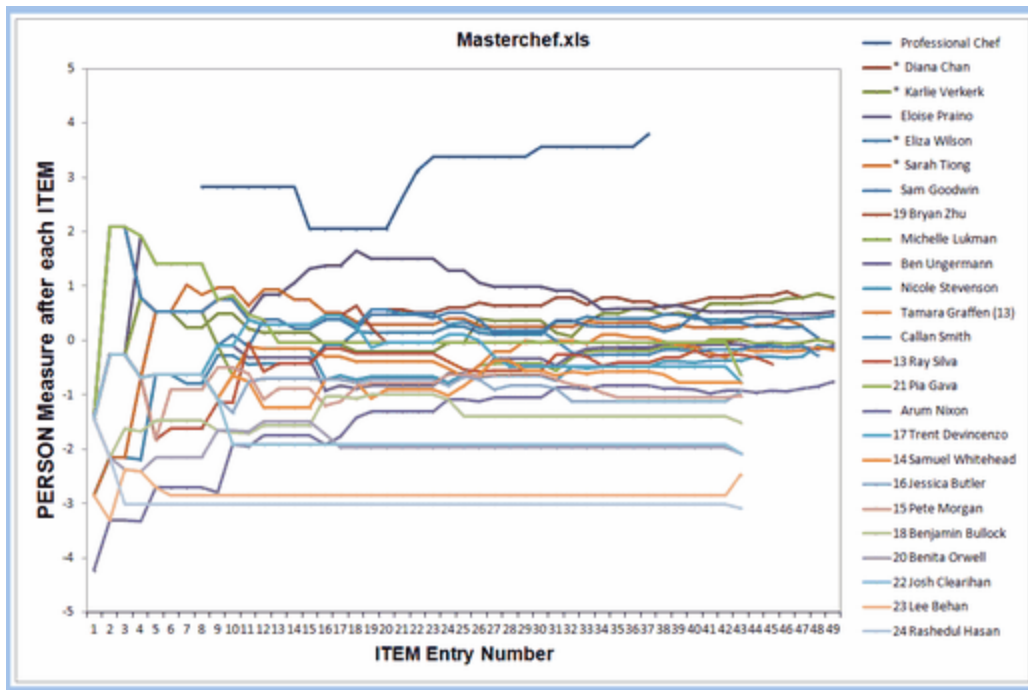
Person Legend field:

x-axis labels for items: Item Entry Number
 Item Label

Table 45:	Plot:
<input checked="" type="checkbox"/> Start at first observation	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> Skip missing observations	<input type="checkbox"/>
<input checked="" type="checkbox"/> End at last observation	<input checked="" type="checkbox"/>
	<input checked="" type="checkbox"/> Omit extreme items

OK Cancel Help

Table 45 Plot of "Masterchef" Measures, 2017. Horizontal lines for lower persons are bridging missing data.



14.53 Table heading

At the top of each output Table is basic information about the analysis:

```
TABLE 1.0 LIKING FOR SCIENCE (Wright & Masters p. ZOU214ws.txt Feb 1 14:38 2005
INPUT: 75 KIDS, 25 ACTS REPORTED: 75 KIDS, 25 ACTS, 3 CATS WINSTEPS 4.5.2
-----
```

TABLE 1.0 identifies the current [Table](#) and sub-table.

LIKING FOR SCIENCE (Wright & Masters is set by [TITLE=](#)

ZOU214ws.txt is the name of the disk file containing this Table.

Feb 1 14:38 2005 is the date and time of this analysis.

INPUT: 75 KIDS, 25 ACTS

75 KIDS gives the number of cases in the data file(s) and the case (row) identification [PERSON=](#)

25 ACTS gives the number of items specified by [N=](#) and the column (item) identification [ITEM=](#)

REPORTED: 75 KIDS

shows how many rows, cases, persons, subjects, objects are currently active for reporting ([PDELETE=](#), [PDFILE=](#), [PSELECT=](#))

25 ACTS

shows how many columns, items, agents are currently active for reporting ([IDDELETE=](#), [IDFILE=](#), [ISELECT=](#)).

3 CATS

shows how many categories are currently active for reporting.. The number of categories is determined by the [ISGROUPS=](#) item groups and the data structure as screened by [CODES=](#). etc. For details of the categories, see [Table 3.2](#)

3 CATS indicates that the Rating Scale Model is being used.

6 CATS indicates that the items are analyzed as two groups of items. Each groups has a rating scale with 3 categories.

75 CATS indicates that each of the 25 items is analyzed with 3 categories, the Partial Credit Model.

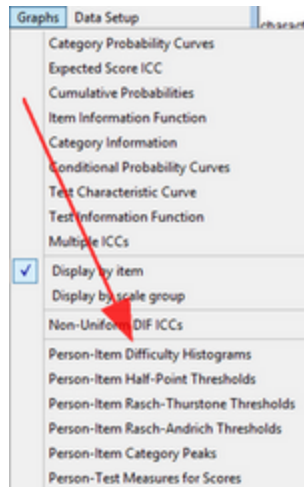
Item Grouping (Rating Scale Model, Grouped Rating Scale Model, Partial Credit Model) is flagged by the extreme right-hand column "G" in item [Table 10.1](#), etc.

In previous versions of Winsteps, ANALYZED was used to mean "used for item analysis", omitting extreme scores.

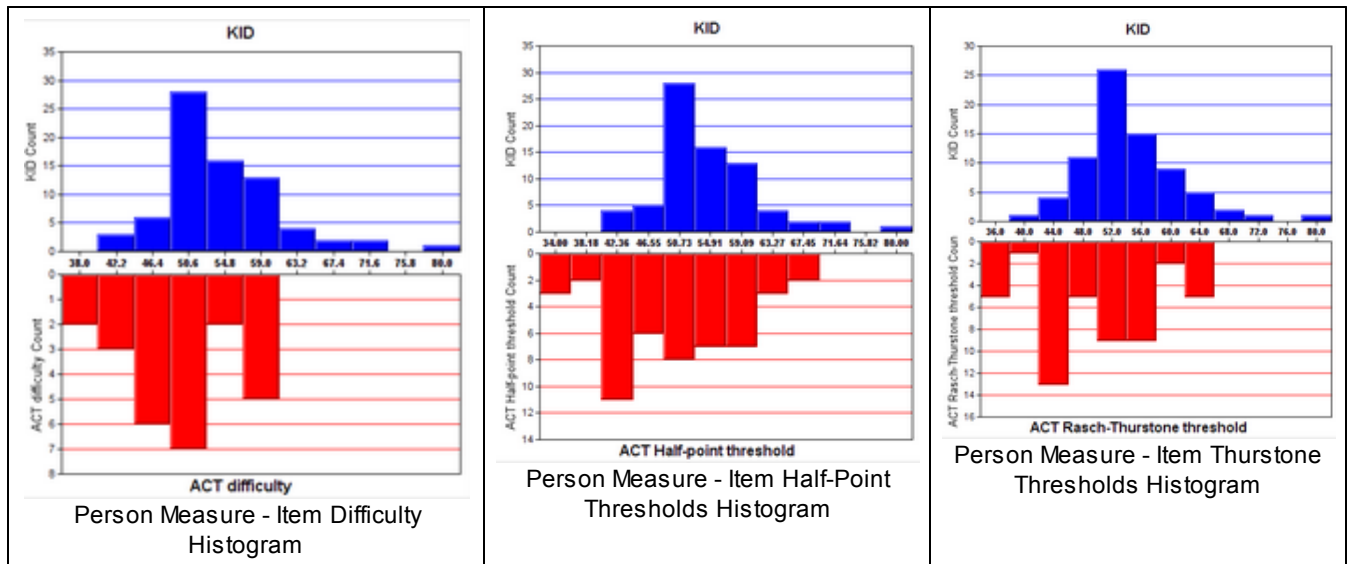
15 Charts

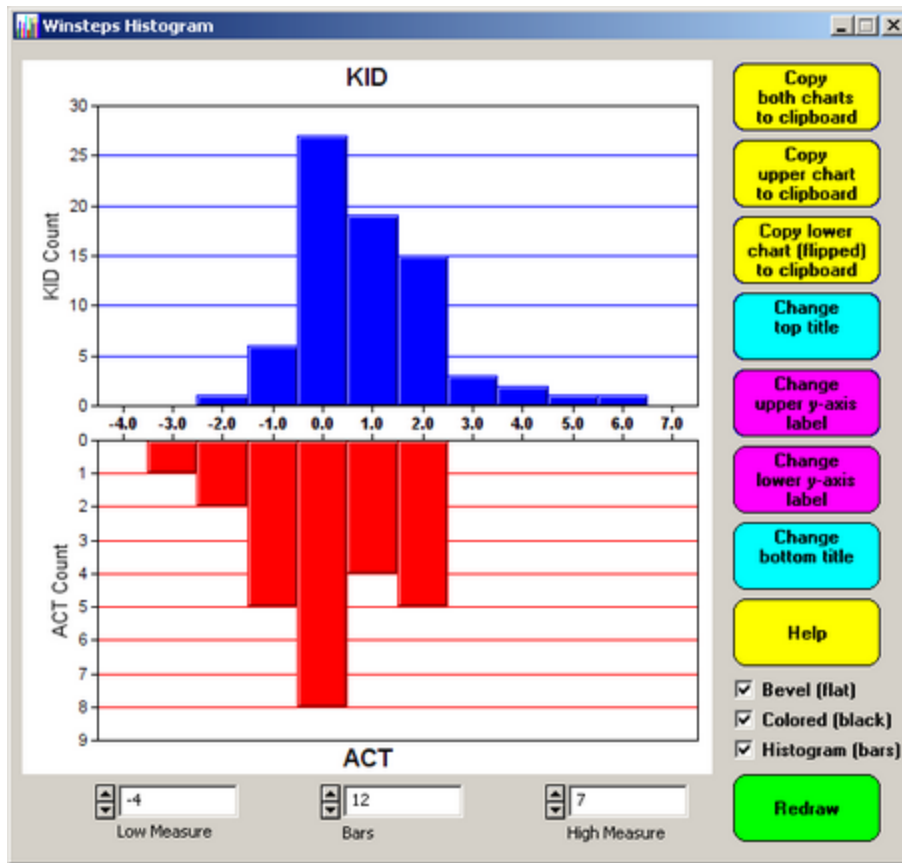
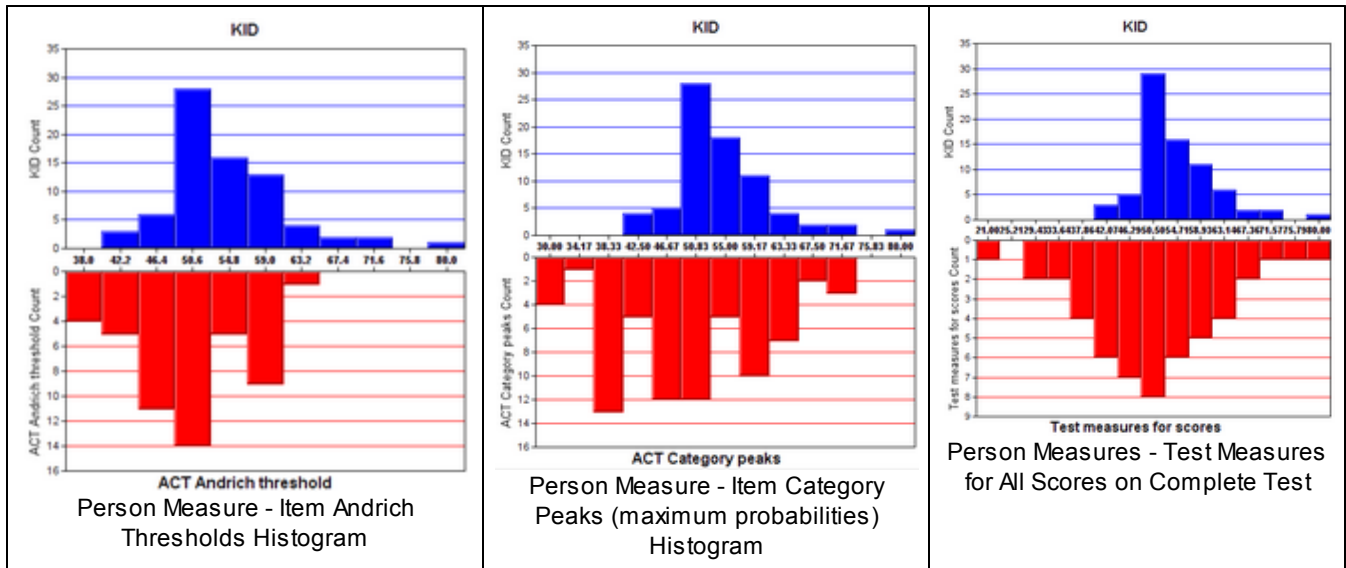
15.1 Person-item histograms or barcharts


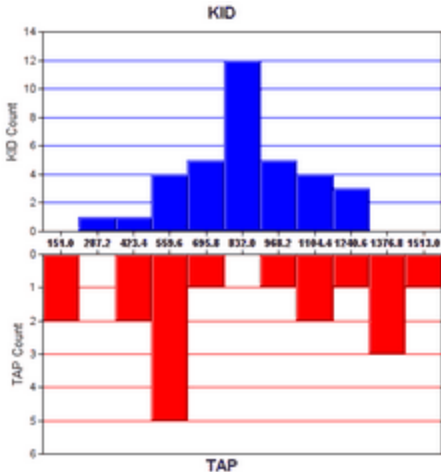
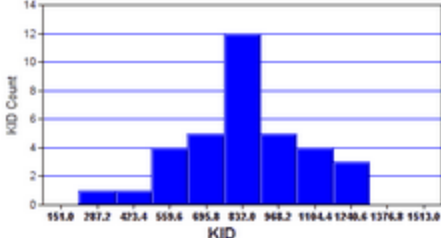
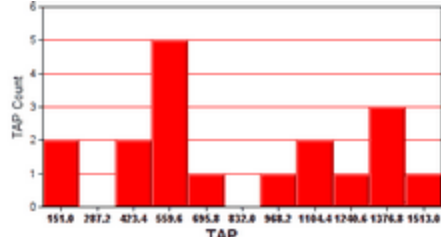



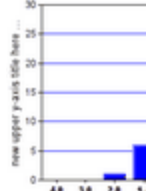

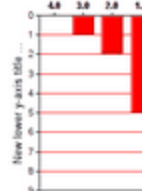
Click on Person-Item Histogram on the [Graphs menu](#):

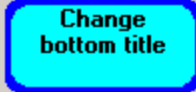









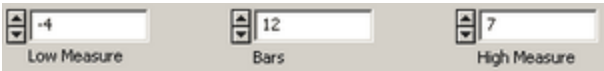
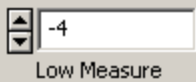
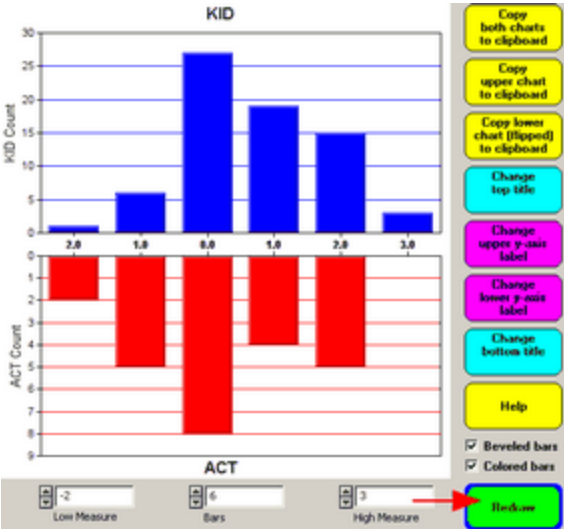
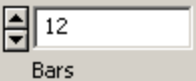



A standard histogram of the person ability distribution (upper, from [PFILE=](#)) and item difficulty distribution (lower, from [IFILE=](#)) displays:



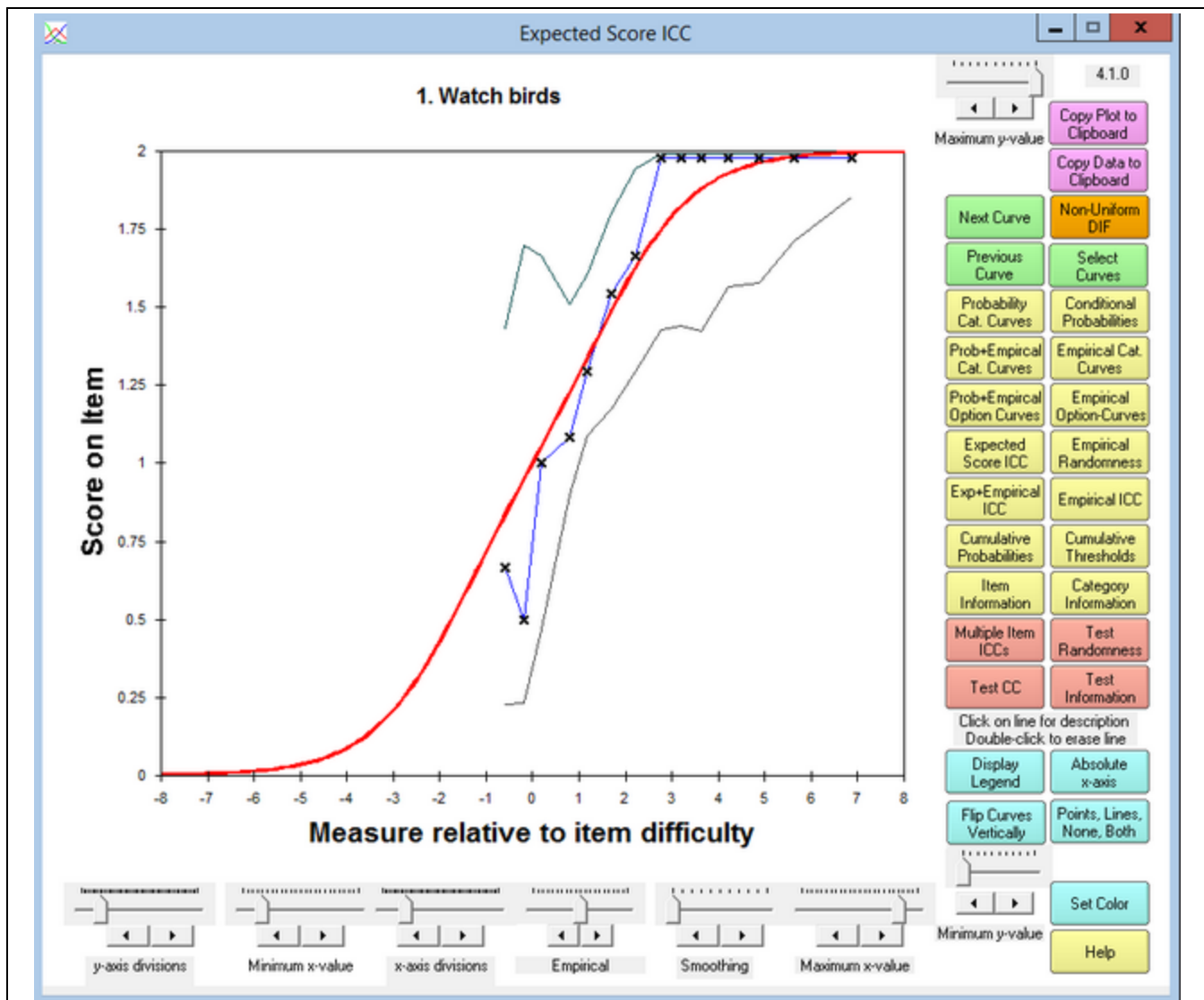


<p>Copy both charts to clipboard</p>	<p>copies both person and item histograms to the Windows clipboard, so that they can be pasted into a Word document or elsewhere.</p> 	
<p>Copy upper chart to clipboard</p>	<p>copies the upper (person) histogram to the Windows clipboard, so that they can be pasted into a Word document or elsewhere.</p> <p>The title is shown below the x-axis.</p>	
<p>Copy lower chart (flipped) to clipboard</p>	<p>copies the lower (item) histogram to the Windows clipboard, so that they can be pasted into a Word document or elsewhere.</p> <p>The bars are drawn upwards.</p>	
<p>Change top title</p>	<p>change the top (person) title</p> 	
<p>Change upper y-axis label</p>	<p>change the upper y-axis (person count) title</p> 	
<p>Change lower y-axis label</p>	<p>change the lower y-axis (item count) title</p> 	

		
<input checked="" type="checkbox"/> Bevel (flat)	Checked in box: beveled: 	No check in box: no bevel: 
<input checked="" type="checkbox"/> Colored (black)	Check in box: colored: 	No check in box: black and white: 
<input checked="" type="checkbox"/> Histogram (bars)	Check in box: histogram-format: 	No check in box: bar-chart-format: 
	Redraw charts after changing charts specifications: 	
	Change lowest displayed value on the x-axis, then click on Redraw button	
	Change the number of bars displayed on the bar chart, then click on Redraw button	
	Change highest displayed value on the x-axis, then click on Redraw button	

16 Graphs

16.1 Graphs window



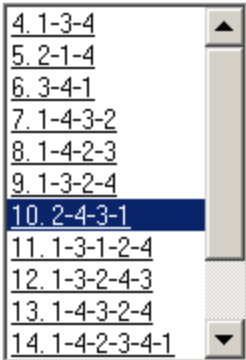
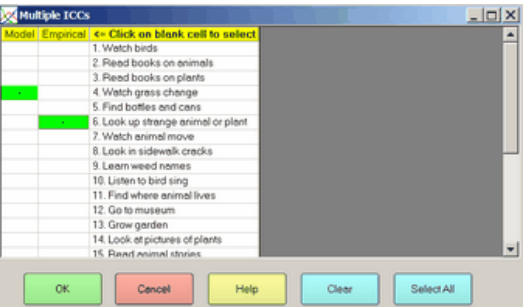

If you don't see all this on your screen, you may have your screen resolution set to 800 x 600 pixels. Try setting it to 1024 x 768. Windows "Start", "Settings", "Control Panel", "Display", "Settings", Move "Screen resolution" slider to the right.

Button or Slider	Function
<i>Graph Box</i>	
Title	Item entry number and label, or depends on type of graph
y-axis label and units	depends on type of graph
x-axis label	"Measure", person measures "Measure relative to item difficulty", (person measures - item difficulty) To switch from "Measure" to "Measure relative to item difficulty", click the blue button "Relative x-axis" To switch from "Measure relative to item difficulty" to "Measure", click the blue button "Absolute x-axis"
x-axis units	logits or user-scaled units
legend below x-axis label	displayed when blue "Display Legend" button is clicked
curve label	click once on curve

hide curve	double-click on curve. If all curves are hidden, then they are all re-displayed.
<i>Right-hand side:</i>	
Maximum y-value	Sets the maximum value on the vertical Y-axis
Copy Plot to Clipboard	<p>Copies the graph to the Windows clipboard. This only copies the part of the plot that is visible on your screen. Maximize the chart window and increase your screen resolution if the entire plot is not visible. To increase screen resolution: Windows "Start", "Settings", "Control Panel", "Display", "Settings" and move the "Screen resolution" slider to the right.</p> <p>Open a graphics or word-processing program, such as Paint or Word, and paste in the image for editing. To save the graphics plot in black-and-white, "Save As", "Black-and-white".</p> <p>Copying many similar graphs can be expedited with a keyboard macro, such as autohotkey: www.autohotkey.com</p>
Copy Data to Clipboard	<p>Copies the graphed numbers to the Windows clipboard (for pasting into Excel, etc.) Use paste special to paste as a picture meta-file, bitmap or as a text listing of the data points.</p> <p>Each row corresponds to a set of points to plot:</p> <ul style="list-style-type: none"> Column A - ignore (internal row identifier) Column B - x-axis for point on first curve: measure relative to item difficulty or relative to latent variable Column C - y-axis for point on first curve: probability, expected score, etc. Column D - x-axis for point on second curve: measure relative to item difficulty or relative to latent variable Column E - y-axis for point on second curve: probability, expected score, etc. <p>etc.</p> <p>Scatterplot:</p> <ul style="list-style-type: none"> Column C (y-axis) against Column B (x-axis) Column E (y-axis) against Column D (x-axis) <p>etc.</p> <p>Combined plots: the curves on two or more Winsteps graphs can be combined into one Excel plot using "Copy Data" from the different graphs into one Excel worksheet.</p>
Next Curve	Displays the same graph for the item with the next higher entry number.
Non-Uniform DIF	Displays the Rasch-model curve for the item and the empirical DIF curves for the DIF classifications groups when "Non-uniform DIF" is clicked on the Graphs Menu ,
Previous Curve	Displays the same graph for the item with the next lower entry number.
Select Curves	Displays the graph for the selected item from the curve-selection-box:
Probability Cat. Curves	Displays the Rasch-model probability-curve for each response-category
Conditional Probabilities	Displays the Rasch-model probability-curve for the relative probability of each adjacent pair of categories. These are dichotomous ogives.
Prob+Empirical Cat. Curves	Displays the Rasch-model probability-curve for each response-category, together with the empirical curve summarizing the data in each empirical interval.
Empirical Cat. Curves	Displays the empirical curve for each response-category, summarizing the data in each empirical interval
Prob+Empirical Option Curves	Displays the Rasch-model probability-curve for each response-option, together with the empirical curve summarizing the data in each empirical interval.
Empirical Option-Curves	Displays the empirical curve for each response-option, summarizing the data in each empirical interval
Expected Score ICC	Displays the expected-score item characteristic curve, the item's Rasch-model logistic ogive.
Empirical Randomness	Displays the fit of the item, summarizing the data in each empirical interval
Exp+Empirical ICC	Displays the expected-score item characteristic curve, the item's Rasch-model logistic ogive, together with the empirical ICC summarizing the data in each empirical interval.

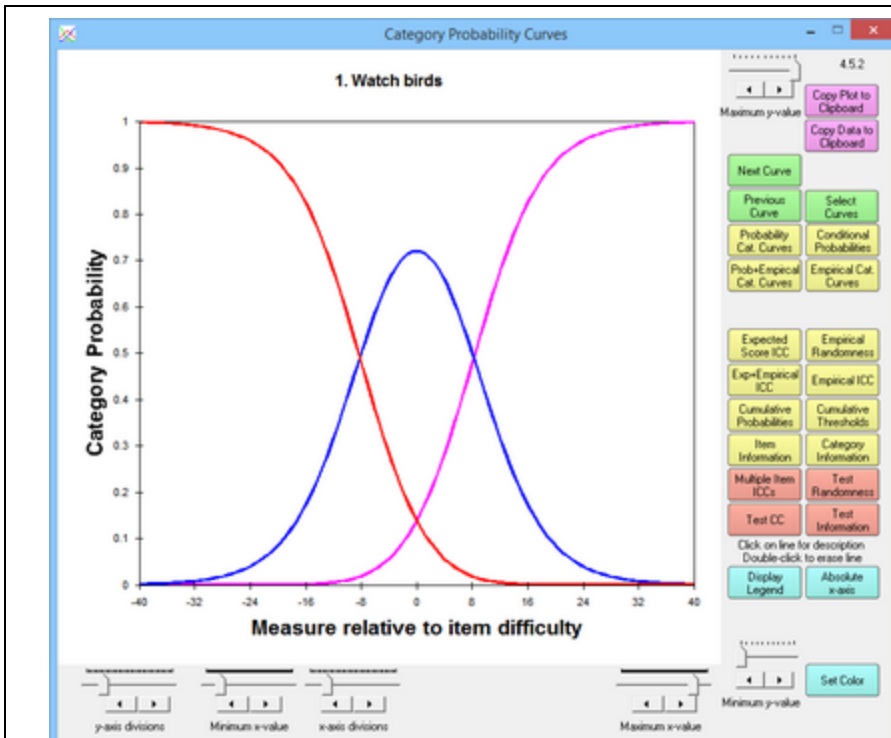
Empirical ICC	Displays the item's empirical item characteristic curve, summarizing the data in each empirical interval.
Cumulative Probabilities	Displays the Rasch-model cumulative probability-curve for each response-category and the categories below it. Click on the "Flip Curves Vertically" button for the probability of each category and the categories above it.
Cumulative Thresholds	Display the probabilities of all categories up to and including each category, and also the probabilities of all the categories above each category. The curves cross on the 0.5 probability line at the cumulative threshold (Rasch-Thurstone threshold)
Item Information	Displays the item-information function according to the Rasch model.
Category Information	Displays the information function for each category according to the Rasch model.
Multiple Item ICCs	Displays multiple Rasch-model and empirical item characteristic curves as selected in the ICC-selection box.
Test Randomness	Displays the fit of all the items in the test, summarizing the data in each empirical interval
Test CC	Displays the Rasch-model test characteristic curve (TCC), showing the relationship between raw scores and Rasch measures on the complete test of all active items.
Test Information	Displays the Test-Information Function (TIF) according to the Rasch model. The inverse square-root of the test information is the standard error of the Rasch person measure.
Click on line for description	Click on a curve to show its description below the graph. You can identify an individual traceline by single-left-clicking on it. Its description will then appear below the plot. Click elsewhere on the plot to remove the selection indicators.
Double-click to erase line	You can remove a traceline by double-left-clicking on it. Click on the command button, e.g., "Probability Curves", to return the plot to its initial appearance.
Display Legend Hide Legend	Click on "Display legend" to show descriptions of all the curves below the graph. Click on "Hide legend" to remove the descriptions.
Absolute x-axis Relative x-axis	Click on "Absolute x-axis" to show the graphs with measures on the latent variable along the x-axis. Click on "Relative x-axis" to show the graphs with the x-axis plotted relative to the item difficulty. Multiple Item ICCs shows what happens for these two choices
Flip Curves Vertically	Reverses the cumulative probability curves on the y-axis
Points, Lines, None, Both	Click on this button to show the empirical curves as both points("x") and lines, points only, lines only, and neither points nor lines.
Adjust minimum Y-value	Sets the minimum value on the vertical Y-axis
Smoothing	Controls the degrees of smoothing of the empirical curve
Set Color	Click on a curve to select it, the click on "Set Color" to change its color using the color-selection-box. White makes the line disappear.
Help	Click on Help to display this page.
<i>Across the bottom: left to right</i>	
y-axis divisions	Sets the number of tick-marks on the Y-axis
Minimum x-value	Sets the minimum value on the horizontal X-axis. To decrease the minimum possible X-value, set EXTREMESCORE= to a small value.
x-axis divisions	Sets the number of tick-marks on the X-axis
Empirical interval	Sets the number of empirical summary-intervals on the X-axis. Each observation, X_{ni} , has an observed (empirical) value, such as 0 or 1 for a dichotomy, and also a measure on the rating scale = Ability of person n - Difficulty of item i. These observations are summarized into empirical intervals on the latent variable. Each summary is plotted as the mean of the observed values in the interval (y-axis) and the mean of the measures (x-axis). The maximum size of the interval is from the lowest measure for any observation to the highest measure for an observation. The

	minimum size of the interval is .01 logits. The difference between the maximum size and the minimum size is divided into roughly 100 increments. Using the empirical interval slider, or left-right arrows, changes the size of the interval. The current size of the interval is shown in "Empirical x.xx Interval" The 1-200 number of the empirical slider is a rough indicator of relative size, and can be ignored.
Smoothing	Smooths the empirical line, which is a series of points joined by line segments.
Maximum x-value	Sets the maximum value on the horizontal X-axis. To increase the maximum possible X-value, set EXTREMESCORE= to a small value.

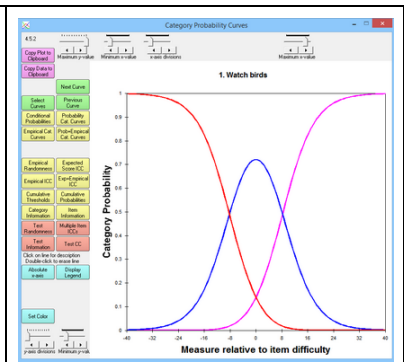
Curve selection Box	 <p>This box displays the item entry numbers and labels. Click on the item which you want displayed.</p>
ICC Selection Box	 <p>To select the curves: click on the cells and they will turn green to indicate which curves are selected for display. Then click OK to display the curves.</p>
Color Selection Box	 <p>Click on desired color, and the OK</p>

16.2 Graphs window - bottom right

If the option is checked in the [Graphs menu](#), the Graphs are displayed in the bottom right of the [Graphs window](#), so that they can be pulled larger. This is useful for increasing the quality of the graphs when their size is reduced.



Standard [Graph Window](#)



Graphs displayed to bottom right

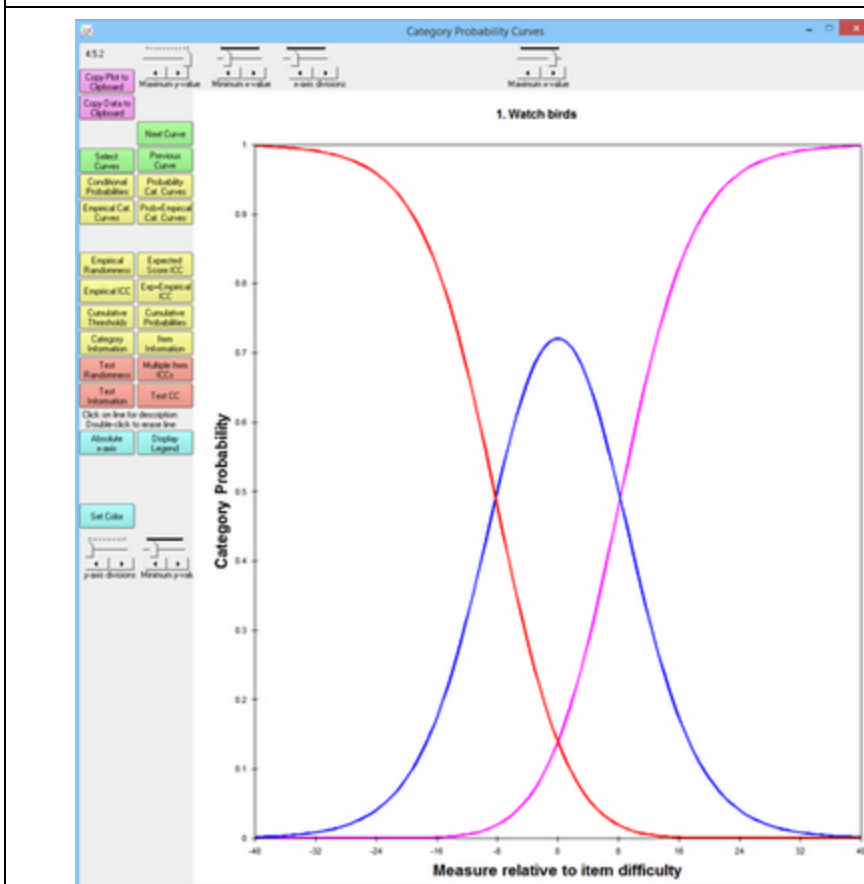


Image pulled to lower-right by dragging on bottom right corner

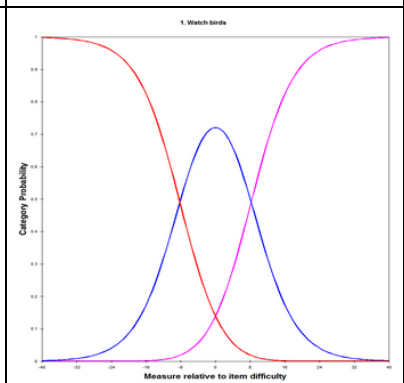
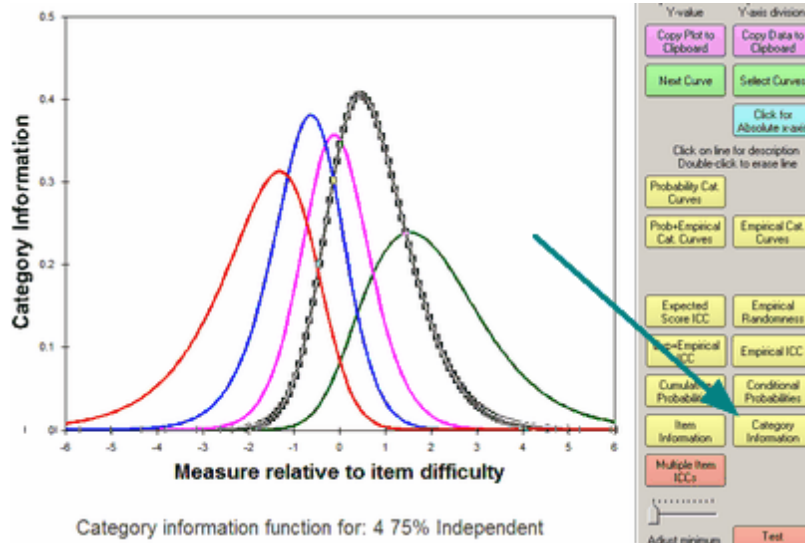


Image resized to standard size.
Notice that the lines are finer.
Use Graphics software to beautify the text.

16.3 Category information function

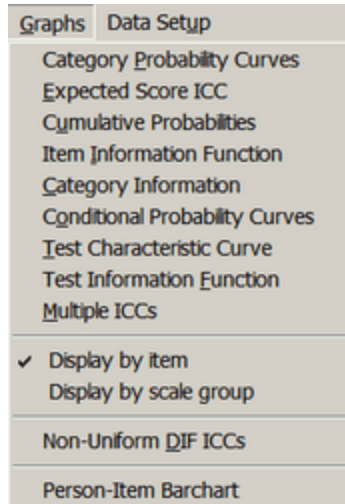


Select by clicking on "Category Information" or from the [Graphs](#) menu. This shows the item information partitioned according to the probability of observing each category.

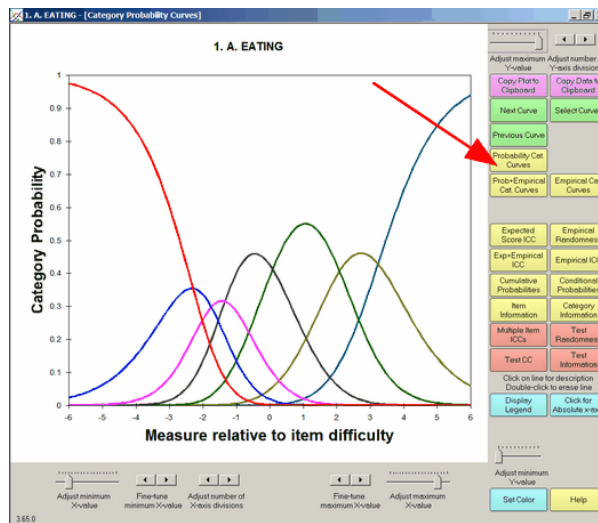
Buttons are described in [Graph window](#).

16.4 Category probability curves (Rasch-Andrich thresholds)

On the [Graphs menu](#), click on "Display by item" or "Display by scale group"

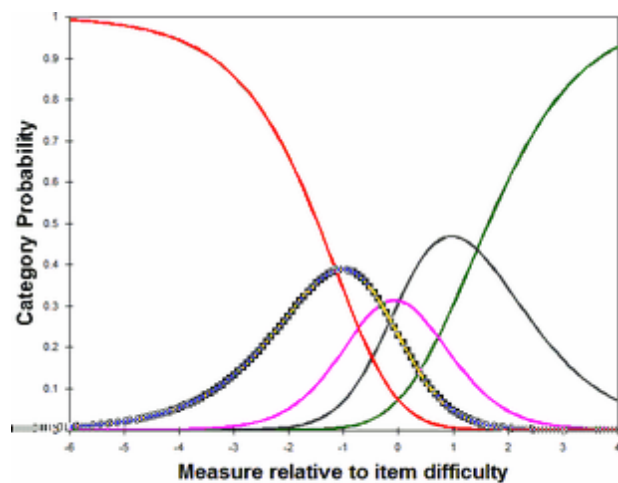


Click on "Category Probability Curves" (Item Response Curves, IRCs)



Select by clicking on "Probability Cat. Curves" in the Graph or from the [Graphs](#) menu. Buttons are described in [Graph window](#).

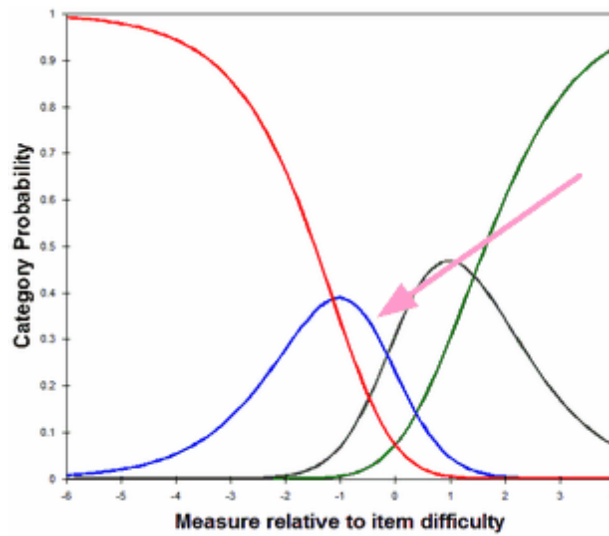
This shows the probability of observing each ordered category according to the Rasch model. [Rasch-Andrich thresholds](#) are the intersections of the curves for adjacent categories. To identify a category, click on it:



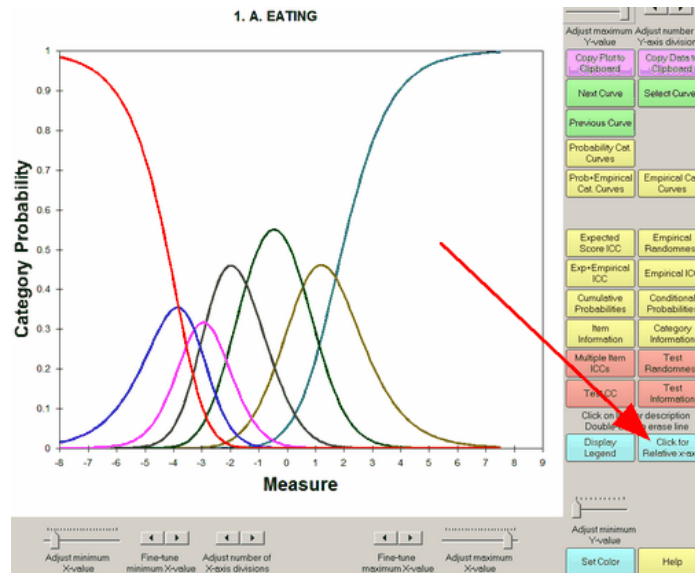
Category probability: 2 25% Independent

The caption can be clicked on and moved. "2" is the category score. "25% independent" is the category description from [CFILE=](#) or [CLFILE=](#). To clear the black highlighting, click somewhere else on the plot.

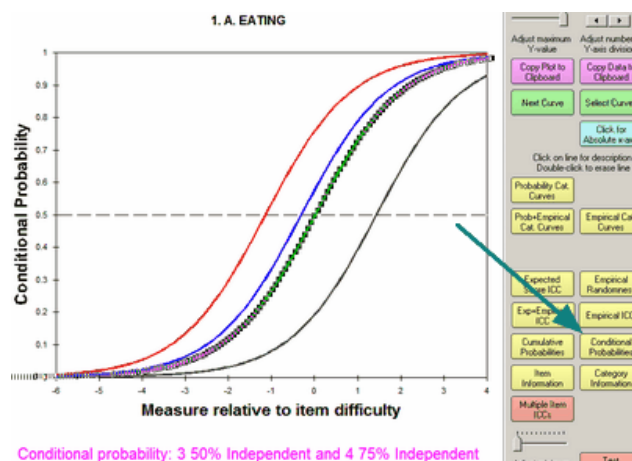
To delete the line corresponding to a category, double-click on it:



For individual items, the horizontal scaling can be changed from relative to item difficulty to relative to the latent variable by clicking on "Click for Absolute x-axis":



16.5 Conditional probability curves



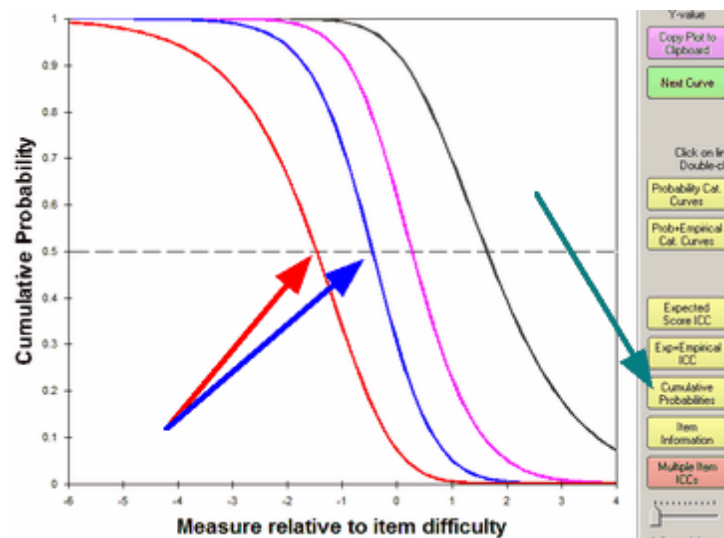
Select by clicking on "Conditional Probabilities" or from the [Graphs](#) menu. Conditional probabilities of observing adjacent categories. These are a series of Rasch dichotomous ogives. The intercepts with 0.5 probability are the Rasch-Andrich thresholds. They can be [disordered](#) relative to the latent variable.

Buttons are described in [Graph window](#).

Polytomous RP50 values are the points where the ogives cross the 0.5 line. Click on "Absolute x-axis" for values relative to the latent variable. "Copy data to clipboard", then paste into Excel to see the numerical value.

Polytomous RP67 values are the points where the ogives cross the invisible 0.67 line. Click on "Absolute x-axis" for values relative to the latent variable. "Copy data to clipboard", then paste into Excel to see the numerical value.

16.6 Cumulative probabilities



Select by clicking on "Cumulative Probabilities" or from the [Graphs](#) menu.

Lines: This shows the modeled category probability curves accumulated so that the left-hand curve (red) is the probability of being observed in the lowest category. The next curve (blue) is the probability of being observed in the lowest or next lowest category. And so on to the right. If the categories are number 1, 2, 3, 4, 5. Then the red line is the probability of observing category 1. The blue line is the probability of observing categories 1 or 2. The pink line is the probability of observing categories 1 or 2 or 3. The black line is the probability of observing categories 1 or 2 or 3 or 4. The probability of observing categories 1 or 2 or 3 or 4 or 5 (which includes every category) is 1.0 and corresponds to the upper side of the graph box.

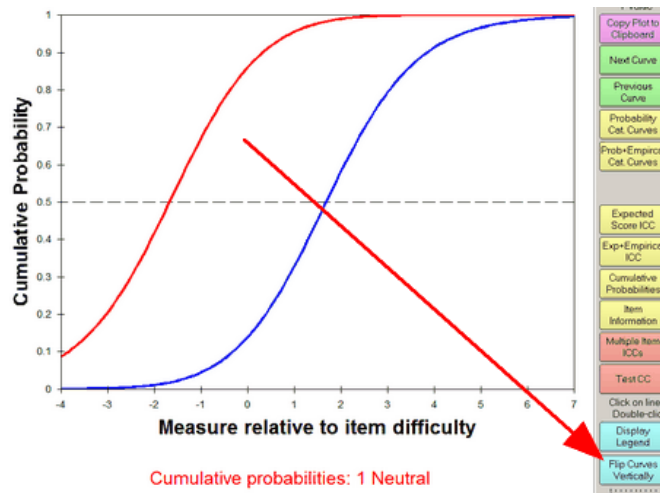
The cumulative probability is the y-axis in this graph, but is the x-axis in G.N. Masters (1988) "The Analysis of Partial Credit Scoring" (Figure 2), *Applied Measurement in Education*, 1:4, 279-297. The cumulative probability accumulates the probabilities across the categories upwards from low category to high category (or *vice versa*) for any ability level.

Arrows: The points of intersection between these curves and the 0.5 probability line are the Rasch-Thurstone thresholds. The points at which being observed in this category (or below) and the category above (or higher) are equal. These curves are always in the order of the category scores. The red arrow points to the Thurstone threshold where the probability of observing category 1 and of observing categories 2 or 3 or 4 or 5 are equal. This is called the Thurstone threshold for Category 2. The blue arrow points to the Thurstone threshold where the probability of observing categories 1 or 2 and of observing categories 3 or 4 or 5 are equal. This is called the Thurstone threshold for Category 3. The interval between the Thurstone Threshold for Category 2 and the Thurstone Threshold for Category 3 is one definition of the interval for category 2 on the latent variable.

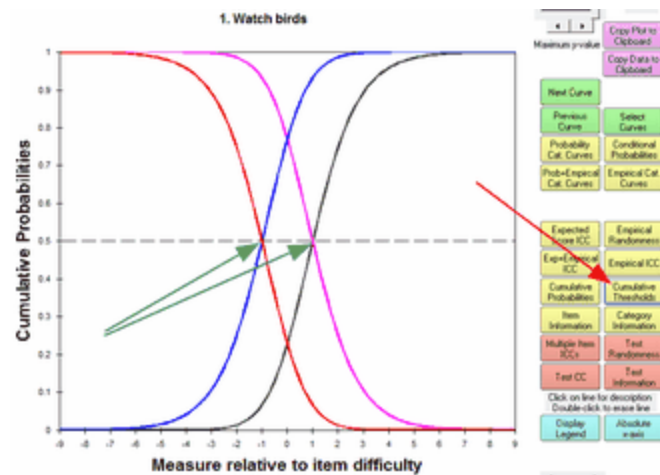
Buttons are described in [Graph window](#).

Click on the "Flip Curves Vertically" button to change the vertical direction of the curves. This reverses the definition of the cumulative probability curves. When flipped the red line in the plot above is the probability of observing categories 2 or 3 or

4 or 5. The flipped blue line in the plot above is the probability of observing categories 3 or 4 or 5. The flipped pink line is the probability of observing categories 4 or 5. The black line is the probability of observing category 5. Flipping the curves does not change the Thurstone Thresholds. The red arrow continues to point to the Thurstone Threshold for category 2.



16.7 Cumulative thresholds (Rasch-Thurstone thresholds)



Select by clicking on "Cumulative Thresholds" in the Graphs window.

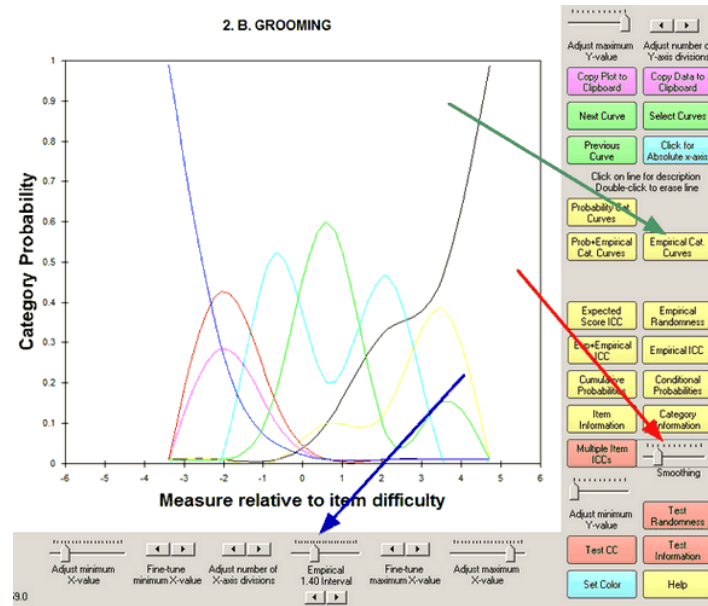
Lines: This shows the modeled category probability curves accumulated in pairs so that the left-hand curve (red) is the probability of being observed in the lowest category. The next curve (blue) is the probability of being observed in categories above the lowest category. Then the next curve (brown) is the probability of being observed in the two lowest categories. The next curve (pink) is the probability of being observed in categories above the two lowest categories. And so on to the right for rating scales with more categories.

The cumulative probability is the y-axis in this graph, but is the x-axis in G.N. Masters (1988) "The Analysis of Partial Credit Scoring" (Figure 2), *Applied Measurement in Education*, 1:4, 279-297. The cumulative probability accumulates the probabilities across the categories upwards from low category to high category (or *vice versa*) for any ability level.

Arrows: The points of intersection between these curves and the 0.5 probability line are the [Rasch-Thurstone thresholds](#) (green arrows). The points at which being observed in this category (or below) and the category above (or higher) are equal. These curves are always in the order of the category scores. The interval between the Thurstone Threshold for Category 2 and the Thurstone Threshold for Category 3 is one definition of the interval for category 2 on the latent variable.

Buttons are described in [Graph window](#).

16.8 Empirical category curves

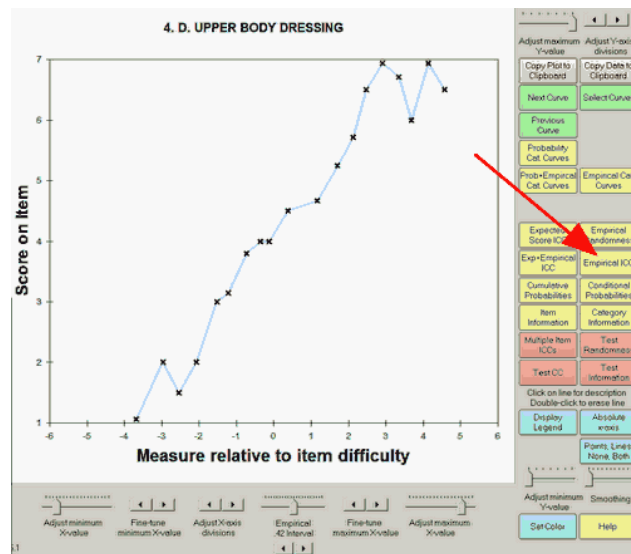


These are the empirical (data-describing) category curves. They are obtained by clicking on "Empirical Cat. Curves" or from the [Graphs](#) menu. The width of each empirical interval can be adjusted by the "Empirical Interval" control. The smoothing of the empirical curves by discrete cubic splines is adjusted by the smoothing control.

Buttons are described in [Graph window](#).

Charles S. Duris (1980). Fortran Routines for Discrete Cubic Spline Interpolation and Smoothing. ACM Algorithm 547.

16.9 Empirical ICC, IRF



Select by clicking on "Empirical ICC" or from the [Graphs](#) menu.

Line: This shows the empirical (data-descriptive) item characteristic curve. Each black "x" represents observations in an interval on the latent variable. The "x" is positioned at the average rating (y-axis) at the average measure (x-axis) for observations close by. "Close by" is set by the empirical slider beneath the plot. The blue lines are merely to aid the eye discern the trend. The curve can be smoothed with the "smoothing" slider. The [Points](#) button controls whether points+lines, points or lines are displayed.

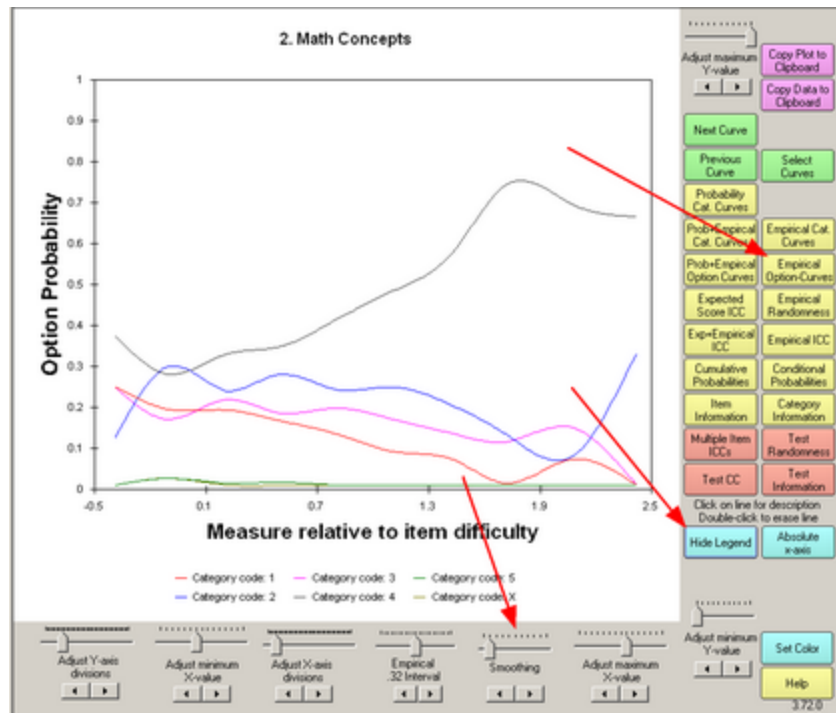
Buttons are described in [Graph window](#).

The markers on the empirical ICCs are the data-points. The lines between the markers are interpolations.

Think of the empirical ICC for an item. This summarizes the scored responses to the item. There is one scored response by each person. Each person has an ability. The abilities are stratified into ability ranges. The size of the range is shown by the slider below the middle of the graph. In each ability range, one marker is plotted. The x-axis is the average ability of the persons in that range. If there is no person, then there is no marker. The y-axis is the average of the scored responses by the persons.

The lines between the markers are to help our eyes see the pattern of the markers more clearly. If each ability range is very wide, then there may be only one or two markers. If the each ability range is very narrow, there could be one marker for each observation. There is an art to adjusting the width of the ability ranges to produce the most useful empirical ICC.

16.10 Empirical option curves



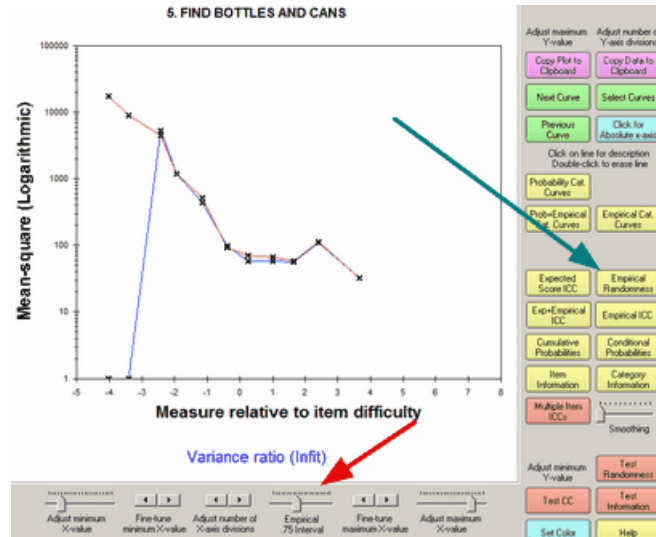
This shows the probability of choosing each option for a multiple-choice item (MCQ) as a trace line. The ascending option should be the correct response. The other options should be the incorrect distractors.

The "Empirical Option Curves" button and curves are active when Winsteps has the required information beyond that displayed in the ["Empirical Category Curves"](#). This happens when

- 1) "Display by item" is checked on the Winsteps [Graphs menu](#).
and
- 2) The data have a [scoring key](#) or other [rescoring](#) commands.

See, for instance, Robert L. Brennan, "Educational Measurement", chapter: "Test Development".

16.11 Empirical randomness

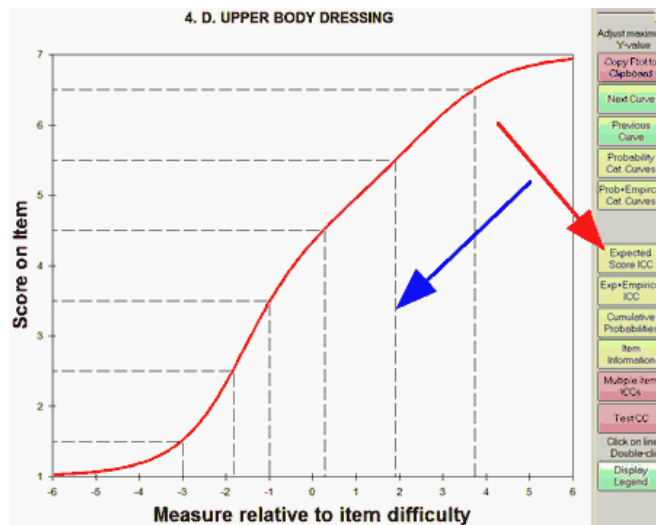


Select by clicking on "Empirical Randomness" or from the [Graphs](#) menu. Empirical intervals are set with the "Empirical Interval" slider.

This displays the local value of the mean-square statistics. The Outfit mean-square statistic (standardized residual chi-square divided by its degrees of freedom) is the red line. The Infit mean-square statistic (ratio of observed to expected residual variance) is the blue line.

Buttons are described in [Graph window](#).

16.12 Expected score ICC, IRF (Rasch-half-point thresholds)



Select by clicking on "Expected Score ICC" or from the [Graphs](#) menu. **Expected Score ICC** plots the model-expected item characteristic curve, also called the Item Response Function, IRF. This shows the Rasch-model prediction for each measure relative to item difficulty. Its shape is always ascending monotonic. The dashed lines indicate the [Rasch-half-point thresholds](#) correspond to expected values of .5 score points. The intervals on the x-axis demarcated by dashed lines are the zones within which the expected score rounds to each observed category. To remove the dashed lines, double-click on them.

The plot is of the "model item characteristic curve (ICC)", also called the "item response function (IRF)". This is used for inference. If we know the score on the item, we can infer the ability of the person responding. If we know the ability of the person responding, we can infer the expected score on the item.

(difficulty of item + person's score on the item on y-axis) -> person's ability on x-axis

(person's ability on x-axis - difficulty of item) -> expected score on the item on y-axis

Item discrimination: the item is more discriminating where the ICC is steeper, and less discriminating where the ICC is flatter.

Buttons are described in [Graph window](#).

Computation of the IRF

For an item with categories $k=0,1,2,\dots,m$ and Andrich Thresholds F_k , with F_0 chosen to be 0, then

At location x on the latent variable (relative to the item difficulty), the probability of observing category k is

$$P_k(x) = \frac{\exp(kx - \sum_{h=0}^k F_h)}{\sum_{k=0}^m \exp(kx - \sum_{h=0}^k F_h)}$$

$$\sum_{k=0}^m P_k(x) = \sum_{k=0}^m \frac{\exp(kx - \sum_{h=0}^k F_h)}{\sum_{k=0}^m \exp(kx - \sum_{h=0}^k F_h)}$$

The expected score y at location x is given by

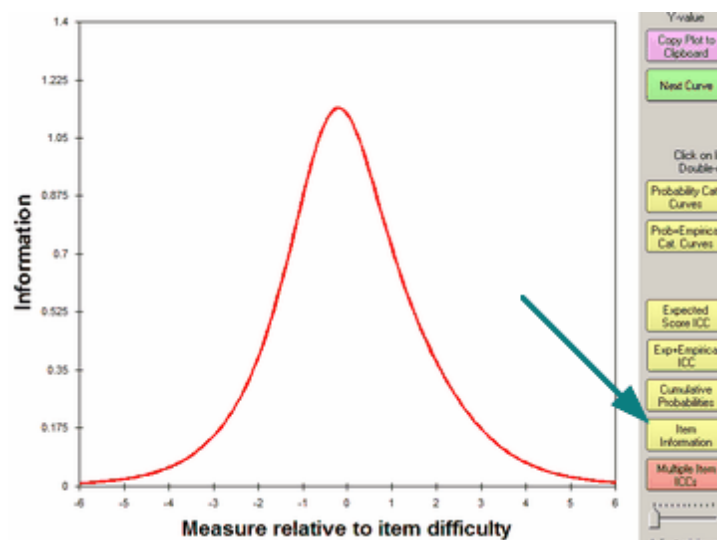
$$y = \sum_{k=0}^m (k \cdot P_k(x))$$

Computation of the Half-Point Threshold

The formula for the half-point threshold is solved iteratively, for category j , the half-point threshold is at point x (relative to the item difficulty) given by

$$j - 0.5 = \sum_{k=0}^m (k \cdot P_k(x))$$

16.13 Item information function

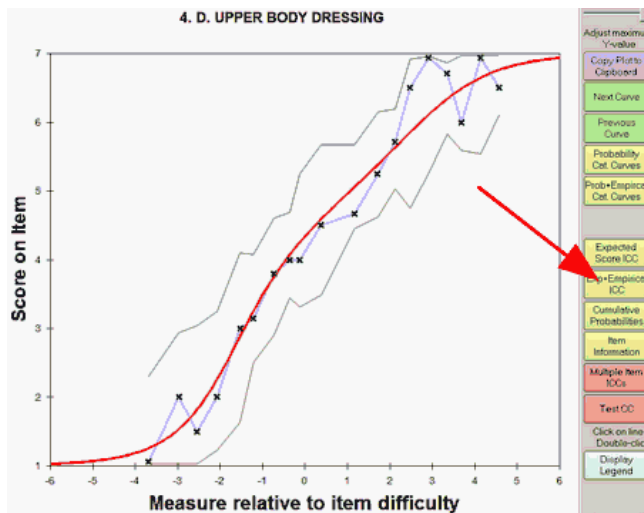


Select by clicking on "Item Information" or from the [Graphs](#) menu. This shows the (Ronald A.) Fisher measurement information in responses made to items. It is the same as the binomial variance (dichotomies) or the multinomial variance (polytomies).

This is the weighting used in [TARGET=Y](#).

Buttons are described in [Graph window](#).

16.14 Model and empirical ICCs



This shows the joint display of the expected and empirical ICCs. The boundary lines indicate the upper and lower 95% two-sided confidence intervals (interpreted vertically). When an empirical point lies outside of the boundaries, then some unmodeled source of variance maybe present in the observations. **Double-click** on a line on this plot to remove it from the display.

The solid red "model" line is generated by the relevant Rasch model. For a test of dichotomous items, these red curves will be the same for every item.

The empirical blue line is interpolated between the average ratings in each interval along the variable, marked by x. The empirical ("x") x- and y-coordinates are the means of the measures and ratings for observations in the interval. The display of the blue line and the x's is controlled by the Points button.

The upper green line (and the lower gray line) are at 1.96 model [standard errors](#) above (and below) the model "red line", i.e., form a two-sided 95% confidence band around the model red line. The vertical distance of these lines from the red line is determined by the number of observations in the interval, not by their fit.

Buttons are described in [Graph window](#).

The confidence bands are read vertically (not horizontally). There are two plotted values:

1. the mean of the observed values in the interval. Indicated by *
 2. the mean of the expected values in the interval. This is the red central line.
 3. the confidence interval is around the mean expected value, the point on the red line vertically aligned with *
- Each expected value has a model variance. These model variances are accumulated in the interval. Then converted into the standard error of the mean, SEM, around the mean expected value. $SEM = \text{square-root}(\text{summed model variance}) / (\text{count of observations in the interval})$. $1.96 * SEM$ vertically from the expected value are the confidence bands.

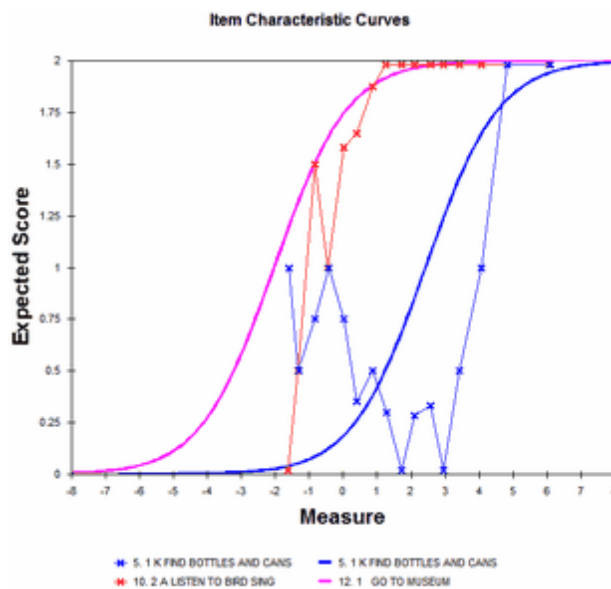
16.15 Multiple item ICCs

Select by clicking on "Multiple Item ICCs" in the [Graphs window](#) or from the [Graphs](#) menu. This enables the display of multiple model and empirical item characteristic curves/item response functions ICCs/IRFs and item information functions IIFs on the same graph.

Graphs	Data Setup
Category Probability Curves	
Expected Score ICC	
Cumulative Probabilities	
Item Information Function	
Category Information	
Conditional Probability Curves	
Test Characteristic Curve	
Test Information Function	
Multiple ICCs	
<input checked="" type="checkbox"/> Display by item	
Display by scale group	
Non-Uniform DIF ICCs	
Person-Item Barchart	

On the Graphs pull-down menu, please check "Display by Item"

Click on the "Model", "Empirical", and "Information" curves you wish to display. Click again on your selection to clear it. Click on "OK" to display the curves.

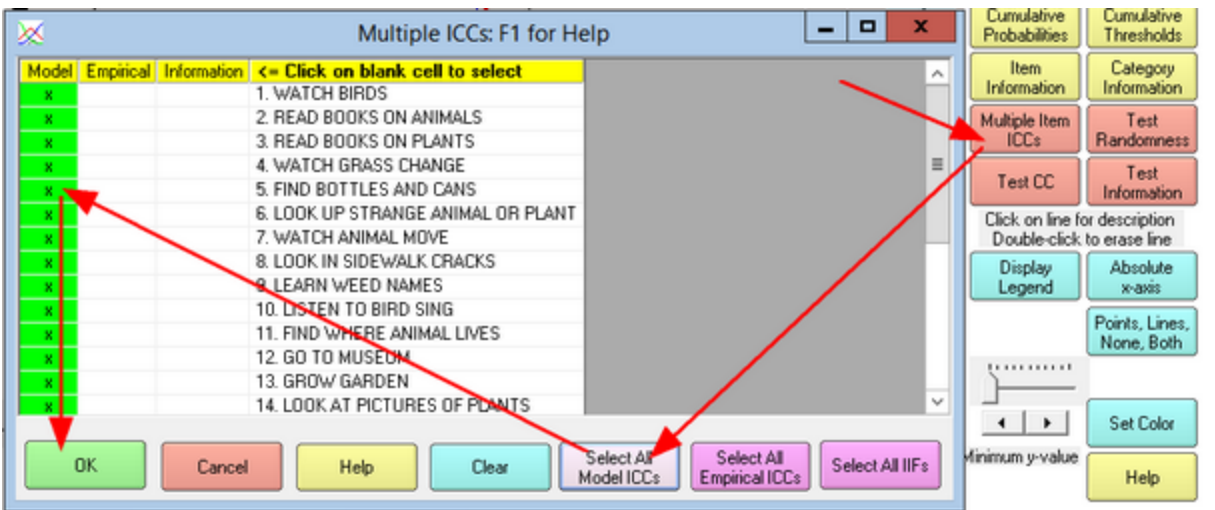


Buttons are described in [Graph window](#). The x-axis can use absolute scaling relative to the latent trait (as in this Figure) or scaling relative to each item difficulty.

Displayed are the selected model and empirical ICCs and IIFs. A similar technique is used to display [DIF item characteristic curves](#) for [non-uniform DIF](#).

Example: Partial credit ICC's - ISGROUPS=0

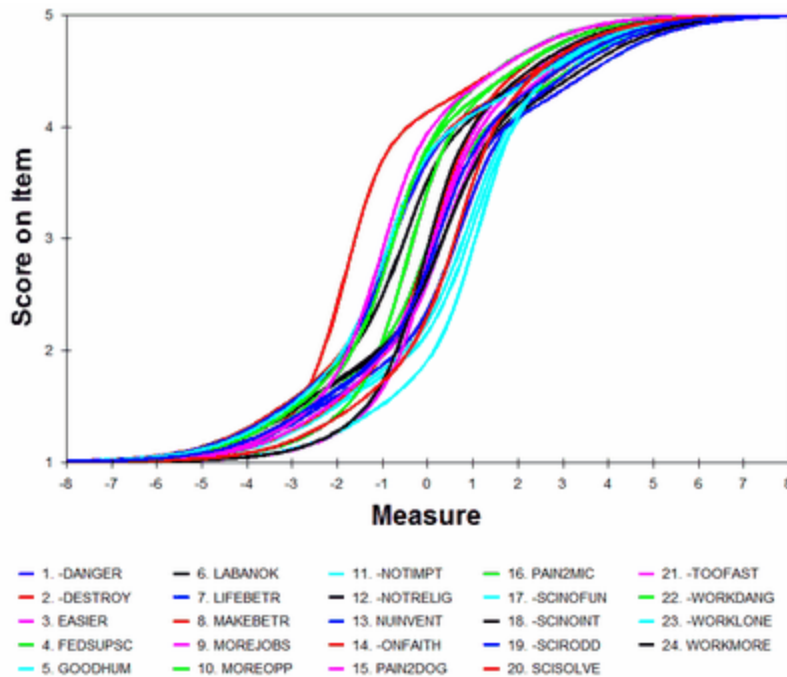
"Graphs" window, "Multiple-item ICCs", "Select All Model ICCs", OK



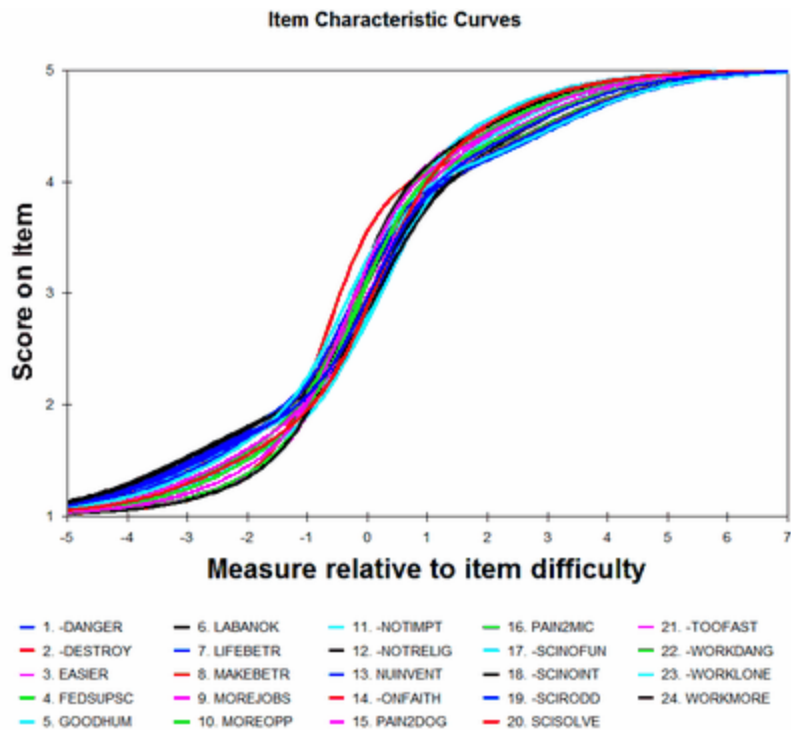
"Absolute x-axis": Model ICCs with absolute scaling relative to the latent variable = shown with item difficulties. The item difficulties spread out, but comparing the slopes of the ICCs is not easy.

When the curves are plotted "Relative x-axis" = "relative to item difficulty", they all center on zero. The curves squash together.

When the curves are plotted "Absolute x-axis" = "relative to the latent variable", each curve centers on its own item difficulty. The curves spread out.

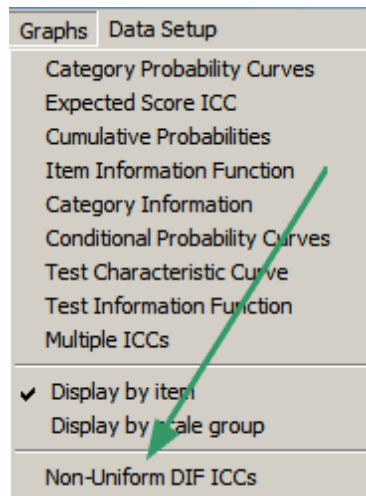


"Relative x-axis:" Model ICCs with relative scaling, relative to the item difficulties. The item difficulties are set to 0, so that comparing the slopes of the ICCs is easy, but comparing the difficulties of the items is impossible.



16.16 Non-uniform DIF ICCs

Non-uniform DIF ICCs (non-uniform Differential Item Functioning Item Characteristics Curves) are also called *Group Empirical IRFs* (Item Response Functions) and *Group Trace Lines*. These are displayed for items. To see non-uniform DIF ICCs, first [transpose](#) the data, then display these with the persons as items.



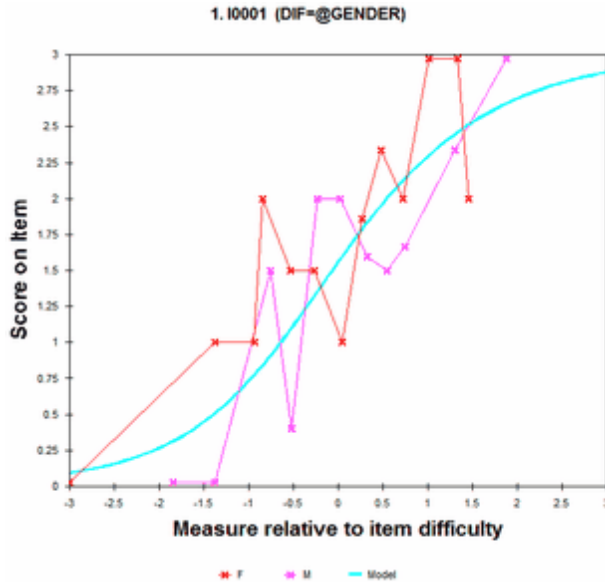
Click on "Display by item". Select the [Non-Uniform DIF](#) option on the [Graphs](#) menu.



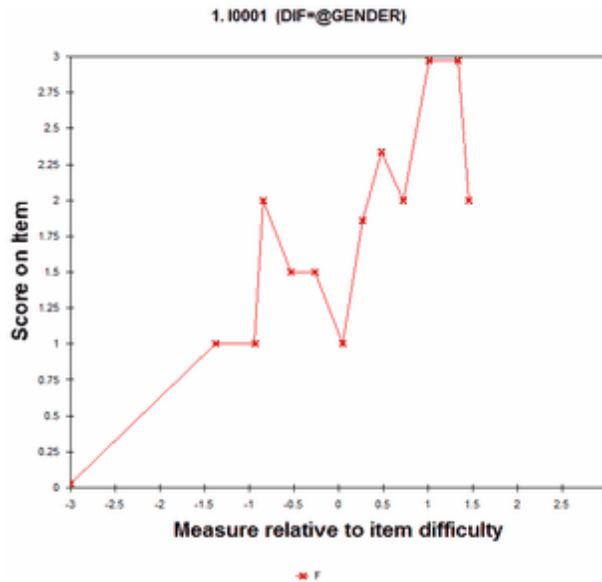
Confirm the columns in the person label of the [DIF=](#) group classification.

```
Building list of classification codes ...  
>-----<  
Collecting empirical data ...  
>-----<
```

Your screen reports that the DIF information is being collected from the data file.



The model ICC (green line) for the item and empirical ICCs for the two classification groups are shown (F=red and M=mauve lines). If you want a model ICC for each classification group, please use the [Split-items](#) procedure, and then [Multiple Item ICCs](#).



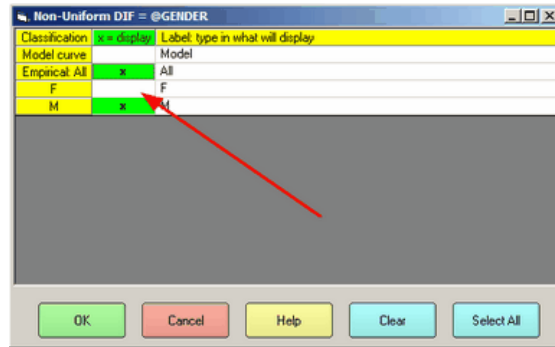
Double-click on unwanted lines to remove them. Here only the empirical ICC for the F group is shown.

Or click on Non-Uniform DIF button to select lines to display:



Buttons are described in [Graph window](#).

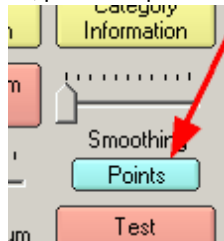
Click on box in the green column to select or deselect a family of lines for all DIF items



[Non-uniform DIF](#) can also be investigated numerically.

16.17 Points and lines

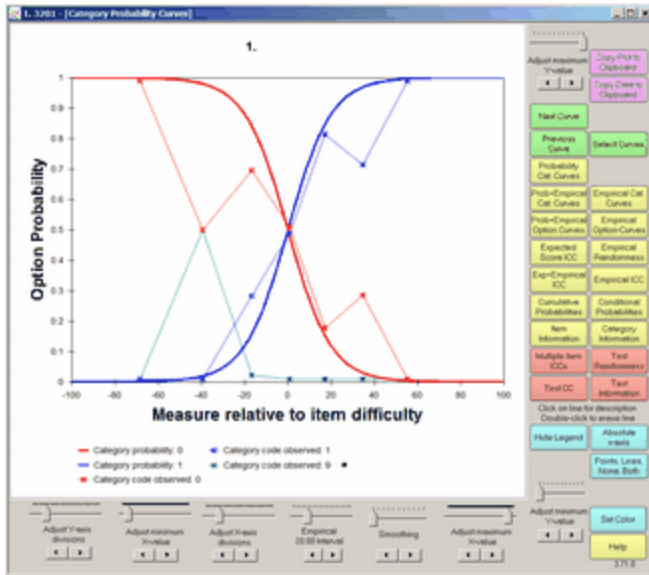
On the empirical curves you can control whether lines, points or points+lines are displayed.



By default, points+lines are displayed. Click on "Points" to display only points. Then click on it again to display only lines. Then click again for points+lines.

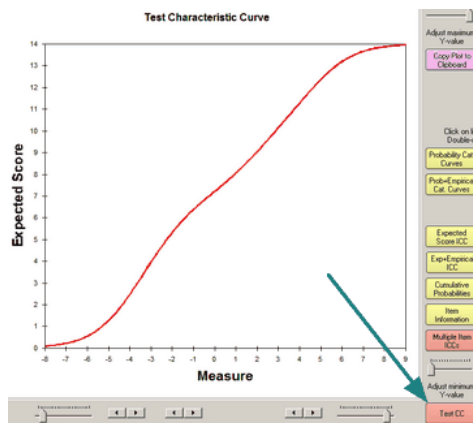
Buttons are described in [Graph window](#).

16.18 Category probability and empirical curves



Enter topic text here.

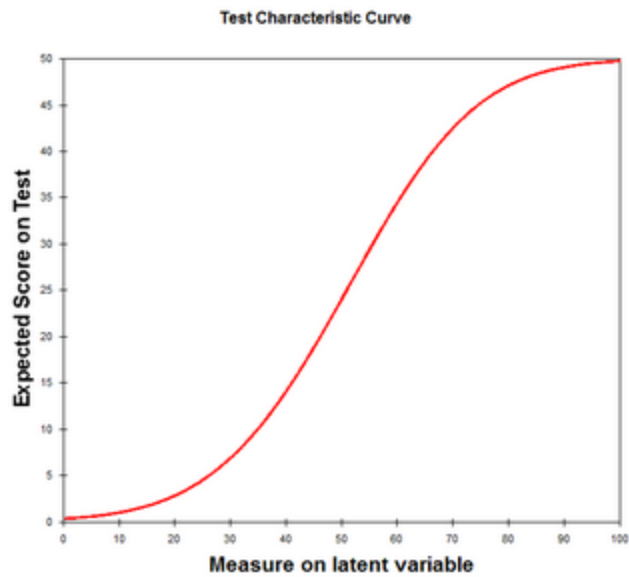
16.19 Test Characteristic Curve TCC



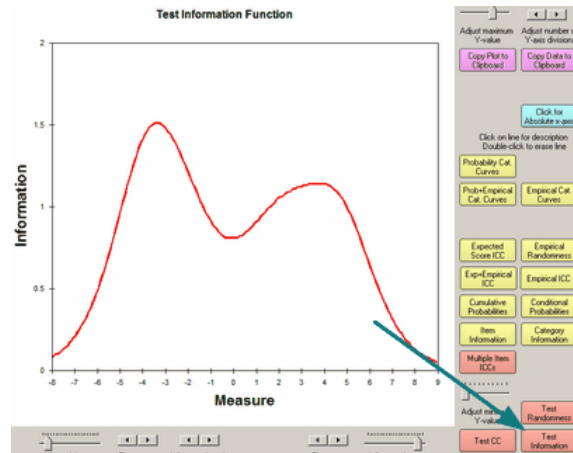
Select by clicking on "Test CC" or from the [Graphs](#) menu. This is the test characteristic curve (TCC), the score-to-measure ogive for this set of items. It is always monotonic ascending. See also [TCCFILE=](#) and [Table 20](#).

Buttons are described in [Graph window](#).

With 0-100 user-friendly rescaling using UIMEAN= and USCALE= from Table 20.1, the TCC for Example0.txt becomes this:



16.20 Test Information Function TIF



Select by clicking on "Test Information" or from the [Graphs](#) menu. This shows the Fisher information for the test (set of items) on each point along the latent variable. The test information function reports the "statistical information" in the data corresponding to each score or measure on the complete test. Since the standard error of the Rasch person measure (in logits) is $1/\sqrt{\text{information}}$, the units of test information would be in inverse-square-logits. See also [TCCFILE=](#) and [Table 20](#).

In practice, the values of the information function are largely ignored, and only its shape is considered.

We usually want the test information function to peak

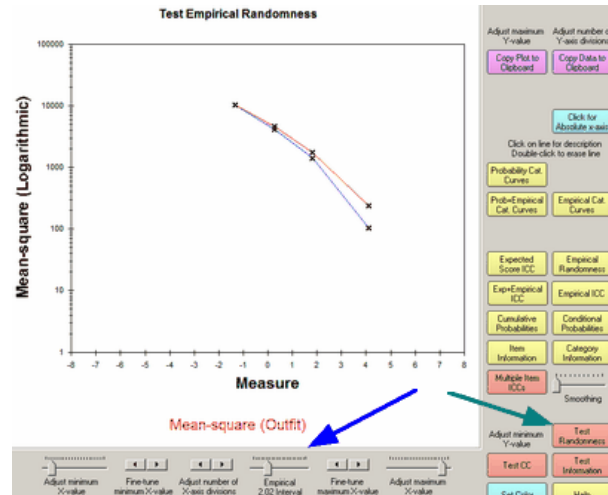
1. where the most important cut-point is (criterion-referenced tests)
2. where the mode of the sample is (norm-referenced tests)

The width of the information function is the effective measurement range of the test.

Buttons are described in [Graph window](#).

To increase the maximum possible Measure range, set [EXTREMESCORE=.01](#) or other small value.

16.21 Test Empirical Randomness



Select by clicking on "Test Randomness" or from the [Graphs](#) menu. Empirical intervals are set with the "Empirical Interval" slider. This displays the local value of the mean-square statistics. The Outfit mean-square statistic (standardized residual chi-square divided by its degrees of freedom) is the red line. The Infit mean-square statistic (ratio of observed to expected residual variance) is the blue line.

The Test Empirical Randomness shows the average mean-square for the subset of responses in a (person ability - item difficulty) interval on the latent variable. For instance,

- Are the noisy responses by people with high abilities relative to the item difficulties (= carelessness) = high mean-squares for high measures?
- Are the noisy responses by people with low abilities relative to the item difficulties (guessing) = high mean-squares for low measures?
- Are the responses too predictable for people with abilities far from the item difficulties (= Binet-style response imputation) = low mean-squares for high measures and for low measures?

In classical test theory, 3-PL IRT, and conventional statistics, low mean-squares are considered good. In Rasch theory, they indicate some redundancy in the responses, but they do no harm.

Buttons are described in [Graph window](#).

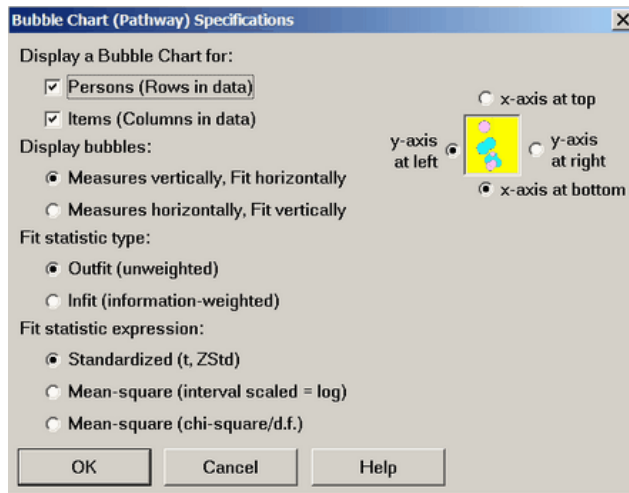
17 Plots

17.1 Bubble charts: Developmental pathways

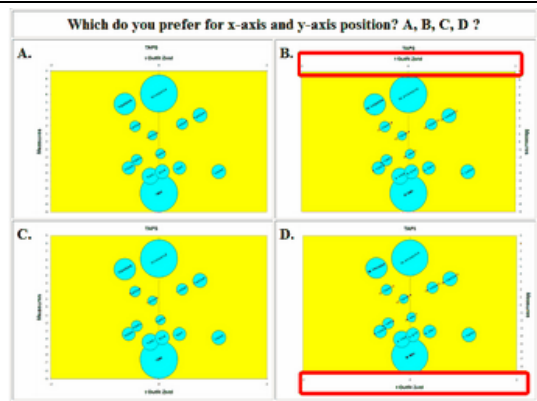
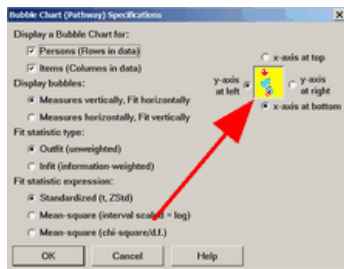
From the [Plots menu](#), Bubble charts show measures and fit values graphically. They are featured in [Bond & Fox](#). For successful operation, Excel must be available on your computer. Its maximum is 65,536 rows.

To produce these charts, Winsteps writes out the requisite values into a temporary file. Excel is then launched automatically. It reads in the temporary file and follows instructions to create a bubble plot. The plot is displayed, and the Excel worksheet becomes available for any alterations that are desired. The Excel worksheet may be made permanent using "Save As". The free Excel Add-in XY Chart Labeler from <http://www.appspro.com/Utilities/ChartLabeler.htm> is useful if you want to change the bubble labels.

Selecting "Bubble Chart" from the [Plots](#) menu:



1. **Positioning the axes and labels:** use the dialog box, or right-click on the axes and axes-titles to reposition them.



Bubble Resizing: The relative bubble sizes are set by the [standard errors](#) of the measures. But their overall absolute size is set by Excel. To change their overall sizes, right click on the edge of a circle.

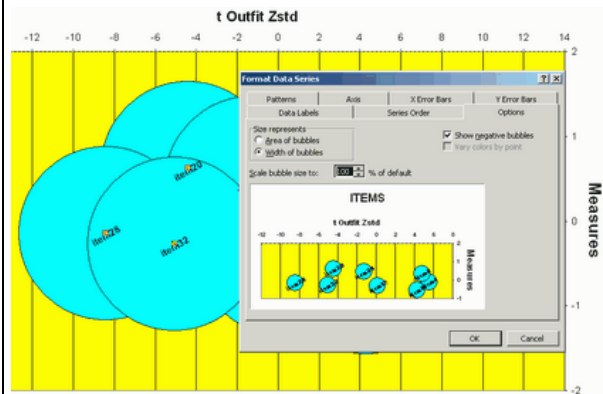
"Format Data Series"

"Options"

"Scale bubble size to:"

If the bubbles are roughly the same size, then their measures will be significantly different ($p \leq .05$) if they do not overlap along the measure axis when the bubble radius is $1.96 * 1.414 * S.E. / 2 = 1.4 * S.E.$

Please determine the actual bubble resizing by comparing the bubble size to the measure-axis scaling.



Bubbles showing Likert categories:

Likert data is shown in vertical bubbles -- such in Bond and Fox (2007) Fig 6.2 p.104 -- but this is not yet produced by Winsteps automatically. Please produce an item bubble-chart. Then, in its Excel worksheet, edit in extra lines for each item. Use values from [ISFILE](#) for the item difficulties.

	A1			ENTRY
	A	B	C	D
1	ENTRY	STATUS	MAX	CAT
2	1	-11	0	0
3	2	-11	0	0
4	3	-11	0	0
5	4	-9	0	0

Autoshape the bubbles:

Excel 2007 and later:

Let us change the shapes of the bubbles for persons and items.

This may take several attempts to produce the bubbles you want.

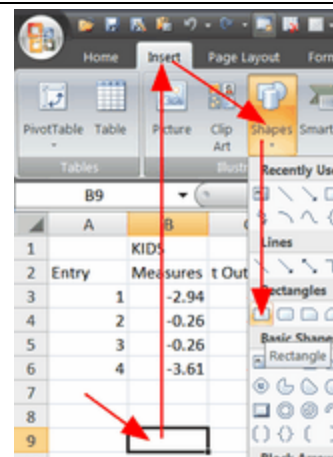


In Excel 2007,

In your worksheet,

Click on an empty cell

Insert
Shapes
Rectangle



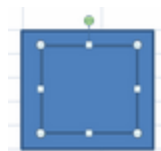
Drag your mouse to draw a square



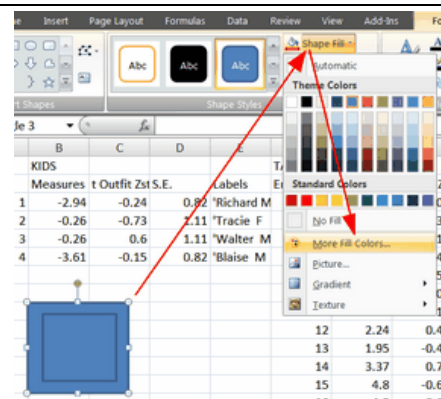
Insert
Shapes
Rectangle (or any other shape)

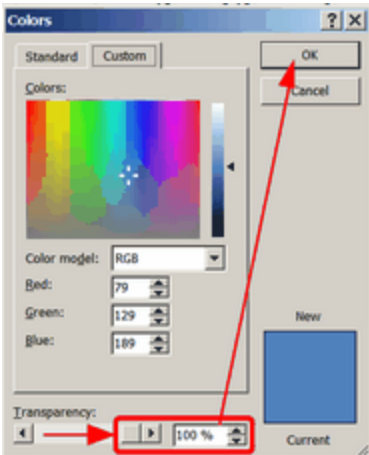
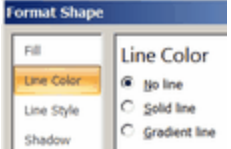

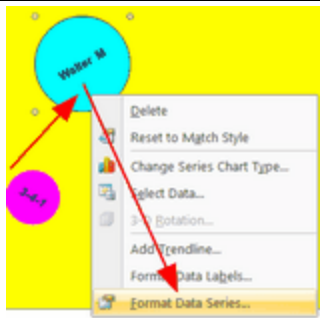
Drag your mouse to draw a square inside the other square

Right-click the new square
Format Shape
Fill
Color (your choice of color)



2. Click on the outer square to select it.
"Shape fill"
"More fill colors"

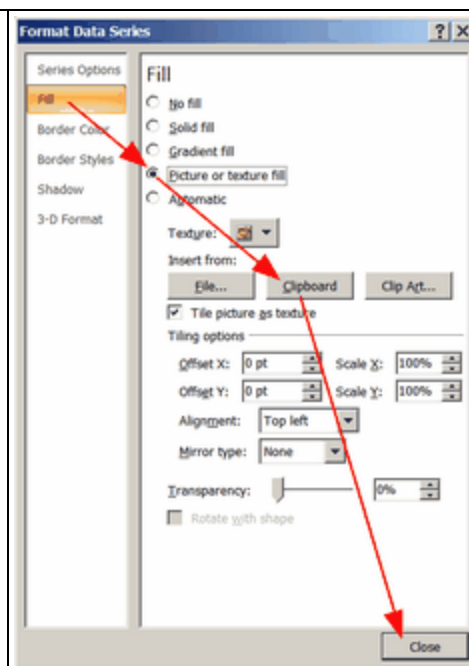


<p>"Colors" Transparency" "100%" OK</p>	
<p>Right-click on box Format shape Line color No line (the box disappears)</p>	
<p>Click on one box ctrl+click on the other box ctrl+c to copy</p>	
<p>On the bubble chart Right-click a child bubble Format data series</p>	

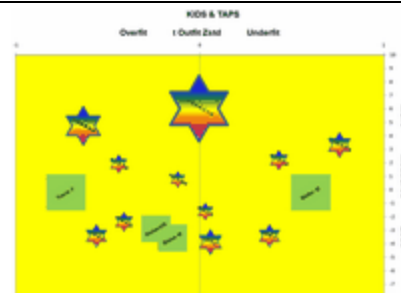
Format Data Series
 Border Color
 No line

Fill
 Picture
 Clipboard
 Close

The bubbles should now look like your colored rectangle.



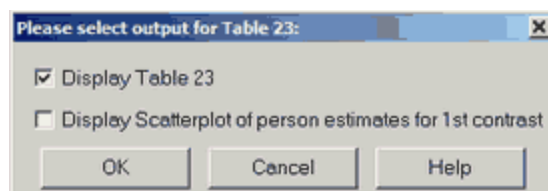
This is a little more creative, using Star shapes and gradient-color fill.



17.2 Cluster Measure Plot - Table 23.6

From the [Plots menu](#), Cluster Measure Plot produces the plotted output corresponding to [Table 23.6](#). It shows the person measures for each cluster of items (perhaps belonging to different dimensions) in [Table 23.1](#).

From the Output Tables or the Plots menu, these dialogs are displayed:



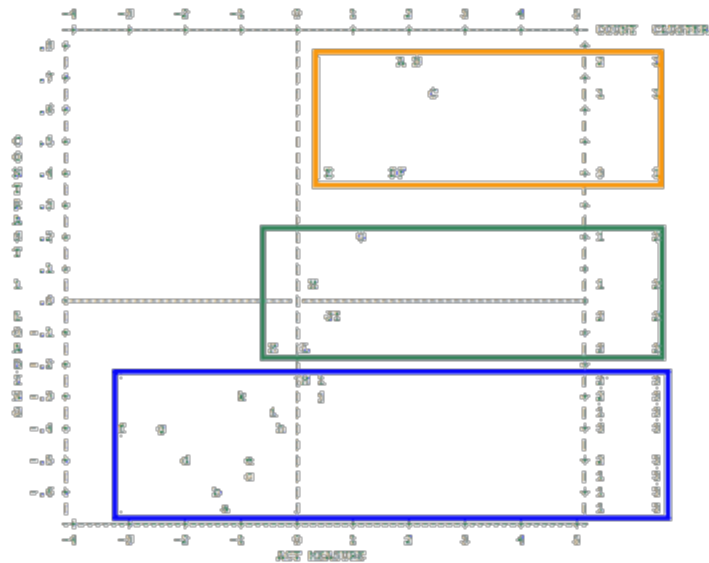
Choose whether to output Table 23 and/or the Excel plot of Table 23.6



When the Excel plot is chosen, then select what [data point labels](#) are to be displayed on the Excel plot. In this case, the person entry number.

Here is Table 23.1

3 clusters of items are identified vertically. The items at the top contrast in some way with the items at the bottom. They could be on different dimensions or strands or item formats or



Here is Table 23.6

This reports the measures for each person on each cluster of items. If the clusters are different dimensions, then we expect the person measures to be considerably different for the different clusters.

TABLE 23.6 LIKING FOR SCIENCE (Wright & Masters ZOU550WS.TXT Apr 1 2015 10:55
 INPUT: 79 KID 25 ACT REPORTED: 77 KID 25 ACT 6 CATS WINSTEPS 3.90.0

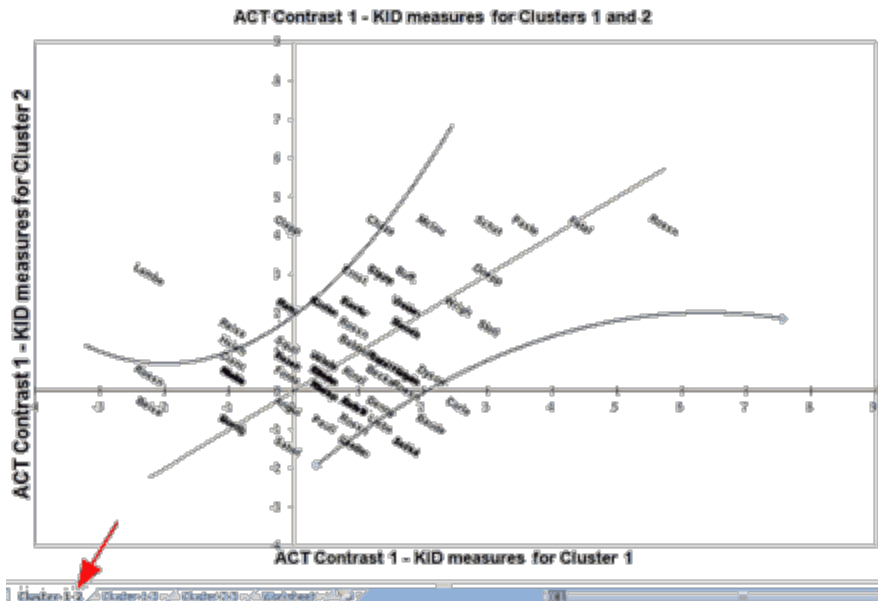
KID	STATUS	CONTRAST 1 CLUSTER MEASURES						KID LABEL
		CLUSTER 1		CLUSTER 2		CLUSTER 3		
		MEASURE	S.E.	MEASURE	S.E.	MEASURE	S.E.	
5	1	.48	.71	.00	.63	1.09	.59	M Rossner, Marc Daniel
6	1	1.35	.63	.77	.62	1.14	.58	M Rossner, Toby G.
7	1	.48	.71	.38	.62	.07	.47	M Rossner, Michael T.
8	1	.94	.66	-.87	.71	-1.15	.45	F Rossner, Rebecca A.

Here is the Excel plot for Clusters 1 and 2.

The person measures, estimated for each cluster of item, are cross-plotted. 95% Confidence Bands are drawn based on the S.E.s of the measures. The confidence bands are averaged across the persons. The diagonal line on the plot is the identity line. All the points in the plot would fall on this line if the person measures on both clusters were the same.

If almost all the points are within the confidence bands, then the two clusters of items are not measuring the persons in a statistically different way. However, individual persons outside the confidence bands do have statistically different performance levels on the clusters.

Each plotted point has its own confidence interval. These are in the Worksheet Column (U,V) and (W,X). Excel plots smooth curves (trend lines) through the end-points of all these intervals using quadratic (2) polynomials. Please right click on the curves to change the polynomials.

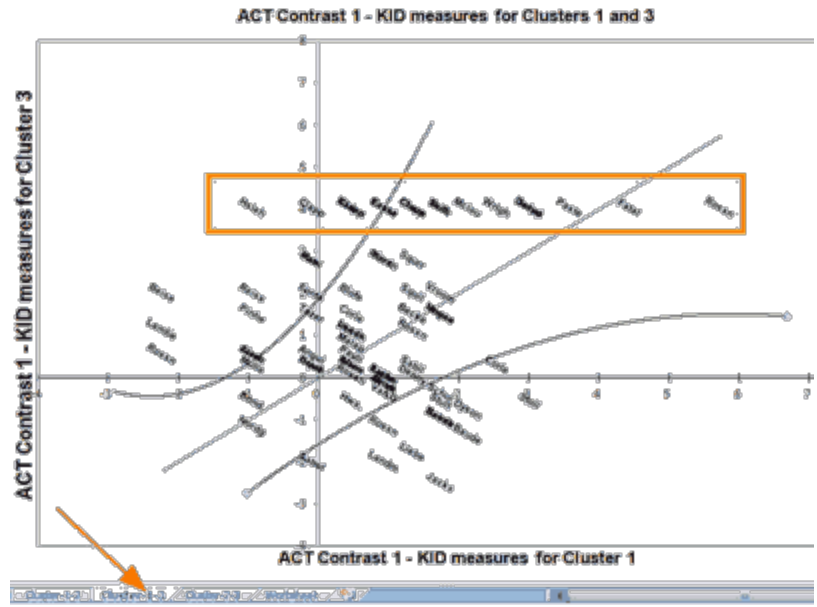


For some plots, quartic (4) polynomials may be better.

Here is the Excel plot for Clusters 1 and 3.

The items in Clusters 1 and 3 are the most different, so we expect that their person measures will also be the most different. Here we see more persons outside the confidence bands, but also many within.

In this plot, the persons in the orange box have maximum scores on the items in Cluster 3, but different score on the items in Cluster 1.



Cluster Measure Worksheet

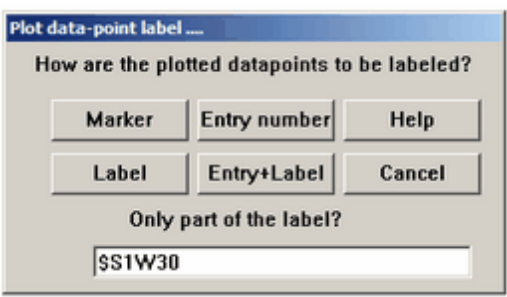
Columns A-I correspond to Table 23.6.

Columns J-X are for drawing the confidence bands. For the formulas for these columns, see [Scatterplot](#).

	A	B	C	D	E	F	G	H	I	J	K	L
1	Table 23.6	ACT Contr	KID Cluste	LIKING FC		4/1/2015	10:55					
2	KID	Status	3 Measure	1 S.E.	2 Measure	2 S.E.	3 Measure	3 S.E.	KID Label	12 Cl.	12 Upper	12 Upper
3	10	1	0.4763	0.7079	-0.0021	0.6273	1.0873	0.3893	Fossen	0.9271	-0.69	1.36
4	11	1	1.3489	0.6309	0.766	0.6227	1.141	0.5829	Fossen	0.8687	0.1892	1.92
5	12	1	0.476	0.7081	0.383	0.6172	0.0731	0.4726	Fossen	0.9206	-0.4911	1.35
6	13	1	0.9378	0.6361	-0.874	0.7088	-1.1514	0.4486	Fossen	0.9473	-0.9154	0.97
7	14	1	1.7432	0.6237	-0.0021	0.6273	-0.952	0.4448	Fossen	0.867	0.0033	1.73
8	15	1	2.5481	0.658	2.1651	0.7997	4.0406	1.8239	Whigh	1.0149	1.3417	3.37
9	16	1	-2.2291	1.8424	2.9924	1.0626	1.1419	0.5837	Lambe	2.0843	-1.7027	3.4
10	17	1	1.742	0.6237	-1.4467	0.8243	-0.3331	0.4531	Schul	1.0083	-0.8377	1.35
11	18	1	-0.0919	0.8114	0.7651	0.6223	0.3031	0.4382	Hsieh	1.0023	-0.8657	1.33
12	19	1	-0.9408	1.0737	1.1668	0.6464	4.0406	1.8239	Hsieh	1.2282	-1.1152	1.94
13	20	1	1.3489	0.6309	-0.874	0.7088	-1.7717	0.4667	Lieber	0.9307	-0.6927	1.36
14	21	1	0.9378	0.6361	-1.4467	0.8243	0.0732	0.4726	Lieber	1.025	-1.2794	0.77
15	22	1	0.476	0.7081	-0.0021	0.6273	-0.5353	0.448	Hwa	0.9272	-0.6903	1.36
16	23	1	2.1346	0.632	0.3834	0.6172	-0.753	0.4452	Dyson	0.8658	0.3932	2.12
17	24	1	1.7417	0.6237	2.9924	1.0626	4.0406	1.8239	Buff	1.2075	1.1595	3.57

17.3 Datapoint labels

Several [plots](#) have options as to how the points are to be labeled.



Click on:

Marker: plotted point is indicated with a symbol, such as ♦

Entry number: plotted point is indicated with its item or person entry number

Label: plotted point is indicated with its item or person label

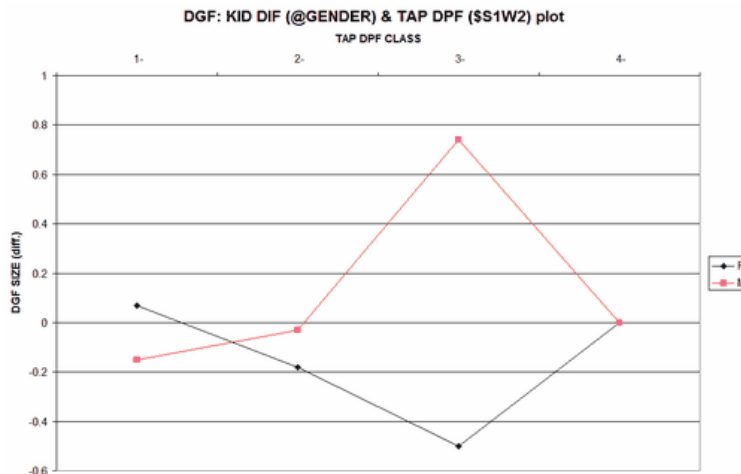
Entry+Label: plotted point is indicated by its entry number and its label

Only part of label? Enter here instructions to select only part of the person or item label, e.g., 2W4 means "starting with column 2 of the label with a width of 4 characters." Then click on "Label"

17.4 DGF DIF-DPF Plot - Table 33.3

From the [Plots menu](#), DGF- DIF-DPF Plot 33 produces the plotted output corresponding to [Table 33.3](#). It shows Differential Group Functioning (DGF) between classes of items and classes of persons.

The [DGF dialog](#) displays when Table 33 is called from the [Output Tables](#) or [Plots](#) menus.



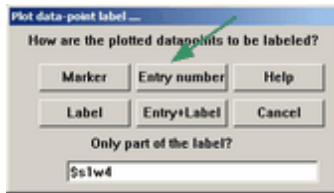
There are six Excel plots. You have the full Excel capabilities to edit the plot, and the Excel worksheet contains the plotted numbers.

DGF SIZE (diff.)-ip = item-person	reports the relative DGF difficulty size of the item classes as points and the person classes as columns
DGF t-value (diff.)-ip	tests the hypothesis that the DIF sizes can be attributed to chance: $t\text{-value} = \text{DGF size} / \text{DGF size S.E.}$ This displays the item classes as points and the person classes as columns
DGF Average Score-points-ip	displays the average difference between the observed score and the expected score for each response on this item CLASS by this person CLASS
DGF SIZE (diff.)-pi = person-item	reports the relative DGF difficulty size of the person classes as points and the item classes as columns
DGF t-value (diff.)-pi	tests the hypothesis that the DIF sizes can be attributed to chance: $t\text{-value} = \text{DGF size} / \text{DGF size S.E.}$ This displays the person classes as points and the item classes as columns
DGF Average Score-points-pi	displays the average difference between the observed response and the expected response for each person CLASS on each item CLASS

17.5 DIF Plot - Table 30.2

From the [Plots menu](#), DIF Plot 30 produces the plotted output corresponding to [Table 30.2](#). It shows differential item functioning (DIF).

From the Output Tables menu, the [DIF dialog](#) is displayed.



Select what [data point labels](#) are to be displayed on the Excel plot. In this case, the item entry number. The person classifications will be identified by their column codes. There are four Excel plots. You have the full Excel capabilities to edit the plot, and the Excel worksheet contains the plotted numbers. The Excel worksheets contain the plotted numbers.

DIF Measure (diff.)

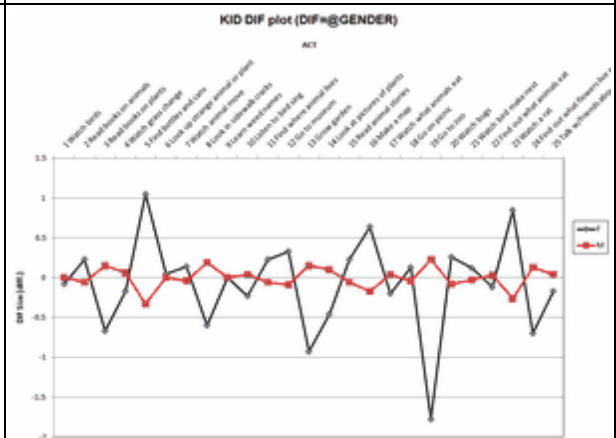
reports the difficulty = (diff.) of the item for each person on classification: DIF "local absolute" Measure = DIF "relative" Size + Overall "baseline" item difficulty

* indicates the baseline measures (no DIF), corresponding to the zero line (x-axis) in the DIF Size plot.



DIF Size (diff.)

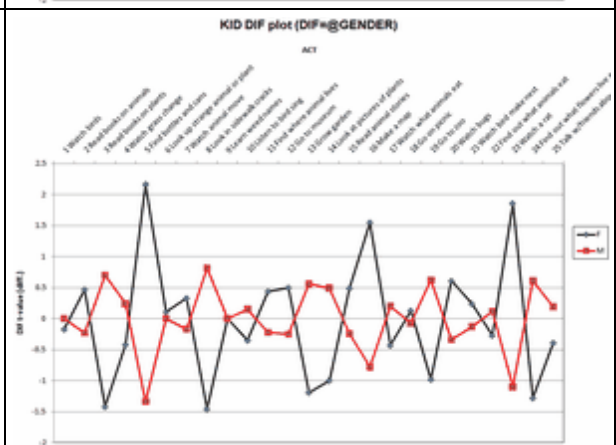
reports the size of the item DIF for the person-classification relative to the overall "baseline" item difficulty.



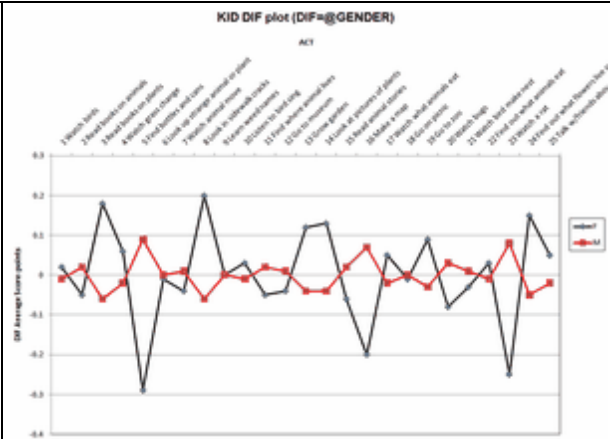
DIF t-value (diff.)

reports a simple approximate t-test of the item DIF against the overall item difficulty. The t-statistic is the statistical probability of the DIF size relative to its measurement error expressed as an approximate unit-normal deviate. A critical values is usually ± 2 . This tests the hypothesis that the DIF size can be attributed to measurement error.

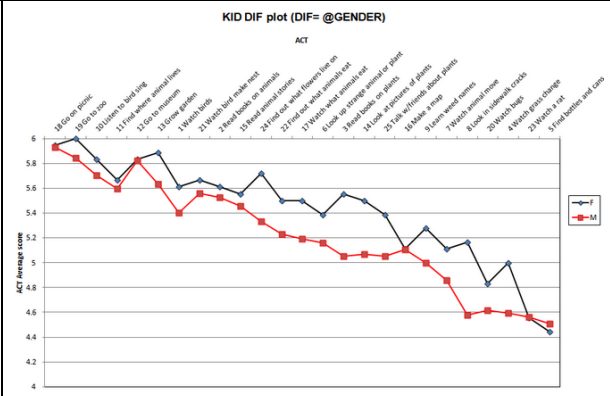
For a test of one group against another group, the joint t-value is usually around 70% of the difference between their t-values on the plot.



DIF Average Score-point Difference
 displays the average difference between the observed response and the expected response for each person CLASS on each item.



DIF Average Score-points
 displays the average of the observed responses for each person CLASS on each item. Here the items are in Measure order: easiest items to hardest items.

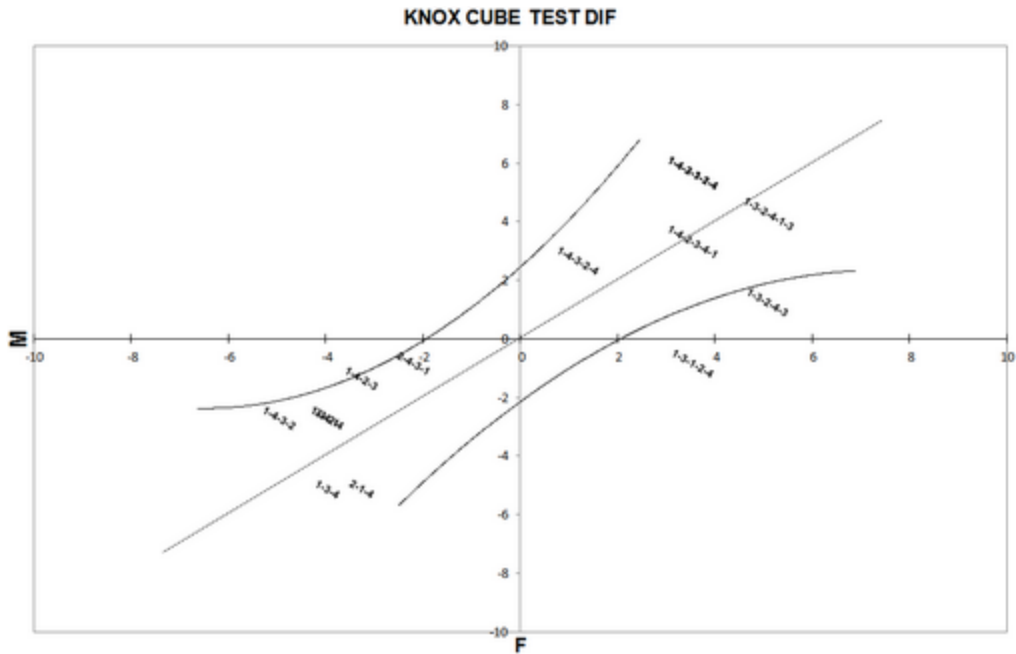


DIF Worksheet
 Sort the section of the worksheet to change the left-to-right order of the items in the plots.
 Edit the worksheet to change the words shown in the plots.
 * is the baseline column for "everyone". These are the item difficulties on [Table 14](#).

	A	B	C	D	E	F
1	KID DIF plc	LIKING FO		12/26/2010 16:34	4	
2	DIF Measu			KID DIF CLASS		
3	Entry	ACT	Entry+ACT	F	M	*
4		1 Watch birc	1 Watch bi	-0.48	-0.4	-0.4
5		2 Read book 2	2 Read boi	-0.48	-0.76	-0.71
6		3 Read book 3	3 Read boi	-0.25	0.57	0.42
7		4 Watch gra 4	4 Watch gy	1.57	1.81	1.75
8		5 Find bottle 5	5 Find bott	3.47	2.09	2.42
9		6 Look up st 6	6 Look up	0.36	0.31	0.31
10		7 Watch ani 7	7 Watch ai	1.24	1.07	1.1
11		8 Look in sid 8	8 Look in s	1.07	1.86	1.67
12		9 Learn wee 9	9 Learn wi	0.71	0.71	0.71

17.6 DIF/DPF/DTF Scatterplots of paired measures - Table

The DIF/DPF measure pairs in [Table 30.1, 31.1](#), are [scatterplotted](#). The confidence bands and empirical slope (correlation) are also shown.

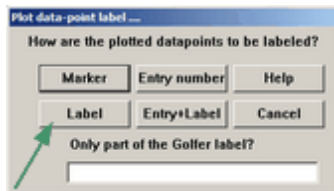


17.7 DPF Plot - Table 31.2

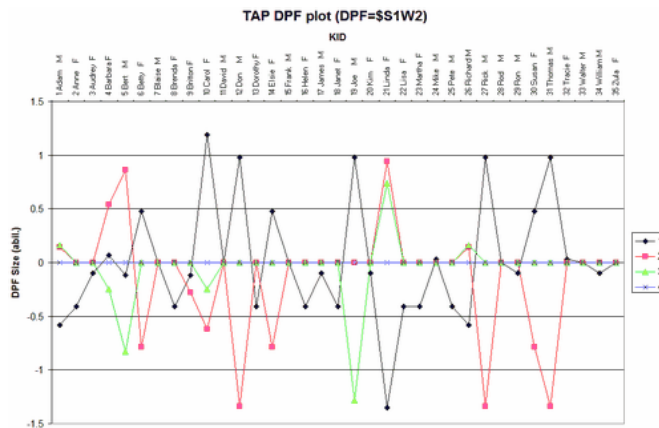
From the [Plots menu](#), DPF Plot 31 produces the plotted output corresponding to [Table 31.2](#). It shows differential person on functioning (DPF).

From the Output Tables menu, the [DPF dialog](#) is displayed.

In the [DPF=](#) box specify the column in the item label that identifies the DPF classification for each person.



Select what [data point labels](#) are to be displayed on the Excel plot. In this case, the person label. The item classifications will be identified by their column codes. You have the full Excel capabilities to edit the plot, and the Excel worksheet contains the plotted numbers. The plotted numbers are contained in the Excel worksheet.



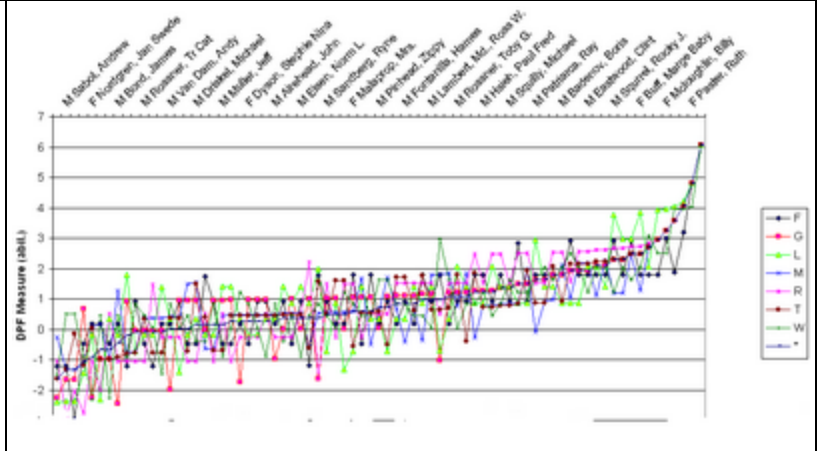
There five Excel plots:

DPF Measure (abil.) "***" indicates the baseline measures (no DPF) , corresponding to the zero line (x-axis) in the DPF Size plot.	reports the ability of the person for each item classification: DPF "local absolute" Measure = DPF "relative" Size + Overall "baseline" person ability
DPF Size (abil.)	reports the size of the person DPF for the item classification relative to the overall "baseline" person ability
DPF t-value (abil.)	reports a simple approximate t-test of the person DPF against overall ability. The t-statistic is the statistical probability of the DPF size relative to its measurement error expressed as an approximate unit-normal deviate. A critical values is usually ± 2 . This tests the hypothesis that the DPF size can be attributed to measurement error.
DPF Average Score-point difference	displays the average difference between the observed response and the expected response for each person on each item CLASS.
DPF Average Score-points	displays the average of the observed responses for each person on each item CLASS.
Worksheet	contains the numbers that are plotted. Sort the section of the worksheet to change the left-to-right order of the persons in the plots. Edit the worksheet to change the words shown in the plots.

Example: In the Excel worksheet, sort the DPF Measure block, Ascending on overall measure (the * column)

	A	B	C	D	E	F	G	H	I	J	K
1	ACT DPF	LUKING FC		#####	4						
2	DPF Meas			ACT DPF CLASS							
3	Entry	KID	Entry+KID	F	G	L	M	R	T	W	*
4	53 M Sabol	453 M Sabc		-1.22	-2.24	-2.39	-0.25	-1.05	-1.61	-1.48	-1.61
5	29 M Landma	29 M Land		-1.22	-1.64	-2.37	-1.32	-2.83	-1.32	0.52	-1.32
6	72 M Jackson	72 M Jack		-2.93	-1.64	-2.37	-1.32	-2.03	-0.14	0.52	-1.32
7	30 F Nordgrer	30 F Nordg		-0.48	0.68	-1.42	-1.14	-2.75	-1.18	-0.88	-1.05
8	12 M Lieberm	12 M Liebe		0.2	-2.25	-0.16	0.08	-1.05	0.04	-2.31	-0.92
9	5 F Rossner	5 F Rossn		0.2	-0.96	-2.29	0.19	-2	-1	0.46	-0.67
10	68 M Bond	468 M Bond		-0.48	-0.96	0.36	-0.96	0.52	-1	-2.24	-0.67
11	49 M Baudet	49 M Bauc		0.2	-2.43	-0.16	1.28	-1.05	-0.9	0.46	-0.43
12	14 F Hwa, Na	14 F Hwa,		-1.22	0.93	1.8	-0.76	-1.05	-0.81	-0.87	-0.2
13	6 M Rossnei	6 M Rossr		0.93	-0.02	-0.73	-0.72	-1.05	-0.76	0.85	-0.08
14	9 M Schutz,	9 M S-hub		-0.48	-0.02	0.16	0.91	1.05	0.34	0.46	-0.08

Then look at the DPF measure plot, the persons are ordered by performance on the items, revealing any trends.



17.8 Incremental person measures - Table 45

See [Table 45](#)

17.9 Keyform plot

See also [Tables 37, 38, 39, 40, 41](#) - Person Keyforms

From the [Plots menu](#), Keyforms are self-measuring and diagnosing forms, such as the KeyMath diagnostic profile, and the KeyFIM. These are a powerful application of Rasch measurement for instantaneous use. These are plotted in horizontal or

vertical format using Excel - but be patient, Excel is somewhat slow to display them. The Keyform can be plotted with either horizontal or vertical orientation. In earlier version of Winsteps, these were specified by KEYFORM=

For an explanation of Keyforms, please see www.rasch.org/memo60.htm - "Instantaneous Measurement and Diagnosis".

The 7 columns in the Excel Worksheet are:

For points in the Keyform:

COLUMN The horizontal location (x-axis) in the vertical layout or vertical location (y-axis) in the horizontal layout.

MEASURE The measure (y-axis) in the vertical layout or (x-axis) in the horizontal

POINT-LABEL The value with which to label the point. Use the Excel point-label add-in at www.winsteps.com/ministep.htm

For column (row headings)

COLUMN The horizontal location (x-axis) in the vertical layout or vertical location (y-axis) in the horizontal layout.

HEAD-MEASURE The top-of-column measure (y-axis) in the vertical layout or end-of-row (x-axis) in the horizontal

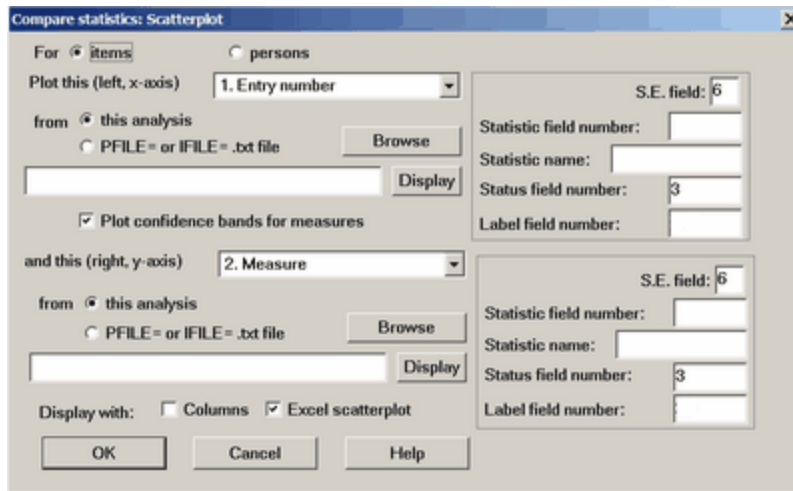
ITEM-ID The item number

ITEM-LABEL The item identifying label

Example: For the first 3 items of the "Liking For Science" Data with Excel, produces vertical plots like: (These can be plotted directly from the [Plots](#) menu.)

	WATCH BIRDS	READ BOOKS ON ANIMALS	READ BOOKS ON PLANTS	Raw Score	Measure	S.E.
		2		6	5	2.02
		-			4	1.38
2	2			5	3	1.38
-	-	1		4	2	1.22
					1	1.22
1	1	-		3	0	1.28
		0		2	-1	1.28
-	-				-2	1.28
0	0			1	-3	1.32
					-4	1.32
				0	-5	1.97

with Excel, produces horizontal plots like: (These can be plotted directly from the [Plots](#) menu.)



Measures, standard errors, fit statistics indicate which statistics are to be compared.

Display with columns generates a line-printer graphical-columns plot. It is displayed as [Table 34.](#)

The first column is the Outfit Mean-Square of this analysis.

The third column is the Outfit Mean-Square of the Right File (exam12lop.txt in this case)

The second column is the difference.

The fourth column is the identification, according to the current analysis.

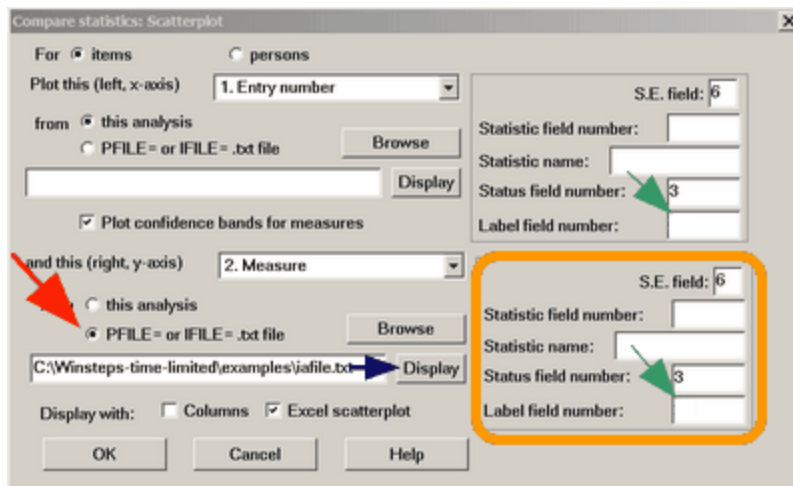
Persons or items are matched and listed by Entry number.

Table 34.1

PERSON	Outfit MnSq	Difference	exam12lop.txt	File	Compa
1	3	-2	0	2	1
NUM	LABEL				
1	*	*	*	1	21101
2	*	*	*	2	21170
3	*	*	*	3	21174
...					
35	*	*	*	35	22693

Display with Excel scatterplot initiates a graphical scatterplot plot. If the statistics being compared are both measures, then a 95% confidence interval is shown. This plot can be edited with all Excel tools.

Comparing with files



One or both sets of statistics can be in a [IFILE=](#) or [PFILE=](#) file (red arrow). Since these files can have different formats, please check that the selected field number matches the correct field in your file by clicking on the Display button (blue arrow). This displays the file. Count across the fields to your selected statistic. If your field number differs from the standard field number, please provide the correct details for your field in the selection box (orange box).

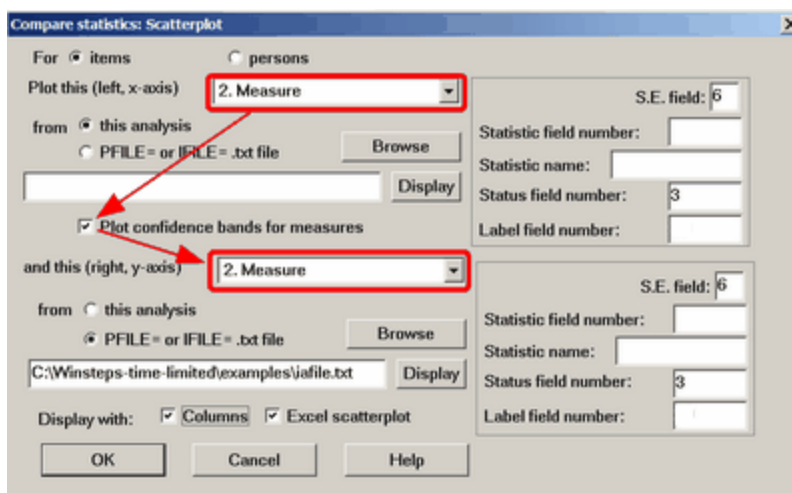
If two files are cross-plotted, please enter the Label field number in one of the files (either of the green arrows). Click on Display for one of the files, and count across the fields to the Label field.

There are several decisions to make:

1. Do you want to plot person (row) or item (column) statistics?
2. Which statistic for the x-axis?
3. Which statistic for the y-axis?
4. Do you want to use the statistic from this analysis or from the [PFILE=](#) or [IFILE=](#) of another analysis?
5. Do you want to display in the statistics as Columns in a Table or as an Excel scatterplot or both?

If you are using the statistic from a [PFILE=](#) or [IFILE=](#) and Winsteps selects the wrong column, then identify the correct column using the "Statistic field number" area.

When two measures are compared, then their standard errors are used to construct confidence bands when "Plot confidence bands" is checked:



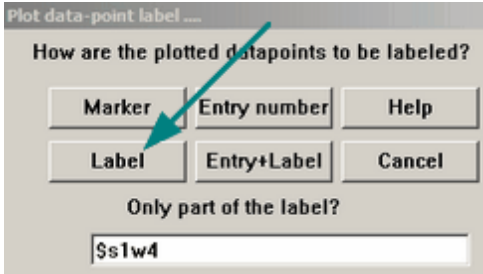
Here the item calibrations in the current analysis are being compared with the item calibrations in file [IFILE=SFIF.txt](#) from another analysis. This is the columnar output:

TABLE 34.1 An MCQ Test: administration was Comput ZOU630WS.TXT Apr 21 2:21 2006
 INPUT: 30 STUDENTS 69 TOPICS REPORTED: 30 STUDENTS 69 TOPICS 2 CATS 3.60.2

Measures		Differences		Measures		Comparison	
				SFIF.txt			
-4	1 -2	5 -3	2	NUM	TOPIC		
*	.	.	*	*	.	1	n101 Month
.	*	*	.	*	.	2	n102 Sign
*	.	.	*	*	.	3	n103 Phone number
*	.	.	*	*	.	4	n104 Ticket
*	.	*	*	*	.	5	n105 building
*	.	*	*	*	.	6	nm01 student ticket
*	.	*	*	*	.	7	nm02 menu
*	.	*	*	*	.	8	nm03 sweater
*	.	*	*	*	.	9	nm04 Forbidden City
*	.	*	*	*	.	10	nm05 public place
*	.	*	*	*	.	11	nm06 post office
*	.	*	*	*	.	12	nm07 sign on wall

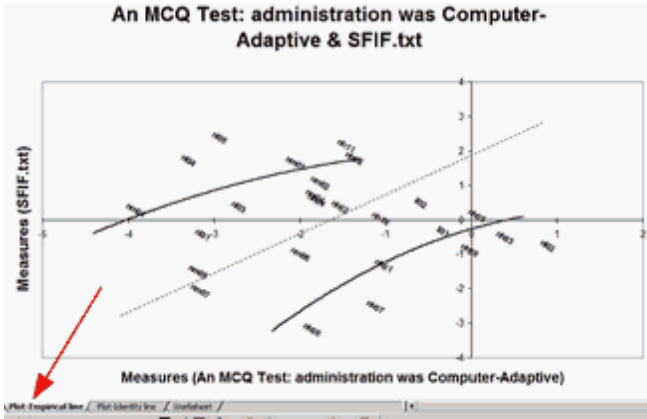
*	.	*	*	.	13	nh01	supermarket
*	.	*	*	*	14	nh02	advertisement
*	.	*	*	*	15	nh03	vending machine
*	.	*	*	*	16	nh04	outside store
*	.	*	*	*	17	nh05	stairway
*	.	*	*	*	18	nh06	gas station
*	.	*	*	*	19	nh07	Taipei
*	.	*	*	*	20	nh08	window at post office
*	.	*	*	*	21	nh09	weather forecast
*	.	*	*	*	22	nh10	section of newspaper
*	.	*	*	*	23	nh11	exchange rate
*	.	*	*	*	24	il01	open
*	.	*	*	*	25	il02	vending machine

and the plotted output:

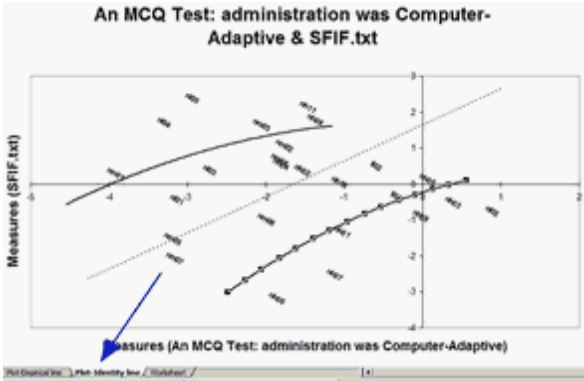


We are selecting only the first 4 characters of the item label, e.g., "nl01" and plotting only the Label:

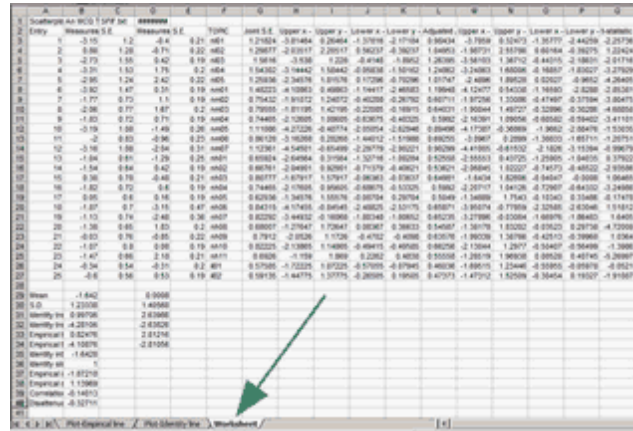
1. Plots with confidence bands:



The points are plotted by their labels by [Excel](#). The curved lines are the approximate 95% two-sided confidence bands (smoothed across all the points). They are not straight because the standard errors of the points differ. In this plot called "Plot-Empirical line" (red arrow), the dotted line is the empirical equivalence (best-fit) line. Right-click on a line to reformat or remove it.



A line parallel to the identity line is shown on the "Plot-Identity line" (blue arrow) by selecting the tab on the bottom of the Excel screen. This line is parallel to the standard identity line (of slope 1), but goes through the origin of the two axes. This parallel-identity line goes through the mean of the two sets of measures (vertical and horizontal).



The plotted points are in the Excel Worksheet (green arrow). You can edit the data points and make any other changes you want to the plots.

Cell and Column Descriptions for Scatterplots of Measures with Confidence Bands		
Cell	Description	
A1	Scatterplot	
B1	TITLE=	
D1	Date and time	
F1	CI= (Confidence Interval is)	
G1	1.96 (normal deviate for 95% 2-sided confidence bands)	
H1	68%=1.00, 90%=1.65, 95%=1.96, 99%=2.58 (Typical normal deviates for 2-sided confidence bands)	
B22 (or similar), B23	Mean of Measure 1 in Column B and its population S.D.	
D22 (or similar), D23	Mean of Measure 2 in Column D and its population S.D.	
Column	Meaning	Formula for Row B
A	Entry	Entry number of Person or Item
B	Measure 1	Measure on y-axis
C	S.E. 1	Standard Error of Measure in column B
D	Measure 2	Measure on y-axis
E	S.E. 2	Standard Error of Measure in column D
F	PERSON or ITEM	Person or Item Label
G	C.I. - Identity (for Identity-line Confidence Band)	$=\text{SQRT}(C3^2+E3^2)*G1*0.5$
H	Upper x - Identity (for upper Confidence Band on Identity-line plot)	$=((B22+B3+D3-D22)/2-G3)$
I	Upper y - Identity	$=((D22+B3+D3-B22)/2+G3)$

J	Lower x - Identity (for lower Confidence Band on Identity-line plot)	$=((B22+B3+D3-D22)/2+G3)$
K	Lower y - Identity	$=((D22+B3+D3-B22)/2-G3)$
L	C.I. - Empirical (for Empirical-line Confidence Band)	$=\text{SQRT}((C3/B23)^2 + (E3/D23)^2)*G1*0.5$
M	Upper x - Empirical (for upper Confidence Band on Empirical-line plot)	$= (B22 + (((B3-B22)/(2*B23)) + ((D3-D22)/(2*D23)) - L3) * B23)$
N	Upper y - Empirical	$= (D22 + (((B3-B22)/(2*B23)) + ((D3-D22)/(2*D23)) + L3) * D23)$
O	Lower x - Empirical	$= (B22 + (((B3-B22)/(2*B23)) + ((D3-D22)/(2*D23)) + L3) * B23)$
P	Lower y - Empirical	$= (D22 + (((B3-B22)/(2*B23)) + ((D3-D22)/(2*D23)) - L3) * D23)$
Q	t-statistic (of difference between Measures relative to their means)	$= ((B3-B22+D22-D3)/\text{SQRT}(C3^2+E3^2))$

Entry	Measure	S.E.	Measure
Mean	0.3812		0.0008
S.D.	1.851301		1.40568
Identity trend	3.63818		3.25778
Identity trend	-2.87578		-3.25618
Empirical trend	4.083802		0.50081
Empirical trend	-3.3214		-0.49921
Identity intercept	0.3804		
Identity slope	1		
Empirical intercept with x-axis	0.375276		
Empirical intercept with y-axis	-0.05068		
Empirical slope	0.135043		
Correlation	0.135043		
Reliability x	0.875797		
Reliability y	0.972948		
Disattenuated Correlation	0.146294		

The relationship between the variables is summarized in the lower cells.

Empirical slope = Correlation between x-values and y-values

Intercept = intersection of the line with empirical slope through the point: mean(x-values), mean(y-values)

Predicted y-value = intercept with y-axis + x-value * empirical slope

Predicted x-value = intercept with x-axis + y-value / empirical slope

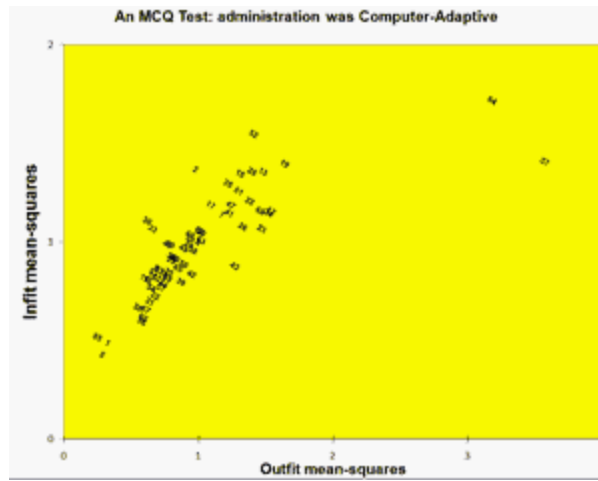
Disattenuated correlation approximates the "true" correlation without measurement error.

Disattenuated correlation = Correlation / sqrt (Reliability(x-values) * Reliability(y-values))

E	F	G	H
	CI=	1.96	68%=1.00
	ACT	Joint S.E.	Upper x -
0.21	Watch bir	0.291045	-0.69105
0.22	Read boo	0.304904	-1.0149

In Row 1, the worksheet allows for user-adjustable confidence bands.

2. Plots without confidence bands



The plot can be edited with the full functionality of Excel.

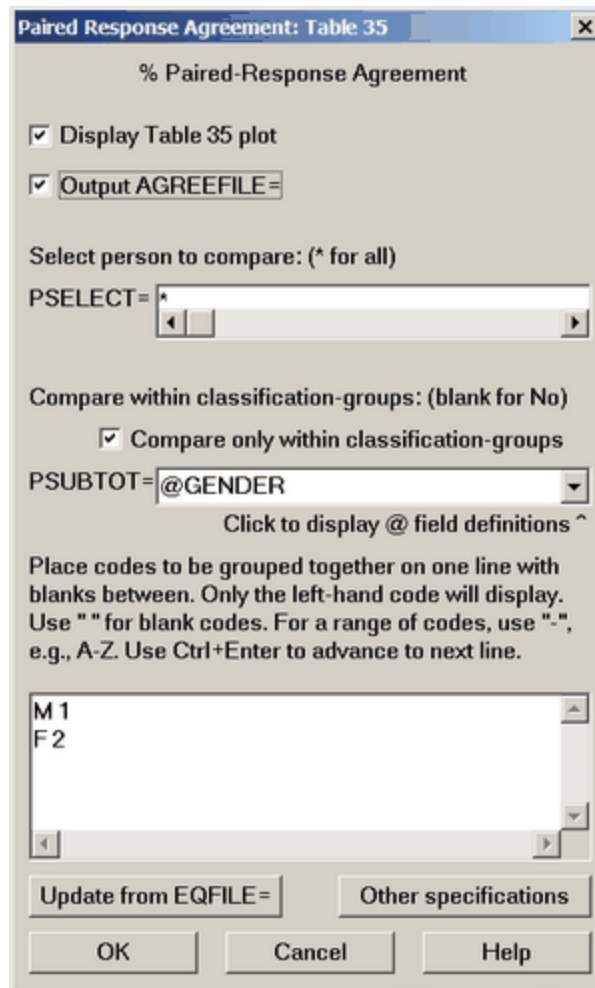
The Worksheet shows the correlation of the points on the two axes.

	A	B	C	D	E	F
1	Scatterplot	An MCQ Test		8/30/08 11:08 PM	Correlation	0.782639
2	Entry	Outfit mean	Infit mean	TOPIC		

18 Output dialogs

18.1 AGREEFILE and Table 35 Specifications

This is the dialog box for [Table 35](#) from the [Output Tables](#) menu, and [AGREEFILE=](#) from [Output Files](#) menu.



% Paired Response Agreement

Display Table 35 plot: check this box to output [Table 35](#), "Paired-person (cheating) agreement"

Output AGREEFILE=: check this box to output the [AGREEFILE=](#) agreement file.

Select persons to compare: enter the [PSELECT=](#) value, or * to select all persons

To omit persons from selection using [PDELETE=](#), click on "Other Specifications" push-button.

Compare only within classification groups:

Compare only within their own classification-groups ([AGREEGROUP=](#)). For the overall plots in [Table 35](#), and [AGREEFILE=](#), check the box if persons are to be compared only with others in their own classification groups, and not across classification-groups.

PSUBTOT=:

enter [PSUBTOTAL=](#) value to specify the groups. Leave blank if all persons are to be part of one classification-group.

Place codes to be grouped together on the same line. This allows multiple classification groups to be combined into one group.

Update from EQFILE= re-initializes the equivalent codes from the [EQFILE=](#) information in the control file.

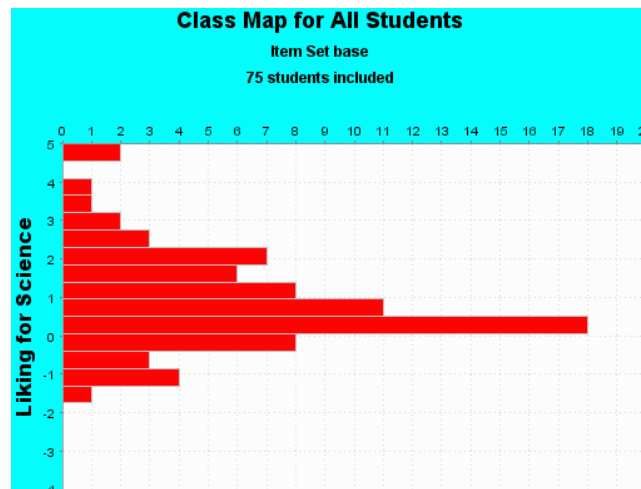
Other Specifications: other control variables can be changed.

OK - launch the paired-agreement computation

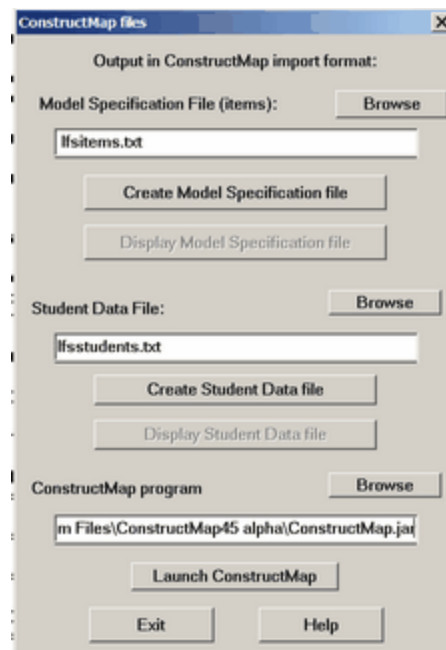
18.2 Compare statistics

See [Scatterplot: Compare statistics](#)

18.3 ConstructMap (GradeMap) interface



To produce this plot from ConstructMap (formerly GradeMap), an analysis and graphical program distributed in association with the book "Constructing Measures: An Item Response Modeling Approach" by Mark Wilson (2004) Mahwah NJ: Lawrence Erlbaum Associates.



The ConstructMap option on the [Output Files](#) menu displays this box which enables a simple conversion of Winsteps control and data files into ConstructMap format. These files can then be imported by ConstructMap - see the ConstructMap User Guide. The files can be displayed after they are created by Winsteps. Excel is used if it is available. You can edit the files if you wish. If changes are made, save the files in tab-separated .txt format - which is expected by ConstructMap.

Example ConstructMap Dialog:

Winsteps: Create Model Specification file: lfsitems.txt

Winsteps: Create Student Data file: lfsstudents.txt

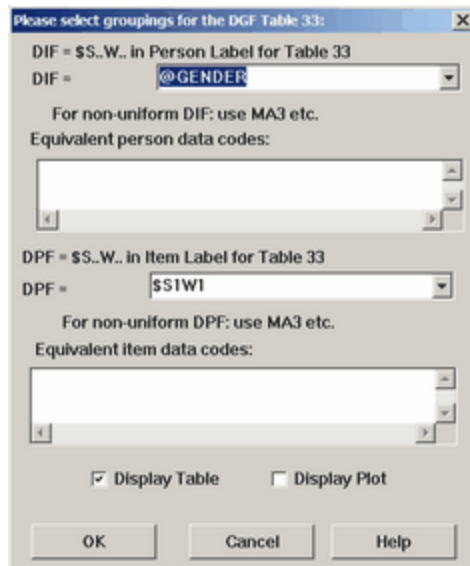
Winsteps: Launch ConstructMap

ConstructMap screen displays:
User name: admin
Password: bear

Please follow the ConstructMap instructions in the documentation at bearcenter.berkeley.edu/software/constructmap

You will notice that Winsteps can analyze datasets which ConstructMap cannot (without editing).
Hint: click on ConstructMap item-column headings to deactivate items causing errors.

18.4 DGF specifications



This is displayed from the [Output Tables](#) or [Plots](#) menus for Differential Group Functioning DGF [Table 33](#) between classification groups of persons and classification groups of items.

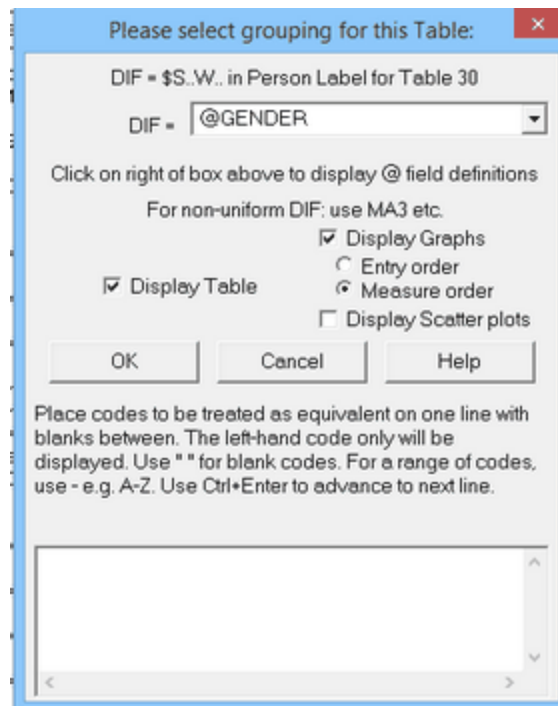
In the [DIF=](#) box specify the column in the **person** label that identifies the DIF classification for each person. Enter the column of the DIF field in the person label. This is in the format [\\$\\$S9W1](#) or [@GENDER](#) or [\\$MA3](#)

In the [DPF=](#) box specify the column in the **item** label that identifies the DPF classification for each person. Enter the column of the DPF field in the item label. This is in the format [\\$\\$S9W1](#) or [@ITEM-TYPE](#) or [\\$MA3](#)

Check "Display Table" for DGF [Table 33](#).
Check "Display Plot" for the [DGF Excel Plot](#).

Please also see [DIF Specifications](#).

18.5 DIF specifications



This is displayed from the [Output Tables](#) or [Plots](#) menus for Differential Item Functioning DIF [Table 30](#).

In the [DIF=](#) box specify the column in the **person** label that identifies the DIF classification for each person. Enter the column of the DIF field in the person label. This is in the format [\\$S9W1](#) or [@GENDER](#) or [\\$MA3](#)

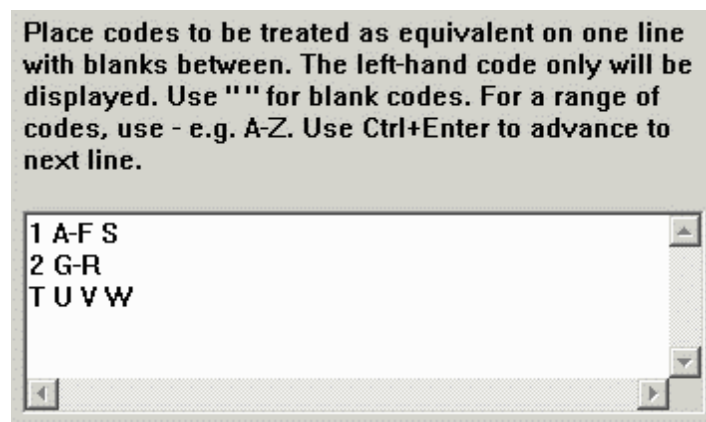
Check "Display Table" for DIF [Table 30](#).

Check "Display Graphs" for the [DIF Excel plots](#).

Graphs can be displayed with the items in Entry order or Measure (difficulty) order

Check "Display Scatter plots" for the [DIF Excel scatter plots](#).

Transformation: Selected [DIF=](#), [DPF=](#), [ISUBTOTAL=](#), [PSUBTOTAL=](#) codes can be transformed into other codes (of the same or fewer characters) for Tables [27](#), [28](#), [30](#), [31](#), [33](#):

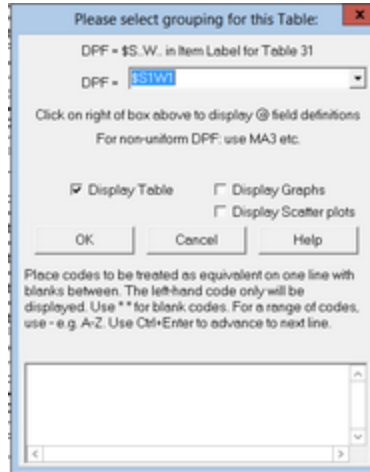


First produce the Table with the transformation box blank. Inspect the reported codes. Transform and combine them using the transformation box in a second run. In this example, codes 1,A,B,C,D,E,F,S will all be converted to 1. 2,G,H,I,J,K,L,M,N,O,P,Q,R will all be converted to 2. T,U,V,W will all be converted to T. Codes X,Y,Z and any others will be

unchanged.

In each line in the transformation box, the code at the beginning (extreme left) of each line is the code into which it and all other blank-separated codes on the line are transformed. Ranges of codes are indicated by -. To specify a blank or hyphen as a code, place them in quotes: " " and "-". Codes are matched to the transformation box starting at the top left and line by line until a match is found, or there is no match.

18.6 DPF specifications



This is displayed from the [Output Tables](#) or [Plots](#) menus for Differential Person Functioning DPF [Table 31](#).

In the [DPF=](#) box specify the column in the **item** label that identifies the DPF classification for each person. Enter the column of the DPF field in the item label. This is in the format [\\$\\$S9W1](#) or [@ITEM-TYPE](#) or [\\$MA3](#)

Check "Display Table" for DPF [Table 31](#).

Check "Display Graphs" for the [DPF Excel plots](#).

Check "Display Scatter plots" for the [DPF Excel scatter plots](#).

Please also see [DIF Specifications](#).

18.7 Edit Initial Settings (Winsteps.ini) file

To change the Winsteps control & data directory/folder from a short-cut icon:



Use an existing short-cut to Winsteps (e.g., on your Start menu or Desktop)

or create a new short-cut to Winsteps by right-clicking on Winsteps.exe, SendTo, Create shortcut

Right-click the short-cut to Winsteps

Click on "Properties"

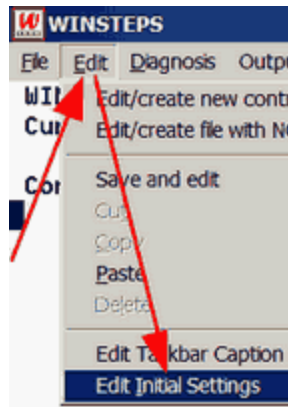
Select "Shortcut"

Type the desired directory into "Start in"

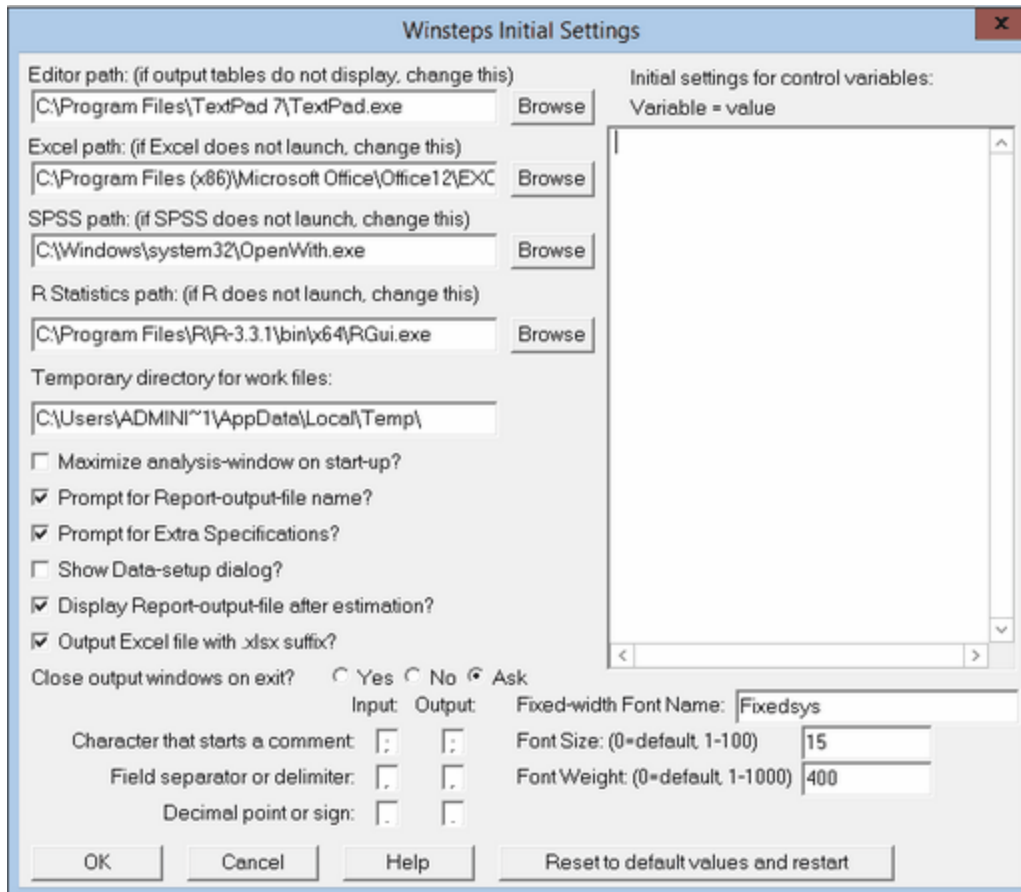
Click on "Apply"

Altering the Initial Settings

1) Pull-down the "Edit" menu



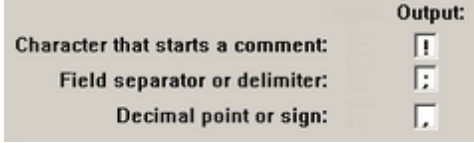
2) Click on "Edit initial settings"



3) The Settings display. If this is too big for your screen, see Display too big.

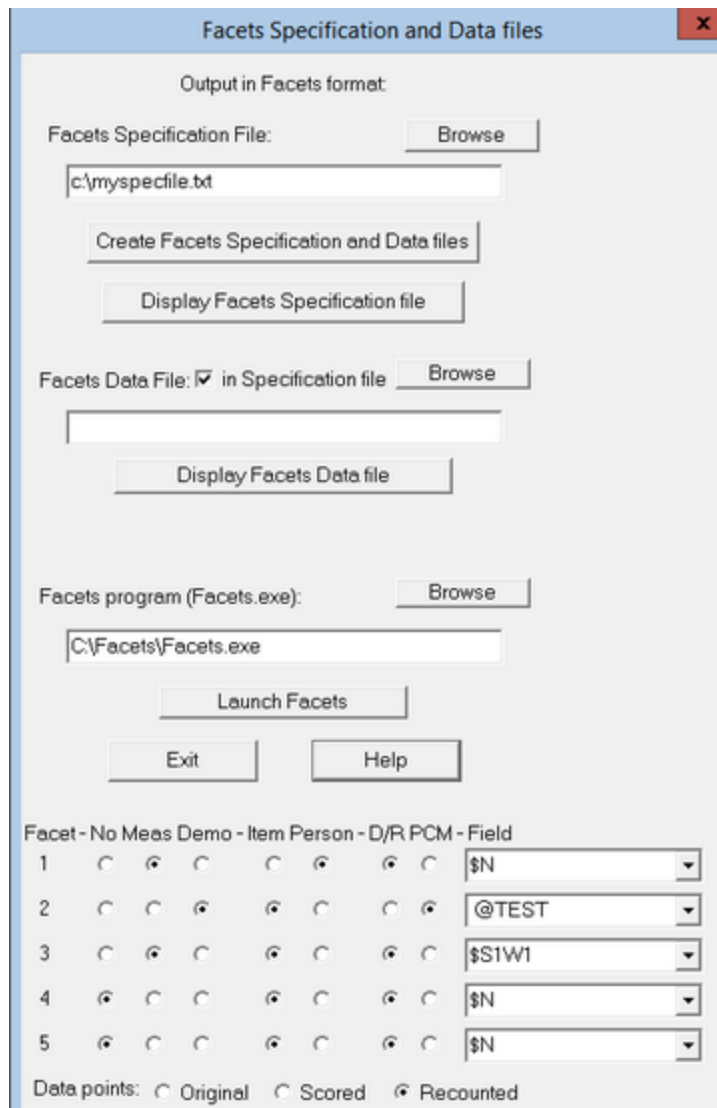
<p>(1) Editor path: Editor=NOTEPAD</p>	<p>This specifies the word processor or text editor that is to be used by Winsteps. It is usually your Windows text editor. If Output Tables do not display, please change this text editor. You can make Winsteps use your own text editor or word processor by changing the Editor in Edit initial settings or editing Winsteps.ini in your Winsteps folder. Put the full pathname of your editor in place of Notepad, e.g., for WordPad it could be: "C:\Program Files\Windows NT\Accessories\wordpad.exe" for Word 6.0, it could be: "C:\MSOFFICE\WINWORD\WINWORD.EXE"</p>
---	--

	<p>You can find the full path to your word processor by doing a "Find" or "Search" for it from the "Start" menu. Useful replacements for Notepad include NotePad++ freeware: notepad-plus-plus.org</p>
<p>(2) Excel path: Excel=C:\Program Files\Microsoft Office\Office12\EXCEL.EXE</p>	<p>This provides a fast link to Excel from the File pull-down menu. The path to any program can be placed here. If Winsteps does not find Excel automatically, please go to the Windows "Start" menu, do a "Find" or "Search" for "Files and Folders" to locate it, and enter its path here. Excel is not part of Winsteps, and may not be present on your computer.</p> <p>If this malfunctions, a) check that the Windows filetypes are correct, by creating a .xls file, then double-clicking on it. Does it open in Excel? If not, then Excel is not properly installed. b) Launch Winsteps c) blank out the Excel path in "Edit menu", "Initial settings" d) click on OK e) "Edit menu", "Initial settings" should now show the correct path to Excel.</p> <p>If the path is wrong or blank, then please update the .xls file association in the Windows Registry - Google can help with this.</p>
<p>(3) SPSS path: SPSS=C:\SPSS.EXE</p>	<p>This provides a fast link to SPSS from the File pull-down menu. The path to any program can be placed here. If Winsteps does not find SPSS automatically, please go to the Windows "Start" menu, do a "Find" or "Search" for "Files and Folders" to locate it, and enter its path here. SPSS is not part of Winsteps, and may not be installed on your computer.</p>
<p>(4) R Statistics path: RSTAT=C:\R\bin\R.exe</p>	<p>This provides a fast link to R Statistics from the File pull-down menu. The path to any program can be placed here. If Winsteps does not find R Statistics automatically, please go to the Windows "Start" menu, do a "Find" or "Search" for "Files and Folders" to locate it, and enter its path here. R Statistics is not part of Winsteps, and may not be installed on your computer.</p>
<p>(5) Temporary directory for work files: Default Windows temporary directory/folder = %temp%</p>	<p>Temporary Output and Table files are placed in the same directory as your Input file (if possible), or in the "Temporary directory". Other temporary files are placed in the "Temporary directory." Files ending "...ws.txt" can be deleted when Winsteps is not running.</p> <p>If the temporary directory does not change when you type in a new one, and click OK: 1. The desired temporary folder does not exist: please create it. 2. You do not have write permission for the desired temporary folder: please give yourself write-permission 3. You do not have write-permission for Winsteps.ini in the Winsteps.exe folder: please give yourself write-permission</p> <p>You could also: 1. Change the Windows temporary folder: how do i change the location of windows temporary files 2. Delete Winsteps.ini in the Winsteps.exe folder 3. Launch Winsteps</p>
<p>(6) Maximize analysis-window on start-up?</p>	<p>When checked, the analysis window is displayed maximized (filling your screen). Unchecked, the window size is normal (big enough for practical purposes).</p>
<p>(7) Prompt for Report-output-file name?</p>	<p>With an unchecked box, this specifies that your standard report output file will be a temporary file. A checked box produces a prompt for a file name on the screen. You can always view the Report output file from the Edit menu, and save it as another file.</p>

(8) Prompt for Extra Specifications:?	unchecked specifies that there will be no extra specifications to be entered.
(9) Show Data-setup dialog?	When the box is checked the Data-setup dialog box displays.
(10) Display the Report-output-file after estimation?	When checked, this automatically shows the report output file if TABLES= or TFILE= is active in the Control file or at Extra Specifications . The Report Output File can always be displayed from the Edit menu .
(11) Output Excel files with .xlsx suffix?	Uncheck this box to output .xls Excel files
(12) Close output windows on exit? Closeonexit=Yes or No or Ask	This choice can be overridden when Winsteps stops .
(13) Character that starts SIGNS=;;,....	This enables the processing of files in accordance with international usages. For instance, it may be more convenient to output files with decimal commas, use semicolons as separators, and indicate comments with exclamation marks: 
(14) Initial settings for control variables: Variable = value	Control variables and their settings entered into this box are pre-pended to the settings in your control file. This enables you to set values which apply to all analyses. They are reported first in Table 0.3
(15) Changing the Font in the Winsteps Analysis window:	Fixed-width Font Name: Courier New is the most widely used. Type in your preferred font, Lucida Console, Consolas, Andale Mono. Font Size: , Font Weight: , OK , then restart Winsteps. Winsteps Font details are maintained in the Windows Registry.
OK:	Click on this to accept the settings. Some settings go into effect the next time you start Winsteps. The changes are permanent until you change them again.
Cancel:	Do not change the previous settings.
Help:	Displays this page of the Help file
Reset to default values and restart:	Resets these settings to their default values, then restart Winsteps. Current settings are in file <i>Winsteps.ini</i> in your Windows AppData area.

18.8 Facets interface: this requires Facets software, available separately

From the [Output Files](#) menu, [Facets](#) Specification and Data files can be constructed after Winsteps estimation completes.



In Facets dialog box	Instructions
Facets Specification File	this text file is created to contain Facets specifications, ready to run, or to be further edited
Create Facets Specification and Data files	create these from the Winsteps control and data files
Display Facets Specification file	this file is displayed when it is created. It can be redisplayed here.
Facets data file: <input checked="" type="checkbox"/> in Specification file	If this is checked, then the Facets data follows the specifications in the specification file. No Facets data file is needed.
Facets data file: <input type="checkbox"/> in Specification file	If this is not checked, then the Facets data is in a separate file. A data file name is needed.
Facets program (Facets.exe)	Path to Facets.exe (or Minifac.exe) on your computer. Facets software is available from www.winsteps.com/winbuy.htm
Launch Facets	Analyze the Facets specification and data files

Exit	Close the window. All current settings are saved, but lost when Winsteps closes.
Facet 1 2 3 4 5	Up to 5 facets can be defined
No	this facet is not used
Meas	this is a measurement facet
Demo	this is a demographic (dummy facet)
Item	this facet's elements are based on the Winsteps items
Person	this facet's elements are based on the Winsteps persons
D/R	this facet is modeled with "?".
PCM	this facet is modeled with "#".
Field	element label construction rules
Data points: Original	facets data points are the original values in the Winsteps data file
Scored	the data values after scoring with KEY1= , IVALUE= , etc.
Recounted	if there are unobserved intermediate categories, and STKEEP=No , then the resulting values.

Example: using Example0.txt

Title= Facets Specification File 4/10/2019 LIKING FOR SCIENCE (Wright & Masters p.18)

Facets= 3

Positive= 1

Noncenter= 1

Models= ?, ?, #, R2

Labels=

1, KID @PLABEL

1= F Buff, Marge Baby

2= F Clapp, Charlie

...

75= M Wright, Benjamin

*

2, KID @GENDER, D ; demographic/dummy

1= F

2= M

*

3, ACT @ILABEL

1= Find bottles and cans

2= Find out what animals eat

...

25= Watch what animals eat

*

Data=

60,2,22, 1

60,2,16, 2

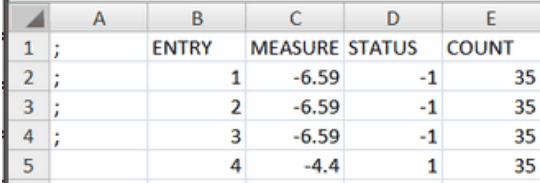
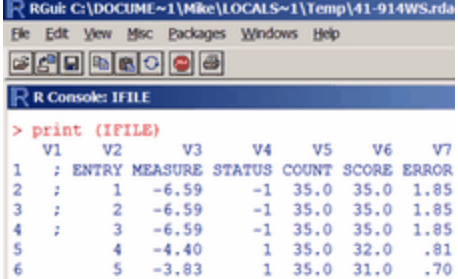
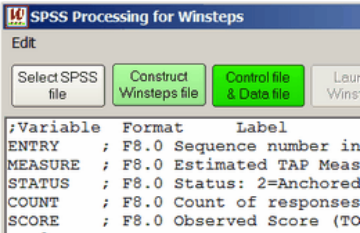
...

54,2,18, 0

18.9 Output file formats

Several output file formats are available from the [Output Files menu](#), or by specifying the output filename with the appropriate suffix in your control file.

For Output Table formats, see [ASCII](#)

<p>Text file - fixed fields: Suffix: .txt CSV=N HLINES=Yes</p>	<pre>; TAP KNOX CUBE TEST Nov 20 2:54 2008 ;ENTRY MEASURE ST COUNT SCORE ERRORF ; 1 -6.59 -1 35.0 35.0 1.85 ; 2 -6.59 -1 35.0 35.0 1.85 ; 3 -6.59 -1 35.0 35.0 1.85 4 -4.40 1 35.0 32.0 .81 5 -3.83 1 35.0 31.0 .70 6 -3.38 1 35.0 30.0 .64 7 -3.83 1 35.0 31.0 .70</pre>																																				
<p>Text file - tab-delimited: Suffix: .txt CSV=T HLINES=Yes ROW1HEADING=No omits the first heading line</p>	<pre>";TAP KNOX CUBE TEST Nov 20 2:54 2 ";" "ENTRY" "MEASURE" "STATU ";" -6.59 -1 35.0 ";" Tabs 2 -6.59 -1 35.0 ";" 3 -6.59 -1 35.0 " " 4 -4.40 1 35.0 " " 5 -3.83 1 35.0 " " 6 -3.38 1 35.0</pre>																																				
<p>Text file - comma-separated: Suffix: .txt CSV=Y HLINES=Yes</p>	<pre>";TAP KNOX CUBE TEST Nov 2 ";" "ENTRY" "MEASURE" "STATU ";" ,1, -6.59, -1, 35.0, 35.0, 1.8 ";" ,2, -6.59, -1, 35.0, 35.0, 1.8 ";" ,3, -6.59, -1, 35.0, 35.0, 1.8 " " ,4, -4.40, 1, 35.0, 32.0, .81, " " ,5, -3.83, 1, 35.0, 31.0, .70,</pre>																																				
<p>Excel file: Suffix: .xls</p>	 <table border="1"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>;</td> <td>ENTRY</td> <td>MEASURE</td> <td>STATUS</td> <td>COUNT</td> </tr> <tr> <td>2</td> <td>;</td> <td></td> <td>1</td> <td>-6.59</td> <td>-1</td> </tr> <tr> <td>3</td> <td>;</td> <td></td> <td>2</td> <td>-6.59</td> <td>-1</td> </tr> <tr> <td>4</td> <td>;</td> <td></td> <td>3</td> <td>-6.59</td> <td>-1</td> </tr> <tr> <td>5</td> <td></td> <td></td> <td>4</td> <td>-4.4</td> <td>1</td> </tr> </tbody> </table>		A	B	C	D	E	1	;	ENTRY	MEASURE	STATUS	COUNT	2	;		1	-6.59	-1	3	;		2	-6.59	-1	4	;		3	-6.59	-1	5			4	-4.4	1
	A	B	C	D	E																																
1	;	ENTRY	MEASURE	STATUS	COUNT																																
2	;		1	-6.59	-1																																
3	;		2	-6.59	-1																																
4	;		3	-6.59	-1																																
5			4	-4.4	1																																
<p>R-Statistic file: Suffix: .rdata Quick R commands: > ls.str() > print(IFILE)</p>	 <pre>R Console: IFILE > print (IFILE) v1 v2 v3 v4 v5 v6 v7 1 ; ENTRY MEASURE STATUS COUNT SCORE ERROR 2 ; 1 -6.59 -1 35.0 35.0 1.85 3 ; 2 -6.59 -1 35.0 35.0 1.85 4 ; 3 -6.59 -1 35.0 35.0 1.85 5 4 -4.40 1 35.0 32.0 .81 6 5 -3.83 1 35.0 31.0 .70</pre>																																				
<p>SPSS file: Suffix: .sav Suitable for input in AMOS</p>	 <pre>SPSS Processing for Winsteps Edit Select SPSS file Construct Winsteps file Control file & Data file ;Variable Format Label ENTRY ; F8.0 Sequence number in MEASURE ; F8.0 Estimated TAP Meas STATUS ; F8.0 Status: 2=Anchored COUNT ; F8.0 Count of responses SCORE ; F8.0 Observed Score (TO</pre>																																				

Webpage (HTML):
Suffix: .htm

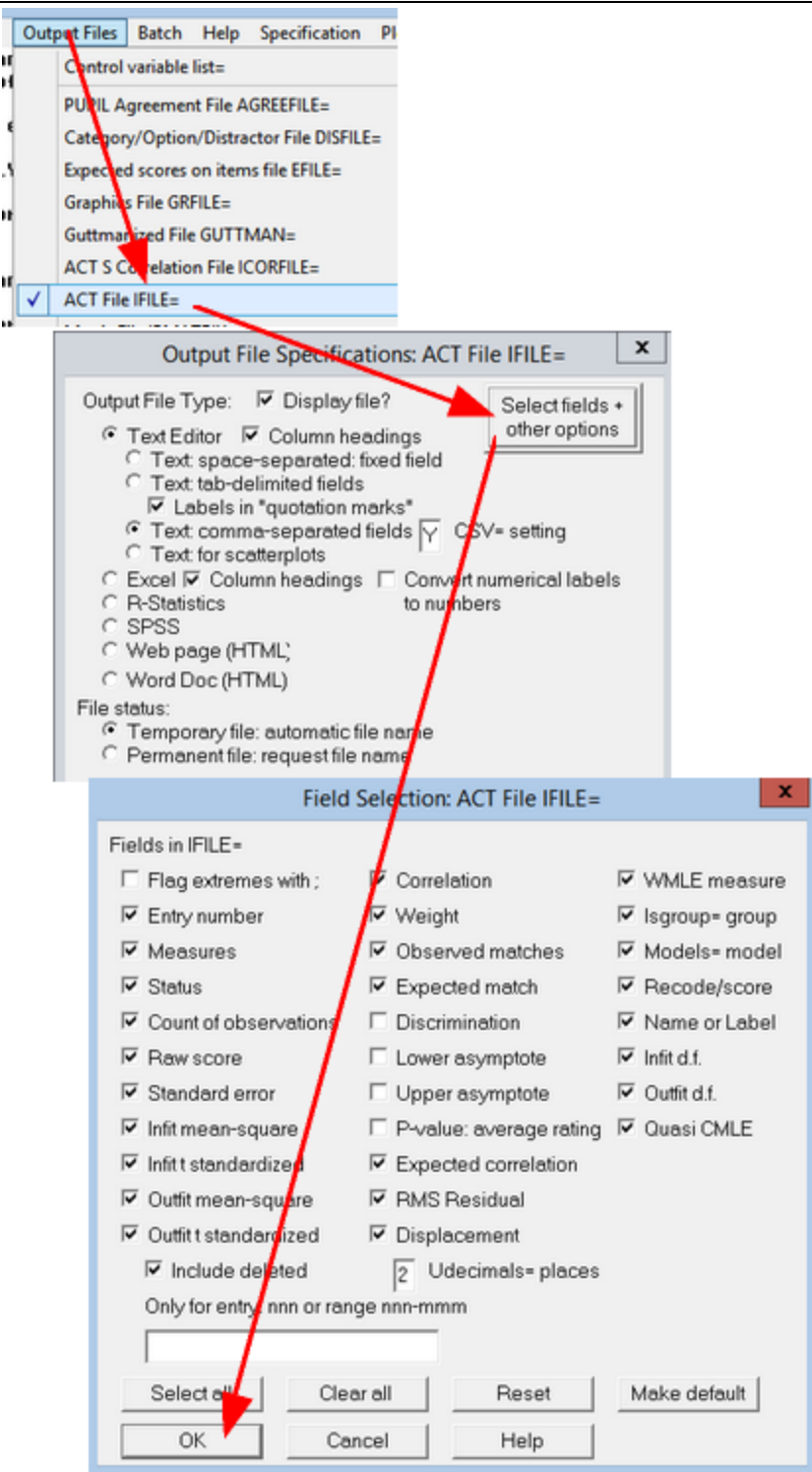
ENTRY	MEASURE	ST	COUNT	SCORE	ERROR	IN
1	-6.59	-1	35.0	35.0	1.85	
2	-6.59	-1	35.0	35.0	1.85	
3	-6.59	-1	35.0	35.0	1.85	
4	-4.40	1	35.0	32.0	.81	
5	-3.83	1	35.0	31.0	.70	
6	-3.38	1	35.0	30.0	.64	
7	-3.83	1	35.0	31.0	.70	

18.10 Output file specifications and Field selection

[Output files](#) can be written in several formats, and called with different programs, so a dialog box is shown when output files are requested from the [Output Files menu](#):

Dialog box for most output files	Dialog box for IFILE= and PFILE=	Dialog box for SIFILE=

Output File Type:	output file formats are shown here.
Display file?	check (tick) if the output file is to be displayed by the relevant program.
Select fields + other options	click on this button to select output fields. It shows the "Field Selection" dialog box, indicating which fields are to be included in the IFILE= or PFILE= .



Text Editor	Output is a .txt file. If <i>Display file?</i> is checked, then the file is automatically displayed by Notepad or your own text editor .
Column headings	check to produce column headings in the output file
Text: space-separated: fixed field	output file is organized in columns. This is usually easiest to look at.
Text: tab-delimited	columns are separated by Tab characters, which is useful for importing into modern software
Text: comma-separated fields	columns are separated by commas or other characters equivalents, which may be required for importing into software

Separator character	character to use instead of , for comma-separated fields
Labels in "quotation marks"	place non-numeric values within quotation marks, required for correct importing of data into some software.
Text: for scatter-plots	Chooses default settings for the IFILE= or PFILE= , suitable for scatter-plotting . These can be overridden using Select fields
Excel	<p>Output is a .xls file. Excel must be installed on your computer.</p> <p>If <i>Display file?</i> is checked, then Excel is also activated on the Windows task bar, and the file may display. This message is in the Analysis Window: <i>"Excel will activate soon. Please continue ..."</i></p> <p>If the Excel file does not display, but you want it to, then please change the Windows Registry (at your own risk):</p> <ol style="list-style-type: none"> Windows: Start >> Run >> <i>regedit</i> Navigate to the following path: <i>HKEY_CURRENT_USER\Control Panel\Desktop</i> Modify the value named <i>ForegroundLock Timeout</i> to 0 If <i>ForegroundLock Timeout</i> does not exist, then create <i>ForegroundLock Timeout</i> with DWORD value 0 <p>Do you want numerical person and item labels to become numbers (not text) in Excel? Then EXCELNUMBER=Yes</p>
Column headings	check to produce column headings in row 1 of the Excel worksheet.
R-Statistics	<p>Output is an R Statistics .rdata file. If <i>Display file?</i> is checked, then the file is automatically displayed by RGui (if available). If this malfunctions, check that the path to R is correct with Edit Initial Settings.</p> <p>When the RGui window opens, type > ls.str() # to see the data file structure > print(IFILE) # IFILE is the name of a data.frame</p>
SPSS	Output is a .sav file. If <i>Display file?</i> is checked, then the file is automatically displayed by SPSS (if available). If this malfunctions, check that the path to SPSS is correct with Edit Initial Settings .
Web page (HTML)	Output is a .htm file. If <i>Display file?</i> is checked, then the file is automatically displayed by your Internet browser.
File status:	
Temporary file: automatic file name	the file will be deleted when Winsteps terminates. "Save as" to make the file permanent.
Permanent file: request file name	the file will not be deleted when Winsteps terminates. Type in the filename of your choice when Winsteps requests it
Buttons:	
OK	Perform the indicated actions
Cancel	Return to Winsteps processing
Help	Display this documentation
Set as default	Make the select actions the permanent selections

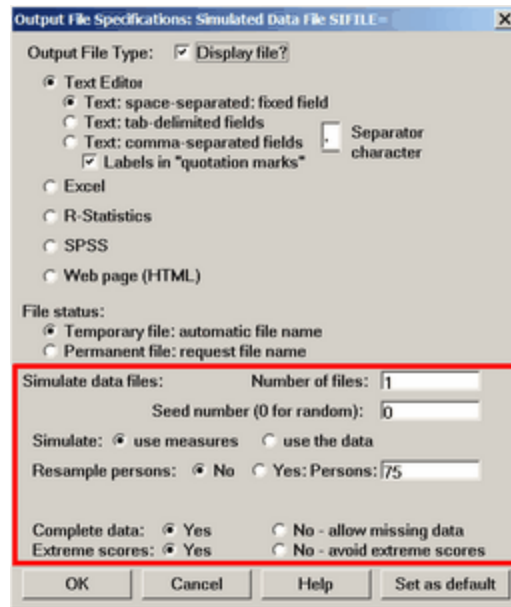
18.11 Simulated file specifications

Winsteps uses two methods to simulate data:

- 1) probabilistically-generated data based on anchored parameter estimates
- 2) re-sampling-with-replacement from the current dataset

Winsteps uses the estimated (or anchored) person, item and Andrich Thresholds or person-response-string re-sampling-with-replacement to simulate a data file equivalent to the raw data. This can be used to investigate the stability of measures, distribution of fit statistics and amount of statistical bias. Each time [SIFILE=](#) is run, or selected from the [Output](#)

[Files](#) pull-down menu, a simulated data file produced. Do simulated analyses with several simulated datasets to verify their overall pattern.



The parts of the dialog box outside the red rectangle are described in [Output file specifications](#). The file format matches the input data file if both are in fixed-field format. When SIFILE= is written with [CSV=Y](#), comma-separated or [CSV=T](#), tab-separated, the item responses precede the person label.

Simulated data files:	invoked with SIFILE=
Number of files:	SINUMBER= , specifies the number of simulated files to produce. If SINUMBER= is greater than 1, then the data file name is automatically incremented, and so is the SISEED= pre-set seed value
Seed number (0 for random):	SISEED= , controls whether the pseudo-random number generator is seeded with the system clock (0 or 1), or with a user-chosen value, (2 and above)
Simulate: use measures or use the data	SIMEASURE= , chooses whether the simulated data is generated from the estimated measures (use measure), or by re-sampling from the observed data (use the data). If you wish to over-ride the estimated measures, then use IAFILE= , PAFILE= and SAFILE=
Re-sample persons: No or Yes: Persons	SIRESAMPLE= , controls whether re-sampling occurs (sampling with or without replacement), and, if it does, how many person records to include in the simulated data file
Complete data: Yes or No - allow missing data	SICOMPLETE= , Yes for complete data. No for missing data patterns to be repeated in the simulated data file
Extreme scores: Yes or No - avoid extreme scores	SIEXTREME= , Yes to allow the simulated data to include extreme (zero, minimum possible or perfect, maximum possible) scores. No to avoid generating extreme scores (when possible)

Winsteps simulates data two ways:

i) the default: using the parameter values (persons, items, Andrich thresholds) from the current analysis. This way generates simulated data according to the probabilistic distributions defined by the Rasch model and the generating Rasch parameters. This is for investigations relating to exact Rasch conditions.

ii) by resampling (with replacement) the data (observations, responses) in the current analysis. This way generates data according to the empirical distribution of the generating data so pervasive misfit, such as [DIF](#), is replicated.

Example 0 . KCT.txt simulated with CSV=N fixed field format (re-sampling response strings):

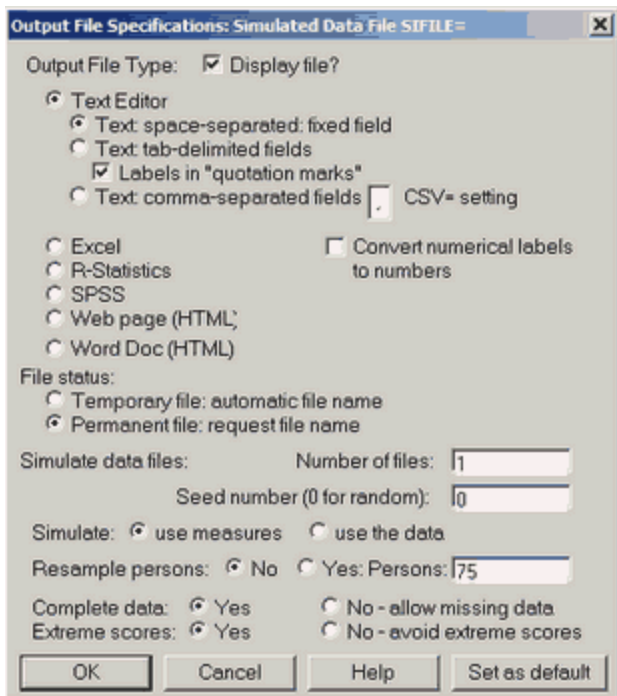
```
Dorothy F 111111111100000000    -.2594      13
Elsie   F 111101111100000000    -1.3696     14
Thomas  M 111111111101000000    -.2594      31
Rick    M 111111111101000000    -.2594      27
```

KCT.txt simulated with comma-separated, CSV=Y, [HLINES=Y](#), [QUOTED=Y](#) format (re-sampling person measures):

```
"1-4", "2-3", "1-2-4", "1-3-4", "2-1-4", . . . , "KID", "Measure", "Entry"
1,1,1,1,1,1,0,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,"Rick M",- .2594,27
1,1,1,1,1,1,1,1,1,0,1,0,0,0,0,0,0,0,0,0,0,0,"Helen F",- .2594,16
1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,"Rod M",1.9380,28
1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,"William M",.9229,34
```

Example 1. It is desired to investigate the stability of Rasch measures.

- (1) Estimate measures from your control and data files (e.g., SF.txt)
- (2) Choose SIFILE= from the Output Files menu.
- (3) Choose to output a permanent file:



- (4) Simulated data filename: SFSIMUL.TXT
- (5) Rerun Winsteps with your Winsteps control file and DATA=SFSIMUL.TXT on the "Extra Specifications" line.
- (6) Compare person, item and Andrich Thresholds.

Example 2. To estimate the measure standard errors in a linked [equating](#) design.

1. Do a concurrent calibration with Winsteps
2. Simulate data files SIFILE= from the [Output Files](#) menu.
Specify "complete data" [SICOMPLETE=](#) as "No" to maintain the same data pattern.
Save 10 simulated sets, SINUMBER=, as S.txt S2.txt
3. Rerun your Winsteps analysis 10 times
Specify in [Extra Specifications](#) "DATA=S.txt PFILE=P1.txt CSV=TAB" etc.
This will produce 10 PFILE=s. Export them in Excel format.
4. Use Excel to compute the standard deviation of the measures for each person based on the 10 person measures

5. These are the model standard errors for the equating design for the measures.
6. Inflate these values by 20%, say, to allow for systematic equating errors, misfit, etc.

Example 3. If you do need estimation-bias-correction ([STBIAS=](#)) that is as accurate as possible with your data set, you will need to discover the amount of bias in the estimates and correct for it:

1. In your control file, [STBIAS=No](#) and [USCALE=1](#)
2. Obtain the Winsteps estimates for your data
3. Simulate many datasets using those estimates. (SIFILE= on the Winsteps Output Files menu).
4. Obtain the Winsteps estimates from the simulated data sets
5. Regress the simulated estimates on your initial estimates. These will give a slope near 1.0.
6. Obtain the Winsteps estimates for your data with [USCALE = 1/slope](#). The set of estimates in 6 is effectively unbiased.

Example 4. You need to simulate data from generating values. You can use Winsteps to simulate a dataset.

1. Obtain the generating item difficulties, person abilities and threshold values. If you need a normal distribution of person abilities, you can generate this with Excel.

a. From your standard analysis, output [IFILE=if.txt](#), [SFILE=sf.txt](#)

b. Use Excel or similar to simulate a normal distribution of person abilities with the mean and S.D. that you want.

In Excel:

Cell A1 = Mean

Cell B1 = S.D.

Cell A2 = =ROW()-1

Cell B2 = =NORMSINV(RAND())*\$B\$1 +\$A\$1

then copy A2, B2 for as many rows as you want the sample size.

c. Copy Columns A and B into a text file, [pf.txt](#). Delete row 1.

For a random uniform distribution of item difficulties, use the Excel formula:

=(item difficulty range)*(RAND() - 0.5) + (mean item difficulty)

d. In your Winsteps control file:

[IAFILE=if.txt](#)

[SAFILE=sf.txt](#)

[PAFILE = pf.txt](#)

[SIFILE= simulated.txt](#)

2. Construct a Winsteps control file including the generating values in [IAFILE=](#), [PAFILE=](#), [SAFILE=](#)

3. Make a rectangular dataset with a row of valid responses (can be the same one) as wide as the number of items and with a column of valid responses (can be the same one) as long as the number of persons.

For instance, number of persons = 2000, number of items =100, then an artificial dichotomous Winsteps control file and dataset can be:

```
ITEM1=1
NAME1=101
NI=100
CODES=01
EDFILE=*
1-2000 1-100 1
*
IAFILE=*
1 2.3
2 1.5
.....
*
PAFILE=*
1 3.1
```

```

2 -2.8
.....
*
SAFILE=*
0 0
1 0
*
&END
END LABELS

```

and nothing else.

4. Run Winsteps. Ignore the results of this analysis. Choose [SIFILE=](#) option from the output files menu. Click on "complete data" to simulate the entire data matrix.

Example 5. Bootstrap estimation of the confidence interval for a reliability.

Bootstrapping is based on generating new datasets using "sampling with replacement" from the current dataset. The new datasets can be generated in Winsteps using:

Simulate: use the data

Re-sample persons: yes, with same number of rows as the original data.

Compute the reliability of each new dataset. See "**Performing multiple simulations in Batch mode**".

After transformation with Fisher's z, the distribution of the linearized reliabilities (mean and S.D.) are the linearized expected value and linearized S.E. of the observed linearized reliability for the original data set. Transform the linearized (mean-S.D.), mean, and (mean+S.D.) back to the original reliability metric by reversing the Fisher's z transformation.

Example 6. Multiple simulations in Batch mode. See [BATCH=](#)

These can construct bootstrap confidence intervals for DIF estimates, etc.

Set up 100 simulations in a batch file, and extract the relevant numbers from the 100 output DIF tables.

[PowerGREP](#) is great software for extracting values from files. For instance:

To pick out lines 10-35 in the output files (after line 9, for 26 lines):

```

Action type: Search
File sectioning: Search and collect sections
Section search: \A([\r\n]*+r\n){9}([\r\n]*+r\n){0,26})
Section collect: \2
Search: the search string: .* for everything

```

Performing multiple simulations in Batch mode

1. Use NotePad to create a text file called "Simulate.bat"
2. In this file:

```

REM - produce the generating values: this example uses example0.txt:
START /WAIT c:\Winsteps\Winsteps BATCH=YES example0.txt example0.out.txt PFILE=pf.txt
      IFILE=if.txt SFILE=sf.txt

REM - initialize the loop counter
set /a test=1
:loop

REM - simulate a dataset - use anchor values to speed up processing (or use SINUMBER= to avoid
      this step)
START /WAIT c:\Winsteps\Winsteps BATCH=YES example0.txt example0%test%.out.txt PFILE=pf.txt
      IFILE=if.txt SAFILE=sf.txt SIFILE=SIFILE%test%.txt SISEED=0

REM - estimate from the simulated dataset
START /WAIT c:\Winsteps\Winsteps BATCH=YES example0.txt data=SIFILE%test%.txt SIFILE%test
      %.out.txt pfile=pf%test%.txt ifile=if%test%.txt sfile=sf%test%.txt TFILE=* 3.1 *
REM - do 100 times

```



```
set /a test=%test%+1
if not "%test%"=="101" goto loop
PAUSE
```

3. Save "Simulate.bat", then double-click on it to launch it.

4. The simulate files and their estimates are numbered 1 to 100.

5. The files of estimates can be combined and sorted using MS-DOS commands, e.g.,

Copy if*.txt combinedif.txt

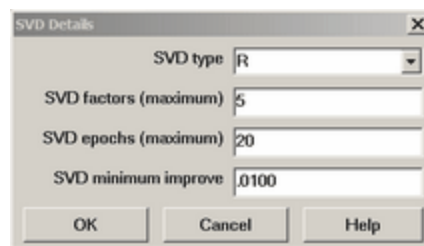
Sort /+(sort column) <combinedif.txt >sortedif.txt

6. Individual lines from the output files can be written to one file using MS-DOS batch commands. For instance, using an MS-DOS batch routine (.bat or .cmd), the same text line can be extracted from many text files and output into a new text file. The new text file can be pasted into Excel. Save these MS-DOS commands as *extract.bat* in the folder that has the files of statistics. Double click on *extract.bat* to execute it.

```
rem replace 2 with the number of lines to skip before the line you want
@echo off
setlocal EnableDelayedExpansion
if exist result.csv del result.csv
for %%f in (*.txt) do (
  echo %%f
  set i=a
  for /F "skip=2 delims=" %%l in (%%f) do (
    if "!i!" == "a" echo %%f, %%l >> result.csv
    set i=b
  )
)
)
notepad result.csv
```

18.12 SVD specifications

This menu is called from the [Output Files menu](#).



For the meaning of these settings, please see [SVDFILE=](#).

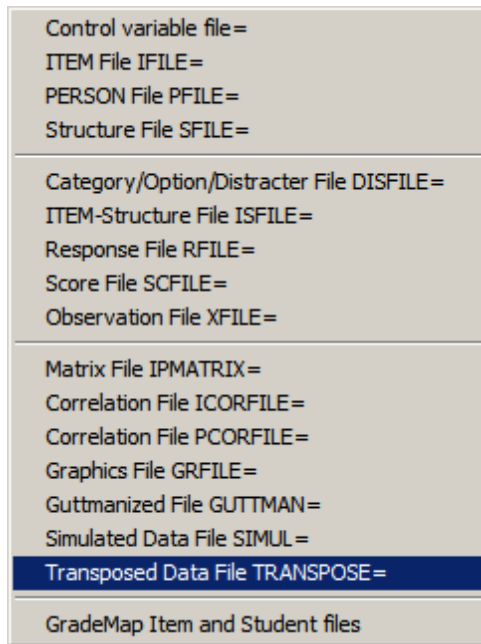
Click on OK to see the [Output file specifications dialog box](#) to produce the [SVDFILE=](#).

18.13 Table/Plot 45 specifications

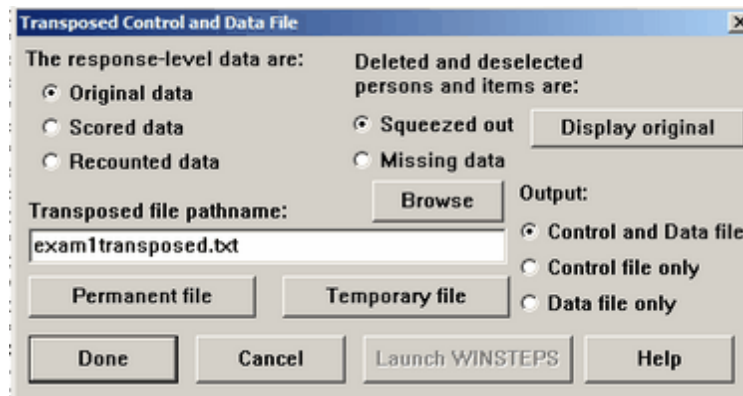
See [Table 45](#)

18.14 Transpose file specifications

To transpose the rows and columns of the data matrix, select Transpose on the [Output Files menu](#) or specify [TRPOFILE=](#) in your control file.



then



This section corresponds to [TRPOTYPE=](#).

The response-level data are:

Original data: data codes are those in the data file.

Scored data: data after applying [NEWSCORE=](#), [VALUE=](#), [KEY=](#), etc.

Recounted data: the data recounted upwards from 0 after applying [STKEEP=No](#), etc.

Deleted and deselect persons and items are: ([IDFILE=](#), [IDELETE=](#), [ISELECT=](#), [PDFILE=](#), [PDELETE=](#), [PSELECT=](#))

Squeezed out: deleted and deselected items and persons are omitted from the transposed file.

Missing data: deleted and deselected items and persons remain in the transposed file, but all their responses are indicated as missing data, "."

Output:

Control and Data file: the transposed file can be analyzed. Additional control instructions may be required in the control section in order for the analysis to proceed correctly.

Control file only: Only the control file for the transposed data is output. The transposed data are not output.

Data file only: Only the transposed data are output. The control instructions are not.

Display original: show the original, un transposed, control file.

Permanent file: file name for the transposed file. Enter in the white box, or use the **"Browse"** function to locate the desired folder.

The transposed file is created and displayed.

Temporary file: use a temporary file: this can be "saved as" a permanent file.

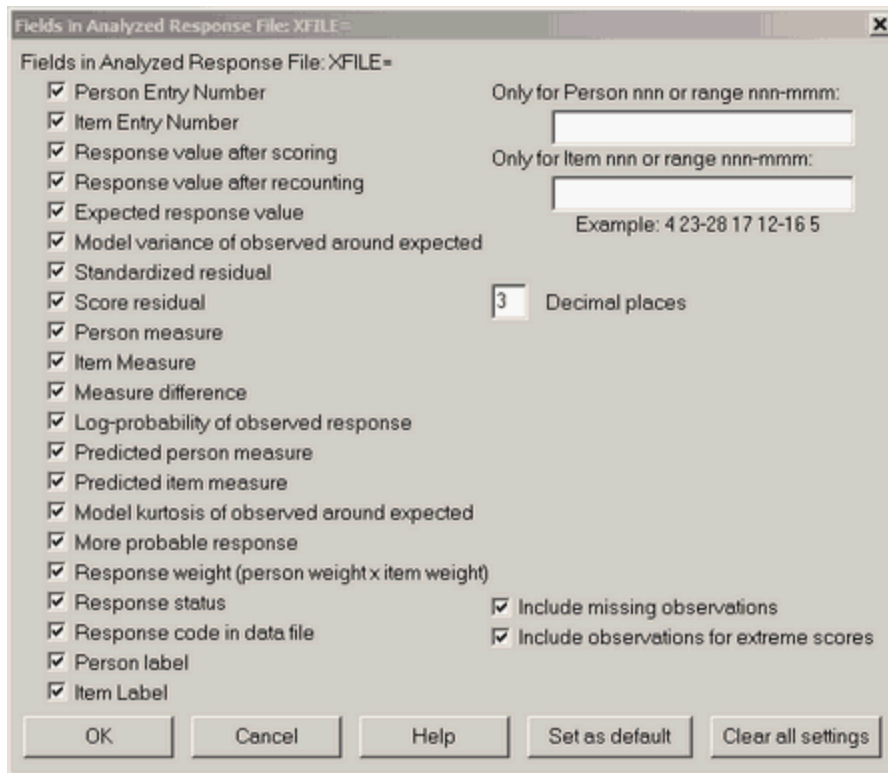
The transposed file is created and displayed.
Done: transposing actions completed
Cancel: exit from this routine
Launch Winsteps: launch Winsteps with the permanent transposed file as its control file.
Help: show this Help file entry

Example: This produces, for [Exam1.txt](#), the equivalent to TRPOFILE=exam1transposed.txt and TRPOTYPE=OSB:

```
; Transposed from: C:\Winsteps\examples\exam1.txt
&INST
TITLE = "TRANSPOSED: KNOX CUBE TEST"
ITEM = KID
PERSON = TAP
NI = 35                ; ACTIVE PERSONS BECOME COLUMNS
;NN = 18              ; ACTIVE ITEMS BECOME ROWS
ITEM1 = 1
XWIDE = 1
NAME1 = 37
NAMLEN = 16          ; ends at 52
CODES = 01
STKEEP = Y
; Add here from original control file: C:\Winsteps\examples\exam1.txt
&END
Richard M  1
..... (more item labels)
Helen   F  35
END NAMES
11111111111111111111111111111111 1-4           1
..... (more data records)
00000000000000000000000000000000 4-1-3-4-2-1-4 18
```

18.15 XFILE= field selection

This is called from the [Output Files menu](#). You can select what fields to include in the [XFILE=](#) file. And you can also specify particular persons and/or items. If this is too big for your screen see Display too big. See [XFILE=](#) for the descriptions of the fields.



<i>Only for Person nnn or range nnn-mmm:</i> <i>Only for Item nnn or range nnn-mmm:</i>	this allows selection of particular items and persons. Example in the "Only for Person" box: 4 11-13 6 20-23 will list those persons in that order in the XFILE=
<i>Decimal places</i>	controls the number of decimal places in the XFILE file.
<i>Set as default</i>	make this field selection the default setting for future Winsteps analyses. If "Set as default" is ignored, then you need to take ownership of file Winsteps.ini in your c:\Winsteps folder. See www.addictivetips.com/windows-tips/windows-7-access-denied-permission-ownership

19 Special Topics

19.1 Advice to novice analysts

Start with the [Diagnosis menu](#) ... Work your way down it.

In test construction, the guiding principle is "all items must be about the same thing, but then be as different as possible"! The central idea is that there is a latent variable which we are attempting to measure people on. The empirical definition of the latent variable is the content of the items. Essentially, we should be able to summarize on the items into a sentence which matches our intended definition of the latent variable. Latent variables can be very broad, e.g., "psychological state" or "educational achievement", or very narrow, e.g., "degree of paranoia" or "ability to do long division".

In other words, all items share something in common, but each item also brings in something that the others don't have.

Of course, this never happens perfectly. So what we need is:

(!) the CODES= statement contains the valid data codes in the data and nothing else.

(a) all items to point in the same direction, so that a higher rating (or "correct" answer) on the item indicates more of the latent variable. The first entry on the [Diagnosis menu](#) displays correlations. Items with negative correlations probably need their scoring reversed with `IVALUE=`.

(b) what the items share overwhelms what they don't share

(c) what the items don't all share, i.e., what is different about each of the items, is unique to each item or shared by only a few items.

What they all (or almost all) share, is usually thought of as the "test construct", the "major dimension", or the "Rasch dimension", or the "first factor in the data". This is what test validation studies focus on. Evaluating or confirming the nature of this construct.

What is unique to each item, or to clusters of only a few items, are "subdimensions", "secondary dimensions", "secondary contrasts", "misfit to the Rasch dimension", etc. We are concerned to evaluate: (i) are they a threat to scores/measures on the major dimension? (ii) do they manifest any useful information?

There are always as many contrasts in a test as there are items (less one). So how do we proceed?

(a) We want the first dimension to be much larger than all other dimensions, and for all items to have a large positive loading on it. This is essentially what the point-biserial correlation tells us in a rough way, and Rasch analysis tells us in a more exact way.

(b) We want so few items to load on each subdimension that we would not consider making that subdimension into a separate instrument. In practice, we would need at least 5 items to load heavily on a contrast, maybe more, before we would consider those items as a separate instrument. Then we crossplot and correlate scores or measures on the subdimension against scores on the rest of the test to see its impact.

(c) When a contrast has 2 items or less heavily loading on it, then it may be interesting, but it is only a wrinkle in this test. For instance, when we look at a two item contrast, we may say, "That is interesting, we could make a test of items like these!" But to make that test, we would need to write new items and collect more data. Its impact on this test is obviously minimal.

In reporting your results, you would want to:

(a) Describe, and statistically support, what most items share: the test construct.

(b) Identify, describe and statistically support, sub-dimensions big enough to be split off as separate tests. Then contrast scores/measures on those subdimensions with scores/measures on the rest of the test.

(c) Identify smaller sub-dimensions and speculate as to whether they could form the basis of useful new tests.

In all this, statistical techniques, like Rasch analysis and factor analysis, support your thinking, but do not do your thinking for you!

In what you are doing, I suggest that you choose the simplest analytical technique that enables you to tell your story, and certainly choose one that you understand!

Question: How do I know when to stop removing badly-fitting items and persons from my analysis?

Answer: You need to ask yourself a practical question: Why am I removing items and persons?

Is this your answer? "To improve the measurement of the persons!"

OK - then here is the strategy.

1. Estimate the person measures from the original analysis.

Remove whatever you see to be really, really bad.

2. Estimate the person measures again.

Cross-plot the two sets of person measures. Are there any noticeable changes that matter to you?

No. Then the really, really bad wasn't so bad after all. Keep everything. Stop here.

Yes. Now remove the really bad.

3. Estimate the person measures again.

Cross-plot the two sets of person measures (2. and 3.). Are there any noticeable changes that matter to you?

No. Then the really bad wasn't so bad after all. Analysis 2 is what you want. Stop here.

Yes. Now remove the somewhat bad.

4. Estimate the person measures again.

Cross-plot the two sets of person measures (3. and 4.). Are there any noticeable changes that matter to you?

No, then the somewhat bad wasn't so bad after all. Analysis 3 is what you want. Stop here.

Yes, (and so on)

It is usual to discover that the recommended fit criteria are much too tight. That is because those criteria were formulated by statisticians concerned about model-fit. They were not formulated by practical people who are concerned about measurement quality.

There is an interesting parallel to this in industrial quality-control. Part of the reason for the huge advance in the quality of Japanese cars relative to American cars was because Japanese quality-control focused on the overall quality of the car (which is what the consumer wants), while American quality-control focused on the quality of individual components (which is what the automotive engineers want).

19.2 Anchored estimation

How many anchor items or anchor persons?

The **percent** of anchor items is less important than the **number** of anchor items. We need enough anchor items to be statistically certain that a large incorrect value of one anchor item will not distort the equating. In my experiments, 10 were needed in order to prevent large accidental deviation in one from distorting the measurement. Please remember that the choice of 10 anchor items is before cleaning. We expect the 10 anchor items to malfunction somewhat. If we did not, only one anchor item would be needed.

Suppose that we have 10 anchor items. Then we discover one anchor item has changed its difficulty in the new test administration. In our analysis of the new test administration, we unanchor that item, so that there now only 9 anchor items. Have we failed the 10-anchor-item criterion? No! The remaining 9 anchor items are now better than the original 10 anchor items.

Anchoring or fixing parameter estimates (measures) is done with [IAFILE=](#) for items, [PAFILE=](#) for persons, and [SAFILE=](#) for response structures.

From the estimation perspective under [JMLE](#), anchored and unanchored items appear exactly alike. The only difference is that anchored values are not changed at the end of each estimation iteration, but unanchored estimates are. JMLE converges when "observed raw score = expected raw score based on the estimates". For anchored values, this convergence criterion is never met, but the fit statistics etc. are computed and reported by Winsteps as though the anchor

value is the "true" parameter value. Convergence of the overall analysis is based on the unanchored estimates. If large displacements are shown for the anchored items or persons, try changing the setting of [ANCESTIME](#).

Using pre-set "anchor" values to fix the measures of items (or persons) in order to equate the results of the current analysis to those of other analyses is a form of "common item" (or "common person") equating. Unlike common-item equating methods in which all datasets contribute to determining the difficulties of the linking items, the current anchored dataset has no influence on those values. Typically, the use of anchored items (or persons) does not require the computation of equating or linking constants. During an anchored analysis, the person measures are computed from the anchored item values. Those person measures are used to compute item difficulties for all non-anchored items. Then all non-anchored item and person measures are fine-tuned until the best possible overall set of measures is obtained. Discrepancies between the anchor values and the values that would have been estimated from the current data can be reported as [displacements](#). The standard errors associated with the displacements can be used to compute approximate [t-statistics](#) to test the hypothesis that the displacements are merely due to measurement error.

Example: I have a study in which an instrument is administered serially over time to each participant. I want to estimate the trait for subsequent occasions treating the baseline parameter estimates as fixed.

Here's an approach:

1. Format the so that each data row is a set of 7 items followed by a person label containing a time-point code (say, 01,02,03,04,...) in the first two columns and a participant number.
2. Analyze each time-point in a separate file or separately in a combined file using PSELECT=01 then PSELECT=02, etc. to verify that the data for each time-point is correctly entered.
3. Analyze the time-point 01 file or separately in a combined file using PSELECT=01. Output IFILE=01if.txt SFILE=01sf.txt
4. Analyze all the data in separate files or combined files with no PSELECT= and IAFILE=01if.txt SAFILE=01sf.txt
Output PFILE= filename.txt or PFILE=filename.xls containing a measure for every time-point + participant using the time-point 01 parameter values.

19.3 Andrich Thresholds and disordered rating categories

Andrich thresholds (step values) are parameters of the Rasch Rating-Scale and Partial-Credit models. They are the points on the latent variable where adjacent categories of the item are equally probable. Their locations are estimated primarily from the category frequencies. So, if the intermediate category has a relatively high frequency, the thresholds will advance, but if the intermediate category has a relatively low frequency, then the thresholds will be reversed (disordered). This is illustrated in a youtube video. https://www.youtube.com/watch?v=Rs3F7a6l8_0

If you need to interpret the partial-credit rating scale, e.g., 0-1-2, as "a person will go from 0 to 1 to 2 as they advance", then you need the items to have ordered thresholds, but if "a person will go from 0 to 1 or 2 to 2 as they advance" is good enough, then the threshold ordering does not matter. This is usually the case with partial-credit items. Regardless of the threshold ordering, we need the average ability of the persons who respond in each category to advance. This indicates that the categories represent advancing levels of the latent variable.

Summary: If the average measures for the person sample advance across categories for an item or rating scale, then the categories support the Rasch axiom that: "higher scores on the item <-> higher measures for the sample".

The chief purpose for collapsing categories is to enable inferences to be made at the item-category-level:

"as person measures increase, each category in turn is more probable than any one of the others to be observed".

If you want to make this statement, please collapse categories adjacent to disordered thresholds (and lose some measurement information). Otherwise, not.

"Disordering" is a contentious issue in Rasch analysis. It is often misunderstood by novice analysts, the type of analysts eager to get papers published and their feet firmly on the academic/professional ladder. More experienced analysts would tend to produce more ambivalent findings, suggesting alternate interpretations and actions. This type of finding is less likely to be submitted/accepted for publication because it appears to be wishy-washy. My own recommendation is usually that "threshold disordering" is a minor problem, (only relevant if category-level inferences are to be drawn from the data about individuals,) provided that "category disordering" (disordering of the substantive meanings of the categories) is not observed in the data. Unfortunately novice analysts may confuse "threshold disordering" with "category disordering" and so make incorrect statements about the data and the rating scales that generate it.

In my experience, category disordering is observed when

(1) raters are asked to rate in more categories than they can discriminate, e.g., "on a scale from 0-100, rate the person's cheerfulness".

(2) category definitions are not clearly ordered, e.g.,

1=never, 2=rarely, 3=occasionally, 4=sometimes, 5=often, 6=frequently, 7=always

(3) arbitrary rules distort the rating-scale, e.g., "if the behavior is not observed by the rater or not allowed or not possible, then rate the person 1=never". So that "does the person use the stairs" is rated "1" for all persons in a facility without stairs.

"Threshold disordering" occurs when a category corresponds to a narrow interval of the latent variable, e.g., "almost always" in

1=never, 2=sometimes, 3=often, 4=almost always, 5=always

Here the threshold between categories 3 and 4 will be disordered, even if the raters can clearly discriminate the 5 different levels.

Disordered Andrich thresholds indicate that a category occupies a narrow interval on the latent variable (usually because the category is observed relatively rarely). This is not a problem unless you need advancing abilities to probably increment smoothly up the categories of the rating scale without skipping narrow categories.

If this is a problem for you, then please collapse categories. This usually requires some experimenting to find the best solution:

- 1) Look at the average measures for each category. Combine categories with disordered average measures
- 2) Look at the category frequency. Combine categories with low frequencies

Disordered Andrich thresholds indicate that some categories on the latent variable are narrow. Disordered Andrich thresholds do not violate Rasch models, but they may impact our interpretation of how the rating scale functions.

Example: Imagine a location on the latent variable that is the boundary between two categories. If there are exactly 1,000 people with measures at that boundary. We would expect to observe 500 of them in categories below the boundary and 500 of them in categories above the boundary. Dichotomous items function exactly this way.

dichotomous item 0-1	
category 0 on latent variable	category 1 on latent variable
1000 persons here →	
500 persons observed here as 0	500 persons observed here as 1
dichotomous item difficulty →	

Polytomous items (RSM, PCM) are more complex. RSM and PCM predict that some of the 1000 will be observed in categories next to the boundary, and some in categories further away, so that there will be 500 in total above the boundary and 500 below the boundary. OK so far?

polytomous item 0-1-2 (ordered categories, ordered Andrich thresholds)		
category 0	category 1 (wide)	category 2
1000 persons here →	←	
500 persons observed here as 0	490 persons observed here as 1	10 persons observed here as 2
← Andrich threshold → (ordered)		

For 3 rating-scale categories 0,1,2 our first boundary is between 0 and 1. If category 1 is very wide, almost all the 500, say 490, will be observed in category 1 and 10 in category 2. The Andrich threshold corresponding to each boundary is basically $\ln(\text{frequency of category below} / \text{frequency of category above})$, so the Andrich threshold for our first boundary 0-1 is

something like $\ln(500/490) = 0.0$ and for the next boundary, 1-2, $\ln(490/10) = 3.9$. The Andrich thresholds are in ascending order.

polytomous item 0-1-2 (ordered categories, disordered Andrich thresholds)		
category 0	category 1 (narrow)	category 2
1000 persons here →	←	
500 persons observed here as 0	10 persons as 1	490 persons observed here as 2
→ Andrich threshold ← (disordered)		

But if category 1 is very narrow, only 10 of the 500 may be observed in category 1, and 490 in category 2 above the narrow category 1. The Andrich threshold for our first boundary 0-1 is something like $\ln(500/10) = 3.9$ and for the next boundary, 1-2, $\ln(10/490) = -3.9$. So the categories are disordered, even though the data fit the Rasch model perfectly!

Disordered Andrich thresholds indicate narrow categories on the latent variable. Statistically, these are no problem, but for practical purposes we may want all the categories to be wide enough that the Andrich thresholds are ordered.

The category boundaries in this example are reported as the "Rasch-Thurstone thresholds".

Effect of collapsing categories on measurement

In general, reducing the number of categories tends to improve the fit of the data to the model at the expense of losing some of the statistical information in the ratings. You will probably see the impact of the loss of information in the Person Reliability value.

For instance, if the person reliability drops from 0.9 to 0.8, we can use the Spearman-Brown Prophecy Formula to tell us what the loss of information is equivalent to in terms of items lost. If the original number of items is, say, 10, then reducing the number of categories in the rating scale is equivalent to reducing the number of items to:

$$\text{Items} = 10 * 0.8 * (1-0.9) / ((1-0.8)*0.9) = 4.4, \text{ so items lost} = 10 - 4.4 = 5.6$$

In general, we expect: $\text{items lost} = (\text{original item count}) * (\text{original category count} - \text{new category count}) / (\text{original category count} - 1)$

This only matters if the effect of the reduction in categories is to make the Person Reliability too small to discriminate the desired number of ability strata in the target person population. www.rasch.org/rmt/rmt63i.htm

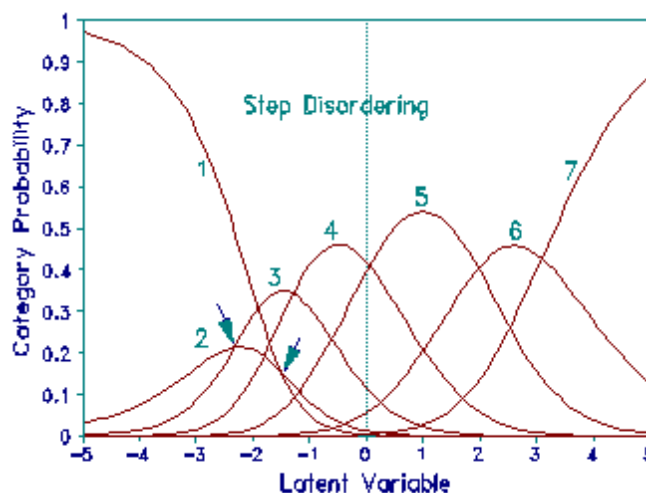
There is considerable debate in the Rasch community about the meaning of rating (or partial credit) scales and polytomies which exhibit "disorder". Look at [Table 3.2](#), distractor/option analysis. Two types of disorder have been noticed:

(i) Disorder in the "[average measures](#)" of the categories can imply disorder in the category definitions.

FIM LEVEL	COUNT	AVERAGE MEASURE	INFIT MNSQ	OUTFIT MNSQ	STEP CALIBRATN
1 (2)	88	-1.97	1.47	1.41	NONE
2 (1)	96	-2.18	.54	.69	-2.08
3	101	-.95	1.05	1.02	-1.49
4	168	-.25	.91	.99	-1.24
5	210	.80	.97	.87	.08
6	146	2.14	.66	.75	1.87
7	101	3.02	.83	.86	2.86

In this example, from Linacre, J.M. (1999) *Category Disordering vs. Step Disordering*, Rasch Measurement Transactions 13:1 p. 675, "FIM Level" categories have been deliberately disordered in the data. It is seen that this results in disordering of the "average measures" or "observed averages", the average abilities of the people observed in each category, and also large mean-square fit statistics. The "Rasch-Andrich thresholds", also called "step calibrations", "step difficulties", "step measures", "deltas", "taus", etc., remain ordered.

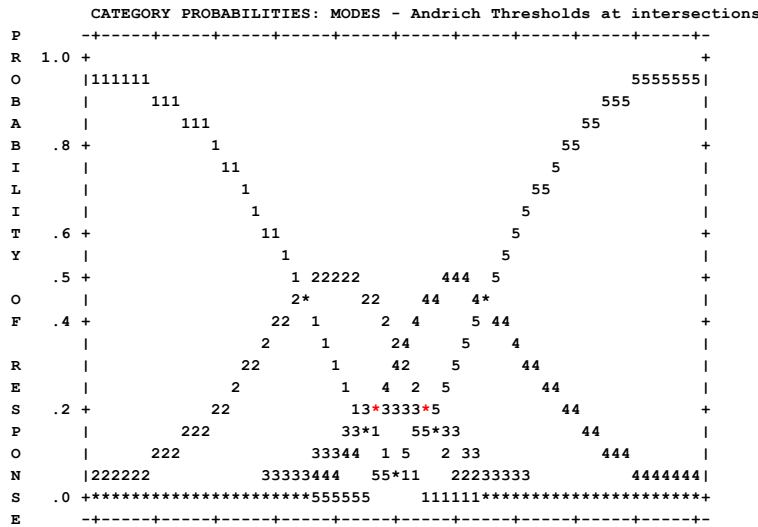
(ii) Disordered Rasch-Andrich thresholds imply less frequently observed intermediate categories, i.e., that they correspond to narrow intervals on the latent variable.



In this example, the FIM categories are correctly ordered, but the frequency of level 2 has been reduced by removing some observations from the data. Average measures and fit statistics remain well behaved. The disordering in the Andrich thresholds now reflects the relative infrequency of category 2. This infrequency is pictured in plot of probability curves which shows that category 2 is never a modal category in these data. The Andrich Threshold values do not indicate whether measurement would be improved by collapsing levels 1 and 2, or collapsing levels 2 and 3, relative to leaving the categories as they stand.

FIM LEVEL	COUNT	AVERAGE MEASURE	INFIT MNSQ	OUTFIT MNSQ	STEP CALIBRATN
1	96	-2.81	.90	.96	NONE
2	44	-1.96	.88	.92	-1.49
3	101	-1.03	1.02	.98	-2.33
4	168	-.30	1.07	1.22	-1.29
5	210	.82	.96	.88	.05
6	146	2.30	.75	.82	1.97
7	101	3.27	.87	.89	3.09

Example: Here are the category probability curves for a Likert rating scale: 1=Strongly disagree, 2=Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree. 3="Neutral" is relatively rarely observed. It has disordered thresholds. The point of equal probability between categories 3 and 4, the Rasch-Andrich threshold for category 4, is less than the The point of equal probability between categories 2 and 3, the Rasch-Andrich threshold for category 3.



There are several options:

1. Do not change the rating scale. We do this when recoding the categories in any way would confuse our audience, or the threshold disordering is not important for our inferences. For example: scoring Olympic Ice-Skating in [Exam 15.txt](#). This has a very long rating-scale for a small sample of skaters, so that there are many disordered thresholds and unobserved categories.
2. Rewrite the rating scale in the attitude survey without category 3: 1=Strongly disagree, 2=Disagree, 3=Agree, 4=Strongly Agree. This changes the meaning of all the categories, so we must then re-administer the survey.
3. Recode "Neutral" as missing (not administered) data. We do this when we think "Neutral" means "I do not want to tell you what I really think."
4. Recode "Neutral" as "Disagree". We do this when we want to be sure that "Agree" means that respondents truly agree.
5. Recode "Neutral" as "Agree". We do this when we want to be sure that "Disagree" means that respondents truly disagree.
6. Recode "Neutral" as the nearer of "Disagree" and "Agree" according to the empirical data. Please look at the "Observed Average Measures" (OBSVD AVRGE) for each category, and recode "Neutral" accordingly. In this example, Neutral (.12) is nearer to Disagree (-.96) than to Agree (1.60)

```

-----
|CATEGORY  OBSERVED|OBSVD | ANDRICH |
|LABEL SCORE COUNT %|AVRGE | THRESHOLD|
|-----+-----+
...
| 2  2    453 32|  -.96 |  -1.37 | 2 Disagree
| 3  3     35  2|  .12  |   0.58 | 3 Neutral
| 4  4    537 37|  1.60 |  -0.42 | 4 Agree
...
-----

```

19.4 Arithmetical expressions

Many input values can be entered as expressions, for instance
USCALE = 100*2

Allowed operators are: + - * /

1) - as a range separator takes precedence over - as a negative.

IAFILE=*

1-10 2-2

*

1-10 is the item range 1 to 10, and then 2-2 is the value 0.

2) no blanks or tabs are allowed between operators and numbers:

valid: 2*2

invalid: 2 * 2

3) error codes include:

1: one of ^*/-+ followed by one of /^

2: one of ^*/-+" followed by *, or ***

4: floating point operators d, e, D, E must be followed by + or -

8: leading or trailing operators not allowed: ^*/-+

9: division by zero /0

10: operator ^*/-+ expected between numbers. No blanks allowed.

11: parenthesis or blank in unexpected place

13: unexpected character, such as A, in your arithmetic expression

19.5 Automating file selection

Use the Winsteps ["Batch"](#) pull-down menu to do this.

Assigning similar names to similar disk files can be automated using Batch commands.

For example, suppose you want to analyze your data file, and always have your output file have suffix "O.TXT", the PFILE have suffix ".PF" and the IFILE have suffix ".IF". Key in your control file and data, say "ANAL1", omitting PFILE= and IFILE= control variables. Then key in the following Batch script file, called, say, MYBATCH.BAT (for Windows-98 etc.) or MYBATCH.COM (for Windows-2000 etc.), using your word processor, saving the file as an ASCII or DOS text file:

```
REM the MYBATCH.BAT batch file to automate Winsteps
START /w ..\Winsteps BATCH=YES %1 %1O.TXT PFILE=%1.PF IFILE=%1.IF
```

For Winsteps, specify BATCH=YES to close all windows and terminate the program when analysis is complete.

To execute this, type at the DOS prompt (or using the Winsteps "Batch" pull-down menu:

```
C:> MYBATCH ANAL1 (Press Enter Key)
```

This outputs the tables in ANAL1O.TXT, PFILE= in ANAL1.PF and IFILE= in ANAL1.IF.

You can also edit the files WINBATCH.BAT or WINBATCH.COM. These can be executed from the DOS prompt or from the Winsteps Batch pull-down menu. See [Running Batch Files](#).

19.6 Average measures, distractors and rating scales

The "average measure" for a category is the average ability of the people who respond in that category or to that distractor (or distractor. The term "distractor" has been in use since at least 1934). This is an empirical value. It is **not** a Rasch-model parameter.

The Rasch-Andrich threshold (step difficulty, step calibration, etc.) is an expression of the log-odds of being observed in one or other of the adjacent categories. This is a model-based value. It is a Rasch-model parameter.

In Table 2.5, 14.3 and similar Tables describing the items, the "observed average" measure is: sum (person abilities) / count (person abilities) for each response option or rating-scale category.

In Table 3.2 and similar Tables describing the response structures, the "observed average" measure is: sum (person abilities - item difficulties) / count (person abilities) for each response option or rating-scale category.

Our theory is that people who respond in higher categories (or to the correct MCQ option) should have higher average measures. This is verified by "average measure".

Often there is also a theory about the rating scale, such as "each category in turn should be the most probable one to be observed as one advances along the latent variable." If this is your theory, then the "step difficulties" should also advance. But alternative theories can be employed. For instance, in order to increase item discrimination one may deliberately over-

categorize a rating scale - visual-analog scales are an example of this. A typical visual analog-scale has 101 categories. If these functioned operationally according to the "most probable" theory, it would take something like 100 logits to get from one end of the scale to the other.

The relationship between "average measure" and Andrich thresholds or "item difficulties" is complex. It is something like:

Andrich threshold = $\log \left(\frac{\text{count in lower category}}{\text{count in higher category}} \right) + (\text{average of the measures across both categories}) - \text{normalizer}$

normalized so that: $\text{sum}(\text{Andrich thresholds}) = 0$

So that,

the higher the frequency of the higher category relative to the lower category, the lower (more negative) the Andrich threshold (and/or item difficulty)

and the higher the average of the person measures across both categories, the higher (more positive) the Andrich threshold (and/or item difficulty)

but the Andrich thresholds are estimated as a set, so that the numerical relationship between a pair of categories is influenced by their relationships with every other category. This has the useful consequence that even if a category is not observed, it is still possible to construct a set of Andrich thresholds for the rating scale as a whole.

Suggestions based on researcher experience:

In general, this is what we like to see:

- (1) More than 10 observations per category (or the findings may be unstable, i.e., non-replicable)
- (2) A smooth distribution of category frequencies. The frequency distribution is not jagged. Jaggedness can indicate categories which are very narrow, perhaps category transitions have been defined to be categories. But this is sample-distribution-dependent.
- (3) Clearly advancing average measures. The average measures are not [disordered](#).
- (4) Average measures near their expected values.
- (5) Observations fit with their categories: Outfit mean-squares near 1.0. Values much above 1.0 are much more problematic than values much below 1.0.

19.7 Batch mode example: Score Tables

Winsteps produces a Score Table in Table 20 which is for responses to every active item. You can use the Batch File feature to automate the production of Score Tables for subtests of items.

(1) Do the calibration run:

```
title="Item calibration"
ni=3           ; these match your data file
item1=1
name1=1
codes=123
ISGROUPS=0
ifile= if.txt  ; item calibrations
sfile= sf.txt  ; rating (or partial credit) scale structure
data=data.txt
&end
.....
END LABELS
```

(2) Set up the control file for the batch run:

```
; This is scon.txt
title="Produce score table"
ni=3           ; match your data file
item1=1
name1=1
codes=123
```

```

ISGROUPS=0
iafile = if.txt ; item anchor file
safile = sf.txt ; rating (or partial credit) scale structure anchor file
pafile=*
1 0 ; dummy person measures
2 0
*
CONVERGE=L ; only logit change is used for convergence
LCONV=0.005 ; logit change too small to appear on any report.
&end
....
END LABELS
121.... ; two lines of dummy data - such that every item has a non-extreme score.
212....

```

(3) Set up the Batch (.bat or .cmd) file and [run it](#). Use IDFILE=* to select the items you want (or ISELECT=)

```

rem this is score.cmd
del sc*.txt
start /w ..\Winsteps sc.txt dummy SCOREFILE=sc001.txt batch=yes idfile=* +3 * title=item3
start /w ..\Winsteps sc.txt dummy SCOREFILE=sc010.txt batch=yes idfile=* +2 * title=item2
start /w ..\Winsteps sc.txt dummy SCOREFILE=sc100.txt batch=yes idfile=* +1 * title=item1
start /w ..\Winsteps sc.txt dummy SCOREFILE=sc011.txt batch=yes idfile=* +2 +3 * title=items23
start /w ..\Winsteps sc.txt dummy SCOREFILE=sc101.txt batch=yes idfile=* +1 +3 * title=item13
start /w ..\Winsteps sc.txt dummy SCOREFILE=sc110.txt batch=yes idfile=* +1 +2 * title=items12
start /w ..\Winsteps sc.txt dummy SCOREFILE=sc111.txt batch=yes idfile=* +1 +2 +3 *
        title=items123
.....
copy sc*.txt scores.txt

```

(4) The Score Tables are in file scores.txt

Batch command to select one line from a file.

Save the code below as oneline.bat

Then

c:>oneline 23 infile.txt >>outfile.txt

selects line 24

```

@echo off
rem select one line from a file
setlocal enabledelayedexpansion
if [%1] == [] goto usage
if [%2] == [] goto usage
for /f "skip=%1 delims=" %%a in (%2) do (
echo %%a
goto exit
)
goto exit
:usage
echo Usage: oneline.bat linebefore infile >>outfile
:exit

```

19.8 Bayesian prior distributions

Winsteps does not support Bayesian prior distributions directly, but the effect of a Bayesian prior on the final estimates can be obtained in other ways. However,

"For the Rasch model, an explicit pattern [= prior distribution] that could be used to obtain more accurate item parameter estimates was not found." (Karadavut, T. (2019). The Uniform Prior for Bayesian Estimation of Ability in Item Response Theory Models. *International Journal of Assessment Tools in Education*, 6(4), 568-579.)

Example 1: You have predicted the item difficulties based on your construct theory. You want to use these predicted difficulties as Bayesian Priors on the estimation of the empirical Item Difficulties

1. Constructing some dummy person records. These persons would respond to your items in the way predicted by your desired Bayesian prior.
2. Analyze the dummy person records by themselves. Add and alter dummy person records until the estimated item difficulties match your Bayesian predictions.
3. Analyze the dummy records with your actual sample. [PWEIGHT=](#) the dummy records until the influence of the dummy records on the final estimates is the same as your desired hypothetical Bayesian prior.

Illustration: There are 6 items. Our theory says: items 1-2 are easy, items 3-4 are moderate, items 5-6 are hard. Dummy data records could be these. Notice that they all have the same raw score. Fit to the Rasch model is irrelevant for dummy data.

```
101010
110100
110100
011001
```

Example 2: You want to impose a Bayesian prior on the person-ability distribution.

There are usually two noticeable effects of a Bayesian prior:

- (i) extreme person scores participate in the estimation of item difficulties in the same way as non-extreme scores.
- (ii) the person-ability estimates are more central with a smoother distribution.

These two effects can be obtained by adding item records to the dataset.

1. Analyze the dataset without the Bayesian prior. Verify that all is correct.
2. Define two more dichotomous items in the control file. These are dummy items. Use [ISGROUPS=](#) to place the dummy items in a separate item-group.
3. Append to each person's response string two responses. "01" for the first half of the person sample. "10" for the second half.

Illustration: the original control file has 100 dichotomous items and 500 persons:

```
CODES=01
```

```
NI=100
```

the Bayesian control file has 102 dichotomous items

```
CODES=01
```

```
NI=102 ; two more items
```

```
ISGROUPS=*
```

```
1-100 A ; all the original dichotomous items in one group
```

```
101-102 D ; the dummy items in their own group
```

```
*
```

```
EDFILE=\* ; using EDFILE=, we don't change the actual data
```

```
1-250 101 0
```

```
1-250 102 1
```

```
251-500 101 1
```

```
251-500 102 0
```

```
*
```

```
IWEIGHT=*
```

```
101 1
```

```
102 1
```

```
*
```

4. Analyze the enlarged dataset. Verify that (i) all previously extreme scores are not shown as extreme, [Table 17](#), and (ii) the person-ability S.D. has reduced. We expect that the fit statistics for all the items, except for the two dummy items, are tending towards overfit (lower mean-squares).

5. Adjust the weighting of the dummy items using [IWEIGHT=](#) until the influence of the dummy records on the final estimates is the same as your desired hypothetical Bayesian prior.

19.9 Between-group and Within-group fit statistics

Richard Smith (1982, 2009) proposes group-level fit statistics. These statistics indicate more or less consistency across CLASSES. Lack of consistency may indicate DIF or DPF or some other interaction across CLASSES or inside of CLASSES.

For each item, the persons are segmented into CLASSES using [DIF=](#)

[Table 30.4](#) - Between-group fit statistic for person CLASSES:

"This fit statistic is based on the difference between an item's observed score on a subset of persons and the score on that subset as predicted by the item's overall difficulty estimate."

[Table 30.5](#) - Within-group outfit statistic for a person CLASS for an item:

The outfit mean-square statistic for the item computed only across persons in the person CLASS.

[Table 30.5](#) - Within-group infit statistic for an item CLASS for a person:

The infit mean-square statistic for the item computed only across persons in the person CLASS.

For each person, the items are segmented into CLASSES using [DPF=](#)

[Table 31.4](#) - Between-group fit statistic for item CLASSES:

"This fit statistic is based on the difference between a person's observed score on a subset of items and the score on that subset as predicted by the person's overall ability estimate."

[Table 31.5](#) - Within-group outfit statistic for an item CLASS for a person:

The outfit mean-square statistic for the person computed only across items in the item CLASS.

[Table 31.5](#) - Within-group infit statistic for an item CLASS for a person:

The infit mean-square statistic for the person computed only across items in the item CLASS.

Smith RM, Hedges LV. Comparison of Likelihood Ratio χ^2 and Pearsonian χ^2 Tests of Fit in the Rasch Model. [Education Research and Perspectives. 1982; 9\(1\):44-54](#)

Smith RM, Plackner C. The family approach to assessing fit in Rasch measurement. [Journal of Applied Measurement. 2009;10\(4\):424-37](#)

19.10 Biserial correlation

If the sample is normally distributed (i.e., conditions for the computation of the biserial exist), then to obtain the biserial correlation from the point-biserial for dichotomous data:

Biserial = Point-biserial * f(proportion-correct-value)

Example: Specify [PTBISERIAL=Yes](#) and [PVALUE=Yes](#). Display [Table 14](#).

ENTRY	RAW		MODEL	INFIT	OUTFIT	PTBSE	P-				
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	VALUE	TAP
8	27	34	-2.35	.54	.59	-1.3	.43	-.2	.65	.77	1-4-2-3

Point-biserial = .65. proportion-correct-value = .77. Then, from the Table below, f(proportion-correct-value) = 1.3861, so Biserial correlation = .65 * 1.39 = 0.90

Here is the Table of proportion-correct-value and f(proportion-correct-value).

p-va	f(p-val)	p-va	f(p-val)
0.99	3.7335	0.01	3.7335
0.98	2.8914	0.02	2.8914
0.97	2.5072	0.03	2.5072
0.96	2.2741	0.04	2.2741
0.95	2.1139	0.05	2.1139
0.94	1.9940	0.06	1.9940
0.93	1.8998	0.07	1.8998

0.92	1.8244	0.08	1.8244
0.91	1.7622	0.09	1.7622
0.90	1.7094	0.10	1.7094
0.89	1.6643	0.11	1.6643
0.88	1.6248	0.12	1.6248
0.87	1.5901	0.13	1.5901
0.86	1.5588	0.14	1.5588
0.85	1.5312	0.15	1.5312
0.84	1.5068	0.16	1.5068
0.83	1.4841	0.17	1.4841
0.82	1.4641	0.18	1.4641
0.81	1.4455	0.19	1.4455
0.80	1.4286	0.20	1.4286
0.79	1.4133	0.21	1.4133
0.78	1.3990	0.22	1.3990
0.77	1.3861	0.23	1.3861
0.76	1.3737	0.24	1.3737
0.75	1.3625	0.25	1.3625
0.74	1.3521	0.26	1.3521
0.73	1.3429	0.27	1.3429
0.72	1.3339	0.28	1.3339
0.71	1.3256	0.29	1.3256
0.70	1.3180	0.30	1.3180
0.69	1.3109	0.31	1.3109
0.68	1.3045	0.32	1.3045
0.67	1.2986	0.33	1.2986
0.66	1.2929	0.34	1.2929
0.65	1.2877	0.35	1.2877
0.64	1.2831	0.36	1.2831
0.63	1.2786	0.37	1.2786
0.62	1.2746	0.38	1.2746
0.61	1.2712	0.39	1.2712
0.60	1.2682	0.40	1.2682
0.59	1.2650	0.41	1.2650
0.58	1.2626	0.42	1.2626
0.57	1.2604	0.43	1.2604
0.56	1.2586	0.44	1.2586
0.55	1.2569	0.45	1.2569
0.54	1.2557	0.46	1.2557
0.53	1.2546	0.47	1.2546
0.52	1.2540	0.48	1.2540
0.51	1.2535	0.49	1.2535
0.50	1.2534	0.50	1.2534

To obtain the biserial correlation from a point-biserial correlation, multiply the point-biserial correlation by $\sqrt{\text{proportion-correct-value} \cdot (1 - \text{proportion-correct-value})}$ divided by the normal curve ordinate at the point where the normal curve is split in the same proportions.

There is no direct relationship between the point-polyserial correlation and the polyserial correlation.

19.11 Bonferroni - Multiple t-tests

The exact statement of your null hypothesis determines whether a Bonferroni correction applies. If you have a list of t-tests and a significant result for even one of those t-tests rejects the null-hypothesis, then Bonferroni correction (or similar).

Let's assume your hypothesis is "this instrument does not exhibit DIF", and you are going to test the hypothesis by looking at the statistical significance probabilities reported for each t-test in a list of t-tests. Then, by chance, we would expect 1 out of every 20 or so t-tests to report $p \leq .05$. So, if there are more than 20 t-tests in the list, then $p \leq .05$ for an individual t-test is a meaningless significance. In fact, if we don't see at least one $p \leq .05$, we may be surprised!

The Bonferroni correction says, "if any of the t-tests in the list has $p \leq .05/(\text{number of t-tests in the list})$, then the hypothesis is rejected".

What is important is the number of tests, not how many of them are reported to have $p \leq .05$.

If you wish to make a [Bonferroni](#) multiple-significance-test correction, compare the reported significance probability with your chosen significance level, e.g., .05, divided by the number of t-tests in the Table. According to Bonferroni, if you are testing the null hypothesis at the $p \leq .05$ level: "There is no effect in this test." Then the most significant effect must be $p \leq .05 / (\text{number of item DIF contrasts})$ for the null hypothesis of no-effect to be rejected.

Question: Winsteps Tables report many t-tests. Should Bonferroni adjustments for multiple comparisons be made?

Reply: It depends on how you are conducting the t-tests. For instance, in [Table 30.1](#). If your hypothesis (before examining any data) is "there is no DIF for this CLASS in comparison to that CLASS on this item", then the reported probabilities are correct.

If you have 20 items, then one is expected to fail the $p \leq .05$ criterion. So if your hypothesis (before examining any data) is "there is no DIF in this set of items for any CLASS", then adjust individual t-test probabilities accordingly.

In general, we do not consider the rejection of a hypothesis test to be "substantively significant", unless it is both very unlikely (i.e., statistically significant) and reflects a discrepancy large enough to matter (i.e., to change some decision). If so, even if there is only one such result in a large data set, we may want to take action. This is much like sitting on the proverbial needle in a haystack. We take action to remove the needle from the haystack, even though statistical theory says, "given a big enough haystack, there will probably always be a needle in it somewhere."

A strict Bonferroni correction for n multiple significance tests at joint level α is α/n for each single test. This accepts or rejects the entire set of multiple tests. In an example of a 100 item test with 20 bad items ($.005 < p < .01$), the threshold values for cut-off with $p \leq .05$ would be: $p \leq .00005$, so that the entire set of items is accepted.

Benjamini and Hochberg (1995) suggest that an incremental application of Bonferroni correction overcomes some of its drawbacks. Here is their procedure:

- i) Perform the n single significance tests.
- ii) Number them in ascending order by probability $P(i)$ where $i=1,n$ in order.
- iii) Identify k , the largest value of i for which $P(i) \leq \alpha * i/n$ where $\alpha = .05$ or $\alpha = .01$
- iv) Reject the null hypothesis for $i = 1, k$

In an example of a 100 item test with 20 bad items (with $.005 < p < .01$), the threshold values for cut-off with $\alpha = .05$ would be: 0.0005 for the 1st item, .005 for the 10th item, .01 for the 20th item, .015 for the 30th item. So that k would be at least 20 and perhaps more. All 20 bad items have been flagged for rejection.

Benjamini Y. & Hochberg Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, 57,1, 289-300.

Example of whether to Bonferroni or not ...

Hypothesis 1: There is no [DIF](#) between men and women on item 1.

This is tested for item 1 in [Table 30.1](#)

or there is no [DPF](#) between "addition items" and "subtraction items" for George

This is tested for George in [Table 31.1](#)

Hypothesis 2: There is no DIF between men and women on the 8 items of this test.

Bonferroni correction:

Look at the 8 pair-wise DIF tests in Table 30.1

Choose the smallest p-value = p

Divide it by $0.5/8 = p/8$

If $p \leq .05/8$ then reject hypothesis 2.

Or there is no DPF between "addition items" and "subtractions items" across the 1000 persons in the sample. - Bonferroni applied to Table 31.1.

Question: "Does this mean that if one or a few t-tests turn out significant, you should reject the whole set of null hypotheses and you cannot tell which items have DIF?"

Answer: You are combining two different hypotheses. Either you want to test the whole set (hypothesis 2) or individual items (hypothesis 1). In practice, we want to test individual items. So Bonferroni does not apply.

Let's contrast items (each of which is carefully and individually constructed) against a random sample from the population.

We might ask: "Is there Differential Person Functioning by this sample across these two types of items?" (Hypothesis 2 - Bonferroni) because we are not interested to (and probably cannot) investigate individuals.

But we (and the lawyers) are always asking "is there Differential Item Functioning on this particular item for men and women?" (Hypothesis 1 - not Bonferroni).

19.12 Bradley-Terry model and paired comparisons

Follow "Paired Comparison with Standard Rasch Software" - www.rasch.org/rmt/rmt113o.htm - see Winsteps command: `PAIRED=Yes`

19.13 Category boundaries and thresholds

Conceptualizing rating scales and partial-credit response structures for communication can be challenging. Rasch measurement provides several approaches. Choose the one that is most meaningful for you.

Look at this edited excerpt of Table 3.2:

DATA			QUALITY CONTROL			Andrich		EXPECTATION		50% Cumulat	
Category	Counts	Cum.	Avg	Exp.	OUTFIT	Thresholds	Measure	at	Probability		
Score	Used	%	%	Meas	Meas	MnSq	Measure	S.E.	Category	Zone	at
0	891	6%	6%	-.04	-.07	1.3			(-2.27)		low
1	383	3%	9%	.47	.51	1.3	-.91	.04	-.68	-1.53	-1.23
2	1017	7%	15%	1.07	1.17	.8	-.15	.04	.63	-.03	-.07
3	12683	85%	100%	2.15	2.14	1.0	1.06	.03	(2.34)	1.55	1.29
----- (Mean) ----- (Median) -----											

Here are three ways of conceptualizing and communicating the transition, threshold, boundary between category 1 and category 2:

1.1) Rasch-half-point thresholds (Zone). Someone at the boundary between "1" and "2" would have an expected rating of 1.5, or 1000 persons at the boundary between "1" and "2" would have an average rating of 1.5. This boundary is the "zone" = "Expected Measure at 1+0.5 or 2-0.5" which is -.03 logits, the Rasch-half-point threshold. To illustrate this, use the model item characteristic curve. The expected score ogive / model ICC (Table 21.2 - second on list in Graphs menu). The CAT+.25, CAT-0.5, AT CAT, and CAT-.25 columns in the `ISFILE=` plot points on this ogive. The expected score ogive relates most directly to the estimation of the Rasch parameters. Since it is only one line, it is also convenient for summarizing performance at any point on the latent variable by one number. Crucial points are the points on the variable corresponding to the lower category value + 0.5, i.e., more than the higher adjacent category value - 0.5. These Rasch-half-point thresholds are "average score thresholds" or "Rasch-ICC thresholds".

1.2) Rasch-Thurstone thresholds (50% Cumulative Probability). Someone at the boundary between "1" and "2" would have a 50% chance of being rated 1 or below, and a 50% chance of being rated 2 or above. This is the Rasch-Thurstone threshold of -.07 logits. To illustrate this, use the cumulative probability curves. The cumulative probability curves (Table 21.3 - and third on list in Graphs menu). The 50%PRB columns in the `ISFILE=` are the crucial points on these curves. and are the Rasch-Thurstone thresholds, useful for identifying whether a person is most likely to respond below, at or above a certain category.

1.3) Rasch-Andrich thresholds. Someone at the boundary between "1" and "2" would have an equal chance of being rated 1 or 2. This is the Rasch--Andrich Threshold of $-.15$ logits. To illustrate this, use the category probability curves. The probability curves (Table 21.1 - and top of list in Graphs menu). The Andrich Threshold in the [ISFILE=](#) gives the point of equal probability between adjacent categories. The points of highest probability of intermediate categories are given by the AT CAT values. These probability curves relate most directly to the Rasch parameter values, also called Rasch-Andrich thresholds. They are at the intersection of adjacent probability curves, and indicate when the probability of being observed in the higher category starts to exceed that of being observed in the adjacent lower one. This considers the categories two at a time, but can lead to misinference if there is Rasch-Andrich threshold [disordering](#).

Here are two ways of conceptualizing the measure corresponding to a category.

2.1) The peak of the category probability curve and

2.2) The point on the latent variable at which the expected score on the item has the value of the category number.

2.1) and 2.2) are the same location on the latent variable. For extreme categories, these locations are infinite, so they are usually replaced by the location for 0.25 score units away from the extreme. For example, the locations on the latent variable for a 1,2,3 rating scale are the measures corresponding to 1.25, 2, 2.75.

3.1) Empirical average measures. For any particular sample, there is the average ability of the people who scored in any particular category of any particular item. This is the "[Average Measure](#)" reported in Table 3.2. This is entirely sample-dependent. It is not reported in ISFILE=. In the Table, the empirical average measure of the persons responding in category 1 is $.47$ logits, and in category 2 is 1.07 logits. An empirical boundary between these two categories could be half-way between those values, which is $.77$ logits.

In summary,

What some authors call the "Category Gap" is the interval on the latent variable in which the category is the most likely category to be observed. Ben Wright used to call this the "modal category interval", because the mode is the most likely value in a distribution to be observed. The interval is between Andrich thresholds, and requires advancing Andrich thresholds, because the threshold is the point on the latent variable at which adjacent categories are equally probable to be observed. This interval is particularly useful if we want to make the interpretation "ability level \Leftrightarrow most likely category on the item".

There are also "mean category intervals", which is the interval on the latent variable in which the expected score (= average score for an ability level) is in the range (category number - 0.5) to (category number + 0.5). The expected scores on the items are used in the Rasch-estimation process. They are also useful when the interpretation is something like "this ability level averages 3 on this item", "ability level \Leftrightarrow average score on the item". The ends of this category interval can be called the half-point thresholds.

The "median category interval" is the interval on the latent variable from the location at which there is a 50% or more chance of being observed in a category below the target category up to a 50% or more chance of being observed above the target category. The ends of this interval are called the Thurstone Thresholds. The Thurstone Thresholds are particularly useful when there are many rarely observed categories in a rating scale. This causes the Andrich thresholds to be jumbled and uninterpretable. The Thurstone Thresholds are always ordered. "ability level \Leftrightarrow more than the categories below, less than the categories above."

19.14 Category mean-square fit statistics

Computation of Category Mean-Square fit statistics:

For all observations in the data:

X_{ni} is the observed value

E_{ni} is the expected value of X_{ni}

W_{ni} is the model variance of the observation around its expectation

P_{nik} is the probability of observing $X_{ni}=k$

Category Outfit Mean-Square statistic for category $k =$

$[\text{sum } ((k-E_{ni})^2/W_{ni}) \text{ for all } X_{ni}=k] / [\text{sum } (P_{nik} * (k-E_{ni})^2/W_{ni}) \text{ for all } X_{ni}]$

Category Infit statistic for category $k =$

$$[\text{sum } ((k-E_{ni})^2) \text{ for all } X_{ni}=k] / [\text{sum } (P_{nik} * (k-E_{ni})^2) \text{ for all } X_{ni}]$$

The expected values of the category mean-squares are 1.0.

For dichotomies, with categories 0 and 1:

For category 0: $k = 0, E_{ni} = P_{ni} = 1 - P_{ni}$

Category Outfit Mean-Square statistic for category 0
 $= [\text{sum } (P_{ni}/(1-P_{ni})) \text{ for all } X_{ni}=0] / [\text{sum } (P_{ni}) \text{ for all } X_{ni}]$

Category Infit statistic for category 0
 $= [\text{sum } (P_{ni}^2) \text{ for all } X_{ni}=0] / [\text{sum } ((1-P_{ni}) * P_{ni}^2) \text{ for all } X_{ni}]$

For category 1: $k = 1$

Category Outfit Mean-Square statistic for category 1
 $= [\text{sum } ((1-P_{ni})/P_{ni}) \text{ for all } X_{ni}=1] / [\text{sum } (1-P_{ni}) \text{ for all } X_{ni}]$

Category Infit statistic for category 1
 $= [\text{sum } ((1-P_{ni})^2) \text{ for all } X_{ni}=1] / [\text{sum } (P_{ni} * (1-P_{ni})^2) \text{ for all } X_{ni}]$

19.15 Chi-squares, d.f., probabilities and Mean-squares

Winsteps reports mean-square fit statistics. These are chi-square fit statistics divided by their degrees of freedom.

Mean-squares: the expected value of a mean-square is 1.0. When the mean-square = 1.0, then the unexplained variance in the data is the same size as the unexplained variance (randomness) predicted by the Rasch model. Mean-squares < 1.0 indicate over-fit of the data to the model = the Rasch model is explaining the data too well. Mean-squares > 1.0 indicate under-fit of the data to the model = there is too much unexplained variance in the data.

Unexpectedness: Winsteps test the hypothesis that the mean-square = 1.0 except for the predicted statistical variability. The outcome of the hypothesis test is reported as a unit-normal deviate, ZSTD. Values outside ± 1.96 indicate $p < .05$ for the double-sided test, and the null hypothesis of "model fit" is rejected. Always investigate underfit (Mean-square > 1.0, ZSTD > 0) before overfit. Underfit is a much greater threat to measurement than overfit.

Chi-squares, d.f., probabilities: instead of mean-squares and ZSTD, you may want to report these.

Chi-square = Mean-square * d.f.

d.f. are not usually integers, because of the ordinal, non-linear nature of the data.

To report probabilities, LOCAL=Prob. For instance, in Table 14:

ENTRY	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	MNSQ	INFIIT PROB	OUTFIT PROB	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	ACT	
1	109	75	-.40	.21	.55	.0005	.49	.0105	.64	.49	77.0	61.7	Watch birds
2	116	75	-.71	.22	.93	.7009	.72	.3107	.58	.46	74.3	64.4	Read books on animals

To report probabilities, LOCAL=Prob, and d.f., IFILE= or PFILE=, and the [Field Selection Dialog Box](#)

ENTRY	MEASURE	ST	COUNT	SCORE	MODLSE	IN.MSQ	IN.PRO	OUT.MS	OUT.PR	INDF	OUTDF	G	M	R	NAME
1	-.40	1	75.0	109.0	.21	.55	.0005	.49	.0105	82.54	34.51	1	R	.	Watch birds
2	-.71	1	75.0	116.0	.22	.93	.7009	.72	.3107	72.81	26.16	1	R	.	Read books on animals

To report Chi-squares in IFILE= or PFILE=, instead of Mean-squares: LOCAL=Prob, Chisquare = Yes

ENTRY	MEASURE	ST	COUNT	SCORE	MODLSE	IN.CHI	IN.PRO	OUT.MS	OUT.PR	INDF	OUTDF	G	M	R	NAME
1	-.40	1	75.0	109.0	.21	45.00	.0005	17.05	.0105	82.54	34.51	1	R	.	Watch birds
2	-.71	1	75.0	116.0	.22	67.60	.7009	18.92	.3107	72.81	26.16	1	R	.	Read books on animals

19.16 CMLE, MMLE, PMLE: Using estimates from other software in Winsteps

Winsteps does Joint Maximum Likelihood Estimation (JMLE). As with all estimation methods, JMLE has strengths and weaknesses. You may prefer CMLE ([eRm R Statistics](#)), MMLE ([ConQuest](#)), PMLE ([RUMM2030](#)) or other estimates, but you also want the comprehensive analysis and reporting features of Winsteps. Here is what to do:

1) Analyze your data with the other software. It may be convenient to pre-process it with Winsteps, then output an [RFILE=](#) or [IPMATRIX=](#) for analysis by the other software.

2) From the other software, output the item difficulties (deltas), person abilities (thetas) and, for polytomies, the Andrich thresholds (taus).

3) Anchor the estimates for Winsteps. Format the item difficulties as an [JAFILE=](#) anchor text file, the person abilities as a [PAFILE=](#) and the thresholds as an [SAFILE=](#).

4) Analyze your data with Winsteps using those anchor files. All the analysis and reporting features of Winsteps are available.

Example: CMLE using R Statistics eRm package (Patrick Mair, *et al.*)

Winsteps control and data file	Item (delta) JMLE estimates	Person (theta) JMLE estimates																														
<pre> title=4x4 ni=4 name1=1 item1=1 codes=01 &END END LABELS 1000 0110 0111 0011 </pre>	<pre> ----- ENTRY TOTAL TOTAL NUMBER SCORE COUNT MEASURE ----- 1 1 4 1.26 2 2 4 .00 3 3 4 -1.26 4 2 4 .00 </pre>	<pre> ----- ENTRY TOTAL TOTAL NUMBER SCORE COUNT MEASURE ----- 1 1 4 -1.25 2 2 4 .00 3 3 4 1.26 4 2 4 .00 </pre>																														
Data are symmetric Estimates are symmetric	Range of JMLE delta estimates = 2.52 logits	Range of JMLE theta estimates = 2.51 logits																														
eRm data file	Item (delta) CMLE (eRm)	Person (theta) AMLE (eRm)																														
<pre> X1 X2 X3 X4 1 1 0 0 0 2 0 1 1 0 3 0 1 1 1 4 0 0 1 1 </pre>	<pre> Estimate delta X1 0.955 delta X2 0.000 delta X3 -0.955 delta X4 0.000 </pre>	<pre> Estimate theta P1 -1.208 theta P2 0.000 theta P3 1.208 theta P4 0.000 </pre>																														
Data are symmetric Estimates are asymmetric	Range of CMLE delta estimates = 1.91 logits JMLE bias correction = 1.91/2.52 = 0.76 predicted bias correction = (4-1)/4 = 0.75	Range of AMLE theta estimates = 2.42 logits Almost the same as the JMLE estimates AMLE = Anchored MLE																														
Response probabilities: CMLE probabilities are almost symmetric AMLE probabilities are obviously biased. The item totals are wrong.	at end of the item CMLE computation <table border="1"> <thead> <tr> <th></th> <th>X1</th> <th>X2</th> <th>X3</th> <th>X4</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td>P1</td> <td>0.077</td> <td>0.201</td> <td>0.521</td> <td>0.201</td> <td>1.000</td> </tr> </tbody> </table>		X1	X2	X3	X4	Total	P1	0.077	0.201	0.521	0.201	1.000	after AMLE computation (eRm) <table border="1"> <thead> <tr> <th></th> <th>X1</th> <th>X2</th> <th>X3</th> <th>X4</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td>P1</td> <td>0.10</td> <td>0.23</td> <td>0.44</td> <td>0.23</td> <td>1.00</td> </tr> <tr> <td>P2</td> <td>0.28</td> <td>0.50</td> <td>0.72</td> <td>0.50</td> <td>2.00</td> </tr> </tbody> </table>		X1	X2	X3	X4	Total	P1	0.10	0.23	0.44	0.23	1.00	P2	0.28	0.50	0.72	0.50	2.00
	X1	X2	X3	X4	Total																											
P1	0.077	0.201	0.521	0.201	1.000																											
	X1	X2	X3	X4	Total																											
P1	0.10	0.23	0.44	0.23	1.00																											
P2	0.28	0.50	0.72	0.50	2.00																											

	<table border="1"> <tr> <td>P2</td> <td>0.22 2</td> <td>0.50 0</td> <td>0.77 8</td> <td>0.50 0</td> <td>2.00 0</td> </tr> <tr> <td>P3</td> <td>0.47 9</td> <td>0.79 9</td> <td>0.92 3</td> <td>0.79 9</td> <td>3.00 0</td> </tr> <tr> <td>P4</td> <td>0.22 2</td> <td>0.50 0</td> <td>0.77 8</td> <td>0.50 0</td> <td>2.00 0</td> </tr> <tr> <td>Total</td> <td>1.00 0</td> <td>2.00 0</td> <td>3.00 0</td> <td>2.00 0</td> <td></td> </tr> </table>	P2	0.22 2	0.50 0	0.77 8	0.50 0	2.00 0	P3	0.47 9	0.79 9	0.92 3	0.79 9	3.00 0	P4	0.22 2	0.50 0	0.77 8	0.50 0	2.00 0	Total	1.00 0	2.00 0	3.00 0	2.00 0		<table border="1"> <tr> <td>P3</td> <td>0.56</td> <td>0.77</td> <td>0.90</td> <td>0.77</td> <td>3.00</td> </tr> <tr> <td>P4</td> <td>0.28</td> <td>0.50</td> <td>0.72</td> <td>0.50</td> <td>2.00</td> </tr> <tr> <td>Total</td> <td>1.22</td> <td>2.00</td> <td>2.78</td> <td>2.00</td> <td></td> </tr> </table>	P3	0.56	0.77	0.90	0.77	3.00	P4	0.28	0.50	0.72	0.50	2.00	Total	1.22	2.00	2.78	2.00	
P2	0.22 2	0.50 0	0.77 8	0.50 0	2.00 0																																							
P3	0.47 9	0.79 9	0.92 3	0.79 9	3.00 0																																							
P4	0.22 2	0.50 0	0.77 8	0.50 0	2.00 0																																							
Total	1.00 0	2.00 0	3.00 0	2.00 0																																								
P3	0.56	0.77	0.90	0.77	3.00																																							
P4	0.28	0.50	0.72	0.50	2.00																																							
Total	1.22	2.00	2.78	2.00																																								
<p>For comparison, from the JMLE estimation - IPMATRIX= expected response values</p> <p>CMLE probabilities are close to JMLE probabilities</p>	<p>from JMLE computation</p> <table border="1"> <thead> <tr> <th></th> <th>X1</th> <th>X2</th> <th>X3</th> <th>X4</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td>P1</td> <td>0.07 5</td> <td>0.22 2</td> <td>0.50 1</td> <td>0.22 2</td> <td>1.00 0</td> </tr> <tr> <td>P2</td> <td>0.22 2</td> <td>0.50 0</td> <td>0.77 8</td> <td>0.50 0</td> <td>2.00 0</td> </tr> <tr> <td>P3</td> <td>0.50 1</td> <td>0.77 8</td> <td>0.92 5</td> <td>0.77 8</td> <td>3.00 0</td> </tr> <tr> <td>P4</td> <td>0.22 2</td> <td>0.50 0</td> <td>0.77 8</td> <td>0.50 0</td> <td>2.00 0</td> </tr> <tr> <td>Total</td> <td>1.00 0</td> <td>2.00 0</td> <td>3.00 0</td> <td>2.00 0</td> <td></td> </tr> </tbody> </table>		X1	X2	X3	X4	Total	P1	0.07 5	0.22 2	0.50 1	0.22 2	1.00 0	P2	0.22 2	0.50 0	0.77 8	0.50 0	2.00 0	P3	0.50 1	0.77 8	0.92 5	0.77 8	3.00 0	P4	0.22 2	0.50 0	0.77 8	0.50 0	2.00 0	Total	1.00 0	2.00 0	3.00 0	2.00 0		<p>Person (theta) CMLE estimates obtained by</p> <ol style="list-style-type: none"> 1) transpose the data 2) perform item CMLE computation 3) in this example, the delta and theta estimates are the same. 						
	X1	X2	X3	X4	Total																																							
P1	0.07 5	0.22 2	0.50 1	0.22 2	1.00 0																																							
P2	0.22 2	0.50 0	0.77 8	0.50 0	2.00 0																																							
P3	0.50 1	0.77 8	0.92 5	0.77 8	3.00 0																																							
P4	0.22 2	0.50 0	0.77 8	0.50 0	2.00 0																																							
Total	1.00 0	2.00 0	3.00 0	2.00 0																																								
<p>Quasi-CMLE estimates from JMLE probabilities</p>	<p>For dichotomous data, CMLE item (delta) estimates can be deduced from the CMLE item probabilities for a person with score 1.</p> <p>Since JMLE probabilities are close to CMLE item probabilities, approximate item deltas, QCMLE estimates, can be calculated from JMLE probabilities</p>	<p>For dichotomous data, CMLE person (theta) estimates can be deduced from the CMLE person probabilities for an item with score 1.</p> <p>Since JMLE probabilities are close to CMLE person probabilities, approximate person thetas, QCMLE estimates, can be calculated from JMLE probabilities</p>																																										

How to do this example with Winsteps and R Statistics, eRm package:

With Winsteps control file:

```

title=4x4
ni=4
name1=1
item1=1
codes=01
lconv=.0001
rconv=.0001
converge=b
&END
END LABELS
1000
0110
0111

```

0011

after the analysis phase:

Output Files menu: IPMATRIX=.

Options:

- 3. Response value after scoring
- Uncheck "Also include Person Entry Number"
- Click on OK

Output File Specifications::

- R Statistics
- Temporary
- OK

R Statistics window opens:

IPMATRIX dataset is loaded, see ls()

[Previously saved workspace restored]

```
>IPMATRIX # let's see what the data look like in R
```

```
  x1 x2 x3 x4
1  1  1  1  0
2  1  1  0  0
3  1  0  0  0
4  0  0  0  1
```

```
> install.packages("eRm") # only do this the first time
```

```
> library(eRm) # activate eRm
```

```
> res <- RM(IPMATRIX) # CMLE estimation of item easinesses
```

```
> coef(res) # or summary(res) # report the items
```

```
      beta X1      beta X2      beta X3      beta X4
-9.550007e-01 -2.322257e-07  9.550012e-01 -2.314972e-07
```

```
> pres <- person.parameter(res) # MLE estimation of person abilities (thetas)
```

```
> coef(pres) # or summary(pres) # report the person estimates
```

```
      P1      P2      P3      P4
-1.208460e+00 -4.741322e-07  1.208459e+00 -4.741322e-07
```

Then use these values as anchor values in a Winsteps analysis:

```
IAFILE=*
```

```
1  0.955 ; item difficulty is -item easiness
2  0.000
3 -0.955
4  0.000
*
```

```
PAFILE=
```

```
1 -1.208
2  0.000
3  1.208
4  0.000
*
```

For true CMLE ability estimates, transpose the data:

```
> transp <- t(IPMATRIX)
```

```
> theta <- RM(transp) # CMLE estimation of person abilities
```

```
> coef(theta)
```


19.17 Column and classification selection and transformation



Selection: Several specifications require or allow the choice of columns in the item label, person label or data record. There are several formats:

Specification=	DGF= , DIF= , DPF= , EQFILE= , IAFILE= , IMAP= , IPMATRIX= , ISORT= , ISUBTOT= , IWEIGHT= , PAFILE= , PKMAP= , PMAP= , PSORT= , PSUBTOT= , PWEIGHT= , @Fieldname=
Specification=S...W... Specification=S...E... Specification=...-...	S or s is the starting column in the person or item label W or w is the width in columns E or e is the ending column in the label - indicates the column range
DPF=\$S1W1 DPF=S1W1 DPF=1W1 DPF=1	means "use the section of the item (or person) label starting in column 1 of the label with a width of 1 column".
DPF=\$S3W2 DPF=S3W2 DPF=3W2	means "use the section of the item (or person) label starting in column 3 of the label with a width of 2 columns".
DPF=\$S3E4 DPF=S3E4 DPF=3E4 DPF=3-4	means "use the section of the item (or person) label starting in column 3 of the label and ending in column 4".
Specification=C...W... Specification=C...E... Specification=c...w... Specification=c...e...	C or c is the starting column in the item label or person data record W or w is the width in columns E or e is the ending column in the data record This always works if the columns are within the person or item label. If the columns referenced are in the data record, but outside the person label, the information may not be available.
PWEIGHT = \$C203W10 PWEIGHT = C203W10 PWEIGHT = C203E212	means "person weighting is in column 203 of the data record, with a width of 10 columns, ending in column 212"
Specification=\$N Specification=@ILABEL Specification=@PLABEL	entry number this matches the length of the longest item label this matches the length of the longest person label
Specification = @Fieldname	selects one or a block of columns as specified by a prior @Fieldname= instruction
@AGE = 2E3 PSUBTOTAL = @AGE	the person's age is in columns 2 and 3 of the person label: subtotal the persons by age classifier
Specification=MA3 Specification=MD2	M = Measure A = Ascending or D = Descending 1 or higher integer: number of strata e.g., MA3 = Measures Ascending in 3 ability strata
Specification=OMA3 Specification=ILD4	I = Infit, O = Outfit M = Mean-square, L = Log-mean-square, Z = ZSTD A = Ascending or D = Descending 1 or higher integer: number of strata

	e.g., FILD2 = Fit - Infit - Log-scaled - Descending -2 fit strata
Strata: MA3, etc.	Strata are equally long divisions of the reported range of the statistic. Stratum number for this value = $\text{Max}(1 + \text{Number of strata} * (\text{Current value} - \text{Lowest reported value}) / (\text{Highest reported value} - \text{Lowest reported value}), \text{Number of strata})$
DIF=MA3 Table 30 PSUBTOTAL=MA3 Table 28 PMAP=MA3 Table 16	Non-uniform DIF for 3 ability strata Output DIF Table Person sub-total for same 3 ability strata Output subtotal Table Show ability strata on the person map Output person map
Specification = selection + selection	+ signs can be used to concatenate selections
IMAP = 3-4 + 8W3	show in Table 12 columns 3, 4, 8, 9, 10 of the item label.
Specification = "..." Specification = '...'	constants in quotes can be included with selection.
PMP = 3 + "/" + 6	show in Table 15 column 3, a slash, and column 6 of the person label, e.g., F/3
Multiple-format specifications:	
IAFILE = *	the item anchor values follow this specification in a list
IAFILE = 3-7	the item anchor values are in columns 3-7 of the item label
IAFILE = filename.txt	the item anchor values are in the file called "filename.txt"

19.18 Combining items into superitems and testlets

See [Testlets](#)

19.19 Comparing estimates with other Rasch software

There are many Rasch-specific software packages and IRT packages which can be configured for Rasch models. Each implements particular estimation approaches and other assumptions or specifications about the estimates. Comparing or combining measures across packages can be awkward. There are three main considerations:

- (a) choice of origin or zero-point
- (b) choice of user-scaling multiplier.
- (c) handling of extreme (zero, minimum possible and perfect, maximum possible) scores.

Here is one approach:

Produce person measures from Winsteps and the other computer program on the same data set. For Winsteps set USCALE=1 and UIMEAN=0.

Cross-plot the person measures with the Winsteps estimates on the x-axis. (This is preferential to comparing on item estimates, because these are more parametrization-dependent.)

Draw a best-fit line through the measures, ignoring the measures for extreme scores.

The slope is the user-scaling multiplier to apply. You can do this with [USCALE=](#) slope.

The intercept is the correction for origin to apply when comparing measures. You can do this with [UIMEAN=](#) y-axis intercept.

The departure of extreme scores from the best-fit line requires adjustment. You can do this with [EXTRSCORE=](#). This may take multiple runs of Winsteps. If the measures for perfect, maximum possible scores are above the best-fit line, and those for zero, minimum possible scores are below, then decrease EXTRSCORE= in 0.1 increments or less. If vice-versa, then increase EXTRSCORE= in 0.1 increments or less.

With suitable choices of UIMEAN=, USCALE= and EXTRSCORE=, the crossplotted person measures should approximate the identity line.

The item estimates are now as equivalent as they can be even if, due to different choice of parametrization or estimation procedure, they appear very different.

You may notice scatter of the person measures around the identity line or obvious curvature. These could reflect differential weighting of the items in a response string, the imposition of prior distributions, the choice of approximation to the logistic function, the choice of parametrization of the Rasch model or other reasons. These are generally specific to each software program and become an additional source of error when comparing measures.

There are technical details at [Estimation](#).

Winsteps (JMLE) vs. [CMLE](#) and RUMM ([PMLE](#)) estimates. CMLE and RUMM estimates are more central than Winsteps estimates. Winsteps estimates can be adjusted using [STBIAS=](#).

Winsteps (JMLE) vs. ConQuest ([MMLE](#)), the estimate differences are primarily because:

1. ConQuest assumes a regular shape to the person-ability distribution. Winsteps does not.
 2. ConQuest includes extreme person scores (zero and perfect) when estimating item difficulties. Winsteps does not.
- There are also other technical differences, but these are usually inconsequential with large datasets.

Maximum Likelihood Estimates (MLE, any type) vs. Warm's Mean Weighted Likelihood Estimates (WLE): Warm estimates are usually slightly more central than MLE estimates.

19.20 Comparing tests or instruments

Is there a way to compare two different instruments in the same Rasch analysis using Winsteps?

Let us assume a sample of persons been administered both instruments:

If so

1. analyze the two instruments separately, and produce [PFILE=](#)s.
2. cross-plot the person measures, e.g., with the Winsteps "[Plots](#)", "[Compare](#)" plot.
3. does the empirical trend line on the plot have the slope of an identity line (i.e., the line is: $y = x + a$) ?

4. If so, all the data can be analyzed together as one long "super test".

To produce comparable results for the two instruments:

Write out anchor files: [IFILE=](#) and [SFILE=](#)

Then anchor [IAFILE=](#) and [SAFILE=](#) and delete the items for each instrument in turn, [IDFILE=](#), to produce comparable results for the instruments.

5. If not, then you are in a "Celsius to Fahrenheit" situation.

The instruments cannot be combined into one analysis.

Use [USCALE=](#) and [UIMEAN=](#) to adjust the measures in one instrument so that the relationship between the measures on analyses of the two instruments becomes $y=x$, the identity line.

Then the person and item hierarchies can be compared directly using their [PFILE=](#) and [IFILE=](#)

19.21 Computer-Adaptive Testing (CAT)

If you are analyzing CAT data with Winsteps, the easiest way to enter the data into Winsteps is with [EDFILE=](#).
EDFILE=filename, e.g., edfile.txt

In edfile.txt, the data lines look like

Person entry number bank number for the item person response

19.22 Convergence considerations

For early runs, set the convergence criteria loosely, or use Ctrl+F to stop the [iterations](#).

If in doubt, set the convergence criteria very tightly for your final report, e.g.,

[CONVERGE=B](#) ; both [LCONV=](#) and [RCONV=](#) apply

[LCONV=](#) .0001 ; largest logit change .0001 logits

[RCONV=](#) .01 ; largest score residual .01 score points

[MJMLE=](#) 0 ; unlimited [JMLE](#) iterations

and be prepared to stop the iterations manually, if necessary, using "Finish Iterating", Ctrl+F, on the [File pull-down menu](#)

Remember that convergence criteria tighter than the reported standard error of a measure are somewhat of a numerical fiction.

The Rasch model is non-linear. This means that estimates cannot be obtained immediately and exactly, as can be done with the solution of simultaneous linear equations. Instead, estimates are obtained by means of a series of guesses at the correct answer.

The initial guess made by Winsteps is that all items are equally difficult and all persons equally able. The expected responses based on this guess are compared with the data. Some persons have performed better than expected, some worse. Some items are harder than expected, some easier. New guesses, i.e., estimates, are made of the person abilities and item difficulties, and also the rating (or partial credit) scale structures where relevant.

The data are again examined. This is an "iteration" through the data. Expectations are compared with observations, and revised estimates are computed.

This process continues until the change of the estimates is smaller than specified in [LCONV=](#), or the biggest difference between the marginal "total" scores and the expected scores is smaller than specified in [RCONV=](#). The precise stopping rule is controlled by [CONVERGE=](#). When the estimates are good enough, the iterative process has "converged". Then iteration stops. Fit statistics are computed and the reporting process begins.

There are standard convergence criteria which are suitable for most small and medium-sized complete data sets. LCONV= is harder to satisfy for small complete data sets and many sparse data designs, RCONV= for large complete data sets.

The "least converged" persons (and items) are those that Winsteps is having the most difficulty estimating. This is not usually because their responses misfit the model but because their scores are extreme. It is more difficult to estimate the measure for a person who scores 1 out of 20 than of a person who scores 10 out of 20 (even if the 10 out of 20 greatly misfit the model). This is because Rasch estimation is done as though the data fit the model. Then, after the estimates have been made, the data are evaluated to determine the extent to which the data do fit the model. It is the same process in linear regression. The parameters are estimated as though the data fit the linear regression model. Then the data are investigated to see whether they do. It is at this later stage that outliers are detected. The Rasch model is a special case of logit-linear regression.

Anchored analyses

Anchor values always misalign somewhat with the current dataset unless they are estimated from it. Thus, the maximum residuals can never reduce to zero. Convergence occurs when the maximum logit change is too small to be meaningful. Accordingly, RCONV= is unproductive and only LCONV= is useful. Suggested specifications are:

CONVERGE = L ; only LCONV is operative

LCONV = .005 ; smaller than visible in standard, two decimal, output.

Missing data

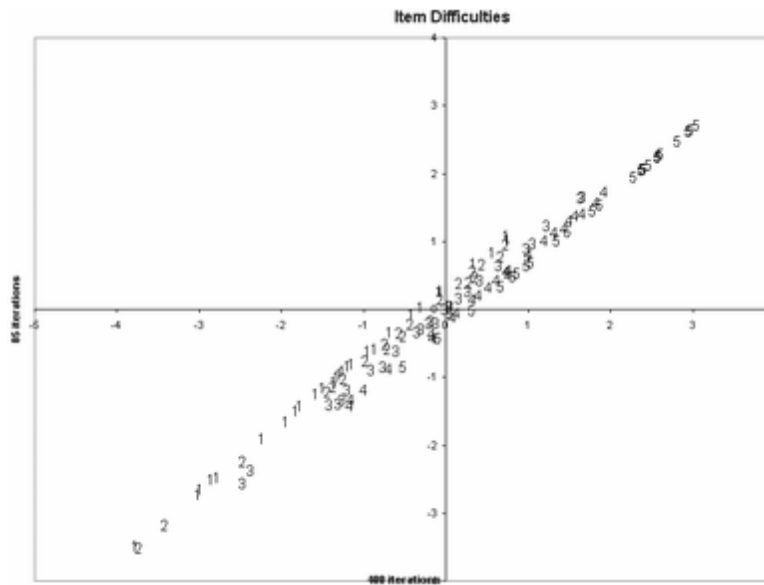
For some data designs much tighter criteria are required, particularly linked designs with large amounts of missing data. For instance, in a vertical [equating](#) design of 5 linked tests, standard convergence occurred after 85 iterations. Tightening the convergence criteria, i.e., using smaller values of LCONV= and RCONV=, convergence occurred after 400 iterations. Further iterations made no difference as the limit of mathematical precision in the computer for this data set had been reached. The plot shows a comparison of the item difficulties for the 5 linked tests estimated according to the standard and strict convergence criteria.

CONVERGE = B

LCONV = .001 ; 10 time stricter than usual

RCONV = .01 ; 10 times stricter than usual

Note that convergence may take many iterations, and may require manual intervention to occur: Ctrl+F.



19.23 Correlations: point-biserial, point-measure, residual

In Rasch analysis, we use item correlations as an immediate check that the response-level scoring makes sense. If the observed correlation is negative, something may have gone wrong (MCQ miskey, reversed survey item, etc.)

In general, correlations are much too difficult to interpret, so we switch over to using mean-squares. The "expected correlation" indicates when conventional rules such as eliminate items with point-biserials less than 0.2 are misleading.

Item correlations are difficult to interpret because they are influenced by:

1. predictability of the data
2. targeting of the item on the person sample
3. distribution of the person sample

In Rasch analysis, we are chiefly concerned about the predictability of the data when assessing item quality, so we examine the predictability directly using the mean-square statistics, rather than indirectly through the correlations.

All correlations are computed as Pearson product-moment correlation coefficients. If you wish to compute other correlations, the required data are in [XFILE= IPMATRIX=](#), [IFILE=](#) or [PFILE=](#). The [Biserial](#) correlation can be computed from the point-biserial.

Point-Biserial and Point-Measure Correlations			
Control Instruction	Observed value	Explanation	Expected value (EXP.)
PTBISERIAL= Yes PTBISERIAL= Exclude	PBSX PTBISERL-EX	Point-biserial (or point-polyserial) correlation excluding the current observation from the raw score. Computes the point-biserial or point-polyserial correlation coefficients, r_{pbis} , for persons and items. This is the Pearson product-moment correlation between the scored responses (dichotomies and polytomies) and the "rest scores", the corresponding total (marginal) scores excluding the scored responses to be correlated. This is a point-biserial correlation for dichotomies, or a point-polyserial correlation for polytomies. Extreme (perfect, maximum possible and zero, minimum possible) scores are included in the computation, but missing observations are omitted pairwise. The Biserial correlation can be computed from the Point-biserial. This correlation loses its meaning when there are missing data or with CUTLO= or CUTHI= . Specify PTBISERIAL=X instead.	PBSX-E PBX-E

PTBISERIAL= All PTBISERIAL= Include	PBSA PTBISERL-AL	Point-biserial correlation for all observations including the current observation in the raw score. Computes the Pearson correlation between the total (marginal) scores including all responses and the responses to the targeted item and person. This is a point-biserial correlation for dichotomies, or a point-polyserial correlation for polytomies. This correlation loses its meaning when there are missing data or with CUTLO= or CUTHI= . Specify PTBISERIAL=N instead.	PBSA-E PBA-E
PTBISERIAL= No PTBISERIAL= RPM PTBISERIAL= Measure	PTMA PTMEASUR- AL	Point-measure correlation for all observations. Computes the Pearson point-measure correlation coefficients, r_{pm} between the observations and the measures, estimated from the raw scores including the current observation or the anchored values. Measures corresponding to extreme scores are included in the computation.	PTMA-E PMA-E
PTBISERIAL= X	PTMX PTMEASUR- EX	Point-measure correlation excluding the current observation from the estimation of the measure. Computes the Pearson point-measure correlation coefficients, r_{pm} between the observations and the measures or anchor values adjusted to exclude the current observation. Measures corresponding to extreme scores are included in the computation.	PTMX-E PMX-E

Point-correlations are always reported for items. Point-correlations are reported for persons when (i) all the items are dichotomies, (ii) all the items have three categories, or (iii) all the items are in the same item group (Andrich Rating Scale Model). Otherwise, the observed and expected correlations are reported as zero in [PFILE=](#).

Example: Here are the point-biserial-type coefficient options for [Example0.txt](#), Item 1 (Watch Birds).

Control Instruction	PTBISERIAL= ALL	PTBISERIAL= YES	PTBISERIAL= NO	PTBISERIAL= X
Calculation: Correlate observation with	raw score	raw score - current observation	Rasch measure	Rasch measure estimated from raw score - current observation
Table 14 , etc., Heading	PTBISERL-AL	PTBISERL-EX	PTMEASURE-A	PTMEASUR-EX
IFILE= / PFILE= Heading	PBSA	PBSX	PTMA	PTMX
Observed correlation	.69	.66	.64	.61
Expected correlation (-E)	.53	.47	.49	.48

In Table 14.1 and other measure Tables: **Point-Biserial Correlations for dichotomies, and Point-Polyserial for polytomies**

when [PTBISERIAL=Yes](#)

PTBSE is the point-biserial correlation between the responses to this item by each person and the total marginal score by each person (omitting the response to this item). This is the "point-biserial corrected for spuriousness". Henrysson, S. (1963). Correction for item-total correlations in item analysis. *Psychometrika*, 28, 211-218.

when PTBISERIAL=All

PTBSA is the point-biserial correlation between the responses to this item by each person and the total marginal score by each person (including the response to this item). This is the conventional point-biserial.

In Table 14.3 and other or distractor Tables:

when PTBISERIAL=Yes or PTBISERIAL=All

PTBSD is the distractor point-biserial correlation between the indicated response to this item (scored 1 and other responses scored 0) by each person and the total marginal score by each person.

There is a closer match between Table 14.1 and Table 14.3 when PTBISERIAL=All

PTBIS=Y or E (indicated by PTBSE): The **point-biserial** correlation r_{pbis} for item i (when $i=1,L$ for persons $n=1,N$) is the correlation between the observation for each person on item i and the total score for each person on all the items excluding item i (and similarly for the point-biserial for each person):

$$r_{pbis_i} = \frac{\sum_{n=1}^N \left(X_{ni} - \sum_{m=1}^N X_{mi}/N \right) \left(\sum_{j=1, j \neq i}^L X_{nj} - \sum_{m=1, j \neq i}^N \sum_{j=1, j \neq i}^L X_{mj}/N \right)}{\sqrt{\sum_{n=1}^N \left(X_{ni} - \sum_{m=1}^N X_{mi}/N \right)^2 \sum_{n=1}^N \left(\sum_{j=1, j \neq i}^L X_{nj} - \sum_{m=1, j \neq i}^N \sum_{j=1, j \neq i}^L X_{mj}/N \right)^2}}$$

PTBIS=All (indicated by PTBSA): All the observations are included in the total score:

$$r_{XY} = \frac{\sum_{n=1}^N \left(X_n - \sum_{m=1}^N X_m/N \right) \left(Y_n - \sum_{m=1}^N Y_m/N \right)}{\sqrt{\sum_{n=1}^N \left(X_n - \sum_{m=1}^N X_m/N \right)^2 \sum_{n=1}^N \left(Y_n - \sum_{m=1}^N Y_m/N \right)^2}}$$

where X_1, \dots, X_N are the responses, and Y_1, \dots, Y_N are the total scores. The range of the correlation is -1 to +1.

Under classical (raw-score) test theory conventions, point-biserial correlations should be 0.3, 0.4 or better. Under Rasch conditions, point-biserial (or point-measure) correlations should be positive, so that the item-level scoring accords with the latent variable, but the size of a positive correlation is of less importance than the fit of the responses to the Rasch model, indicated by the mean-square fit statistics.

Point-Measure Correlations

PTBIS=No (indicated by PTMEA): The correlation between the observations and the Rasch measures:

$$r_{XY} = \frac{\sum_{n=1}^N \left(X_n - \sum_{m=1}^N X_m/N \right) \left(Y_n - \sum_{m=1}^N Y_m/N \right)}{\sqrt{\sum_{n=1}^N \left(X_n - \sum_{m=1}^N X_m/N \right)^2 \sum_{n=1}^N \left(Y_n - \sum_{m=1}^N Y_m/N \right)^2}}$$

where X_1, \dots, X_N are the responses by the persons (or on the items), and Y_1, \dots, Y_N are the person measures (or the item easinesses = - item difficulties). The range of the correlation is -1 to +1.

Jaspens, N. (1946) "Serial Correlation", *Psychometrika*, 11, 23-30.

Olsson, U., Drasgow, F. and Dorans, N.J. (1982) "The Polyserial Correlation Coefficient", *Psychometrika*, 47, 337-347.

The Expected Value of a Correlation

Interpreting an observed value is made easier if we can compare it with its expected value. Is it much higher than expected or much lower than expected?

The general formula for a Pearson correlation coefficient is shown above. Suppose that X_n is X_{ni} the observation of person n on item i . Y_n is B_n the ability of person n , then the point-measure correlation is:

$$r_{pmi} = \frac{\sum_{n=1}^N (X_{ni} - \bar{X}_i)(B_n - \bar{B})}{\sqrt{\sum_{n=1}^N (X_{ni} - \bar{X}_i)^2 \sum_{n=1}^N (B_n - \bar{B})^2}}$$

According to the Rasch model, the expected value of X_{ni} is E_{ni} and its model variance around the expectation is W_{ni} . For dichotomies, E_{ni} is the Rasch probability of success and $W_{ni} = E_{ni}(1-E_{ni})$. For polytomies, E_{ni} and W_{ni} are given by RSA, p. 100 also at www.rasch.org/rmt/rmt34e.htm. For JMLE estimates, $\text{Sum}(E_{ni}) = \text{Sum}(X_{ni})$ for $n=1, N$. Thus an estimate of the expected value of the point-measure correlation is given by the Rasch model proposition that: $X_{ni} = E_{ni} \pm \sqrt{W_{ni}}$. Other variance terms are much smaller.

$$E(r_{pmi}) \approx \frac{\sum_{n=1}^N (E_{ni} \pm \sqrt{W_{ni}} - \bar{X}_i)(B_n - \bar{B})}{\sqrt{\sum_{n=1}^N (E_{ni} \pm \sqrt{W_{ni}} - \bar{X}_i)^2 \sum_{n=1}^N (B_n - \bar{B})^2}}$$

$\pm \sqrt{W_{ni}}$ is a random residual with mean 0 and variance W_{ni} . Its cross-product with any other variable is modeled to be zero. Thus, simplifying,

$$E(r_{pmi}) \approx \frac{\sum_{n=1}^N (E_{ni} - \bar{X}_i)(B_n - \bar{B})}{\sqrt{\sum_{n=1}^N ((E_{ni} - \bar{X}_i)^2 + W_{ni}) \sum_{n=1}^N (B_n - \bar{B})^2}}$$

and similarly for the point-biserial correlations. Here is an example point-measure correlation and its expected value computed with an Excel spreadsheet:

X_{ni}	$X_{ni} - \bar{X}_i$	$(X_{ni} - \bar{X}_i)^2$	B_n	$B_n - \bar{B}$	$(B_n - \bar{B})^2$	$\frac{(X_{ni} - \bar{X}_i)(B_n - \bar{B})}{(B_n - \bar{B})}$	E_{ni}	$E_{ni} - \bar{X}_i$	$\frac{(E_{ni} - \bar{X}_i)(B_n - \bar{B})}{(B_n - \bar{B})}$	$E_{ni}(1-E_{ni}) = W_{ni}$	$\frac{(E_{ni} - \bar{X}_i)^2}{W_{ni}} +$
0	-0.50	0.25	-1.33	-1.66	2.74	0.83	0.16	-0.34	0.56	0.13	0.25
1	0.50	0.25	-0.03	-0.36	0.13	-0.18	0.41	-0.09	0.03	0.24	0.25
0	-0.50	0.25	1.33	1.01	1.01	-0.50	0.73	0.23	0.23	0.20	0.25
1	0.50	0.25	1.33	1.01	1.01	0.50	0.73	0.23	0.23	0.20	0.25
average $\bar{X}_i = 0.50$		(i) $\sum = 1.00$	average $\bar{B} = 0.33$		(ii) $\sum = 4.89$	(iii) $\sum = 0.65$			(iv) $\sum = 1.05$		(v) $\sum = 1.00$
Observed point-measure Correlation = (iii) / $\sqrt{((i)(iii))} =$						0.29					
Expected point-measure Correlation = (iv) / $\sqrt{((v)(ii))} =$									0.47		
Difficulty of item $i = 0.35$ logits											

Table courtesy of Peter Karaffa

Point-Correlations and the Partial Credit Model or Grouped Rating Scale Model: [ISGROUPS=](#)

The usual computations for correlation coefficients assume that all scores or measures within each of the two variables is in the same metric. For instance, when correlating age with height for children, the ages are all in months and the heights are all in centimeters. If we mix ages in months with ages in years, or height in centimeters with height in inches, then the age-height correlations can become meaningless.

The same situation can arise when we use the Partial Credit Model (ISGROUPS=0) or Grouped Rating Scale Model (ISGROUPS=AABB..). In particular, the Expected value of a Point-Correlation can be reported as negative. This is a warning is that the reported correlations are meaningless.

Disattenuated correlation coefficients

"Attenuated" means "reduced". "Disattenuation" means "remove the attenuation".

The observed correlation between two variables is attenuated (reduced toward zero) because the variables are measured with error. So, when we remove the measurement error (by a statistical operation), the resulting correlation is disattenuated. Disattenuated correlations are always further from zero.

Algebraically:

{A} and {B} are the "true" values of two variables. Their true (disattenuated) correlation is

Disattenuated ("true") correlation = $r(A,B)$

But the observed values of the variables are measured with error $\{A \pm a\}$, $\{B \pm b\}$, so the observed correlation is

observed correlation = $r(A,B) * \sqrt{(\text{var}(A) * \text{var}(B)) / ((\text{var}(A) + a^2) * (\text{var}(B) + b^2))}$.

"Disattenuation" reverses this process.

If the reliability of {A} is RA , and the reliability of {B} is RB , then the disattenuated correlation between {A} and {B} is:
disattenuated $r(A,B) = r(\text{observed}) / \sqrt{RA * RB}$.

Point-Correlations in [Table 14.3](#), etc. and [DISFILE=](#)

For the item point-biserial correlations, [PTBISERIAL=A](#) and [PTBISERIAL=E](#);, scored observations are correlated with their corresponding person raw scores. For the point-measure correlations, [PTBISERIAL=M](#), scored observations are correlated with their corresponding Rasch measures.

Each response code (scored or "missing") is correlated only with scored response codes. With [PTBISERIAL=Yes](#) or [All](#), the correlation is between the occurrence and the person raw score, indicated by [PTBSD CORR](#). When this correlation is high positive for a correct MCQ option, then the item exhibits convergent validity. When this correlation is low or negative for incorrect MCQ options, then the item exhibits discriminant validity. Krus, D. J. & Ney, R. G. (1978) Convergent and discriminant validity in item analysis. Educational and Psychological Measurement, 38, 135-137. Each scored response code is correlated only with the other scored codes: "1" for target code "0" for other scored codes. . Missing responses are correlated with all responses.

[PTBSA CORR](#) is [PTBIS = All](#) (include the current item in the person score)

[PTBSE CORR](#) is [PTBIS = Yes](#) (exclude the current item in the person score)

[PTMEA CORR](#) is [PTBIS = No](#) (correlate with the person measure)

The "Missing" point-measure correlation is the correlation between the raw scores for the persons and the responses to the item with "missing" scored 1 and the other responses scored 0.

The category (or option) correlation is always computed the same way:

1. We focus on a particular response to a particular item.
2. Everyone who chose that response to the item is scored "1", everyone else is scored "0". Missing data are ignored.
3. The scores of "1" or "0" are correlated with the person measures.
4. We do this for every response to every item.

Example: Point-Measure Correlations of Person Measures with Item 6 Options (a,b,c,d):

<i>Person measure</i>	<i>Person's response to item 6</i>	<i>Category a</i>	<i>b</i>	<i>c</i>	<i>d</i>
0.5	d	0	0	0	1
0.8	c	0	0	1	0

0.9	b	0	1	0	0
0.4	d	0	0	0	1
1.0	a	1	0	0	0
1.2	b	0	1	0	0
	Correlation of measure with category:	0.32	0.64	0.00	-0.89

For the options reported in Table 14.3 and DISFILE=

For each item and each option, a correlation is computed for each code in CODES=

If the response in the data file is the code, then a value of 1 is assigned

If the response in the data file is not the code and in CODES=, then a value of 0 is assigned

If the response in the data file is not in CODES= and scored with MISSING-SCORED=, then a value of 0 is assigned.

If the response in the data file is not in CODES= and not scored with MISSING-SCORED=, then the response is ignored.

For each value of 0 or 1, an average response score is computed = (raw score / count scored observations)

The correlation is between the (weighted) values and the average response scores or Rasch measures

For each item, one correlation is computed for all codes in the data not in CODES=

If the response in the data file is not in CODES=, then a value of 1 is assigned

If the response in the data file is in CODES=, then a value of 0 is assigned

For each value of 0 or 1, an average response score is computed = (raw score / count scored observations)

The correlation is between the weighted values and the average response scores or Rasch measures

If your data file contains unscored data or missing data, then the item point-correlations will differ from the option point-correlations. If you want the correlations to be the same, then force missing data to be scored 0:

1. Do your usual analysis
2. Output IFILE=if.txt, PFILE=pf.txt, SFILE=sf.txt
3. Add to your control file:
 - IAFILE=if.txt ; anchor everything at their estimated values
 - PAFILE=pf.txt
 - SAFILE=sf.txt
 - CODES=your codes + all other codes in the data file. For example, if your codes are A,B,C,D and missing data is blank, then CODES=" ABCD "
4. Do the analysis again. Output Table 14.3, DISFILE=. The values should now agree.
5. Do NOT use the fit statistics estimated from this analysis.

Suggestion: since the point-biserial is not defined when there is unscored missing data, please use the point-measure correlation, PTBISERIAL=M.

Statistical Significance of a Correlation

Null hypothesis is that the correlation is 0.00.

Student's significance test: $t = r / \sqrt{(1-r^2) / (N-2)}$ with N-2 d.f. where N is the count of correlated cases.

Online calculator at <http://vassarstats.net/rsig.html>

Null hypothesis is that the correlation is another value, not 1 or -1

1. Use Fisher z' transformation to linearize hypothesized correlation and observed correlation:

$$z' = .5[\ln(1+r) - \ln(1-r)]$$

2. normal $z = (z'(\text{observed}) - z'(\text{hypothesized})) * \sqrt{N-3}$

Null hypothesis is that the correlation is 1 or -1

Use a number that means the same thing: 0.99 or -0.99

19.24 Decimal, percentage and continuous data

Winsteps analyzes ordinal data, not decimal data. Typical decimal data is over-precise. Its numerical precision is greater than its substantive precision. Example: I can measure and report my weight to the nearest gram, but my "true" weight has a precision of about 500 grams.

One solution is to use [SFUNCTION=](#) with the data transformed to positive integers.

Another solution to this is to discover the precision in the data empirically.

1. Dichotomize the data for each item around the median decimal value into 0 = below median, 1 = above median
2. Analyze those data.
3. If the analysis makes sense, then dichotomize each subset of the data again, so that it is now scored 0,1, 2,3
4. Analyze those data.
5. If the analysis makes sense, then dichotomize each subset of the data again, so that it is now scored 0,1, 2,3, 4,5, 6,7
6. Analyze those data.
7. If (and so on).

From a Rasch perspective, the relationship between a continuous variable (such as time to run 100 meters) and a Rasch latent variable (such as physical fitness) is always non-linear. Since we do not know the form of the non-linear transformation, we chunk the continuous variable into meaningful intervals, so that the difference between the means of the intervals is greater than the background noise. With percents, the intervals are rarely smaller than 10% wide, with special intervals for 0% and 100%. These chunked data can then be analyzed with a rating-scale or partial-credit model. We can then transform back to continuous-looking output using the item characteristic curve or the test characteristic curve.

Winsteps analyzes ordinal data expressed as integers, cardinal numbers, in the range 0-32,767, i.e., 32,768 ordered categories.

Example: The data are reported as 1.0, 1.25, 1.5, 1.75, 2.0, 2.5,

Winsteps only accepts integer data, so multiply all the ratings by 4.

If you want the score reports to look correct, then please use [IWEIGHT=](#)

IWEIGHT=*

1-100 of items you have 0.25 ; 100 is the number of items

*

Percentage and 0-100 observations:

Observations may be presented for Rasch analysis in the form of percentages in the range 0-100. These are straightforward computationally but are often awkward in other respects.

A typical specification is:

```
XWIDE = 3
CODES = " 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19+
        + 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39+
        + 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59+
        + 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79+
        + 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99100"
STKEEP = Yes ; to keep intermediate unobserved categories
```

Since it is unlikely that all percentages will be observed, the rating (or partial credit) scale structure will be difficult to estimate. Since it is even more unlikely that there will be at least 10 observations of each percentage value, the structure will be unstable across similar datasets.

It is usually better from a measurement perspective (increased person "test" reliability, increased stability) to collapse percentages into shorter rating (or partial credit) scales, e.g., 0-10, using IREFER= and IVALUE= or NEWSCORE=.

Alternatively, model the 0-100 observations as 100 binomial trials. This imposes a structure on the rating scale so that unobserved categories are of no concern. This can be done by anchoring the Rasch-Andrich thresholds at the values: $F_j = C * \ln(j/(101-j))$, or more generally, $F_j = C * \ln(j / (m-j+1))$ where the range of observations is 0-m. Adjust the value of the constant C so that the average mean-square is 1.0.

Decimal observations:

When observations are reported in fractional or decimal form, e.g., 2.5 or 3.9, multiply them by suitable multipliers, e.g., 2 or 10, to bring them into exact integer form.

Specify STKEEP=NO, if the range of observed integer categories includes integers that cannot be observed.

Continuous and percentage observations:

These are of two forms:

(a) Very rarely, observations are already in the additive, continuous form of a Rasch variable. Since these are in the form of the measures produced by Winsteps, they can be compared and combined with Rasch measures using standard statistical techniques, in the same way that weight and height are analyzed.

(b) Observations are continuous or percentages, but they are not (or may not be) additive in the local Rasch context. Examples are "time to perform a task", "weight lifted with the left hand". Though time and weight are reported in additive units, e.g., seconds and grams, their implications in the specific context is unlikely to be additive. "Continuous" data are an illusion. All data are discrete at some level. A major difficulty with continuous data is determining the precision of the data for this application. This indicates how big a change in the observed data constitutes a meaningful difference. For instance, time measured to .001 seconds is statistically meaningless in the Le Mans 24-hour car race - even though it may decide the winner!

To analyze these forms of data, segment them into ranges of observably different values. Identify each segment with a category number, and analyze these categories as rating scales. It is best to start with a few, very wide segments. If these produce good fit, then narrow the segments until no more statistical improvement is evident. The general principle is: if the data analysis is successful when the data are stratified into a few levels, then it may be successful if the data are stratified into more levels. If the analysis is not successful at a few levels, then more levels will merely be more chaotic. Signs of increasing chaos are increasing misfit, categories "average measures" no longer advancing, and a reduction in the sample "test" reliability.

May I suggest that you start by stratifying your data into 2 levels? (You can use Excel to do this.) Then analyze the resulting the 2 category data. Is a meaningful variable constructed? If the analysis is successful (e.g., average measures per category advance with reasonable fit and sample reliability), you could try stratifying into more levels.

Example 1: My 20 items are in an Excel file: 0-0.5 for dichotomies and 0-0.5-1.0 for partial credit items. What should I do?

i) Winsteps only analyzes integers, so multiply all the items by 2 in Excel.

ii) In order for the raw scores to look right, in Winsteps

IWEIGHT=*

1-20 0.5

*

Do this to verify that the data are being transformed correctly

iii) But, since in the original data, a dichotomous item is 0 - 0.5, instead of the usual 0 - 1,

in order for the Rasch measures to have the correct standard errors and the standardized fit statistics to be valid,

omit IWEIGHT=

iv) Yes, for these data, the partial credit model, which will also process the dichotomous items correctly:

ISGROUPS=0

Example 2: My dataset contains negative numbers such as "-1.60", as well as positive numbers such as "2.43". The range of potential responses is -100.00 to +100.00.

Winsteps expects integer data, where each advancing integer indicates one qualitatively higher level of performance (or whatever) on the latent variable. The maximum number of levels is 0-32767. There are numerous ways in which data can be recoded. One is to use Excel. Read your data file into Excel. Its "Text to columns" feature in the "Data" menu may be useful. Then apply a transformation to the responses, for instance,

$$\text{recoded response} = \text{integer} \left(\frac{(\text{observed response} - \text{minimum response}) * 100}{(\text{maximum response} - \text{minimum response})} \right)$$

This yields integer data in the range 0-100, i.e., 101 levels. Set the Excel column width, and "Save As" the Excel file in ".prn" (formatted text) format. Or you can do the same thing in SAS or SPSS and then use the Winsteps [SAS/SPSS menu](#).

Example 3: We want to construct Rasch measures from the values of the indicators to produce a 'ruler'.

There are two approaches to this problem, depending on the meaning of the values:

1. If you consider that the values of the indicators are equivalent to "item difficulties", in the Rasch sense, then it is a matter of finding out their relationship to logits. For this, one needs some ordinal observational data of the data. Calibrate the observational data, then cross plot the resulting indicator measures against their reference values. The best-fit line or simple curve gives the reference value to logit conversion.
- or 2. If the values are the observations (like weights and heights), then it is a matter of transforming them into ordinal values, and then performing a Rasch analysis on them. The approach is to initially score the values dichotomously high-low (1-0) and see if the analysis makes sense. If so, stratify the values into 4: 3-2-1-0. If the results still make good sense, then into 6, then into 8, then into 10. At about this point, the random noise in the data will start to overwhelm the categorization so that their will be empty categories and many "category average measures" out of sequence. So go back to the last good analysis. The model ICC will give the relationship between values and logits.

19.25 Demographic codes - changing

Question: the demographic codes in my Winsteps control file are gender=1, 2. I want gender = F, M to display in reports.

Answer: Use rectangular copy-and-paste in NotePad++ or TextPad.

1. Open the Winsteps data file in NotePad++ (freeware)
2. Rectangular cut the gender column (use alt+mouse for rectangles)
3. Paste into a new NotePad++ document
4. Global replace 1 with M, 2 with F in the new document
5. Rectangular copy the edited column
6. Paste into the correct column of the Winsteps control file: it now has M, F instead of 1,2

19.26 Dependency and unidimensionality - Repeated measures - Longitudinal studies

There are three approaches to constructing repetition-bias-free Rasch measures of persons being re-measured.

1. The situations are such that the person being re-measured is substantively different from the original person. Any dependency between the pairs of measures of the persons is below the noise level caused by other activities. For instance, when measuring a child entering the educational system at age 6 and then measuring the child again at age 18, any specific dependency between the two measures will be at the noise level. All person records can be analyzed together.
2. The first of each person measurement is the benchmark. The persons are measured, and the item difficulties and responses structures estimated. For all subsequent time-points, the items are anchored (IAFILE=) at their values for the first time point, and similarly for the response structures (SAFILE=). Since the measurement framework is anchored, no dependency between the measurements biases the measurements. Since the analysis is anchored, all time-points can be analyzed together in one analysis.
3. All measurements of each person are equally important, but it is thought that local dependency between the measurements may bias the measurements or the findings. Then, randomly select the observations at one time-point for each person. Construct a data file with only these observations. Analyze this data set. The random set of person records are measured, and the item difficulties and responses structures estimated. For all other time-points, the items are anchored (IAFILE=) at these "random" values, and similarly for the response structures (SAFILE=). Since the measurement framework is anchored, no dependency between the measurements biases the measurements. Since the analysis is anchored, all time-points can be analyzed together in one analysis.

Question: To calibrate item difficulty, I am using data from 75 subjects. Most of the subjects have been tested repeatedly, between two and 9 times each. The reason for this was that I assumed that by training and time (with natural development) the subjects ability was different between different testing situations. Now the referee has asked me to verify that "the requirement of local independence is not breached". How can I check this?

Unidimensionality can be violated in many different ways. If you run all known statistical tests to check for violations (even with your subjects tested only once), your data would undoubtedly fail some of them - (for technical details of some of these tests see Fischer & Molenaar, "Rasch Models", chapter 5.) Consequently, the question is not "are my data perfectly unidimensional" - because they aren't. The question becomes "Is the lack of unidimensionality in my data sufficiently large to threaten the validity of my results?"

Pre-Test - Post-Test dependency (or any two tests with the same persons)

1. [Stack](#) the data: all pre-test data records. Below them, all post-test data records, in the same sequence.
2. Rasch-analyze these data.
3. Output the [IPMATRIX](#) of standardized residuals to Excel
4. For each item, correlate the pre-test standardized residuals with the post-test standardized residuals.
5. Noticeable positive correlations indicate dependency for those items between pre-test and post-test.

Imagine that you accidentally entered all your data twice. Then you know there is a lack of local independence. What would happen? Here is what happened when I did this with the dataset *exam 12/lo.txt*:

Data in once:

```

SUMMARY OF 35 MEASURED PERSONS
+-----+
|          RAW          MODEL          INFIT          OUTFIT          |
|          SCORE        COUNT        MEASURE        ERROR        MNSQ        ZSTD        MNSQ        ZSTD        |
+-----+-----+-----+-----+-----+-----+-----+-----+
| MEAN          38.2          13.0          -.18          .32          1.01          -.1          1.02          .0          |
| P. SD          10.1           .0           .99          .06          .56          1.4          .57          1.3          |
| MAX.           54.0          13.0          1.44          .59          2.36          2.9          2.28          2.5          |
| MIN.           16.0          13.0          -2.92         .29          .23          -2.9         .24          -2.3         |
+-----+-----+-----+-----+-----+-----+-----+-----+
| REAL RMSE     .36 TRUE SD     .92 SEPARATION  2.55 PERSON RELIABILITY .87 |
| MODEL RMSE    .33 TRUE SD     .94 SEPARATION  2.85 PERSON RELIABILITY .89 |
| S. E. OF PERSON MEAN = .17 |
+-----+-----+-----+-----+-----+-----+-----+
PERSON RAW SCORE-TO-MEASURE CORRELATION = .99
CRONBACH ALPHA (KR-20) PERSON RAW SCORE RELIABILITY = .89

```

```

SUMMARY OF 13 MEASURED ITEMS
+-----+
|          RAW          MODEL          INFIT          OUTFIT          |
|          SCORE        COUNT        MEASURE        ERROR        MNSQ        ZSTD        MNSQ        ZSTD        |
+-----+-----+-----+-----+-----+-----+-----+-----+
| MEAN          102.9          35.0           .00          .20          1.08          -.2          1.02          -.2          |
| P. SD          23.6           .0           .93          .03          .58          2.3          .53          2.0          |
| MAX.           145.0          35.0          2.45          .31          2.16          3.9          2.42          4.3          |
| MIN.           46.0          35.0          -1.65         .18          .31          -4.2         .39          -3.3         |
+-----+-----+-----+-----+-----+-----+-----+-----+
| REAL RMSE     .24 TRUE SD     .90 SEPARATION  3.81 ITEM RELIABILITY .94 |
| MODEL RMSE    .20 TRUE SD     .91 SEPARATION  4.53 ITEM RELIABILITY .95 |
| S. E. OF ITEM MEAN = .27 |
+-----+-----+-----+-----+-----+-----+-----+

```

Data in twice:

SUMMARY OF 70 MEASURED PERSONS

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	38.2	13.0	-.18	.32	1.01	-.1	1.02	.0
P.SD	10.1	.0	.99	.06	.56	1.4	.57	1.3
MAX.	54.0	13.0	1.44	.59	2.36	2.9	2.28	2.5
MIN.	16.0	13.0	-2.92	.29	.23	-2.9	.24	-2.3
REAL RMSE	.36	TRUE SD	.92	SEPARATION	2.55	PERSON RELIABILITY	.87	
MODEL RMSE	.33	TRUE SD	.94	SEPARATION	2.85	PERSON RELIABILITY	.89	
S.E. OF PERSON MEAN = .12								

PERSON RAW SCORE-TO-MEASURE CORRELATION = .99
 CRONBACH ALPHA (KR-20) PERSON RAW SCORE RELIABILITY = .89

SUMMARY OF 13 MEASURED ITEMS

	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	205.8	70.0	.00	.14	1.08	-.3	1.02	-.4
P.SD	47.2	.0	.93	.02	.58	3.2	.53	2.9
MAX.	290.0	70.0	2.45	.22	2.16	5.4	2.42	6.1
MIN.	92.0	70.0	-1.65	.13	.31	-6.0	.39	-4.7
REAL RMSE	.17	TRUE SD	.92	SEPARATION	5.48	ITEM RELIABILITY	.97	
MODEL RMSE	.14	TRUE SD	.92	SEPARATION	6.48	ITEM RELIABILITY	.98	
S.E. OF ITEM MEAN = .27								

There is almost no difference in the person report. Person reliability does not change merely because the sample size becomes larger. Person reliability changes when the person distribution changes.

The biggest impact the lack of local independence has in this situation is to make the item [standard errors](#) too small. Consequently you might report item results as statistically significant that aren't.

So, with your current data, you could adjust the size of the item standard errors to their biggest "worst case" size:

Compute $k = \text{number of observations in your data} / \text{number of observations if each person had only been tested once}$

Adjusted item standard error = reported item standard error * sqrt(k).

This would also affect item [Reliability](#) computations:

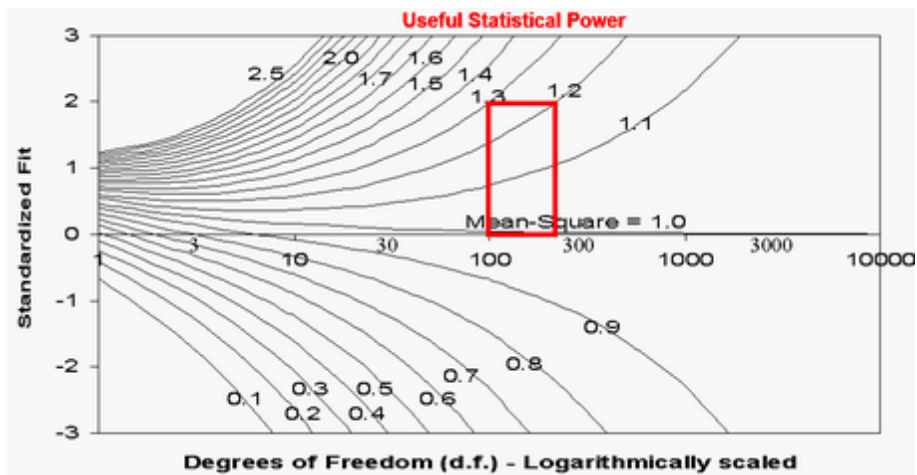
Adjusted item separation coefficient = reported item separation coefficient / sqrt(k)

Adjusted item Reliability (separation index) = $\text{Rel.} / (k + \text{Rel.} - \text{Rel.} * k) = \text{TRUE Sep}^{**2} + \text{Adj.} + \text{Adj. Sep.}^{**2}$

The size of the item mean-square fit statistics does not change, but you would also need to adjust the size of the item t standardized fit statistics (if you use them). This is more complicated. It is probably easiest to read them off the plot from Rasch Measurement Transactions 17:1 shown below.



Look at your current item mean-square and significance. Find the point on the plot. Go down to the x-axis. Divide the value there by k. Go to the same mean-square value contour. The "worst case" lower statistical significance value is on the y-axis.

Another noticeable aspect of your current data could be that there are misfitting subjects who were tested 9 times, while fitting persons are tested only twice. This would introduce a small distortion into the measurement system. So, arrange all the Tables in fit order, and look at each end, do some subjects appear numerous times near the end of a Table? If so, drop out those subjects and compare item calibrations with and without those subjects. If there is no meaningful difference, then those subjects are merely at the ends of the probabilistic range predicted by the Rasch model.



19.27 Dichotomous mean-square fit statistics

For a general introduction, see [Diagnosing Misfit](#), also [Polytomous mean-square fit statistics](#).



Person Responses to Items: Easy--Items--Hard	Diagnosis of Pattern	Person OUTFIT MnSq	Person INFIT MnSq	Point-Measure Correlation
111 0110110100 000	Modeled/Ideal	1.0	1.1	0.62
111 1111100000 000	Guttman/Deterministic	0.3	0.5	0.87
000 0000011111 111	Miscoded	12.6	4.3	-0.87
011 1111100000 000	Carelessness/Sleeping	3.8	1.0	0.65
111 1111000000 001	Lucky Guessing	3.8	1.0	0.65
101 0101010101 010	Response set /Miskey	4.0	2.3	0.11
111 1000011110 000	Special knowledge	0.9	1.3	0.43
111 1010110010 000	Imputed outliers *	0.6	1.0	0.62
Right Transition Wrong				
 high - low - high	OUTFIT sensitive to outlying observations	>>1.0 unexpected outliers	>>1.0 disturbed pattern	
 low - high - low	INFIT sensitive to pattern of inlying observations	<<1.0 overly predictable outliers	<<1.0 Guttman pattern	
* as when a tailored test (such as a Binet intelligence test) is scored by imputing all "right" responses to an administered easier items and all "wrong" responses to an administered harder items. The imputed responses are indicated by italics				

The exact details of these computations have been lost, but the items appear to be uniformly distributed about 0.4 logits apart, extracted from [Linacre, Wright \(1994\) Rasch Measurement Transactions 8:2 p. 360](#)

The ZSTD Z-score standardized Student's [t-statistic](#) report, as unit normal deviates, how likely it is to observe the reported mean-square values, when the data fit the model. The term Z-score is used of a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's [t-statistic](#) value has been adjusted to a unit normal value.

When there is **guessing by a person**, we expect

- (1) the person to be on the less-able side
- (2) the Outfit Mean-square to be large (much greater than 1.0)
- (3) the point-measure correlation to be much less than expected
- (4) the estimated lower asymptote for the person to be away from 0.0
- (5) individual responses to be among the worst fitting.

Item Responses by Persons: High-Person-Ability-Low	Diagnosis of Pattern	Item OUTFIT MnSq	Item INFIT MnSq	Point-Measure Correlation
111 0110110100 000	Modeled/Ideal	1.0	1.1	0.62
111 1111100000 000	Guttman/Deterministic	0.3	0.5	0.87
000 0000011111 111	Miscoded	12.6	4.3	-0.87
011 1111110000 000	Carelessness/Sleeping	3.8	1.0	0.65
111 1111000000 001	Lucky Guessing Response set	3.8	1.0	0.65
000 0010010001 110	Miskey	4.0	2.3	0.11
111 1010110010 000	Imputed outliers *	0.6	1.0	0.62
111 0101010101 000	Low discrimination	1.5	1.6	0.46
111 1110101000 000	High discrimination	0.5	0.7	0.79
111 1111010000 000	Very high discrimination	0.3	0.5	0.84
Right Transition Wrong				
 high - low - high	OUTFIT sensitive to outlying observations	>>1.0 unexpected outliers	>>1.0 disturbed pattern	
 low - high - low	INFIT sensitive to pattern of inlying observations	<<1.0 overly predictable outliers	<<1.0 Guttman pattern	
* as when a tailored test (such as a Binet intelligence test) is scored by imputing all "right" responses to un administered easier items and all "wrong" responses to un administered harder items. The imputed responses are indicated by italics				

When there is **guessing on an item**, we expect

- (1) the item to be on the difficult side
- (2) the Outfit Mean-square to be large (much greater than 1.0)
- (3) the point-measure correlation to be much less than expected
- (4) the estimated lower asymptote to be away from 0.0
- (5) the empirical ICC/IRF to have noticeably high lower tail. One glance at this, and we know we are in trouble!

19.28 DIF - DPF - bias - interactions concepts

How to run a DIF analysis

1. Put a classification-group in the person label. In Exam1.txt this "M" or "F" in column 9
2. Specify the classification-group column in your Winsteps control file DIF = 9
3. Output Table 30
In your control file:
TFILE=*

or from the Output Tables menu.

4. If you strongly suspect DIF in some items, and do not want them to influence the person measures until you have checked those items, then weight those items with 0 using IWEIGHT= until you have done your DIF analysis.

The [DIF plot](#) is a useful guideline where to look. [Table 30.1](#) is decisive. Look at the two "Person Classes" of interest to you (columns 1 and 4). And at the item (row) of interest. This gives you the exact size and significance of the DIF.

Then (see Table below)

- 1) we need the DIF to be big enough in size to be noticeable (usually size difference > 0.5 logits, but look at your item map to see what difference between the difficulties of two items indicates a real change in performance level).
- 2) we need the DIF to be significant enough not to have happened by chance (usually $|t| > 2.0$)

But both (1) and (2) are general guidelines, which should be adjusted to your situation, in particular the control you have over your data. Most DIF rules are constructed for tightly controlled educational tests. But QOL data is much less controlled, so we must adjust our DIF criteria accordingly. We need to consider whether the DIF is likely to continue into other samples, or is the DIF only an accident of this dataset?

Winsteps reports two sets of DIF/DPF statistics:

Winsteps implements two DIF methods. First, the widely used Mantel-Haenszel and Mantel methods which are (log-)odds estimators of DIF size and significance from cross-tabs of observations of the two groups. Second, a logit-difference (logistic regression) method, which estimates the difference between the Rasch item difficulties for the two groups, holding everything else constant.

1) Mantel tests. The Mantel-Haenszel test is used for dichotomies. The Mantel test is used for polytomies. These methods are generally considered most authoritative, but require complete data in their original form. Winsteps stratifies by theta, instead of raw score, to overcome this limitation.

2) Rasch-Welch (logistic regression) t-test: this estimates a Rasch difficulty for the item for each person group (or person for each item group). The estimation process models the item difficulty according to the type of item (dichotomous, Rating Scale, Partial Credit, etc.) The estimation process does not re-estimate the Andrich thresholds. Only the overall item difficulty is re-estimated using a logistic regression model:

log-odds(data for the classification group) = (previously estimated Rasch effects) + DIF effect for the classification group
 DIF MEASURE is the same doing a full analysis of the data, outputting PFILE=pf.txt and SFILE=sf.txt, then doing another analysis with PAFILE=pf.txt and SAFILE=sf.txt and PSELECT=@DIF=code

DPF MEASURE is the same doing a full analysis of the data, outputting IFILE=if.txt and SFILE=sf.txt, then doing another analysis with IAFILE=if.txt and SAFILE=sf.txt and ISELECT=@DPF=code

What is DIF?

Differential Item Functioning, DIF, indicates that one group of respondents is scoring better than another group of respondents on an item (after adjusting for the overall scores of the respondents.)

This could mean

1. One group is performing at its usual "attitude/ability" level on the item, the other is performing better than usual.
2. One group is performing at its usual "attitude/ability" level on the item, the other is performing worse than usual.
3. The item has its usual difficulty for one group, but is more difficult than usual for the other.
4. The item has its usual difficulty for one group, but is easier than usual for the other.
5. etc.

The statistics cannot distinguish between these meanings. Andrich and Hagquist (2012) suggest that "The item with the largest DIF is the one with the real (as opposed to artificial) DIF" (Andrich & Hagquist, 2012). However Luppescu (1993) points out situations in which this is not true.

David Andrich and Curt Hagquist (2012) Real and Artificial Differential Item Functioning. Journal of Educational and Behavioral Statistics, 37, 387-416.
 Lupescu S (1993) DIF detection examined. [Rasch Measurement Transactions, 7:2 p.285-6](#)

Unfairness comes in when it is thought that the items are selected (accidentally or deliberately) to favor one group or disadvantage the other. This can happen for any of reasons 1-5 and more. When presenting DIF results in a political context, DIF is usually presented as 3. with the "other" group being the politically-disadvantaged group.

DIF statistical significance is influenced by:
 1. Size of DIF effect.
 2. Size of classification groups.
 It is largely uninfluenced by model fit.

ETS DIF Category	with DIF Contrast and DIF Statistical Significance	
C = moderate to large	 DIF ≥ 0.64 logits	prob(DIF ≤ 0.43 logits) ≤ .05 (2-sided) <i>approximately: DIF > 0.43 logits + 2 * DIF S.E.</i>
B = slight to moderate	 DIF ≥ 0.43 logits	prob(DIF = 0 logits) ≤ .05 (2-sided) <i>approximately: DIF > 2 * DIF S.E</i>
A = negligible	-	-
C-, B- = DIF against focal group; C+, B+ = DIF against reference group		
ETS (Educational Testing Service) use Delta δ units. 1 logit = 2.35 Delta δ units. 1 Delta δ unit = 0.426 logits.		
Zwick, R., Thayer, D.T., Lewis, C. (1999) An Empirical Bayes Approach to Mantel-Haenszel DIF Analysis. . Journal of Educational Measurement, 36, 1, 1-28 More explanation at www.ets.org/Media/Research/pdf/RR-12-08.pdf pp. 3,4		

Suggestion: If your research will be evaluated by a reviewer and you have complete dichotomous data, then report the Mantel-Haenszel statistics .

Minimum Sample Size

"When there was no DIF [for 4-category items], type I error rates were close to 5%. Detecting moderate uniform DIF in a two-item scale required a sample size of 300 per group for adequate (>80%) power. For longer scales, a sample size of 200 was adequate. Considerably larger sample sizes were required to detect nonuniform DIF, when there were extreme floor effects or when a reduced type I error rate was required."

N. Scott, P. Fayers, N. Aaronson, A. Bottomley, A. de Graeff, M. Groenvold, C. Gundy, M. Koller, M. Petersen, M. Sprangers. (2009). A simulation study provided sample size guidance for differential item functioning (DIF) studies using short scales. Journal of Clinical Epidemiology, Volume 62, Issue 3, Pages 288-295. www.jclinepi.com/article/S0895-4356%2808%2900158-3/abstract

This suggests that for dichotomous items, the sample size of each group needs to be around 1,000.

If we want to compare reasonably robust item difficulties for each sub-group in a somewhat informal way, then 30 per group-classification: www.rasch.org/rmt/rmt74m.htm

What if your sample size is smaller? This is not unusual in many content areas, such as medical research into rare diagnoses, but the findings are always statistically tentative. In fact, DIF findings, even with large samples, are notoriously difficult to replicate. Small sample sizes put your work in the category of "pilot test" rather than "large-scale field testing". However, even so, pilot tests are intended to indicate problematic areas before huge resources are expended on them, and items with DIF are a conspicuous problem area.

So my recommendation would be to abandon statistical tests and focus instead on the size of the DIF effects. If there are large effects, even if they could have happened by chance, the sizes of the effects indicate that actions or precautions should be taken when using the assessment. It is the same as driving a car and hearing a strange sound. The sound

could be caused by chance, such as a stone from the road, but it could also indicate the start of a major problem. After hearing a strange sound, we are all extra vigilant and may even stop to inspect the car. What DIF size is enough for caution? This depends on your latent variable. How big a movement along your latent variable is important? For many educational variables, one logit = one year, so anything above 0.5 logits is a big jump (= 6 months growth).

The ETS guidelines suggest 0.64 logits, see, for instance, the Zwick table in <https://www.rasch.org/rmt/rmt32a.htm>

Computation

Winsteps anchors the person estimates at their values from the main analysis. The item difficulties are unanchored. The the item difficulties for each DIF group are estimated. So this would be an "anchor person method" or the IRT folks would say an "anchor theta method"

Winsteps also reports Mantel-Haenszel statistics. The MH method is based on stratifying by person raw scores (or by person estimates when there are missing data). This is equivalent to anchoring the person estimates, so MH is also an "anchor theta method".

The DIF (differential item functioning) or DPF (differential person functioning) analysis proceeds with all items and persons, except the item or person currently targeted, anchored at the measures from the main analysis (estimated from all persons and items, including the currently targeted ones). The item or person measure for the current classification is then computed, along with its S.E. Mathematically, it is unlikely that no bias effects will be observed, or that bias sizes will cancel out exactly. The DIF contrast in [Table 30](#) and 31 is the difference between the DIF sizes, and is a log-odds estimate, equivalent to a [Mantel-Haenszel](#) DIF size. The t is the DIF contrast divided by the joint S.E. of the two DIF measures. It is equivalent to the Mantel-Haenszel significance test, but has the advantage of allowing missing data. This analysis is the same for all item types supported by Winsteps (dichotomies, rating (or partial credit) scales, etc.).

To replicate this with Winsteps yourself:

From a run of all the data, produce a PFILE=pf.txt and a SFILE=sf.txt

Then for each person classification of interest:

PAFILE=pf.txt

SAFILE=sf.txt

PSELECT=?????X; to select only the person classification of interest

IFILE = X.txt ; item difficulties for person classification on interest

CONVERGE=L ; only logit change is used for convergence

LCONV=0.005 ; logit change too small to appear on any report.

Do this for each class.

The [IFILE=](#) values should match the values shown in [Table 30.2](#)

To graph the ICCs for different [non-uniform DIF](#) for different classification groups on the same plot, see [DIF item characteristic curves](#).

Classification sizes

There is no minimum size, but the smaller the classification size (also called reference groups and focal groups), the less sensitive the DIF test is statistically. Generally, results produced by classifications sizes of less than 30 are too much influenced by idiosyncratic behavior to be considered dependable.

Do DIF effects sum to zero?

If the DIF person-classification groups have the same average ability and roughly the same size (see [Table 28 - PSUBTOTAL=](#)), then the DIF effects will sum to near 0.0. But if the DIF groups differ in average ability or differ in size, then the DIF terms will not cancel out. This is because the baseline measure is the average of the observed performances.

Example: There are 10 boys and 1000 girls. The baseline measure will be close to the girls' DIF measure. The boys' DIF measure can be considerably different.

Effect of imprecision in person or item estimates

This computation treats the person measures (for DIF) or the item measures (for DPF) as point estimates (i.e., exact). You can inflate the reported standard errors to allow for the imprecision in those measures. Formula 29 of Wright and Panchapakesan (1969), www.rasch.org/memo46.htm, applies. You will see there that, for dichotomies, the most by which imprecision in the baseline measures can inflate the variance is 25%. So, if you multiply the DIF/DPF point estimate S.E. by

sqrt(1.25) = 1.12 (and divide the t by 1.12), then you will be as conservative as possible in computing the DIF/DPF significance.

Impact on Person/Item Measurement

Unless DIF/DPF is large and mostly in one direction, the impact of DIF/DPF on person/item measurement is generally small. Wright & Douglas (1976) Rasch Item Analysis by Hand. "In other work we have found that when [test length] is greater than 20, random values of [item calibration mis estimation] as high as 0.50 have negligible effects on measurement."

Wright & Douglas (1975) Best Test Design and Self-Tailored Testing. "They allow the test designer to incur item discrepancies, that is item calibration errors, as large as 1.0. This may appear unnecessarily generous, since it permits use of an item of difficulty 2.0, say, when the design calls for 1.0, but it is offered as an upper limit because we found a large area of the test design domain to be exceptionally robust with respect to independent item discrepancies."

If DIF is large, then [splitting](#) the DIF item may be productive.

DIF/DPF statistical significance

[Table 30.1](#) shows pair-wise test of the statistical significance of DIF across classes. [Table 30.2](#) shows statistical significance of DIF for a class against the average difficulty. A statistical test for DIF for multiple classes on one item is a "fixed effects" chi-square of homogeneity. For L measures, D_i , with standard errors SE_i , a test of the hypothesis that all L measures are statistically equivalent to one common "fixed effect" apart from measurement error is a chi-square statistics with L-1 d.f. where $p > .05$ (or $> .01$) indicates statistically equivalent estimates.

$$\chi^2 = \sum_{i=1}^L \frac{D_i^2}{SE_i^2} - \left(\left(\sum_{i=1}^L \frac{D_i}{SE_i^2} \right)^2 / \sum_{i=1}^L \frac{1}{SE_i^2} \right)$$

Non-Uniform DIF or DPF

To investigate this with the Winsteps, include in the item or person label a stratification variable, indicating, low, middle or high performers (or item difficulties). Use this as the classification variable for [DIF=](#) or [DPF=](#). Also view the Graphs menu, [Non-uniform DIF](#).

DIF and Extreme Scores

DIF: For non-extreme-score items, extreme-score persons are excluded.
 For extreme-score items, everyone is included, but no DIF is computed:

KID	OBSERVATIONS	BASELINE	DIF	DIF	DIF	DIF	DIF	TAP				
CLASS	COUNT	AVERAGE	EXPECT	MEASURE	SCORE	MEASURE	SIZE	S.E.	t	Prob.	Number	Name
Item 1 is extreme (everyone succeeded on it). All persons are included in the DIF report												
M	17	1.00	1.00	-6.52	.00	-6.52	.00	.00	.00	1.000	1	1-4
Item 4 is not extreme. Extreme persons (one) is excluded:												
M	16	1.00	.96	-4.38	.04	-4.85	-.47	1.57	-.30	.7695	4	1-3-4

DPF: For non-extreme-score persons, extreme-score items are excluded.
 For extreme-score persons, every items is included, but no DPF is computed:

The Mathematics of Winsteps DIF and DPF Estimation

A first approximation is:
 size of interaction (in logits) = (sum of observational residuals) / (sum of model observational variances)
 For each observation, its residual and its model variance are reported in the [XFILE=](#)

The DIF and DPF are estimated as minor effects in a logit-linear procedure. The major effects are the person abilities, item difficulties, and rating scale structures. The approach in Winsteps parallels the use of logit models in Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. Journal of Educational Statistics, 7, 105-107; Van Der Flier, H., Mellenbergh, G. J., Adèr, H. J. & Wijn, M. (1984). An iterative item bias detection method. Journal of Educational Measurement, 21, 131-145; Kok, F. G., Mellenbergh, G. J. & Van Der Flier, H. (1985). Detecting experimentally induced item bias using the iterative logit method. Journal of Educational Measurement, 22, 295-303.

Algebraically, the general model is in two stages:

Stage 1: $\text{Log} (P_{nij} / P_{ni(j-1)}) = B_n - D_{gi} - F_{gj}$

Where B_n is the ability of person n , D_{gi} is the difficulty of person i in classification g , F_{gj} is the Rasch-Andrich threshold measure of category j relative to category $j-1$ for items in item-grouping g .

For the Rasch dichotomous model, all items are in the same item-grouping (so that g is omitted), and there are only two categories, with $F_1=0$.

For the Andrich rating-scale model, all items are in the same item-grouping (so that g is omitted), and there are more than two categories, with $\text{sum}(F_j)=0$.

For the Masters' partial-credit model, each item is in its own item-grouping ($g=i$), and there are more than two categories, with $\text{sum}(F_{ij})=0$. To re-parameterize into the conventional partial-credit model formulation, $D_i + F_{ij} = D_{ij}$.

Estimates of B_n , D_{gi} and F_{gj} are obtained.

Stage 2: [Table 30](#): For person-sub-sample (ps) DIF: $\text{Log} (P_{nij} / P_{ni(j-1)}) = B_n - D_{gi} - F_{gj} - \text{DIF}(ps)_i$

[Table 31](#): For item-sub-sample (is) DPF: $\text{Log} (P_{nij} / P_{ni(j-1)}) = D_n - D_{gi} - F_{gj} + \text{DPF}(is)_n$

[Table 33](#): For person-sub-sample item-sub-sample (ps)(is) DIPF: $\text{Log} (P_{nij} / P_{ni(j-1)}) = B_n - D_{gi} - F_{gj} + \text{DIPF}(ps)(is)$

Estimates of b_n , d_{gi} and f_{gj} anchored (fixed) from stage 1. The estimates of DIF, DPF or DIPF are the maximum likelihood estimates for which the marginal residuals for the sub-samples from the stage 1 analysis are the sufficient statistics. All these computations are as sample-distribution-free as is statistically possible, except when the sub-sampling is based on the sample-distribution (e.g., when persons are stratified into sub-samples according to their ability estimates.)

Different forms of DIF detection

The "comparison of item locations with two separate analyses" is what I call "Differential Test Functioning" (DTF). For instance, if the high-performers are more careful in the way they respond to the items than the low performers, then the test will have higher Test Discrimination for the high-performers than the low performers. The spread of the item difficulties for the high performers will be wider than for the low performers. If the mean item difficulty is the same in both analyses, then the highest and lowest difficulty items are most likely to be reported as significantly different. DTF is illustrated at www.rasch.org/rmt/rmt163g.htm

In Winsteps, DIF investigates each item by itself (like Mantel-Haenszel, MH). In MH, the person raw-scores are held constant. In Winsteps, the Rasch ability estimates from the analysis of all the data are held constant (also Andrich Thresholds). The abilities provide the common logit measurement scale. The assumption in this method is that performance on all the items with all the data is the best item-independent indicator of each person's ability measure. The difficulty of each item is then estimated for each group on that common "ability" scale. Advantages for a Rasch approach to DTF or DIF are that the Rasch approach allows for missing data and high-low ability group-splits, but MH does not. Of course, MH has the authority of ETS and that is decisive in legal situations.

DTF: a cross-plot of item difficulties derived from independent runs of the focal and reference classifying-groups, is basically reporting "Is the instrument working differently for the two sample classifications?", and, if so, "Where are the most conspicuous differences?" In the old days, when much analysis was done by hand, this would identify which items to choose for more explicitly constructed DIF tests, such as [Mantel-Haenszel](#). From these plots we can get approximate DIF t-tests. This approach is obviously useful - maybe more useful than the item-by-item DIF tests. But it allows DIF in an item to change the person measures, and to alter the difficulties of other items and to change the rating (or partial credit) scale structure. To apply this "Differential Test Functioning" approach to DIF detection, perform independent analyses of each sample class, produce [FILE=](#) and cross-plot the measures using the [Compare Statistics](#) plot.

But, it is the item-by-item DIF tests that have traditional support. So, for these, we need to hold everything else constant while we examine the DIF of each item. This is what Mantel-Haenszel does (using person raw scores), or the Winsteps DIF Table does (using person measures).

The Winsteps DIF table is equivalent to constructing a "ruler" based on the persons, and measuring the items on it, first for one person-group, then for the other person-group. The equivalent procedure is:

- (a) The joint run of all person-group classifications, producing anchor values for person abilities and rating (or partial credit) scale structure.
- (b) The classification A run with person abilities and rating (or partial credit) scale structure anchored at their joint values to produce person-group classification A item difficulties.
- (c) The classification B run with person abilities and rating (or partial credit) scale structure anchored at their joint values to produce person-group classification B item difficulties.
- (d) Pairwise item difficulty difference t-tests between the two sets of item difficulties (for person-group classifications A and B).

Lord's Chi-square DIF method takes a different approach, automatically looking for a core of stable items, but it is accident-prone and appears to over-detect DIF. In particular, if items were slightly biased, 50% against boys and 50% against girls, it would be accidental which set of items would be reported as "unbiased" and which as "biased".

Mantel-Haenszel method. See [Mantel and Mantel-Haenszel DIF statistics](#)

ANOVA method. This can be facilitated by Winsteps.

- (1) Identify the relevant demographic variable in the person label, and set ITEM1= at the variable, and NAMLEN=1.
- (2) Perform a standard Winsteps analysis
- (3) Use USCALE=, UMEAN= and UDECIM= to transform the person measures into convenient "class intervals": integers with lowest value 1, and highest value 10 for 10 class intervals (or a different number of intervals, as you prefer).
- (4) Write out an XFILE= selecting only:
 - person measure (class interval)
 - standardized residual
 - person label (demographic variable)
- (5) Read this file into your statistics software.
- (6) Transform the demographic variable into values such as 1 and 2 (whatever is most convenient for your statistical software)
- (7) Perform the "fully randomized" ANOVA with standardized residual as the dependent variable, and person measure and person label as the independent variables.

Two-Stage or Multiple-Stage DIF

Zenisky, A. L., Hambleton, R. K., & Robin, F. (2003). Detection of differential item functioning in large-scale state assessments: A study evaluating a two-stage approach. *Educational and Psychological Measurement*, 63, 51-64.

Stage 1: perform a standard DIF analysis Winsteps [Table 30](#).

Stage 2: for items identified with DIF in Stage 1, say items 8, 10, 14, give them a weight of zero using [IWEIGHT=](#), [IWEIGHT=*](#)

```
8 0
10 0
14 0
*
```

reanalyze the data and produce Winsteps Table 30.

This will show the DIF for all items (include 8, 10, 14) but using the purified scoring.

Resolving DIF

There are several tactics when DIF is discovered.

- 1. Ignore it as inherent in the measurement system. For instance, in a test of English as a Second Language, different items will exhibit DIF for speakers of different native languages, depending on the relationship of those languages to English. When constructing such tests, content experts should try to balance the DIF across the major groups, but this can never be done perfectly.
- 2. Remove the item (perhaps rewriting it). This is when the DIF is seen to be a confounding factor which overwhelms the intention of the item. For instance, a mathematics word problem which uses a technical cooking or car-mechanics term better known to girls or boys. [DELETE=](#)
- 3. Treat the data for one DIF group as missing data. For instance, if a generally good item is misunderstood by speakers of a specific dialect of English, then make the item missing data for them. [IPRFILE=](#)

4. Split the item. Make the item two items, each with active data for one group and missing data for the other group. This maintains the same raw scores, but produces different measures for each group. [FORMAT=](#) and [IPRFILE=](#)

The decision about which approach to use is often driven by political and legal considerations. So the literature either focuses on the mathematics of DIF remedies or contains case studies of the resolution of DIF under specific conditions.

Most important in resolving DIF is that we can communicate our DIF-resolution process simply, clearly and precisely to our audience. This will remove any feeling of unfairness or bias. "Mathe-magical" approaches to DIF resolution may make some in the audience wonder whether something underhanded is being perpetrated upon them.

19.29 DIF for two person groups

You have two person-groups in your DIF analysis. For your two groups on item 1.

[Table 30.1](#)

PERSON	DIF	DIF	PERSON	DIF	DIF	DIF	JOINT	Rasch-Welch	MantelHanzl	ITEM			
CLASS	MEASURE	S.E.	CLASS	MEASURE	S.E.	CONTRAST	S.E.	t	d.f.	Prob.	Prob.	Size	Number
1	-.47	.16	2	-1.57	.51	1.10	.53	2.07	62	.0422	.1986	+	1

The DIF measure (item difficulty on the latent variable) for Person group 1 on item 1 is -.47 logits
 The DIF measure (item difficulty on the latent variable) for Person group 2 on item 1 is -1.57 logits

DIF Contrast = difference between item difficulties for the two groups = -.47 - -1.57 = 1.10 logits

[Table 14.1](#)

ENTRY	TOTAL	TOTAL	MODEL	INFI	OUTFIT	PT-MEASURE	EXACT MATCH					
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%
1	394	83	-.76	.16	1.07	.3	.77	-.3	.27	.24	81.9	82.1

The baseline difficulty (from Table 14.1) is -0.76 logits.

[Table 30.2](#)

PERSON	OBSERVATIONS	BASILINE	DIF	DIF	DIF	DIF	DIF	DIF	ITEM		
CLASS	COUNT	AVERAGE	EXPECT	MEASURE	SCORE	MEASURE	SIZE	S.E.	t	Prob.	Number
1	42	4.57	4.77	-.76	-.19	-.47	.29	.16	1.83	.0751	1
2	41	4.93	4.73	-.76	.20	-1.57	-.81	.51	-1.61	.1158	1

DIF Size = DIF Measure - baseline difficulty

The DIF Size (relative to the baseline item difficulty) for Person group 1 on item 1 is .29 logits = -.47 - - .76
 The DIF measure (relative to the baseline item difficulty) for Person group 2 on item 1 is -.81 logits = -1.57 - -.76

DIF Contrast = difference between item difficulties for the two groups = .29 - -.81 = 1.10 logits (same as above) = paired DIF effect for the two groups.

It is easiest to look at the [DIF Size](#) plot. The difference between the two lines is the DIF Contrast. You can plot the difference from the Worksheet column D - column E.

19.30 Dimensionality

See [Multidimensionality](#)

19.31 Dimensionality investigation - an example

Michael Commons asked: "What in Winsteps tells how unidimensional a fit is?"

The Rasch measures estimated by unidimensional Rasch models are forced to be unidimensional. Off-dimensional aspects of the data are in the part of the data not explained by the Rasch measures, i.e., the Rasch residuals. The Rasch residuals decompose into (a) the randomness predicted by the Rasch model, and (b) components on dimensions other than the unidimensional Rasch variable, (c) off-dimensional noise, such as random guessing.

In empirical data, (a) and (c) usually dominate (b), so that item-level or person-level fit statistics tend to be insensitive to multidimensionality, as R.P. McDonald (1985) reports. Accordingly we must focus on techniques that quantify (b), such as PCA of residuals. If the eigenvalues reported by PCA approximate the size predicted by the Rasch model, then the data are effectively unidimensional. Otherwise, the bigger the eigenvalues, the less unidimensional are the data.

Roderick P. McDonald, Factor Analysis and related methods. Hillsdale, NJ: Lawrence Erlbaum Associates, 1985.

Multidimensionality is complicated, because it depends on the purpose of the instrument.

Here is a multidimensional instrument. It has 5 items: 1 geography item. 1 history item. 1 cooking item. 1 carpentry item. 1 arithmetic item.

This instrument is multidimensional, but none of its dimensions is big enough to make a unidimensional instrument. We cannot split this multidimensional instrument into unidimensional instruments, but that does not make the instrument unidimensional. However, we may decide that this instrument is measuring a general "education" variable, and declare that this instrument is **unidimensional for our purposes**.

For instance, an arithmetic test (addition, subtraction, multiplication, division) is unidimensional from the perspective of school administrators deciding whether a child should advance to the next grade-level, but the same test is multidimensional from the perspective of the school psychologist diagnosing learning difficulties. For instance, learning difficulties with subtraction in young children may indicate social maladjustment.

Also

Here is an example. We can proceed as follows:

a. Compare the Raw Variance explained by items (19.8%) with the Unexplained variance in 1st contrast (7.1%). Is this ratio big enough to be a concern? In your analysis, the Rasch dimension dominates (almost 3 times the secondary dimension), but the secondary dimension is noticeable.

b. Is the secondary dimension bigger than chance? Eigenvalue = 2.8. This is the strength of 3 items. We do not expect a value of more than 2 items by chance. www.rasch.org/rmt/rmt191h.htm - and we would also need at least 2 items to think of the situation as a "dimension" and not merely an idiosyncratic item.

c. Does the secondary dimension have substance? Looking at your plot, we can see that items ABCDE are separated vertically (the important direction) from the other items. They are the core of the secondary dimension. 5 items are enough items that we could split them into a separate instrument (exactly as we could with "subtraction" on an arithmetic test). Is this secondary dimension important enough, and different enough, that we would consider reporting two measures (one for ABCDE and one for the other items) rather than one measure for all items combined? The content of ABCDE appears to be psycho-social (e.g., one item includes the word "anxious" in this example). The other items are more physical (e.g., one item includes the word "walking" in this example). Consider the purpose of the instrument. Is "anxious" important or not? Is it part of the central purpose for the instrument? Would the instrument be improved or degraded (from a usefulness perspective) if items ABCDE were omitted? Would the instrument be improved or degraded (from a usefulness perspective) if a separate measure was reported for items like ABCDE?

d. Rasch-analyze the sample on the ABCDE items and then on the other items. Cross-plot the person measures. Look at the correlation of the two sets of person measures (and the correlation disattenuated for measurement error). Is the correlation noticeably low? In this example, the disattenuated correlation was 0.82, indicating that the dimensions share explains about 67% of the person measure variance.

We expect most people to lie along a statistical diagonal. Who is off-diagonal? (Perhaps the people with social problems.) Are they important enough to merit a separate measurement system? For instance, on an English-language test, native-speakers of English, and second-language speakers usually have different profiles. Native speakers speak relatively better. Second-language speaker may spell relatively better. But two measures of English-language-proficiency are rarely reported.

If you decide that the secondary dimension is important enough to merit two measures, or the secondary dimension is off-dimension enough to merit omitting its items, then the instrument is multidimensional (in practice). If not, then the instrument is unidimensional (in practice), no matter what the statistics say.

Tentative guidelines based on the % of the sample are sampling dependent. If you are planning to apply a criterion such as "5% of the sample", then verify that your sample matches the intended target population of the instrument. In general, 5% seems very low. Would we institute a special measurement system for 1 child in a classroom of 20 children? Unlikely? We would probably need at least 4 children = 20% before we would consider reporting (and acting on) two measures.

In the USA, African-Americans comprise 13% of the population, and there is a debate about whether or not they should have special measurement systems. In some situations they do. And, similarly, whether there should be special provision for Spanish-speakers (15% of the USA population). In some situations there are. These percentages suggest that a threshold of about "10% of the sample" may be reasonable for separate measurement procedures.

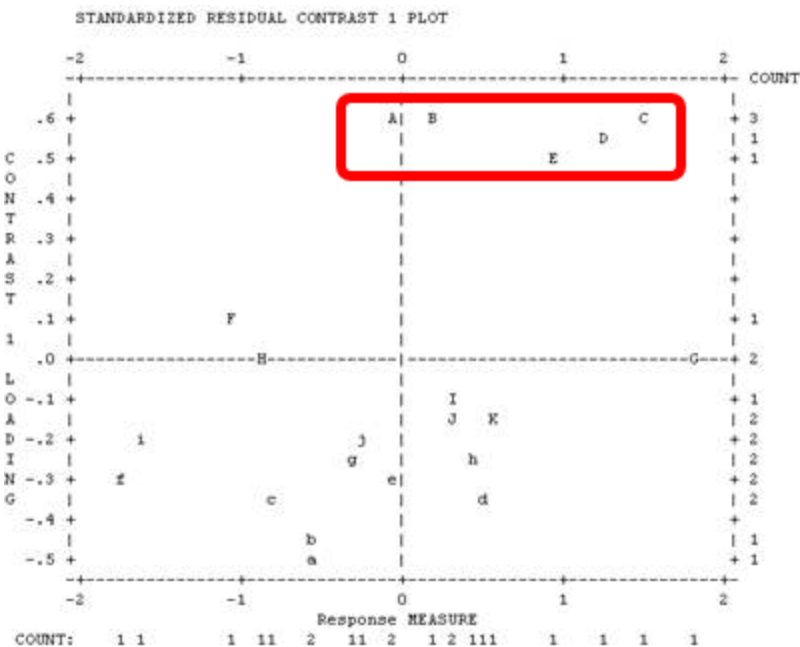
My conclusion about this instrument (knowing nothing about its practical purpose) would be that the instrument is multidimensional and that items ABCDE should be omitted (or rewritten or replaced to emphasize their physical rather than their psychological aspects).

"Unidimensionality" is a choice based on the circumstances, so, if you are writing a paper, then please include a discussion of why (or why not) you decided that the instrument is multidimensional. This would be helpful to other researchers.

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

		-- Observed --		Expected
Total raw variance in observations	=	39.8	100.0%	100.0%
Raw variance explained by measures	=	18.8	47.2%	48.0%
Raw variance explained by persons	=	10.9	27.4%	27.8%
Raw Variance explained by items	=	7.9	19.8%	20.1%
Raw unexplained variance (total)	=	21.0	52.8%	52.0%
Unexplned variance in 1st contrast	=	2.8	7.1%	13.5%
Unexplned variance in 2nd contrast	=	2.6	6.5%	12.3%
Unexplned variance in 3rd contrast	=	2.1	5.4%	10.2%
Unexplned variance in 4th contrast	=	1.7	4.4%	8.3%
Unexplned variance in 5th contrast	=	1.6	4.0%	7.6%

1st contrast:



CON-	TRAST	LOADING	MEASURE	INFINIT	OUTFIT	
			MNSQ	MNSQ	MNSQ	N
1	.62	-.03	1.17	.98		A
1	.62	.17	1.30	1.22		B
1	.61	1.53	1.23	1.64		C
1	.57	1.26	1.32	1.34		D
1	.51	.96	1.07	1.20		E
1	.12	-1.08	.85	.46		F
1	.01	1.82	1.32	1.49		G
1	-.49	-.57	.93	.72		a
1	-.44	-.53	.84	.72		b
1	-.36	-.78	1.03	.97		c
1	-.34	.50	1.04	.97		d
1	-.30	-.09	.79	.72		e
1	-.28	-1.75	.76	.83		f
1	-.27	-.31	.77	.63		g
1	-.24	.45	.99	.91		h
1	-.22	-1.60	.66	.43		i
1	-.19	-.26	.96	.85		j
1	-.16	.56	.67	.66		k
1	-.14	.34	1.22	1.29		J
1	-.10	.32	.99	1.22		I
1	-.02	-.90	.75	.74		H

Content here

Question: My instrument has 200 items and my First Contrast has an eigenvalue of 9.2. Could this be explained by the large number of items analyzed?

Answer: Yes, the more items there are, then the more likely that a random cluster of items are inter-correlated by accident.

So, the first step with the potentially 9 or so items in a secondary dimension is to look at the plot in Winsteps [Table 23.1](#). Is there a cluster of items at the top or bottom of the plot that share a content area that differs noticeably from the other items?

If so, this could be a secondary dimension. So, the next thing is to look at the table of correlations in [Table 23.0](#). This reports the correlations of the person measures on the different clusters of items identified in the first contrast. Is the disattenuated correlation between the person measures on the suspect cluster of items and the person measures on the other items low or negative?

If so, the suspect cluster of items is measuring something different. It is a different dimension. When items on this secondary dimension are included with the items on the dominant dimension, then responses to the secondary items look like random noise from the viewpoint of the dominant dimension.

If not, the suspect cluster of items is probably a secondary strand in the content area, similar to "word problems" on an arithmetic test. Performance profiles of the person across strands will not be even, but will also not be contradictory.

The "variance explained" can be somewhat misleading because it is dominated by the variance of the item difficulties and the variance of the person abilities. We may be contrasting a wide range of abilities and/or difficulties on the dominant dimension against a narrow range of abilities and/or difficulties on the secondary dimension.

19.32 Dimensionality: contrasts & variances

Please explain Rasch PCA of Residuals in plain English for people like me who are not particularly good at mathematical terminology.

Rasch "PCA of residuals" looks for patterns in the part of the data that does not accord with the Rasch measures. This is the "unexpected" part of the data. We are looking to see if groups of items share the same patterns of unexpectedness. If they do, then those items probably also share a substantive attribute in common, which we call a "secondary dimension". Then our questions are:

1. "What is the secondary dimension?" - to discover this we look at the contrast between the content of the items at the top, A,B,C, and the bottom, a,b,c, of the contrast plot in [Table 23.2](#). For instance, if the items at the top are "physical" items and the items at the bottom are "mental" items. Then there is a secondary dimension with "physical" at one end, and "mental" at the other.

2. "Is the secondary dimension big enough to distort measurement?" - usually the secondary dimension needs to have the strength of at least two items to be above the noise level. We see the strength (eigenvalue) in the first column of numbers in [Table 23.0](#).

3. "What do we do about it?" - often our decision is "nothing". On a math test, we will get a big contrast between "algebra" and "word problems". We know that they are conceptually different, but they are both part of math. We don't want to omit either of them, and we don't want separate "algebra" measures and "word problem" measures. So the best we can do is to verify that the balance between the number of algebra items and the number of word-problem items is in accordance with our test plan.

But we may see that one end of the contrast is off-dimension. For instance, we may see a contrast between "algebra items using Latin letters" and "algebra items using Greek letters". We may decide that knowledge of Greek letters is incidental to our purpose of measuring the understanding of algebraic operations, and so we may decide to omit or revise the Greek-letter items.

In summary: Look at the content (wording) of the items at the top and bottom of the contrast: items A,B,C and items a,b,c. If those items are different enough to be considered different dimensions (similar to "height" and "weight"), then split the items into separate analyses. If the items are part of the same dimension (similar to "addition" and "subtraction" on an arithmetic test), then no action is necessary. You are seeing the expected co-variance of items in the same content area of a dimension.

For more discussion see [Dimensionality: when is a test multidimensional?](#)

Principal-Components Analysis of Residuals is not interpreted in the same way as Common-Factor Analysis of the original data.

In Common-Factor analysis, we try to optimize the commonalities, maximization, rotation and obliqueness to give the strongest possible factor structure, where the unstandardized "raw" factor loadings are interpreted as the correlations with the latent factors.

In PCA of residuals, we are trying to falsify the hypothesis that the residuals are random noise by finding the component that explains the largest possible amount of variance in the residuals. This is the "first contrast" (or first PCA component in the correlation matrix of the residuals). If the eigenvalue of the first contrast is small (usually less than 2.0) then the first contrast is at the noise level and the hypothesis of random noise is not falsified in a general way. If not, the loadings on the first contrast indicate that there are contrasting patterns in the residuals. The absolute sizes of the loadings are generally inconsequential. It is the patterns of the loadings that are important. We see the patterns by looking at the plot of the loadings in Winsteps Table 23.2, particularly comparing the top and bottom of the plot.

So, if we notice that the 1st contrast has an eigenvalue of 3 (the strength of 3 items), and we see on the contrast plot that there is a group of 3 items (more or less) outlying toward the top or the bottom of the plot, then we attribute the first contrast to the fact that their pattern contrasts with the pattern of the other items. We then look at the content of those items to discover what those items share that contrasts with the content of the other items.

Please do not interpret Rasch-residual-based Principal Components Analysis (PCAR) as a usual factor analysis. These components show contrasts between opposing factors, not loadings on one factor. Criteria have yet to be established for when a deviation becomes a dimension. So PCA is indicative, but not definitive, about secondary dimensions.

In conventional factor analysis, interpretation may be based only on positive loadings. In the PCA of Residuals, interpretation must be based on the contrast between positive and negative loadings.

The "first factor" (in the traditional Factor Analysis sense) is the Rasch dimension. By default all items (or persons) are in the "first factor" until proven otherwise. The first contrast plot shows a contrast within the data between two sets of items orthogonal to the Rasch dimension. We usually look at the plot and identify a cluster of items at the top or bottom of the plot which share most strongly some substantive off-Rasch-dimension attribute. These become the "second factor".

Winsteps is doing a PCA of residuals, not of the original observations. So, the first component (dimension) has already been removed. We are looking at secondary dimensions, components or contrasts. When interpreting the meaning of a component or a factor, the conventional approach is only to look at the largest positive loadings in order to infer the

substantive meaning of the component. In Winsteps PCA this method of interpretation can be misleading, because the component is reflecting opposing response patterns across items by persons. So we need to identify the opposing response patterns and interpret the meaning of the component from those. These are the response patterns to the items at the top and bottom of the plots.

Sample size: A useful criterion is 100 persons for PCA of items, and 100 items for PCA of persons, though useful findings can be obtained with 20 persons for PCA of items, and 20 items for PCA of persons.

Arrindell, W. A., & van der Ende, J. (1985). An empirical test of the utility of the observations-to-variables ratio in factor and components analysis. *Applied Psychological Measurement*, 9, 165 - 178.

Compatibility with earlier computation: The Winsteps algorithm was changed to align more closely with the usual practice in statistics of explaining raw-score variance (parallel to Outfit). The earlier method in Winsteps was explaining the statistical-information variance (parallel to Infit). Since the outlying observations have high raw-score variance, but low statistical-information variance, the previous computation showed Rasch explaining a higher proportion of the variance.

If you want to do a more conventional interpretation, then output the [ICORFIL=](#) correlation matrix from the Output Files menu. You can read this into a factor analysis program, such as SAS or SPSS. You can then do a PCA or CFA (common factor analysis) of the correlation matrix, with the usual obliquenesses, rotations etc.

In Winsteps, you can also do a PCA of the original observations by specifying [PRCOMP=Obs](#)

Example from Table 23.0 from [Example0.txt](#):

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)				
		-- Observed --		Expected
Total raw variance in observations	=	51.0	100.0%	100.0%
Raw variance explained by measures	=	26.0	51.0%	50.8%
Raw variance explained by persons	=	10.8	21.2%	21.2%
Raw Variance explained by items	=	15.1	29.7%	29.6%
Raw unexplained variance (total)	=	25.0	49.0%	100.0%
Unexplnd variance in 1st contrast	=	4.6	9.1%	18.5%
Unexplnd variance in 2nd contrast	=	2.9	5.8%	11.8%
Unexplnd variance in 3rd contrast	=	2.3	4.5%	9.2%
Unexplnd variance in 4th contrast	=	1.7	3.4%	6.9%
Unexplnd variance in 5th contrast	=	1.6	3.2%	6.5%

The Rasch dimension explains 51.0% of the variance in the data: good! The largest secondary dimension, "the first contrast in the residuals" explains 9.1% of the variance - somewhat greater than around 4% that would be observed in data like these simulated to fit the Rasch model. Check this by using the [SIMUL=](#) option in Winsteps to simulate a Rasch-fitting dataset with same characteristics as this dataset. Then produce this Table 23 for it. Also see: www.rasch.org/rmt/rmt191h.htm

In these data, the variance explained by the items, 29.7% is only three times the variance explained by the first contrast 9.1%, so there is a noticeable secondary dimension in the items. The eigenvalue of the first contrast is 4.6 - this indicates that it has the strength of about 5 items (4.6 rounded to 5, out of 25), somewhat bigger than the strength of two items (an eigenvalue of 2), the smallest amount that could be considered a "dimension". Contrast the content of the items at the top and bottom of the plot in Table 23.2 to identify what this secondary dimension reflects.

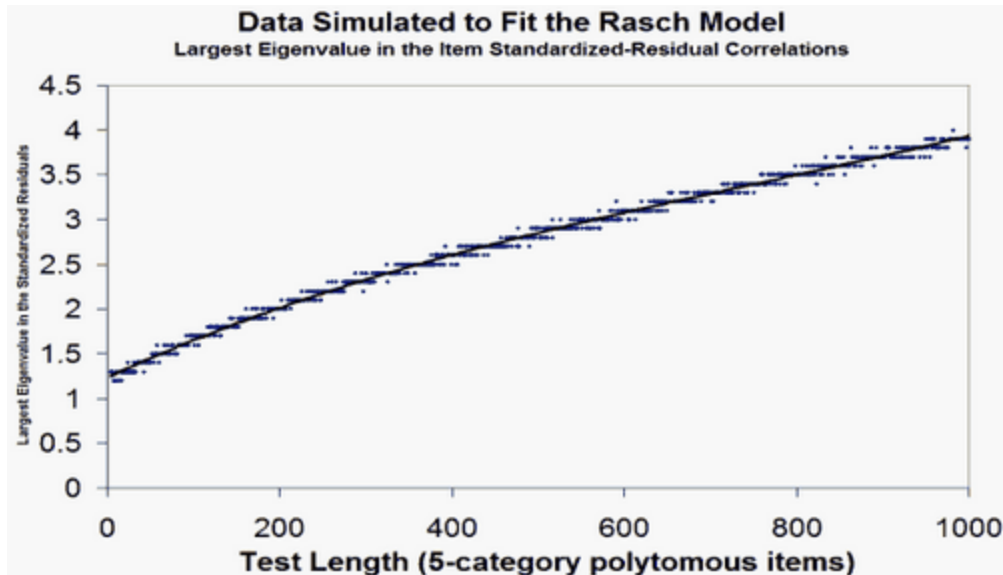
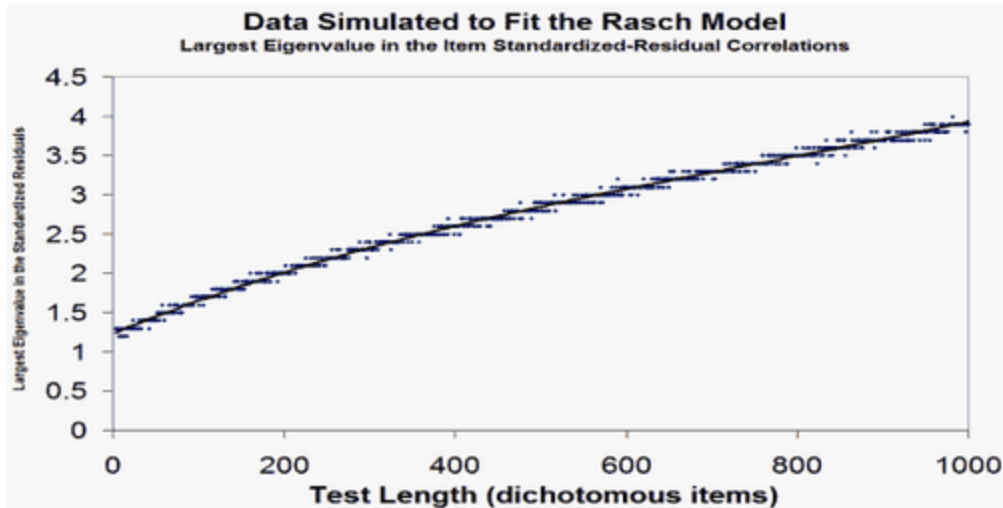
Tentative guidelines:

1. Is your person measure S.D. in [Table 3](#) what you expect (or better)?
Yes, then your "variance explained by persons" is also good.

2. 1. Is your item difficulty S.D. in [Table 3](#) what you expect (or better)?
Yes, then your "variance explained by the items" is also good.

3. Is your unexplained variance explained by 1st contrast (eigenvalue size) < 2.0 ?

Yes, then your biggest secondary dimension has the strength of less than 2 items. Good. For the expected size of the unexplained variance:



4. Now, please [simulate](#) Rasch-fitting data like yours, and compare [Table 23](#) and [Table 24](#) with it.

There is a paradox: "more variance explained" @ "more unidimensional" in the Guttman sense - where all unexplained variance is viewed as degrading the perfect Guttman uni-dimension. But "more unidimensional" (in the stochastic Rasch sense) depends on the size of the second dimension in the data, not on the variance explained by the first (Rasch) dimension. This is because most unexplained variance is hypothesized to be the random noise predicted by the Rasch model, rather than a degradation of the unidimensionality of the Rasch measures

Analytical Note:

Winsteps performs an unrotated "principal components" factor analysis. (using Hotelling's terminology). If you would like to rotate axes, have oblique axes, or perform a "common factor" factor analysis of the residuals, Winsteps can write out the matrices of residual item (or person) correlations, see the "Output Files" pull down menu or [ICORFIL=](#) and [PCORFIL=](#). You can import these into any statistics software package.

The purpose of PCA of residuals is not to construct variables (as it is with "common factor" analysis), but to explain variance. First off, we are looking for the contrast in the residuals that explains the most variance. If this contrast is at the

"noise" level, then we have no shared second dimension. If it does, then this contrast is the "second" dimension in the data. (The Rasch dimension is hypothesized to be the first). Similarly we look for a third dimension, etc. Rotation, oblique axes, the "common factor" approach, all reappportion variance, usually in an attempt to make the factor structure more clearly align with the items, but, in so doing, the actual variance structure and dimensionality of the data is masked.

In Rasch analysis, we are trying to do the opposite of what is usually happening in factor analysis. In Rasch analysis of residuals, we want not to find contrasts, and, if we do, we want to find the least number of contrasts above the noise level, each, in turn, explaining as much variance as possible. This is exactly what unrotated PCA does.

In conventional factor analysis of observations, we are hoping desperately to find shared factors, and to assign the items to them as clearly and meaningfully as possible. In this endeavor, we use a whole toolbox of rotations, obliquenesses and choices of diagonal self-correlations (i.e., the "common factor" approach).

But, different analysts have different aims, and so Winsteps provides the matrix of [residual correlations](#) to enable the analyst to perform whatever factor analysis is desired!

The Rasch Model: Expected values, Model Variances, and Standardized Residuals

The Rasch model constructs additive measures from ordinal observations. It uses disordering of the observations across persons and items to construct the additive frame of reference. Perfectly ordered observations would accord with the ideal model of Louis Guttman, but lack information as to the distances involved.

Since the Rasch model uses disordering in the data to construct distances, it predicts that this disordering will have a particular ideal form. Of course, empirical data never exactly accord with this ideal, so a major focus of Rasch fit analysis is to discover where and in what ways the disordering departs from the ideal. If the departures have substantive implications, then they may indicate that the quality of the measures is compromised.

A typical Rasch model is:

$$\log (P_{nik} / P_{ni(k-1)}) = B_n - D_i - F_k$$

where

P_{nik} = the probability that person n on item i is observed in category k , where $k=0,m$

$P_{ni(k-1)}$ = the probability that person n on item i is observed in category $k-1$

B_n = the ability measure of person n

D_i = the difficulty measure of item i

F_k = the structure calibration from category $k-1$ to category k

This predicts the observation X_{ni} . Then

$$X_{ni} = E_{ni} \pm \text{sqrt}(V_{ni})$$

where

$$E_{ni} = \sum (kP_{nik}) \text{ for } k=0,m.$$

This is the expected value of the observation.

$$V_{ni} = \sum (k^2P_{nik}) - (E_{ni})^2 \text{ for } k=0,m.$$

This is the model variance of the observation about its expectation, i.e., the predicted randomness in the data.

The Rasch model is based on the specification of "local independence". This asserts that, after the contribution of the measures to the data has been removed, all that will be left is random, normally distributed, noise. This implies that when a residual, $(X_{ni} - E_{ni})$, is divided by its model standard deviation, it will have the characteristics of being sampled from a unit normal distribution. That is:

$$(X_{ni} - E_{ni}) / \text{sqrt}(V_{ni}), \text{ the standardized residual of an observation, is specified to be } N(0,1)$$

The bias in a measure estimate due to the misfit in an observation approximates $(X_{ni} - E_{ni}) * S.E.^2(\text{measure})$

Principal Components Analysis of Residuals

"Principal Component Analysis (PCA) is a powerful technique for extracting structure from possibly high-dimensional data sets. It is readily performed by solving an eigenvalue problem, or by using iterative algorithms which estimate principal components [as in Winsteps]. ... some of the classical papers are due to Pearson (1901); Hotelling (1933); ... PCA is an orthogonal transformation of the coordinate system in which we describe our data. The new coordinate values by which we represent the data are called principal components. It is often the case that a small number of principal components is sufficient to account for most of the structure in the data. These are sometimes called factors or latent variables of the

data." (Schölkopf, D., Smola A.J., Müller K.-R., 1999, Kernel Principal Component Analysis, in Schölkopf et al. "Advances in Kernel Methods", London: MIT Press).
 Pearson, K. (1901) On lines and planes of closest fit to points in space. Philosophical Magazine, 2:559-572.
 Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24:417-441, 498-520.

The standardized residuals are modeled to have unit normal distributions which are independent and so uncorrelated. A PCA of Rasch standardized residuals should look like a PCA of random normal deviates. Simulation studies indicate that the largest component would have an eigenvalue of about 1.4 and they get smaller from there. But there is usually something else going on in the data, so, since we are looking at residuals, each component contrasts deviations in one direction ("positive loading") against deviation in the other direction ("negative loading"). As always with factor analysis, positive and negative loading directions are arbitrary. Each component in the residuals only has substantive meaning when its two ends are contrasted. This is a little different from PCA of raw observations where the component is thought of as capturing the "thing".

Loadings are plotted against Rasch measures because deviation in the data from the Rasch model is often not uniform along the variable (which is actually the "first" dimension). It can be localized in easy or hard items, high or low ability people. The Wright and Masters "Liking for Science" data is an excellent example of this.

Total, Explained and Unexplained Variances

The decomposition of the total variance in the data set proceeds as follows for the standardized residual, $PRCOMP=S$ and raw score residual $PRCOMP=R$, option.

(i) The average person ability measure, b , and the average item difficulty measure, d , are computed.

(ii) The expected response, E_{bd} , by a person of average ability measure to an item of average difficulty measure is computed. (If there are multiple rating or partial credit scales, then this is done for each rating or partial credit scale.)

(iii) Each observed interaction of person n , of estimated measure B_n , with item i , of estimated measure D_i , produces an observation X_{ni} , with an expected value, E_{ni} , and model variance, V_{ni} .

The raw-score residual, Z_{ni} , of each X_{ni} is $Z_{ni} = X_{ni} - E_{ni}$.

The standardized residual, Z_{ni} , of each X_{ni} is $Z_{ni} = (X_{ni} - E_{ni}) / \sqrt{V_{ni}}$.

Empirically:

(iv) The piece of the observation available for explanation by B_n and D_i is approximately $X_{ni} - E_{bd}$.

In raw-score residual units, this is $C_{ni} = X_{ni} - E_{bd}$

In standardized residual units, this is $C_{ni} = (X_{ni} - E_{bd}) / \sqrt{V_{ni}}$

The total variance sum-of-squares in the data set available for explanation by the measures is: $V_{Available} = \sum(C_{ni}^2)$

(v) The total variance sum of squares predicted to be unexplained by the measures is: $V_{Unexplained} = \sum(Z_{ni}^2)$

(vi) The total variance sum of squares explained by the measures is: $V_{Explained} = V_{Available} - V_{Unexplained}$

If $V_{Explained}$ is negative, see below.

Under model conditions:

(vii) The total variance sum of squares explained by the measures is:

Raw-score residuals: $V_{MExplained} = \sum((E_{ni} - E_{bd})^2)$

Standardized residuals: $V_{MExplained} = \sum((E_{ni} - E_{bd})^2 / V_{ni})$

(ix) The total variance sum of squares predicted to be unexplained by the measures is:

Raw score residuals: $V_{MUnexplained} = \sum(V_{ni})$

Standardized residuals: $V_{MUnexplained} = \sum(V_{ni} / V_{ni}) = \sum(1)$

x) total variance sum-of-squares in the data set predicted to be available for explanation by the measures is: $V_{MAvailable} = V_{MExplained} + V_{MUnexplained}$

Negative Variance Explained

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

Total variance in observations = 20.3 100.0%

Variance explained by measures = -23.7 -116.2%

According to this Table, the variance explained by the measures is less than the theoretical minimum of 0.00. This "negative variance" arises when there is unmodeled covariance in the data. In Rasch situations this happens when the randomness in the data, though normally distributed when considered overall, is skewed when partitioned by measure difference. A likely explanation is that some items are reverse-coded. Check that all correlations are positive by viewing the [Diagnosis Menu](#), Table A. If necessary, use [IREFER=](#) to recode items. If there is no obvious explanation, please email your control and data file to www.winsteps.com

Principal Components Analysis of Standardized Residuals

(i) The standardized residuals for all observations are computed. Missing observations are imputed to have a standardized residual of 0, i.e., to fit the model.

(ii) Correlation matrices of standardized residuals across items and across persons are computed. The correlations furthest from 0 (uncorrelated) are reported in Tables [23.99](#) and [24.99](#).

(iii) In order to test the specification that the standardized residuals are uncorrelated, it is asserted that all randomness in the data is shared across the items and persons. This is done by placing 1's in the main diagonal of the correlation matrix. This accords with the "Principal Components" approach to Factor Analysis. ("General" Factor Analysis attempts to estimate what proportion of the variance is shared across items and persons, and reduces the diagonal values from 1's accordingly. This approach contradicts our purpose here.)

(iv) The correlation matrices are decomposed. In principal, if there are L items (or N persons), and they are locally independent, then there are L item components (or N person components) each of size (i.e., eigenvalue) 1, the value in the main diagonal. But there are expected to be random fluctuations in the structure of the randomness. However, eigenvalues of less than 2 indicate that the implied substructure (dimension) in these data has less than the strength of 2 items (or 2 persons), and so, however powerful it may be diagnostically, it has little strength in these data.

(v) If items (or persons) do have commonalities beyond those predicted by the Rasch model, then these may appear as shared fluctuations in their residuals. These will inflate the correlations between those items (or persons) and result in components with eigenvalues greater than 1. The largest of these components is shown in Table [23.2](#) and [24.3](#), and sequentially smaller ones in later subtables.

(vi) In the Principal Components Analysis, the total variance is expressed as the sum of cells along the main diagonal, which is the number of items, L, (or number of persons, N). This corresponds to the total unexplained variance in the dataset, VUnexplained.

(vii) The variance explained by the current contrast is its eigenvalue.

Sample size: The more, the better ,....

"There are diminishing returns, but even at large subject to item ratios and sample sizes (such as 201 ratio or N > 1000) and with unrealistically strong factor loadings and clear factor structures, EFA and PCA can produce error rates up to 30%, leaving room for improvement via larger samples." Osborne, Jason W. & Anna B. Costello (2004). Sample size and subject to item ratio in principal components analysis. [Practical Assessment, Research & Evaluation, 9\(11\)](#)

Example: Item Decomposition

From Table [23.2](#): The Principal Components decomposition of the standardized residuals for the items, correlated across persons. Winsteps reports:

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)				
		Empirical		Modeled
Total variance in observations	=	1452.0	100.0%	100.0%
Variance explained by measures	=	1438.0	99.0%	98.6%
Unexplained variance (total)	=	14.0	1.0%	1.4%
Unexpl var explained by 1st contrast	=	2.7	.2%	

The first contrast has an eigenvalue size of 2.7 This corresponds to 2.7 items.

There are 14 active items, so that the total unexplained variance in the correlation matrix is 14 units.

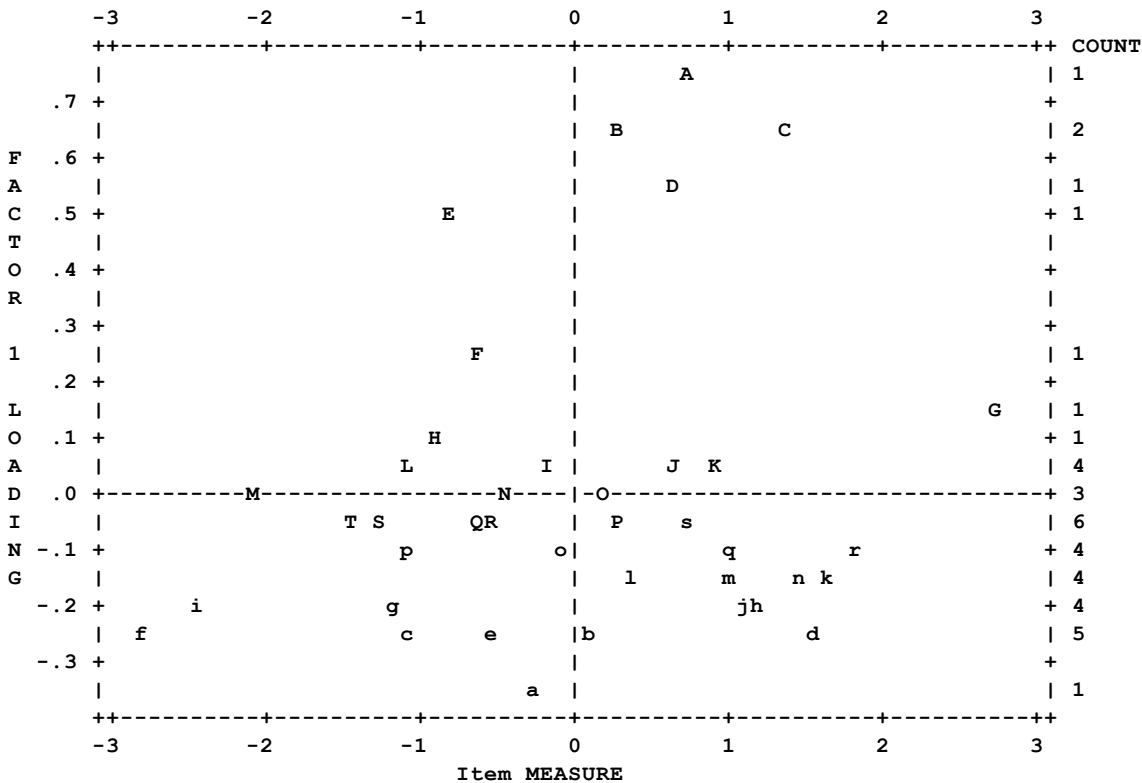
The "Modeled" column shows what this would have looked like if these data fit the model exactly.

Conclusion: Though this contrast has the strength of 3 items, and so might be independently constructed from these data, its strength is so small that it is barely a ripple on the total measurement structure.

Caution: The 1st contrast may be an extra dimension, or it may be a local change in the intensity of this dimension:

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

		Empirical		Modeled
Total variance in observations	=	97.1	100.0%	100.0%
Variance explained by measures	=	58.1	59.8%	59.0%
Unexplained variance (total)	=	39.0	40.2%	100.0%
Unexpl var explained by 1st contrast	=	2.8	2.9%	7.2%



CON-	TRAST	RAW LOADING	MEASURE	INFIT MNSQ	OUTFIT MNSQ	ENTRY NUMBER
1	A	.74	.71	.78	.70	26
1	B	.65	.26	.79	.68	23
1	C	.64	1.34	.87	.84	25
1	D	.56	.64	.85	.80	24
1	E	.51	-.85	.84	.60	22

The first contrast comprises items A-E. But their mean-squares are all less than 1.0, indicating they do not contradict the Rasch variable, but are rather too predictable. They appear to represent a local intensification of the Rasch dimension, rather than a contradictory dimension.

Comparison with Rasch-fitting data

Winsteps makes it easy to compare empirical PCA results with the results for an equivalent Rasch-fitting data set.

From the Output Files menu, make a "Simulated Data" file, call it, say, test.txt

From the Files menu, Restart Winsteps. Under "Extra specifications", type in "data=test.txt".

Exactly the same analysis is performed, but with Rasch-fitting data. Look at the Dimensionality table:

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

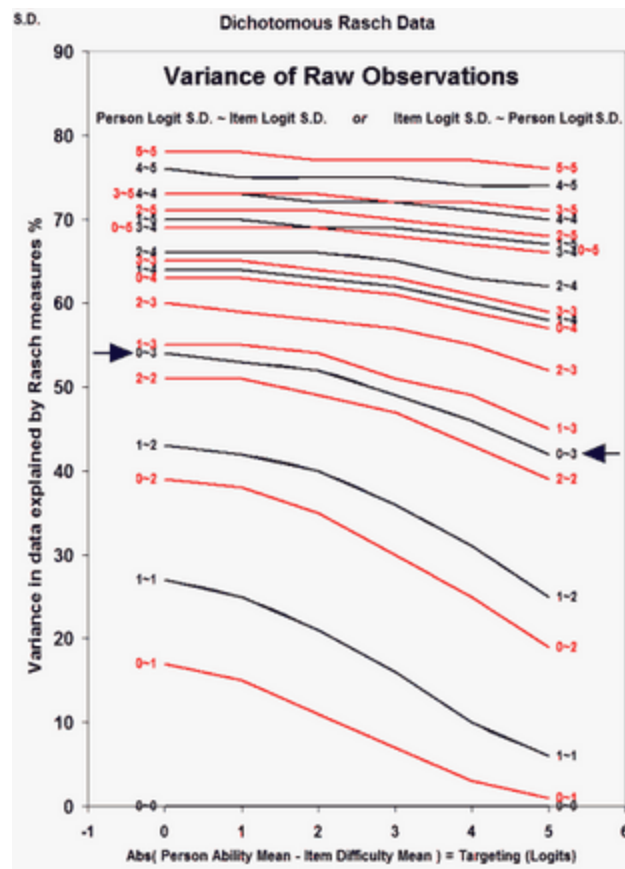
		Empirical	Modeled
Total variance in observations	=	576.8	100.0%
Variance explained by measures	=	562.8	97.6%
Unexplained variance (total)	=	14.0	2.4%
Unexpl var explained by 1st contrast	=	2.2	.4%

Repeat this process several times, simulating a new dataset each time. If they all look like this, we can conclude that the value of 2.7 for the 1st contrast in the residuals is negligibly bigger than the 2.2 expected by chance.

General Advice

A question here is "how big is big"? Much depends on what you are looking for. If you expect your instrument to have a wide spread of items and a wide spread of persons, then your measures should explain most of the variance. But if your items are of almost equal difficulty (as recommended, for instance, by G-Theory) and your persons are of similar ability (e.g., hospital nurses at the end of their training) then the measures will only explain a small amount of the variance.

Winsteps [Table 23.6](#) implements Ben Wright's recommendation that the analyst split the test into two halves, assigning the items, top vs. bottom of the first component in the residuals. Measure the persons on both halves of the test. Cross-plot the person measures. If the plot would lead you to different conclusions about the persons depending on test half, then there is a multidimensionality. If the plot is just a fuzzy straight line, then there is one, perhaps somewhat vague, dimension.



Tentative guidelines:

Please look the nomogram above from <http://www.rasch.org/rmt/rmt221j.htm> - it indicates the expected relationship between "explained variance" and "unexplained variance". 40%-50% are typical values.

In the unexplained variance, a "secondary dimension" must have the strength of at least 3 items, so if the first contrast has "units" (i.e., eigenvalue) less than 3 (for a reasonable length test) then the test is probably unidimensional. (Of course, individual items can still misfit).

Negative variance can occur when the unexpectedness in the data is not random. An example is people who flat-line an attitude survey. Their unexpected responses are always biased towards one category of the rating (or partial credit) scale.

Simulation studies indicate that eigenvalues less than 1.4 are at the random level. Smith RM, Miao CY (1994) Assessing unidimensionality for Rasch measurement. Chapter 18 in M. Wilson (Ed.) Objective Measurement: Theory into Practice. Vol. 2. Norwood NJ: Ablex.) On occasion, values as high as 2.0 are at the random level. (Critical Eigenvalue Sizes in Standardized Residual Principal Components Analysis, Raïche G., Rasch Measurement Transactions, 2005, 19:1 p. 1012.)

To compute these numbers yourself ...

Let's assume we have a complete dichotomous dataset of N persons and L items. Then, in summary,

1. Do the Rasch analysis. Write out the raw observations, the raw residuals, and the standardized residuals
2. Compute the observed variance of the raw observations = $OV = 100\%$ of the variance in the data
3. Compute the unexplained variance = mean-square (raw residuals) = $UV = \text{Unexplained } \%$
4. Compute the explained variance in the raw observations = $OV - UV = EV = \text{Explained } \%$ (can go negative!)
5. The eigenvalue of the item inter-correlation matrix of the standardized residuals = $L = UV$ (rescaled) = $\text{Unexplained } \%$
6. In the decomposition of the correlation matrix: Eigenvalue of the component (factor) = $G = \text{"strength in item units"} = \text{Unexplained } \% * G / L$

For the plots, the item difficulties are the x-axis, and the component loadings as the y-axis. This is a reminder that the Rasch dimension is the implicit first dimension in the original observations.

You will notice that some of this is statistically approximate, but more exact statistics have produced less meaningful findings. Together with Albert Einstein, trust the meaning more than the numbers!

19.33 Dimensionality: when is a test multidimensional?

For more discussion see [dimensionality and contrasts](#).

Beware of:

1. Accidents in the data generating spurious dimensions
2. Content strands within a bigger content area ("addition" and "subtraction" within "arithmetic") generating generally inconsequential dimensions.
3. Demographic groups within the person sample differential influencing item difficulty, so making a reference-group item dimension and a focal-group item dimension (native and second-language speakers on a language test).

"Variance explained" depends on the spread of the item and person measures. Please see <http://www.rasch.org/rmt/rmt221j.htm> - For dimensionality analysis, we are concerned about the "Variance explained by the first contrast in the residuals". If this is big, then there is a second dimension at work. Infit and Outfit statistics are too local (one item or one person at a time) to detect multidimensionality productively. They are too much influenced by accidents in the data (e.g., guessing, [response sets](#)), and generally do not detect the more subtle, but pervasive, impact of a second dimension (unless it is huge).

Question: "I can not understand the residual contrast analysis you explained. For example, in Winsteps, it gave me the five contrasts' eigenvalues: 3.1, 2.4, 1.9, 1.6, 1.4. (I have 26 items in this data). The result is the same as when I put the residuals into SPSS."

Reply: Unidimensionality is never perfect. It is always approximate. The Rasch model constructs from the data parameter estimates along the unidimensional latent variable that best concurs with the data. But, though the Rasch measures are always unidimensional and additive, their concurrence with the data is never perfect. Imperfection results from multidimensionality in the data and other causes of misfit.

Multidimensionality always exists to a lesser or greater extent. The vital question is: "Is the multi-dimensionality in the data big enough to merit dividing the items into separate tests, or constructing new tests, one for each dimension?"

The unexplained variance in a data set is the variance of the residuals. Each item is modeled to contribute 1 unit of information (= 1 eigenvalue) to the principal components decomposition of residuals. So the eigenvalue of the total

unexplained variance is the number of items (less any items with extreme scores). So when a component (contrast) in the decomposition is of size 3.1, it has the information (residual variance) of about 3 items.

In your example, the first contrast has eigenvalue of 3.1. Its expected value is near 2.0 - www.rasch.org/rmt/rmt191h.htm. This means that the contrast between the strongly positively loading items and the strongly negatively loading items on the first contrast in the residuals has the strength of about 3 items. Since positive and negative loading is arbitrary, you must look at the items at the top and the bottom of the contrast plot. Are those items substantively different? Are they so different that they merit the construction of two separate tests?

It may be that two or three off-dimension items have been included in your 26 item instrument and should be dropped. But this is unusual for a carefully developed instrument. It is more likely that you have a "fuzzy" or "broad" dimension, like mathematics. Mathematics includes arithmetic, algebra, geometry and word problems. Sometimes we want a "geometry test". But, for most purposes, we want a "math test".

If in doubt,

1. Split your 26 items into clusters (subtests), based on positive and negative loadings on the first residual contrast. Winsteps does this in [Table 23.1](#) with "cluster" numbers.
2. Measure everyone on each of the clusters. [Table 23.6](#)
3. What is the correlation of the person measures for pairs of clusters? [Table 23.1](#)
4. Do the clusters display two versions of the same story about the persons, or are they different stories?
"The correlation coefficient corrected for attenuation between two tests x and y is the correlation between their true scores [or true measures]. If, on the basis of a sample of examinees, the corrected coefficient is near unity, the experimenter concludes that the two tests are measuring the same trait." (p. 117) in Joreskog, K.G. (1971) Statistical analysis of sets of congeneric tests, *Psychometrika* 36, 109-133.
5. Copy the person measures from [Table 23.6](#) into Excel, and cross-plot the numbers. Which people are off-diagonal? Is that important? If only a few people are noticeably off-diagonal, or the off-diagonal deviance would not lead to any action, then you have a substantively unidimensional test. You may have a "Fahrenheit-Celsius" [equating](#) situation if the best fit line on the plot departs from a unit slope.

You can do a similar investigation for the second contrast of size 2.4, and third of size 1.9, but each time the motivation for doing more than dropping an off-dimension item or two becomes weaker. Since random data can have eigenvalues of size 1.4, there is little motivation to look at your 5th contrast.

Question: Why are my adaptive-test observations always reported to be unidimensional?

Reply: Unobserved data are modeled at their Rasch-predicted values. In an adaptive test these overwhelm the observed data, so the data are reported as unidimensional using standard criteria.

Solution: use the Winsteps "[Simulate data](#)" function (Complete Data - No) to obtain a baseline for the unidimensional eigenvalues for your data. These will be much lower than the standard criterion of 2.

"Variance explained" is a newly developing area in Rasch methodology. We learn something new about it every month or so. Perhaps you will contribute to this. So there are no rules, only tentative guidelines based on the current state of theory and empirical experience.

1. Originally Winsteps implemented 3 algorithms for computing variance-explained. Most people used the default algorithm (based on standardized residuals). User experience indicates that one of the other two algorithms was much more accurate in apportioning explained and unexplained variance. So, in the current version of Winsteps, this other algorithm (based on raw residuals) had become the algorithm for this part of the computation. The three algorithms are still implemented for the decomposition of the unexplained variance into contrasts (raw residuals, standardized residuals and logit residuals), and the default remains the standardized residuals for this part of the computation.

2. www.rasch.org/rmt/rmt221j.htm shows the expected decomposition of raw variance into explained variance and unexplained variance under different conditions.

Since the rules are only guidelines, please always verify their applicability in your particular situation. A meaningful way of doing this is to compute the person measures for each of what might be the biggest two dimensions in the data, and then to cross-plot those measures. Are the differences between the measures big enough, and pervasive enough, to be

classified as "two dimensions" (and perhaps reported separately) or are they merely a minor perturbation in the data. For instance, in arithmetic, word-problems, abstract-problems and concrete-problems have different response profiles (and so noticeable contrasts), but they are rarely treated as different "dimensions".

19.34 Disjoint strings of responses

When the responses are not arranged in one continuous string in the record, instruct Winsteps to skip over or ignore the gaps.

Example: The 18 item string is in columns 40 to 49 and then 53 to 60 of your data file. The person-id is in columns 11-30. Data look like:

```
xxxxxxxxPocahontas Smith, Jrxxxxxxxx1001001110xxx11001110
```

Method a: Delete unwanted "items" in columns 50, 51, 52 using an item delete file, IDFILE=.

```
NAME1 =11 in original record
NAMLEN=20 length in original record
ITEM1 =40 in original record
NI =21 include deleted items
IDFILE =DEL5052 file of deletions
The contents of DEL5052 are:
11 - 13 Cols 50-52 are items 11-13
```

Method b: Rescore "items" in columns 50, 51, 52 as missing values with RESCORE=.

```
NAME1 =11 in original record
NAMLEN=20
ITEM1 =40
NI =21 include rescored items
RESCORE=000000000011100000000 rescore 50-52
CODES =01 (the standard)
NEWSCORE=XX non-numerical characters specify "missing"
```

Method c: Make the items form one continuous string in a new record created with FORMAT=. Then the item string starts in the 21st column of the new record. Reformatted record looks like: Pocahontas Smith, Jr100100111011001110

```
FORMAT=(T11,20A,T40,10A,T53,8A) reformatting
NAME1 =1 in the formatted record
ITEM1 =21 in the formatted record
NI =18 the actual number of items
```

19.35 Displacement measures

DISPLACE column should only appear with anchored or **TARGET=** runs. Otherwise its appearance indicates lack of convergence. If small displacements are being shown, try tightening the convergence criteria, [LCONV=](#).

Anchored analyses, **I**AFILE=, **P**AFILE=: if large displacements are shown for the anchored items or persons, try changing the setting of [ANCESTIM=](#).

The displacement is an estimate of the amount to add to the MEASURE to make it conform with the data.

Positive displacement for a person ability indicates that the observed person score is higher than the expected person score based on the reported measure (usually an anchor value).

Positive displacement for an item difficulty indicates that the observed item score is lower than the expected item score based on the reported measure (usually an anchor value).

The DISPLACE value is the size of the change in the parameter estimate that would be observed in the next estimation iteration if this parameter was free (unanchored) and all other parameter estimates were anchored at their current values.

For a parameter (item or person) that is anchored in the main estimation, DISPLACE indicates the size of disagreement between an estimate based on the current data and the anchor value.

For an unanchored item, if the DISPLACE value is large enough to be of concern, then the convergence criteria are not tight enough LCONV=, RCONV=, CONVERGE=, MJMLE=

It is calculated using Newton-Raphson estimation.

Person: DISPLACE logits = (observed marginal score - expected marginal score)/(model variance of the marginal score)

Item: DISPLACE logits = - (observed marginal score - expected marginal score)/(model variance of the marginal score)

DISPLACE approximates the displacement of the estimate away from the statistically better value which would result from the best fit of your data to the model. Each DISPLACE value is computed as though all other parameter estimates are exact. Only meaningfully large values are displayed. They indicate lack of convergence, or the presence of anchored or targeted values. The best fit value can be approximated by adding the displacement to the reported measure or calibration. It is computed as:

$$\text{DISPLACE} = (\text{observed score} - \text{expected score based on reported measure}) / (\text{Rasch-model-derived score variance}).$$

The "observed score" is the raw score for the person or item.

The "expected score" is the raw score that the Rasch model expects based on the current values of person abilities and item difficulties.

The "Rasch-model-derived score variance" is the inverse of the standard error of the person or item, squared.

This value is the Newton-Raphson adjustment to the reported measure to obtain the measure estimated from the current data. In *BTD*, p. 64, equation 3.7.11: $di(j)$ is the anchor value, $di(j+1)$ is the value estimated from the current data, and $di(j+1) - di(j)$ is the displacement, given by the right-hand term of the estimation equation, also in step 6 of www.rasch.org/rmt/rmt102t.htm. In *RSA*, p. 77, equation 4.4.6, $di(t)$ is the anchor value, $di(t+1)$ is the value estimated from the current data, and $di(t+1) - di(t)$ is the displacement, given by the right-hand term of the estimation equation, also in step 6 of www.rasch.org/rmt/rmt122q.htm

Note: Instead of Newton-Raphson, Winsteps uses a logistic-curve-fitting approach to estimation. This has proved more robust than Newton-Raphson for badly-behaved data (long rating scales with missing categories, sparse datasets, etc.). Essentially, we know that all the underlying functions in Rasch estimation are logistic ogives, so we can take advantage of this to predict the next (improved) estimate based on the current estimate, its expected score, and the observed score. For computing displacements, logistic-curve-fitting can be replicated in Excel. You have the expected scores for the persons on the anchored item from the XFILE=. Then compute the expected scores for the person on the anchored item difficulty + 1 logit using the person and item measures in the XFILE=. This gives two total expected item scores. For each total expected item score and the total observed item score, compute the log odds: $\ln(\text{value}/(\text{max possible item score} - \text{value}))$. This gives 3 log odds: E0, E1 and O. Then displacement is $(O - E0)/(E1 - E0)$ logits. This value of the displacement is usually very close to the Newton-Raphson value.

Standard Error of the Displacement Measure

ENTRY	RAW		MODEL	INFIT	OUTFIT	PTMEA					
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	DISPLACE	TAP
3	35	35	2.00A	.74	.69	-.6	.22	.5	.00	-3.90	1-2-4

Since the reported "measure" is treated as a constant when "displacement" is computed, the S.E. of the reported "measure" actually is the same as the S.E. of the displacement. The DISPLACE column shows the displacement in the same units as the MEASURE. This is logits when USCALE=1, the default. If the anchored measure value is considered to be exact, i.e., a point-estimate, then the S.E. standard error column indicates the standard error of the displacement. The statistical significance of the Displacement is given by

$$t = \text{DISPLACE} / \text{S.E. with approximately COUNT degrees of freedom.}$$

This evaluates how likely the reported size of the displacement is, if its "true" size is zero. But both the displacements and their standard errors are estimates, so the t-value may be slightly mis-estimated. Consequently allow for a margin of error when interpreting the t-values.

If the anchored measure value has a standard error obtained from a different data set, then the standard error of the displacement is:

$$\text{S.E. (Displacement)} = \text{Sqrt}(\text{S.E.}^2 + \text{S.E.}^2(\text{anchor value from original data}))$$

When does large displacement indicate that an item or person should be unanchored or omitted?

This depends on your purpose. If you are anchoring items in order to measure three additional people to add to your measured database of thousands, then item displacement doesn't matter.

Anchor values should be validated before they are used. Do two analyses:

- (a) with no items anchored (i.e., all items floating), produce person and item measures.
- (b) with anchored items anchored, produce person and item measures.

Then cross-plot the item difficulties for the two runs, and also the person measures. The person measures will usually form an almost straight line.

For the item difficulties, unanchored items will form a straight-line. Some anchored items may be noticeably off the line. These are candidates for dropping as anchors. The effect of dropping or un-anchoring a "displaced" anchor item is to realign the person measures by roughly (displacement / (number of remaining anchored items)).

Random displacements of less than 0.5 logits are unlikely to have much impact in a test instrument.

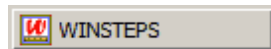
"In other work we have found that when [test length] is greater than 20, random values of [discrepancies in item calibration] as high as 0.50 [logits] have negligible effects on measurement." (Wright & Douglas, 1976, "Rasch Item Analysis by Hand")

"They allow the test designer to incur item discrepancies, that is item calibration errors, as large as 1.0 [logit]. This may appear unnecessarily generous, since it permits use of an item of difficulty 2.0, say, when the design calls for 1.0, but it is offered as an upper limit because we found a large area of the test design domain to be exceptionally robust with respect to independent item discrepancies." (Wright & Douglas, 1975, "Best Test Design and Self-Tailored Testing.")

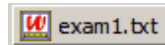
Most DIF work seems to be done by statisticians with little interest in, and often no access to, the substantive material. So they have no qualitative criteria on which to base their DIF acceptance/rejection decisions. The result is that the number of items with DIF is grossly over-reported (Hills J.R. (1989) Screening for potentially biased items in testing programs. Educational Measurement: Issues and practice. 8(4) pp. 5-11).

19.36 Edit taskbar caption

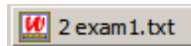
The Windows taskbar is usually at the bottom of your screen. It displays the program icons with short captions. When Winsteps starts it displays:



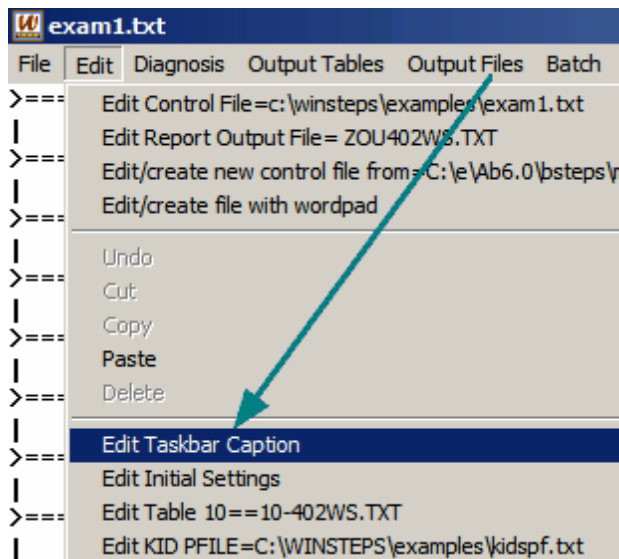
After a control file has been selected, e.g., exam1.txt, the control file name is displayed on the task bar:



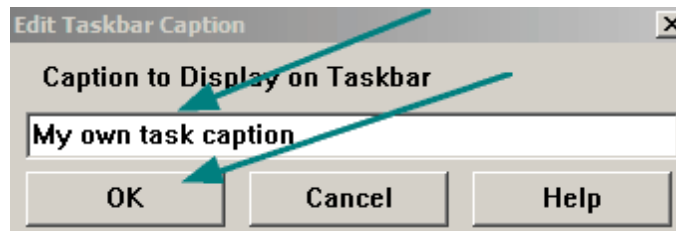
If two runs use the same control file name at the same time, then the runs are numbered uniquely:



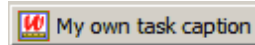
The automatic caption can be replaced by one chosen by you. On the [Edit menu](#), select Edit Taskbar Caption



The following dialog displays:



You can type in your own caption, and then click OK. Your caption displays on the bottom of your screen. This is useful for reminding yourself about a particular analysis you have running.



19.37 Equating and linking tests

Test Equating and linking are usually straightforward with *Winsteps*, but do require clerical care. The more thought is put into test construction and data collection, the easier the equating will be.

Imagine that Test A (the more definitive test, if there is one) has been given to one sample of persons, and Test B to another. It is now desired to put all the items together into one item hierarchy, and to produce one set of measures encompassing all the persons.

Initially, **analyze each test separately**. Go down the "[Diagnosis](#)" pull-down menu. If the tests don't make sense separately, they won't make sense together.

There are several equating methods which work well with *Winsteps*. Test equating is discussed in Bond & Fox "Applying the Rasch model", and earlier in Wright & Stone, "Best Test Design", George Ingebo "Probability in the Measure of Achievement" - all available from www.rasch.org/books.htm

We want to put Test B onto Test A's scale. This is the same as putting Fahrenheit (Test B) temperatures onto a Celsius (Test A) scale:

New $USCALE =$ for Test B = (Old $USCALE =$ for Test B)*(S.D. of relevant persons or items for Test A)/(S.D. of relevant persons or items for Test B)

New $U_{I\text{MEAN}}$ or $U_{P\text{MEAN}}$ for Test B = (Mean of relevant persons or items for Test A) - ((Mean of relevant persons or items for Test B - Old $U_{I\text{MEAN}}$ or $U_{P\text{MEAN}}$ for Test B)*(S.D. of relevant persons or items for Test A)/(S.D. of relevant persons or items for Test B))

Rescaled Test B measure = ((Old Test B measure - Old $U_{I\text{MEAN}}$ or $U_{P\text{MEAN}}$ for Test B)/(Old USCALE for Test B))*(New USCALE for Test B) + (New $U_{I\text{MEAN}}$ or $U_{P\text{MEAN}}$ for Test B)

Example: suppose we have freezing and boiling points in modified Fahrenheit and modified Celsius.

Test A: Celsius*200 + 50: USCALE=200, $U_{I\text{MEAN}}$ =50, freezing: 50, boiling: 20050, range = 20050-50 = 20000, mean = 10050

Test B: Fahrenheit*10+1000: USCALE=10, $U_{I\text{MEAN}}$ =1000, freezing: 1320, boiling: 3120, range = 3120-1320=1800, mean = 2220

For convenience, we will substitute the observed range for the S.D. in the equations above:

New Test B modified Fahrenheit USCALE = $10 * 20000 / 1800 = 111.1111$

Then 1320, 3120 becomes $(1320/10) * 111.111 = 34666.7$, $(3120/10) * 111.111 = 14666.7$, so that the rescaled Test B range becomes $34666.7 - 14666.7 = 20000 =$ the Test A range

New Test B Fahrenheit $U_{I\text{MEAN}}$ = $10050 - (2220 - 1000) * 20000 / 1800 = -3505.56$

Thus, rescaled Test B freezing: 1320 becomes $((1320 - 1000) / 10) * 111.1111 - 3505.56 = 50 =$ Test A freezing
rescaled Test B boiling: 3120 becomes $((3120 - 1000) / 10) * 111.1111 - 3505.56 = 20050 =$ Test A boiling

Concurrent or One-step Equating

All the data are entered into one big array. This is convenient but has its hazards. Off-target items can introduce noise and skew the equating process, C_{UTLO} and C_{UTHI} may remedy targeting deficiencies. Linking designs forming long chains require much tighter than usual convergence criteria. Always cross-check results with those obtained by one of the other equating methods.

Concurrent equating of two tests, A and B, is usually easier than equating separate analyses. M_{FORMS} can help set up the data. But always analyze test A and test B separately first. Verify that test A and test B are functioning correctly. Then scatterplot the item difficulties for the common items from the separate analysis of Test A against the item difficulties from a separate analysis of Test B. Verify that they are on an approximately straight line approximately parallel to the identity line. Then do the concurrent analysis. You can also do a [DIF analysis](#) on the common items.

Common Item Equating

This is the best and easiest equating method. The two tests share items in common, preferably at least 5 spread out across the difficulty continuum. There is no specific minimum number of common items. We usually want as many common items as possible, but are constrained by test design considerations. For instance, <https://www.rasch.org/rmt/rmt51h.htm> recommends 20 common items in each test. Great for tests of 200+ items, but what if the test only has 20 items in total? How low can we go? Only one common item is obviously too fragile. We can probably get away with 3 common items, spread across the ability range, but what if one or two of those items malfunction? So that leads us to 5 common items as a practical minimum. In a 20-item test (common in classroom tests), 5 items is one-quarter of the test. We would probably not want to go higher than one-quarter in order to limit item exposure and increase test security.

In the Winsteps analysis, indicate the common items with a special code in column 1 of the item labels.

Example: there are 6 items. Items 1 and 2 are common to the two tests. Items 3 and 4 are only in Test A. Items 5 and 6 are only in Test B

....
&END
C Item 1
C Item 2
A Item 3
A Item 4
B Item 5

"Partial credit" values are much less stable than dichotomies. Rather than trying to [equate](#) across the whole partial credit structure, one usually needs to assert that, for each item, a particular "threshold" or "step" is the critical one for equating purposes. Then use the difficulties of those thresholds for equating. This relevant threshold for an item is usually the transition point between the two most frequently observed categories - the Rasch-Andrich threshold - and so the most stable point in the partial credit structure.

Stocking and Lord iterative procedure

The Stocking and Lord (1983) present an iterative common-item procedure in which items exhibiting DIF across tests are dropped from the link until no items exhibiting inter-test DIF remain. A known hazard is that if the DIF distribution is skewed, the procedure trims the longer tail and the equating will be biased. To implement the Stocking and Lord procedure in Winsteps, code each person (in the person id label) according to which test form was taken. Then request a DIF analysis of item x person-test-code ([Table 30](#)). Drop items exhibiting DIF from the link, by coding them as different items in different tests.

Stocking and Lord (1983) Developing a common metric in item response theory. Applied Psychological Measurement 7:201-210.

Fuzzy Common-Item Equating

Two instruments measure the same trait, but with no items or persons in common.

1. Identify roughly similar pairs of items on the two instruments, and cross-plot their measures. We expect the plot to be fuzzy but it should indicate an equating line.
2. Virtual equating or pseudo-common-item equating (see below): Print the two item hierarchy maps and slide them up-and-down relative to each other until the overall item hierarchy makes the most sense. The relative placement of the local origins (zero points) of the two maps is the equating constant.

Choosing Common Items

1. Cross-plot the item difficulties of the pairs of possible common items from your original analyses.
2. In the scatterplot, there should be a diagonal line of items parallel to the identity line. These will be the best common items.

Common-item equating. Using Excel to construct a bank of item numbers for use with MFORMS=. Two test forms, 97Up and 97Down have some items in common. The item labels are in two Winsteps data files.

1. Copy the item labels from the Winsteps data files into this worksheet: Columns B and F
2. Put in sequence numbers in column A using =(cell above+1)
3. Copy Column A and paste special (values) into Columns C and G
4. In Column D, VLOOKUP Column F in Columns B, C. If #N/A then make it 9999
`=IF(ISNA(VLOOKUP(F17,B17:C1747,2,FALSE))),9999,VLOOKUP(F17,B17:C1747,2,FALSE))`
5. Copy column D and paste special (values) into Column E
6. Copy columns E F G into I J K
7. Sort IJK on columns I and K
8. Copy K into L
9. Go down column I to first 9999 (the first item in 97Down not in 97Up)
10. Place the first number after the last number in column A into column I
11. Sequence up from that number to the end of the items
12. The MFORMS= item-bank numbers for the first test are in column 1. They are the same as the original entry numbers.
13. The MFORMS= entry numbers for the second test are in column K. The bank numbers are in column I

	A	B	C	D	E	F	G	H	I	J	K
16	Seque	97UP	Entry/Banked	Combined	Combined value	97Down	Entry		Banked	97Down	Entry
17	1	497850414F	1	1670	1670	497600156F	1		1	4978504	776
18	2	497850438F	2	938	938	497232351M	2		3	4978502	789
19	3	497850294F	3	9999	9999	497750169M	3		4	4978503	792
20	4	497850309F	4	1320	1320	497712090M	4		5	4978501	778
21	5	497850127F	5	1612	1612	497220413M	5		6	4978502	787
22	6	497850282M	6	1371	1371	497750066M	6		7	4978502	793

16	Sequence	97UP	Entr	Combined	Comt	97Down	Entr	Bank	97Down	Entry
17	1	497850414F	1	=IF(ISNA(VLOOKUP(F17,\$B\$17:\$C\$1747,2,FALSE)),9999,VLOOKUP(F17,\$B\$17:\$C\$1747,2,FALSE))	1670	497600156F	1	1	497850414F	776
18	=(A17+1)	497850438F	2	=IF(ISNA(VLOOKUP(F18,\$B\$17:\$C\$1747,2,FALSE)),9999,VLOOKUP(F18,\$B\$17:\$C\$1747,2,FALSE))	938	497232351M	2	3	497850294F	789
19	=(A18+1)	497850294F	3	=IF(ISNA(VLOOKUP(F19,\$B\$17:\$C\$1747,2,FALSE)),9999,VLOOKUP(F19,\$B\$17:\$C\$1747,2,FALSE))	9999	497750169M	3	4	497850309F	792
20	=(A19+1)	497850309F	4	=IF(ISNA(VLOOKUP(F20,\$B\$17:\$C\$1747,2,FALSE)),9999,VLOOKUP(F20,\$B\$17:\$C\$1747,2,FALSE))	1320	497712090M	4	5	497850127F	778
21	=(A20+1)	497850127F	5	=IF(ISNA(VLOOKUP(F21,\$B\$17:\$C\$1747,2,FALSE)),9999,VLOOKUP(F21,\$B\$17:\$C\$1747,2,FALSE))	1612	497220413M	5	6	497850282M	787

Common Person Equating

Some persons have taken both tests, preferably at least 5 spread out across the ability continuum.

Step 1. From the separate analyses, [crossplot](#) the abilities of the common persons, with Test B on the y-axis and Test A on the x-axis. The slope of the best-fit line i.e., the line through the point at the means of the common persons and through the (mean + 1 S.D.) point should have slope near 1.0. If it does, then the intercept of the line with the x-axis is the equating constant.

First approximation: Test B measures in the Test A frame of reference = Test B measure + x-axis intercept.

Step 2. Examine the scatterplot. Points far away from the joint-best-fit trend-line indicate persons that have behaved differently on the two occasions. You may wish to consider these to be no longer common persons. Drop the persons from the plot and redraw the joint-best-fit trend-line.

Step 3a. If the best-fit slope remains far from 1.0, then there is something systematically different about Test A and Test B. You must do "Celsius - Fahrenheit" equating. Test A remains as it stands.

The slope of the best fit is: slope = (S.D. of Test B common persons) / (S.D. of Test A common persons)

Include in the Test B control file:

USCALE = the value of 1/slope

UMEAN = the value of the x-intercept

and reanalyze Test B. Test B is now in the Test A frame of reference, and the person measures from Test A and Test B can be reported together.

Step 3b. The best-fit slope is near to 1.0. Suppose that Test A is the "benchmark" test. Then we do not want responses to Test B to change the results of Test A.

From a Test A analysis produce PFILE=

Edit the PFILE= to match Test B person numbers

Use it as a PAFILE= in a Test B analysis.

Test B is now in the same frame of reference as Test A, so the person measures and person difficulties can be reported together

Step 3c. The best-fit slope is near to 1.0. Test A and Test B have equal priority, and you want to use both to define the common persons.

Use your text editor or word processor to append the common persons' Test B responses after their Test A ones, as in the design below. Then put the rest of the Test B responses after the Test A responses, but aligned in columns with the common persons's Test B responses. Perform an analysis of the combined data set. The results of that analysis will have Test A and Test B persons and persons reported together.

```

      Test A items                Test B items
|-----|
|-----|
|-----|
|-----|-----| Common Person
|-----|
|-----|-----| Common Person
|-----|
|-----|
|-----|-----| Common Person
|-----|
|-----|
|-----|-----| Common Person
|-----|
|-----|

```


Common-person equating. Using Excel to construct a data file for use with [Excel input](#). Two test forms, 97Up and 97Down have some person in common. The person labels are data are in Excel worksheets.

In a new worksheet

- Copy test 1 responses and person labels to, say, columns B to AZ. The person labels are in AZ
- Copy test 2 responses and person labels to, say, columns BC to DA. The person labels are in DA
- Put in sequence numbers in column A using =(cell above+1)
- Copy "values" column A to columns BA, DB
- In column BB, VLOOKUP column AZ in column DA. If found BA, If not, 9999
=IF(ISNA(VLOOKUP(AZ15,\$DA\$15:\$DB\$1736,2,FALSE)),9999,BA15)
- In column DC, VLOOKUP column DA in column AZ. If found the look-up value, If not, 9999
=IF(ISNA(VLOOKUP(DA15,\$AZ\$15:\$BA\$1736,2,FALSE)),9999,VLOOKUP(DA15,\$AZ\$15:\$BA\$1736,2,FALSE))
- Sort B-BB on column BB, BA
- Sort BC-DC on column DC, DB
- Copy and paste the 9999 rows for Test 2 below the last row for Test 1
- Use Winsteps Excel input to create a Winsteps control and data file for the combined data

Steps 1-6.

	A	B	AY	AZ	BA	BB	BC	CZ	DA	DB	DC
14	Sequence			97Up	Entry	Match			97Down	Entry	Match
15		1	1	4 497850414F	1	1	1	4	497600156F	1	1670
16		2	1	2 497850438F	2	9999	1	4	497232351M	2	938
17		3	1	1 497850294F	3	3	3	4	497750169M	3	9999
18		4	1	2 497850309F	4	4	3	3	497712090M	4	1320

Steps 7-10.

14	Sequence			97Up	Entry	Match			97Down	Entry	Match
15		1	1	4 497850414F	1	1	1	4	497850414F	776	1
16		2	1	1 497850294F	3	3	1	4	497850294F	789	3
17		3	1	2 497850309F	4	4	1	4	497850309F	792	4

Virtual Equating of Test Forms

The two tests share no items or persons in common, but the items cover similar material.

Step 1. Identify pairs of items of similar content and difficulty in the two tests. Be generous about interpreting "similar" at this stage. These are the pseudo-common items.

Steps 2-4: simple: The two item hierarchies (Table 1 using short clear item labels) are printed and compared, equivalent items are identified. The sheets of paper are moved vertically relative to each other until the overall hierarchy makes the

most sense. The value on Test A corresponding to the zero on Test B is the UMEAN= value to use for Test B. If the item spacing on one test appear expanded or compressed relative to the other test, use USCALE= to compensate.

Or:

Step 2. From the separate analyses, [crossplot](#) the difficulties of the pairs of items, with Test B on the y-axis and Test A on the x-axis. The slope of the best-fit line i.e., the line through the point at the means of the common items and through the (mean + 1 S.D.) point should have slope near 1.0. If it does, then the intercept of the line with the x-axis is the equating constant.

First approximation: Test B measures in the Test A frame of reference = Test B measure + x-axis intercept.

Step 3. Examine the scatterplot. Points far away from the joint-best-fit trend-line indicate items that are not good pairs. You may wish to consider these to be no longer paired. Drop the items from the plot and redraw the trend line.

Step 4. The slope of the best fit is: $\text{slope} = (\text{S.D. of Test B common items}) / (\text{S.D. of Test A common items})$
= Standard deviation of the item difficulties of the common items on Test B divided by Standard deviation of the item difficulties of the common items on Test A

Include in the Test B control file:

USCALE = the value of 1/slope

UMEAN = the value of the x-intercept

and reanalyze Test B. Test B is now in the Test A frame of reference, and the person and item measures from Test A and Test B can be reported together.

Random Equivalence Equating

The samples of persons who took both tests are believed to be randomly equivalent. Or, less commonly, the samples of items in the tests are believed to be randomly equivalent.

Step 1. From the separate analyses of Test A and Test B, obtain the means and sample [standard deviation](#) of the Rasch person-ability measures for two person samples (including extreme scores).

Step 2. To bring Test B into the frame of reference of Test A, adjust by the difference between the means of the Rasch person-ability measures for the person samples and user-rescale by the ratio of their sample standard deviations.

Include in the Test B control file:

USCALE = value of (sample S.D. person sample for Test A) / (sample S.D. person sample for Test B)

UMEAN = value of (mean for Test A) - (mean for Test B * USCALE)

and reanalyze Test B.

Check: Test B should now report the same sample mean and sample standard deviation as Test A for the Rasch person-ability measures.

Test B is now in the Test A frame of reference, and the person measures from Test A and Test B can be reported together.

Paired Item Equating - Parallel Items - Pseudo-Common Items

When constructing tests with overlapping content.

Step 1. From your list of items (or your two tests), choose 20 pairs of items. In your opinion, the difficulties of the two items in each pair should be approximately the same.

Step 2. One item of each pair is in the first test, and the other item is in the second test. Be sure to put codes in the item labels to remind yourself which items are paired.

Step 3. Collect your data.

Step 4. Analyze each test administration separately.

Step 5. Subtotal the difficulties of the paired items for each test administration ([Table 27](#)). The difference between the pair subtotals is the equating constant. Is this a reasonable value? You can use this value in [UJMEAN=](#) of one test to align the two tests.

Step 6. [Cross-plot](#) the difficulties of the pairs of items (with confidence bands computed from the standard errors of the pairs of item estimates). The plot will confirm (hopefully) your opinion about the pairs of item difficulties. If not, you will learn a lot about your items!!

Linking Tests with Common Items

Here is an example:

- A. The first test (50 items, 1,000 students)
- B. The second test (60 items, 1,000 students)
- C. A linking test (20 items from the first test, 25 from the second test, 250 students)

Here is a typical Rasch approach. It is equivalent to applying the "common item" linking method twice.

(a) Rasch analyze each test separately to verify that all is correct.

(b) Cross-plot the item difficulties for the 20 common items between the first test and the linking test. Verify that the link items are on a statistical trend line parallel to the identity line. Omit from the list of linking items, any items that have clearly changed relative difficulty. If the slope of the trend line is not parallel to the identity line (45 degrees), then the test discrimination has changed. The test linking will use a "best fit to trend line" conversion:

Corrected measure on test 2 in test 1 frame-of-reference =

$$\begin{aligned} & ((\text{observed measure on test 2} - \text{mean measure of test 2 link items}) * (\text{SD of test 1 link items}) / (\text{SD of test 2 link items})) \\ & + \text{mean measure of test 1 link items} \end{aligned}$$

(c) Cross-plot the item difficulties for the 25 common items between the second test and the linking test. Repeat (b).

(d1) If both trend lines are approximately parallel to the identity line, then all three tests are equally discriminating, and the simplest equating is "concurrent". Put all 3 tests in one analysis. You can use the [MFORMS=](#) command to put all items into one analysis. You can also selectively delete items using the [Specification](#) pull-down menu in order to construct measure-to-raw score conversion tables for each test, if necessary.

Or you can use a direct arithmetical adjustment to the measures based on the mean differences of the common items: www.rasch.org/memo42.htm "Linking tests".

(d2) If best-fit trend lines are not parallel to the identity line, then tests have different discriminations. Equate the first test to the linking test, and then the linking test to the second test, using the "best fit to trend line" conversion, shown in (b) above. You can also apply the "best fit to trend" conversion to [Table 20](#) to convert every possible raw score.

19.38 Estimation bias correction - warnings

Winsteps uses JMLE (= Joint Maximum-Likelihood Estimation), implemented with iterative-curve fitting, rather than Newton-Raphson estimation, because iterative curve-fitting is more robust against awkward data patterns.

Every estimation method has strengths and weaknesses. The primary weakness of JMLE is that estimates have statistical bias. This is most obvious in a test of two dichotomous items (Andersen, 1973). In such a test, the difference between the item difficulties of the two items will be estimated to be twice its true value. In practical situations, the statistical bias is usually less than the standard errors of the estimates. However, the advantages of JMLE far outweigh its disadvantages. JMLE is estimable under almost all conditions including arbitrary and accidental patterns of missing data, arbitrary anchoring (fixing) of parameter estimates, unobserved intermediate categories in rating scales, and multiple different Rasch models in the same analysis.

Andersen, E. B. Conditional inference for multiple-choice questionnaires. *British Journal of Mathematical and Statistical Psychology*, 1973, 26, 31-44.

In the psychometric literature, the terms "bias" and "inconsistency" are usually used in the context of the estimation of the difficulties of dichotomous items in a fixed length test administered to a sample of persons. Each measure-parameter is imagined to have a true value, and it is the purpose of the estimation procedure is to estimate that value from the available data. We can never be sure that we have exactly estimated the true value.

If the sample size is infinite, and the resulting item estimates are their true values, then the estimates are consistent. If the sample size is finite, and the expectations of the possible item estimates are their true values, then the estimates are unbiased.

Winsteps implements JMLE. JMLE estimates are inconsistent and biased. They are less central than the true values. For instance, if the test consists of two dichotomous items, then, with an infinite sample, the JMLE estimate of the difference between the two item difficulties will be twice the true value. In this situation, an immediate solution is [PAIRED=Yes](#).

Ben Wright and Graham Douglas discovered that the multiplier $(L-1)/L$ is an approximate correction for JMLE item bias, where L is the test length. For a two-item test this correction would be $(2-1)/2 = 0.5$. It is implemented in Winsteps with [STBIAS=YES](#).

Winsteps uses the raw scores as sufficient statistics for its estimates. The parameter estimates reported by Winsteps are the values for which "the observed raw score = the model-expected raw score".

"Statistical Consistency" (as usually conceptualized in the psychometric literature) relates to an infinite sample size with a finite test length. Under these conditions, Winsteps estimates are statistically inconsistent (i.e., are not the "true" parameter values even with an infinite amount of data) because inestimable extreme scores are included in the estimation space. "Conditional" estimation methods, [CMLE](#), remove ("condition out") extreme scores from the estimation space,

The practical concern is "estimation bias", "departure of estimates from their true values with a finite amount of data". Winsteps estimates do have estimation bias. The Winsteps estimates are less central than they should be. But, as the likelihood of observing extreme scores reduces, the bias in the Winsteps estimates also reduces. Published studies indicate that when the test length is longer than 20 dichotomous items and the sample size is greater than 20 cases, then the Winsteps estimation bias is inconsequentially small. Estimation bias is usually only of concern if exact probabilistic inferences are to be made from logit measures obtained from small samples or short tests. But such inferences are imprecise irrespective of the size of the estimation bias.

There are techniques which correct for estimation bias under specific conditions. One such condition is when the data correspond to pairwise observations (such as a basketball league or chess competition). Winsteps has the [PAIRED=YES](#) option for this situation.

At least two sources of estimation error are reported in the literature.

An "estimation bias" error. This is usually negligibly small after the administration of 10 dichotomous items (and fewer rating scale items). Its size depends on the probability of observing extreme score vectors. For a two item test, the item measure differences are twice their theoretical values, reducing as test length increases. This can be corrected. [STBIAS=](#) does this approximately, but is only required if exact probability inferences are to be made from logit measure differences.

A "statistical inflation" error. Since error variance always adds to observed variance, individual measures are always reported to be further apart (on average) than they really are. This cannot be corrected, in general, at an individual-measure level, because, for any particular measurement it cannot be known to what extent that measurement is biased by measurement error. However, if it is hypothesized that the persons, for instance, follow a normal distribution of known mean and standard deviation, this can be imposed on the estimates (as in [MMLE](#)) and the global effects of the estimate dispersion inflation removed. This is done in some other Rasch estimation software.

Estimation Bias

All Rasch estimation methods have some amount of estimation bias (which has no relationship with demographic bias). The estimation algorithm used by Winsteps, [JMLE](#), has a slight bias in measures estimated from most datasets. The effect of the bias is to spread out the measures more widely than the data indicate. In practice, a test of more than 20 dichotomous items administered to a reasonably large sample will produce measures with inconsequential estimation bias. Estimation bias is only of concern when exact probabilistic inferences are to be made from short tests or small samples. Ben Wright opted for JMLE in the late 1960's because users were rarely concerned about such exact inferences, but they were concerned to obtain speedy, robust, verifiable results from messy data sets with unknown latent parameter

distributions. Both of the identifiable sources of error are reduced by giving longer tests to bigger samples. With short tests, or small samples, other threats to validity tend to be of greater concern than the inflationary ones.

If estimation bias would be observed even with an infinitely large sample (which it would be with JMLE), then the estimation method is labeled "statistically inconsistent" (even though the estimates are predictable and logical). This sounds alarming but the inconsistency is usually inconsequential, or can be easily corrected in the unlikely event that it does have substantive consequences.

The JMLE joint likelihood estimation algorithm produces estimates that have a usually small statistical bias. This bias increases the spread of measures and calibrations, but usually less than the [standard error](#) of measurement. The bias quickly becomes insignificantly small as the number of persons and items increases. The reason that JMLE is statistically inconsistent under some conditions, and noticeably biased for short tests or small samples, is that it includes the possibility of extreme scores in the estimation space, but cannot actually estimate them. Inconsistency doesn't really matter, because it asks "if we have infinite data, will the estimation method produce the correct answer?" Estimation bias, also called statistical bias, is more important because it asks "How near to correct are the estimates with finite data?" In practice, JMLE bias is smaller than the other sources of noise in the data. See Ben Wright's comments at www.rasch.org/memo45.htm

For paired comparisons and very short tests, estimation can double the apparent spread of the measures, artificially inflating test reliability. This can be eliminated by specifying [PAIRED=YES](#).

Correcting for bias may be helpful when it is desired to draw exact probabilistic inferences for small, complete datasets without anchoring.

Correcting for bias may be misleading, or may be suppressed by Winsteps, in the presence of missing data or anchored persons or items.

Bias correction can produce apparently inconsistent measures if bias-corrected measures, estimated from an unanchored analysis, are then used to anchor that same dataset.

Estimation correction methods:

[STBIAS=YES](#) implements a variant of the simple bias correction proposed in Wright, B.D. and Douglas, G.A. (1977). Best procedures for sample-free item analysis. *Applied Psychological Measurement*, 1, 281-294. With large samples, a useful correction for bias is to multiply the estimated measures by $(L-1)/L$, where L is the smaller of the average person or item response count, so, for paired comparisons, multiply by 0.5. This is done automatically when [PAIRED=YES](#).

Other Rasch programs may or may not attempt to correct for estimation bias. When comparing results from other programs, try both STBIAS=Y and STBIAS=N to find the closest match.

[Estimation methods](#) with less estimation bias under some circumstances include CMLE and MMLE, but these have other limitations or restrictions which are deemed to outweigh their benefits for most uses.

Technical information:

Statistical estimation bias correction with JMLE is relevant when you wish to make exact probabilistic statements about differences between measures for short tests or small samples. The $(L-1)/L$ correction applies to items on short dichotomous tests with large samples, where L is the number of non-extreme items on a test. For long dichotomous tests with small samples, the correction to person measures would be $(N-1)/N$. Consequently Winsteps uses a bias correction on dichotomous tests for items of $(L-1)/L$ and for persons of $(N-1)/N$

The reason for this correction is because the sample space does not match the estimation space. The difference is extreme score vectors. Estimation bias manifests itself as estimated measures which are more dispersed than the unbiased measures. The less likely an extreme score vector, the smaller the correction to eliminate bias. Extreme score vectors are less likely with polytomies than with dichotomies so the bias correction is smaller. For example, if an instrument uses a rating scale with m categories, then Winsteps corrects the item measures by $(m-1)(L-1)/((m-1)(L-1)+1)$ and person measures by $(m-1)(N-1)/((m-1)(N-1)+1)$ - but these are rough approximations.

With most Rasch software using CMLE, PMLE or MMLE bias correction of item measures is not done because the estimation bias in the item difficulties is generally very small. Bias correction of person abilities is not done though estimation bias exists.

Interaction terms are computed in an artificial situation in which the abilities and difficulties estimates are treated as known. Estimation bias is a minor effect in the interaction estimates. It would tend to increase very slightly the probability that differences between interaction estimates are reported as significant. So this is another reason to interpret DIF tests conservatively. If the number of relevant observations for an interaction term is big enough for the DIF effect to be regarded as real, and not a sampling accident, then the estimation bias will be very small. In the worst case, the multiplier would be of the order of $(C-1)/C$ where C is the number of relevant observations.

Comparing Estimates

Bigsteps and Winsteps should produce the same estimates when

(a) they are run with very tight convergence criteria, e.g.,
RCONV=.00001
LCONV=.00001
MJMLE=0

(b) they have the same statistical bias adjustment
STBIAS=YES ; estimates will be wider spread
or
STBIAS=NO ; estimates will be narrower

(c) they have the same extreme score adjustment
EXTRSC=0.5

The item estimates in BTM were produced with statistical bias adjustment, but with convergence criteria that would be considered loose today. Tighter convergence produces a wider logit spread. So the BTM item estimates are slightly more central than Winsteps or Bigsteps.

Winsteps and Bigsteps are designed to be symmetric. Transpose persons and items, and the only change is the sign of the estimates and an adjustment for local origin. The output reported in BTM (and by most modern Rasch programs) is not symmetric. So the person measure estimates in BTM are somewhat different.

Do-it-yourself estimation-bias correction

Correcting for estimation-bias in Winsteps estimates has both advantages and disadvantages. Corrected estimates are usually slightly more central than uncorrected estimates. The only conspicuous advantage of bias correction is for making inferences based on the exact logit distance between the Rasch estimates. Since, with small data sets, the bias correction is usually less than the standard error of the Rasch estimates, bias correction may be of doubtful statistical utility.

STBIAS= will not correct for bias accurately with missing data, IWEIGHT= or PWEIGHT=. It may over- or under- correct for estimation bias.

If you do need estimation-bias correction that is as accurate as possible with your data set, you will need to discover the amount of bias in the estimates, and then use USCALE= to perform your own estimation-bias correction.

Here is a procedure using simulated data sets:

1. In your control file, STBIAS=No and USCALE=1
2. Obtain the Winsteps estimates for your data
3. Simulate many datasets using those estimates. (SIFILE= on the Winsteps Output Files menu).
4. Obtain the Winsteps estimates from the simulated data sets
5. Regress the simulated estimates on your initial estimates. These will give a slope near 1.0.
6. Obtain the Winsteps estimates for your data with USCALE = 1/slope

The set of estimates in 6 is effectively unbiased.

19.39 Estimation methods: JMLE, PROX, WMLE, CMLE

The Joint Maximum Likelihood Estimation (JMLE) equations in Winsteps are similar to www.rasch.org/rmt/rmt122q.htm for dichotomies, and www.rasch.org/rmt/rmt102t.htm for polytomies, enhanced to allow estimation of both person abilities and item difficulties simultaneously.

The dichotomous estimation equations are implemented in the Excel spreadsheet at www.rasch.org/moulton.htm and the polytomous estimation equations are implemented in the Excel spreadsheet at www.rasch.org/poly.xls

Winsteps implements two methods of estimating Rasch parameters from ordered qualitative observations: JMLE and PROX. Estimates of the Rasch measures are obtained by iterating through the data. Initially all unanchored parameter estimates (measures) are set to zero. Then the PROX method is employed to obtain rough estimates. Each iteration through the data improves the PROX estimates until they are usefully good. Then those PROX estimates are the initial estimates for JMLE which [fine-tunes](#) them, again by iterating through the data, in order to obtain the final JMLE estimates. The iterative process ceases when the [convergence criteria](#) are met. These are set by [MJMLE=](#), [CONVERGE=](#), [LCONV=](#) and [RCONV=](#). Depending on the data design, this process can take hundreds of iterations ([Convergence: Statistics or Substance?](#)). When only rough estimates are needed, force convergence by pressing Ctrl+F or by selecting "Finish iterating" on the [File](#) pull-down menu.

Extreme scores: (perfect, maximum possible scores, and zero, minimum possible scores) are dropped from the main estimation procedure. Their measures are estimated separately using [EXTRSC=](#).

Missing data: most Rasch estimation methods do not require that missing data be imputed, or that there be case-wise or list-wise omission of data records with missing data. For datasets that accord with the Rasch model, missing data lower the precision of the measures and lessen the sensitivity of the fit statistics, but do not bias the measure estimates.

Likelihood: Using the current parameter estimates (Rasch measures), the probability of observing each data point is computed, assuming the data fit the model. The probabilities of all the data points are multiplied together to obtain the likelihood of the entire data set. The parameter estimates are then improved (in accordance with the estimation method) and a new likelihood for the data is obtained. The values of the parameters for which the likelihood of the data has its maximum are the "maximum likelihood estimates" (Ronald A. Fisher, 1922).

JMLE "Joint Maximum Likelihood Estimation" is also called UCON, "Unconditional maximum likelihood estimation". It was devised by Wright & Panchapakesan, www.rasch.org/memo46.htm. In this formulation, the estimate of the Rasch parameter (for which the observed data are most likely, assuming those data fit the Rasch model) occurs when the observed raw score for the parameter matches the expected raw score. "Joint" means that the estimates for the persons (rows) and items (columns) and rating scale structures (if any) of the data matrix are obtained simultaneously. The iterative estimation process is described at [Iteration](#).

Advantages - these are implementation dependent, and are implemented in Winsteps:

- (1) independence from specific person and item distributional forms.
- (2) flexibility with missing data
- (3) the ability to analyze test lengths and sample sizes of any size
- (4) symmetrical analysis of person and item parameters so that transposing rows and columns does not change the estimates
- (5) flexibility with person, item and rating scale structure anchor values
- (6) flexibility to include different variants of the Rasch model in the same analysis (dichotomous, rating scale, partial credit, etc.)
- (7) unobserved intermediate categories of rating scales can be maintained in the estimation with exact probabilities.
- (8) all non-extreme score estimable (after elimination of extreme scores and rarely-observed Guttman subsets)
- (9) all persons with the same total raw scores on the same items have the same measures; all items with the same raw scores across the same persons have the same measures.

Disadvantages:

- (11) measures for extreme (zero, perfect) scores for persons or items require post-hoc estimation.

(12) estimates are statistically inconsistent. Infinite data produces usually only slightly statistically-incorrect estimates. This is seen as estimation bias for finite samples..

(13) estimation bias, particularly with small samples or short tests, inflates the logit distance between estimates. The estimation bias is minuscule for large datasets, and almost always less than the standard error of the estimates. The measure-order of the estimates is correct. Estimation bias is easy to correct when required. [STBIAS=](#).

(14) chi-squares reported for fit tests (particularly global fit tests) may be somewhat inflated, exaggerating misfit to the Rasch model to a very small degree.

Comment on (8): An on-going debate is whether measures should be adjusted up or down based on the misfit in response patterns. With conventional test scoring and Rasch JMLE, a lucky guess counts as a correct answer exactly like any other correct answer. Unexpected responses can be identified by fit statistics. With the three-parameter-logistic item-response-theory (3-PL IRT) model, the score value of an unexpected correct answer is diminished whether it is a lucky guess or due to special knowledge. In Winsteps, responses to off-target items (the locations of lucky guesses and careless mistakes) can be trimmed with [CUTLO=](#) and [CUTHI=](#), or be diminished using [TARGET=](#) Yes.

Comment on (13): JMLE exhibits some estimation bias in small data sets, but this rarely exceeds the precision (model standard error of measurement, SEM) of the measures. Estimation bias is only of concern when exact probabilistic inferences are to be made from short tests or small samples. Estimation bias can be exactly corrected for paired-comparison data with [PAIRED=](#) Yes. For other data, it can be approximately [corrected](#) with [STBIAS=](#) Yes, but, in practice, this is not necessary (and sometimes not advisable).

JMLE estimation details:

Newton-Raphson is an all-purpose estimation procedure and relatively easy to implement, so Winsteps once used this method. However, Newton-Raphson works best when there is a clear maximum in the likelihood function. This usually happens with complete data with dichotomies, but the story is different for long rating scales, partial credit, and incomplete data. Then we can get several local maxima, and Newton-Raphson has difficulty choosing between them. This situation became more common as Winsteps was applied to datasets that were not originally envisioned. For instance, it was analysis of DNA strings that motivated the increase in items in Winsteps to 60,000.

Rasch does have a useful feature. The underlying functions are all monotonic logistic curves. We can take advantage of this to refine the estimation process. So, in Winsteps, the estimation process for each parameter is like this:

step 1.

current value of the parameter estimate -> compute expected score

current value of the parameter estimate + a little bit -> compute expected score

compute the logistic ogive between the two current values and the two expected scores

from the logistic ogive, predict the parameter value that matches the observed score. This is the new current value for this parameter. Do this once.

step 2. Then do the same thing for the next parameter and all the other parameters.

step 3. return to step 1 while the biggest change in any parameter value is big, or the biggest difference between any observed and expected score is big. "Big" is defined by the convergence criteria.

step 4. the estimates have converged. We have the best values of the parameters. Rejoice!

PROX is the Normal Approximation Algorithm devised by Cohen (1979). This algorithm capitalizes on the similar shapes of the logistic and normal ogives. It models both the persons and the items to be normally distributed. The [variant of PROX](#) implemented in Winsteps allows missing data. The form of the estimation equations is:

$$\begin{aligned} \text{Ability of person} = & \text{Mean difficulty of items encountered} + \\ & \log \left(\frac{\text{observed score} - \text{minimum possible score on items encountered}}{\text{maximum possible score on items encountered} - \text{observed score}} \right) \\ & * \text{square-root} \left(1 + \frac{\text{variance of difficulty of items encountered}}{2.9} \right) \end{aligned}$$

In Winsteps, PROX iterations cease when the variance of the items encountered does not increase substantially from one iteration to the next.

Advantages - these are implementation dependent, and are implemented in Winsteps: (2)-(9) of JMLE

Computationally the fastest estimation method.

Disadvantages

- (1) Person and item measures assumed to be normally distributed.
- (11)-(14) of JMLE

Other estimation methods in common use (but not implemented in Winsteps):

Gaussian least-squares finds the Rasch parameter values which minimize the overall difference between the observations and their expectations, $\sum((X_{ni} - E_{ni})^2)$ where the sum is overall all observations, X_{ni} is the observation when person encounters item i , and E_{ni} is the expected value of the observation according to the current Rasch parameter estimates. For Effectively, off-target observations are down-weighted, similar to [TARGET=Yes](#) in Winsteps.

Minimum chi-square finds the Rasch parameter values which minimize the overall statistical misfit of the data to the model, $\sum((X_{ni} - E_{ni})^2 / V_{ni})$ where V_{ni} is the modeled binomial or multinomial variance of the observation around its expectation. Effectively off-target observations are up-weighted to make them less improbable.

Gaussian least-squares and Minimum chi-square:

Advantages - these are implementation dependent:

- (1)-(8) All those of JMLE.

Disadvantages:

- (9) persons with the same total raw scores on the same items generally have different measures; items with the same raw scores across the same persons generally have different measures.
- (11)-(13) of JMLE
- (14) global fit tests uncertain.

[CMLE](#). Conditional maximum likelihood estimation. Item difficulties are structural parameters. Person abilities are incidental parameters, conditioned out for item difficulty estimation by means of their raw scores. The item difficulty estimates are those that maximize the likelihood of the data given the person raw scores and assuming the data fit the model. The item difficulties are then used for person ability estimation using a JMLE approach.

Advantages - these are implementation dependent:

- (1), (6)-(9) of JMLE
- (3) the ability to analyze person sample sizes of any size
- (5) flexibility with item and rating scale structure anchor values
- (12) statistically-consistent item estimates
- (13) minimally estimation-biased item estimates
- (14) exact global fit statistics

Disadvantages:

- (2) limited flexibility with missing data
- (3) test length severely limited by mathematical precision of the computer
- (4) asymmetric analysis of person and item parameters so that transposing rows and columns changes the estimates
- (5) no person anchor values
- (11) of JMLE
- (13) estimation-biased of person estimates small but uncertain

[QCMLE](#). Quasi-Conditional Maximum Likelihood Estimation. This estimates the CMLE values from the JMLE probability matrix. The marginal totals of the CMLE and JMLE probability matrices are the same, and the cell values are similar. QCMLE gives a useful indication of the difference between JMLE and CMLE estimates, and so the size of the estimation bias in JMLE estimates. This is usually small to negligible.

EAP. Expected A Posteriori estimation derives from Bayesian statistical principles. This requires assumptions about the expected parameter distribution. An assumption is usually normality, so EAP estimates are usually more normally distributed than Winsteps estimates (which are as parameter-distribution-free as possible). EAP is not implemented in Winsteps.

MMLE. Marginal maximum likelihood estimation. Item difficulties are structural parameters. Person abilities are incidental parameters, integrated out for item difficulty estimation by imputing a person measure distribution. The item difficulties are then used for person ability estimation using a JMLE approach.

Advantages - these are implementation dependent:

- (3), (6)-(9) of JMLE
- (1) independence from specific item distributional forms.
- (2) flexibility with missing data extends to minimal length person response strings
- (5) flexibility with item and rating scale structure anchor values
- (11) extreme (zero, perfect) scores for persons are used for item estimation.
- (12) statistically-consistent item estimates
- (13) minimally estimation-biased item estimates
- (14) exact global fit statistics

Disadvantages:

- (1) specific person distribution required
- (4) asymmetric analysis of person and item parameters so that transposing rows and columns changes the estimates
- (5) no person anchor values
- (11) measures for extreme (zero, perfect) scores for specific persons or items require post-hoc estimation.
- (13) estimation-biased of person estimates small but uncertain

PMLE. Pairwise maximum likelihood estimation. Person abilities are incidental parameters, conditioned out for item difficulty estimation by means of pairing equivalent person observations. The item difficulties are then used for person ability estimation using a JMLE approach.

Advantages - these are implementation dependent:

- (1), (3), (6), (7) of JMLE
- (5) flexibility with item and rating scale structure anchor values
- (8) all persons with the same total raw scores on the same items have the same measure
- (12) statistically-consistent item estimates

Disadvantages:

- (11) of JMLE
- (2) reduced flexibility with missing data
- (4) asymmetric analysis of person and item parameters so that transposing rows and columns changes the estimates
- (5) no person anchor values
- (8) items with the same total raw scores across the same persons generally have different measures.
- (13) estimation-biased or item and person estimates small but uncertain
- (14) global fit tests uncertain.
- (15) uneven use of data in estimation renders standard errors and estimates less secure

Thomas Warm's (1989) Weighted Mean Likelihood Estimation

WMLE (also called WLE) estimates, reported in [IFILE=](#) and [PFILE=](#), are usually slightly more central than Winsteps estimates. Standard MLE estimates are the maximum values of the likelihood function and so statistical modes. Thomas Warm shows that the likelihood function is skewed, leading to an additional source of estimation bias. The mean likelihood estimate is less biased than the maximum likelihood estimate. Warm suggests an unbiasing correction that can be applied, in principle, to any MLE method, but there are computational constraints. Even when feasible, this fine tuning appears to be less than the relevant standard errors and have little practical benefit. The WMLE procedure can over-correct for the estimation bias in measures estimated from almost-extreme scores or very few observations.

Cohen Leslie. (1979) Approximate Expressions for Parameter Estimates in the Rasch Model, *The British Journal of Mathematical and Statistical Psychology*, 32, 113-120

Fisher R.A. On the mathematical foundations of theoretical statistics. *Proc. Roy. Soc.* 1922 Vol. CCXXII p. 309-368

Warm T.A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54, 427-450

19.40 Exact Match: OBS% and EXP%

ENTRY	RAW	MODEL	INFIT	OUTFIT	PTMEA	EXACT MATCH						
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	OBS%	EXP%	KID
72	14	25	-1.32	.37	2.02	2.9	5.16	5.7	.04	60.0	65.8	JACKSON, SOLOMON

Suppose your dataset consists of observations, $\{X_{ni}\}$, of person n on item i . Based on the Rasch parameters (measures), there is an expected value Eni corresponding to each observation X_{ni} . Eni is obtained by a calculation from the Rasch model.

When the absolute value of $(X_{ni}-Eni)$ is less than 0.5 then the observed data point is within 0.5 score points of its expected value, so the match is the closest possible. Thus, across all observations of item i ,

Count ($|X_{ni}-Eni| < 0.5$) = A - these observations are of the closest categories to their expectations

Count ($|X_{ni}-Eni| = 0.5$) = B - these observations are on the borderline of matching their expectations

Count ($|X_{ni}-Eni| > 0.5$) = C - these observations are at least one category away from their expectations

So that $A+B+C = \text{Count}(X_{ni})$

$$\text{OBS\%} = \text{Observed \%} = 100 * (A + B/2) / (A+B+C)$$

$B/2$: consider the situation when person ability = item difficulty with dichotomous observations. Then every observation is 1 or 0, and every expectation is 0.5, so $B = 100\%$. Under these conditions, it would be reasonable to say that 50% ($B/2$) of the observations matched their expectations.

Each possible value of X_{ni} has a probability according to the Rasch model. Based on these, the expected value of OBS% can be computed, this is the EXP%. So, if the possible values of X_{ni} are $j=0,1,2,\dots,m$, with probabilities P_{nij} , then

$$A = \sum ((|j-Eni| < 0.5) * P_{nij})$$

$$B = \sum ((|j-Eni| = 0.5) * P_{nij})$$

$$C = \sum ((|j-Eni| > 0.5) * P_{nij})$$

So that $A+B+C = \text{Count}(X_{ni})$

$$\text{EXP\%} = \text{Expected \%} = 100 * (A + B/2) / (A+B+C)$$

If $\text{OBS\%} < \text{EXP\%}$ then the local data are more random than the model predicts.

If $\text{OBS\%} > \text{EXP\%}$ then the local data are more predictable than the model predicts.

19.41 Excel plotting

First try the "[Plots](#)" pull-down menu.

Plotting: This is conveniently and flexibly done with Excel:

(A) Check that your copy of Excel works.

(B) Download the free Chart-Labeler add-in from www.appspro.com/Utilities/ChartLabeler.htm

(C) Run XYChartLabeler.exe

The Excel add-in "XY Chart Labels" is added to the Excel Tools pull-down menu.

(D) To plot

Write PFILE= or IFILE= from two analyses, or copy from Output Tables.

Copy and paste each of these into Excel:

Use "Data" "Text to Columns" to put values in columns

Put the columns to be plotted next to each other:

x-axis values to left of y-axis values.

Highlight numbers to be cross-plotted.

To make a plot:

Click on "Chart Wizard" or "Insert" ribbon

Click on "XY Scatter"

"Next"

"Next"

Fill in "Chart Title" and "Value" names

"Next"

Click "A new sheet"

"Finish"

On the plot:

Click "Series 1"

"Clear"

Right click a data point

Click "Format data"

Click "Marker none" (the data points will soon disappear!! - Don't worry!)

Click "OK"

Right click a grid-line

"Clear"

Add point labels:

Click on "Chart" tab

Click "Tools"

Click "XY Chart labels"

Click "XY Add"

Click "Centered"

Click "Select label _"

Click "Sheet 1" tab

Highlight point labels

Click red marker

Click "OK"

Point-labeled XY plot appears.

Use the plot to communicate:

Click on plot

Use handles to make the plot square

If drawing toolbar is not active:

Right click a toolbar

Click on Drawing

Click on "line" tool

Draw in a useful line.

19.42 Excel-formatted output Tables

Exporting Winsteps Tables to Excel is easy.

Winsteps control file, [Extra specifications](#), or [Specification menu: BOXSHOW=No](#)

Winsteps control file or [Output Tables menu](#): the Table you want, such as [Table 6.1](#), misfitting persons.

Output Table in NotePad (or your text editor):

"Select" with the mouse the part you want (Usually column heading lines to the bottom of the Table)
then right-click "Copy"

Launch Excel

Paste into Excel - top left cell. The Winsteps Table will go into the first column.

Excel: "Data", "Text to columns"

Excel: Delimited with spaces, or Fixed Width

Excel usually gets everything exactly right, so that each Winsteps table column is in a separate Excel column.

Done!

You can now use Excel to sort and select the Table columns.
Then paste the Excel table into Word (or wherever).

19.43 Extra Specifications prompt

Winsteps expects to find the control variables in your control file. You may, however, specify one or more control variables on the "Extra specifications" line. These variables supersede instructions in the control file. This is useful for making temporary changes to the control variables. There are special rules for blanks (see Example 3). You can turn off the Extra Specifications prompt from the [Edit Initial Settings](#) menu.

Example 1: You want to verify that your data is correctly formatted, so you only want to do one [JMLE](#) iteration this time, i.e., you want to set [MJMLE=1](#) for this run only:

```
Please enter name of Winsteps control file: SF.TXT(Enter)
Please enter name of report output file: SFO.TXT(Enter)
Extra specifications? (e.g., MJMLE =1), or press Enter:
MJMLE=1(Enter)
```

Note:

```
Extra specifications? (e.g., MJMLE=1), or press Enter:
MJMLE = 1(Enter)
is invalid because there are blanks in MJMLE = 1.
```

Example 2: You want to produce the fit plot in Table 4 with specially chosen ranges on the axes:

```
Please enter name of Winsteps control file: SF.TXT(Enter)
Please enter name of report output file: SFO.TXT(Enter)
Extra specifications? (e.g., MJMLE=1), or press Enter:
TABLES=0001 MRANGE=3 FRANGE=4(Enter)
```

Example 3: To put blanks in an Extra Specification. Put the whole specification within " " (double quotes). Put the argument within ' ' (single quotes). E.g., You want the title [TITLE=](#) to be: Analysis B, and [UMEAN=50](#).

```
Please enter name of Winsteps control file: SF.TXT(Enter)
Please enter name of report output file: SFO.TXT(Enter)
Extra specifications? (e.g., MJMLE=1), or press Enter:
"Title = 'Analysis B' " UMEAN=50
```

Example 4: Sub-lists.

In the Control file:

```
ISGROUPS=*
1-32 A
33-64 B
65-98 C
99-110 D
*
```

At the Extra specifications prompt: Use a comma , instead of a blank for the internal separator. Use a blank for the end-of-line separator.

```
ISGROUPS=* 1-32,A 33-64,B 65-98,C 99-110,D *
```

19.44 Extreme scores: what happens

Estimation:

Extreme scores are the lowest and highest possible scores for persons on items, or for items by persons. They include zero and perfect scores. They are shown in the Tables as MINIMUM ESTIMATE MEASURE and MAXIMUM ESTIMATE MEASURE.

Mathematically, they correspond to infinite or indefinite measures on the latent variable and so are not directly estimable. Accordingly persons or items with extreme scores are dropped for the duration of the measurement estimation process. The extreme persons are dropped casewise. The extreme items are dropped listwise.

Sometimes the effect of dropping extreme items and persons is to make other items and persons extreme. If so, these are also dropped. If the data have a Guttman pattern, ultimately all items and persons are dropped and the measures for that data set are reported as inestimable.

After the measures of all non-extreme items and persons have been estimated, then the extreme scores are reinstated. Reasonable extreme measures are imputed for them (using a Bayesian approach), so that all persons and items have measures. This is done by making small score adjustments using [EXTREMESCORE=](#).

With [TARGET=](#), [CUTLO=](#), [CUTHI=](#) or [ISGROUPS=](#), it may not be possible to estimate a measure for an extreme score. There are reported as [INESTIMABLE](#).

Fit computation:

Item Infit and Outfit: In the [IFILE=](#), these are reported for each item and summarize fit across all scored responses excluding responses in extreme person scores.

In the [DISFILE=](#), these are reported for each scored response and summarize fit excluding responses in extreme person scores.

Let's set [PDROPEXTREME=YES](#), so that extreme scores for persons are not counted. Then, here is an example:

```
IFILE:   Infit and Outfit mean-squares for 16237 responses scored 1 or 0 = 1.0873 and 1.1317
DISFILE: Infit and Outfit mean-squares for 10537 responses scored 1      = 1.0733 and 1.0916
         Infit and Outfit mean-squares for 5700 responses scored 0 have different values for
         each incorrect response code.
```

Responses in extreme person scores always have perfect fit (infit and outfit mean-squares = 0), so they are excluded from fit computations everywhere.

Extreme person scores do have estimated measures, so they are included in the item and item-option response counts, average measures, and correlations.

Example 1: We want statistics that exclude responses in all person and item extreme scores.

[PDROPEXTREME=](#) Yes

[IDROPEXTREME=](#) Yes

19.45 Fit diagnosis: infit outfit mean-square standardized

Remember that our purpose is to measure the persons, not to optimize the items and raters. A good approach is to compute the person measures based on all the different item selections that you think are reasonable. Start with all the items, and then reduce to smaller sets of items. Cross-plot the person measures. If the person measures are collinear, use the larger set of items. If the person measures are not collinear, use the set of items which produces the more meaningful set of person measures.

What do Infit Mean-square, Outfit Mean-square, Infit Zstd (z-standardized), Outfit Zstd (z-standardized) mean?

Every observation contributes to both infit and outfit. But the weighting of the observations differs. On-target observations contribute less to outfit than to infit.

Outfit: outlier-sensitive fit statistic. This is based on the conventional chi-square statistic. This is more sensitive to unexpected observations by persons on items that are relatively very easy or very hard for them (and vice-versa).

Infit: inlier-pattern-sensitive fit statistic. This is based on the chi-square statistic with each observation weighted by its statistical information (model variance). This is more sensitive to unexpected patterns of observations by persons on items that are roughly targeted on them (and vice-versa).

Outfit = $\sum (\text{residual}^2 / \text{information}) / (\text{count of residuals}) = \text{average} ((\text{standardized residuals})^2) = \text{chi-square/d.f.} = \text{mean-square}$

The standardized residual is also called the *Pearson* residual.

Infit = $\sum ((\text{residual}^2 / \text{information}) * \text{information}) / \sum(\text{information}) = \text{average} ((\text{standardized residuals})^2 * \text{information}) = \text{information-weighted mean-square}$

Mean-square: this is the chi-square statistic divided by its degrees of freedom. Consequently its expected value is close to 1.0. Values greater than 1.0 (underfit) indicate unmodeled noise or other source of variance in the data - these degrade measurement. Values less than 1.0 (overfit) indicate that the model predicts the data too well - causing summary statistics, such as reliability statistics, to report inflated statistics. See further [dichotomous](#) and [polytomous](#) mean-square statistics. The mean-square Outfit statistic is also called the *Reduced* chi-square statistic. For computations, see

If the mean-squares average much below 1.0, then the data may have an almost-Guttman pattern. Please use much tighter [convergence criteria](#).

Z-Standardized: these report the statistical significance (probability) of the chi-square (mean-square) statistics occurring by chance when the data fit the Rasch model. "Standardized" means "transformed to conform to a unit-normal distribution". The values reported are unit-normal deviates, in which .05% 2-sided significance corresponds to 1.96. Overfit is reported with negative values. These are also called [t-statistics](#) reported with infinite degrees of freedom.

ZSTD probabilities: two-sided unit-normal deviates	
1.00	p= .317
1.96	.050
2.00	.045
2.58	.01
3.00	.0027
4.00	.00006
5.00	.0000006

Infit was an innovation of Ben Wright's (G. Rasch, 1980, Afterword). Ben noticed that the standard statistical fit statistic (that we now call Outfit) was highly influenced by a few outliers (very unexpected observations). Ben need a fit statistic that was more sensitive to the overall pattern of responses, so he devised Infit. Infit weights the observations by their statistical information (model variance) which is higher in the center of the test and lower at the extremes. The effect is to make Infit less influenced by outliers, and more sensitive to patterns of inlying observations.

Ben Wright's Infit and Outfit statistics (e.g., [RSA](#), p. 100, www.rasch.org/rmt/rmt34e.htm) are initially computed as mean-square statistics (i.e., chi-square statistics divided by their degrees of freedom). For Outfit the d.f. is the count of observations. For Infit the d.f. is the sum of the information in the observations = 1 / item or person logit S.E.**2. Their likelihood (significance) is then computed. This could be done directly from chi-square tables, but the convention is to report them as unit normal deviates (i.e., [t-statistics](#) corrected for their degrees for freedom). I prefer to call them z-statistics, but the Rasch literature has come to call them [t-statistics](#), so now I do to. It is confusing because they are not strictly Student [t-statistics](#) (for which one needs to know the degrees of freedom) but are random normal deviates.

General guidelines:

First, investigate **negative point-measure or point-biserial correlations**. Look at the Distractor Tables, e.g., [10.3](#). Remedy miskeys, data entry errors, etc.

Then, the general principle is:

- Investigate outfit before infit,**
- mean-square before t standardized,**
- high values before low or negative values.**

There is an asymmetry in the implications of out-of-range high and low mean-squares (or positive and negative [t-statistics](#)). High mean-squares (or positive [t-statistics](#)) are a much greater threat to validity than low mean-squares (or negative fit statistics).

Poor fit does not mean that the Rasch measures (parameter estimates) aren't additive. The Rasch model forces its estimates to be additive. Misfit means that the reported estimates, though effectively additive, provide a distorted picture of the data.

The fit analysis is a report of how well the data accord with those additive measures. So a MnSq >1.5 suggests a deviation from unidimensionality in the data, not in the measures. So the unidimensional, additive measures present a distorted picture of the data.

High outfit mean-squares may be the result of **a few random responses** by low performers. If so, drop with [PDFILE=](#) these performers when doing item analysis, or use [EDFILE=](#) to change those response to missing.

High infit mean-squares indicate that the items are mis-performing for the people on whom the items are targeted. This is a bigger threat to validity, but more difficult to diagnose than high outfit.

Mean-squares show the size of the randomness, i.e., the amount of distortion of the measurement system. 1.0 are their expected values. Values less than 1.0 indicate observations are too predictable (redundancy, model overfit). Values greater than 1.0 indicate unpredictability (unmodeled noise, model underfit). Mean-squares usually average to 1.0, so if there are high values, there must also be low ones. Examine the high ones first, and temporarily remove them from the analysis if necessary, before investigating the low ones.

Zstd are t-tests of the hypotheses "do the data fit the model (perfectly)?" ZSTD (standardized as a z-score) is used of a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's [t-statistic](#) value has been adjusted to a unit normal value. They show the improbability (significance). 0.0 are their expected values. Less than 0.0 indicate too predictable. More than 0.0 indicates lack of predictability. If mean-squares are acceptable, then Zstd can be ignored. They are truncated towards 0, so that 1.00 to 1.99 is reported as 1. So a value of 2 means 2.00 to 2.99, i.e., at least 2. For exact values, see Output Files. If the test involves less than 30 observations, it is probably too insensitive, i.e., "everything fits". If there are more than 300 observations, it is probably too sensitive, i.e., "everything misfits".

Interpretation of parameter-level mean-square fit statistics:	
>2.0	Distorts or degrades the measurement system.
1.5 - 2.0	Unproductive for construction of measurement, but not degrading.
0.5 - 1.5	Productive for measurement.
<0.5	Less productive for measurement, but not degrading. May produce misleadingly good reliabilities and separations.

In general, mean-squares near 1.0 indicate little distortion of the measurement system, regardless of the Zstd value. Evaluate high mean-squares before low ones, because the average mean-square is usually forced to be near 1.0. Mean-square fit statistics will average about 1.0, so, if you accept items (or persons) with large mean-squares (low discrimination), then you must also accept the counter-balancing items (or persons) with low mean-squares (high discrimination).

Outfit mean-squares: influenced by outliers. Usually easy to diagnose and remedy. Less threat to measurement.
 Infit mean-squares: influenced by response patterns. Usually hard to diagnose and remedy. Greater threat to measurement.

Extreme scores always fit the Rasch model exactly, so they are omitted from the computation of fit statistics. If an extreme score has an anchored measure, then that measure is included in the fit statistic computations.

Anchored runs: Anchor values may not exactly accord with the current data. To the extent that they don't, the fit statistics may be misleading. Anchor values that are too central for the current data tend to make the data appear to fit too well. Anchor values that are too extreme for the current data tend to make the data appear noisy.

Question: Are you contradicting the usual statistical advice about model-data fit?

Statisticians are usually concerned with "how likely are these data to be observed, assuming they accord with the model?" If it is too unlikely (i.e., significant misfit), then the verdict is "these data don't accord with the model." The practical concern is: "In the imperfect empirical world, data never exactly accord with the Rasch model, but do these data deviate seriously enough for the Rasch measures to be problematic?" The builder of my house followed the same approach (regarding Pythagoras theorem) when building my bathroom. It looked like the walls were square enough for his practical purposes. Some years later, I installed a full-length rectangular mirror - then I discovered that the walls were not quite square enough for my purposes (so I had to make some adjustments) - so there is always a judgment call. The table of mean-squares is my judgment call as a "builder of Rasch measures".

Question: My data contains misfitting items and persons, what should I do?

Let us clarify the objectives here.

A. www.rasch.org/rmt/rmt234g.htm is aimed at the usual situation where someone has administered a test from somewhere to a sample of people, and we, the analysts, are trying to rescue as much of that data as is meaningful. We conservatively remove misfitting items and persons until the data makes reasonable sense. We then anchor those persons and items to their good measures. After reinstating whatever misfitting items and persons we must report, we do the final analysis.

B. A pilot study is wonderfully different. We want to optimize the subset of items. The person sample and the data can be tailored to optimize item selection. Accordingly,

First, even before data analysis, we need to arrange the items into their approximately intended order along the latent variable. With 89 items, item can be conceptually grouped into clusters located at 5 or more levels of the latent variable, probably more than 5. This defines what we want to measure. If we don't know this order, then we will not know whether we have succeeded in measuring what we intended to measure. We may accidentally construct a test that measures a related variable. This happened in one edition of the MMPI where the test constructors intended to measure depression, but produced a scale that measured "depression+lethargy".

Second, we analyze the data and inspect the item hierarchy. Omit any items that are locating in the wrong place on the latent variable. By "omit", I mean give a weight of zero with $IWEIGHT=$. Then the item stays in the analysis, but does not influence other numbers. This way we can easily reinstate items, if necessary, knowing where they would go if they had been given the usual weight of 1.

Third, reanalyze the data with the pruned item hierarchy. Omit all persons who severely underfit the items, these are contradicting the latent variable. Again, "omit" means $PWEIGHT=$ 0. Also omit persons whose "cooperation" is because they have an overfitting response set, such as the middle category of every item.

Fourth, analyze again. The data should be coherent. Items in the correct order. Persons cooperating. So apply all the other selection criteria, such as content balancing, DIF detection, to this coherent dataset.

Question: Should I report Outfit or Infit?

A chi-square statistic is the sum of squares of standard normal variables. Outfit is a chi-square statistic. It is the sum of squared standardized residuals (which are modeled to be standard normal variables). So it is a conventional chi-square, familiar to most statisticians. Chi-squares (including outfit) are sensitive to outliers. For ease of interpretation, this chi-square is divided by its degrees of freedom to have a mean-square form and reported as "Outfit". Consequently I recommend that the Outfit be reported unless there is a strong reason for reporting infit.

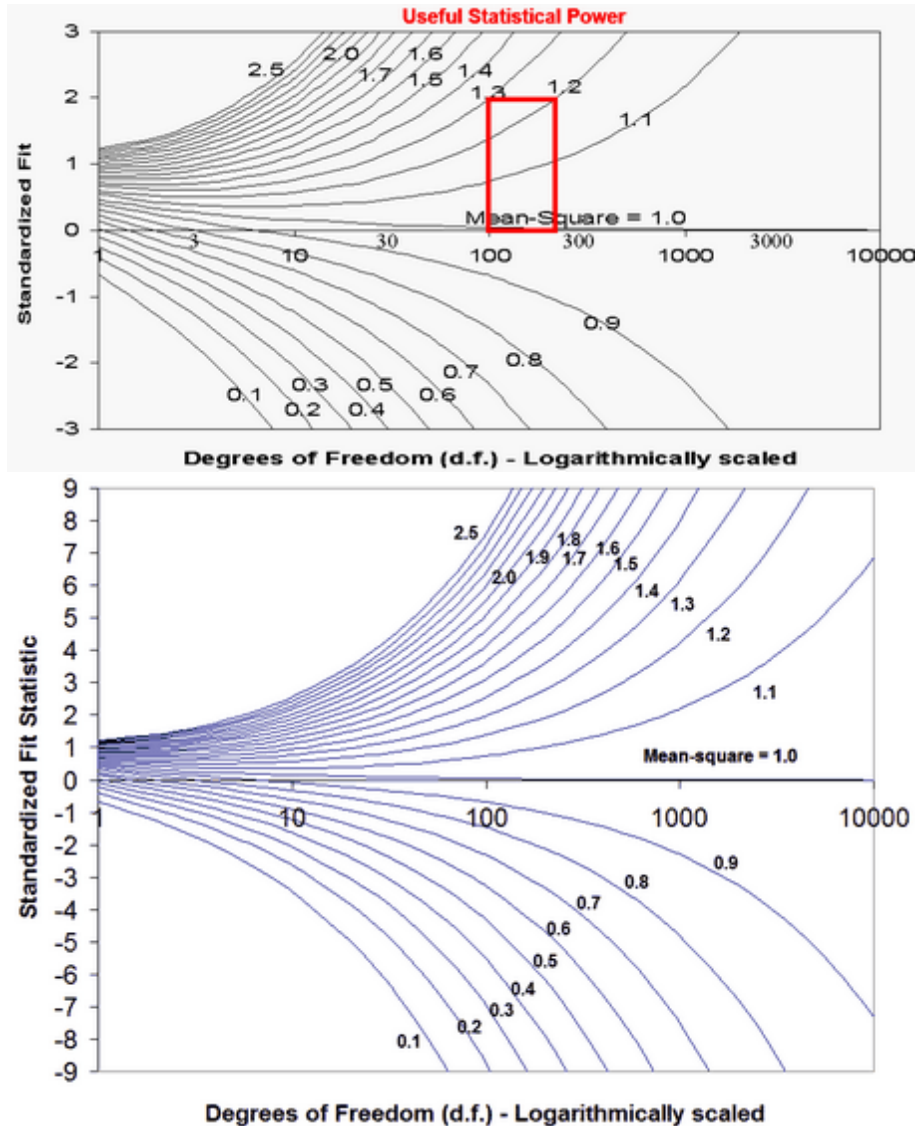
In the Rasch context, outliers are often lucky guesses and careless mistakes, so these outlying characteristics of respondent behavior can make a "good" item look "bad". Consequently, Infit was devised as a statistic that down-weights outliers and focuses more on the response string close to the item difficulty (or person ability). Infit is the sum of (squares of standard normal variables multiplied by their statistical information). For ease of interpretation, Infit is reported in mean-square form by dividing the weighted chi-square by the sum of the weights. This formulation is unfamiliar to most statisticians, so I recommend against reporting Infit unless the data are heavily contaminated with irrelevant outliers.

Question: Are mean-square values, >2 etc, sample-size dependent?

The mean-squares are corrected for sample size: they are the chi-squares divided by their degrees of freedom, i.e., sample size. The mean-squares answer "how big is the impact of the misfit". The [t-statistics](#) answer "how likely are data like these to be observed when the data fit the model (exactly)." In general, the bigger the sample the less likely, so that t-

statistics are highly sample-size dependent. We eagerly await the theoretician who devises a statistical test for the hypothesis "the data fit the Rasch model usefully" (as opposed to the current tests for perfectly).

The relationship between mean-square and z-standardized [t-statistics](#) is shown in this plot. Basically, the standardized statistics are insensitive to misfit with less than 30 observations and overly sensitive to misfit when there are more than 300 observations.



Question: For my sample of 2400 people, the mean-square fit statistics are close to 1.0, but the Z-values associated with the INFIT/OUTFIT values are huge (over 4 to 9.9). What could be causing such high values?

Your results make sense. Here is what has happened. You have a sample of 2,400 people. This gives huge statistically power to your test of the null hypothesis: "These data fit the Rasch model (exactly)." In the nomographs above, a sample size of 2,400 (on the right-hand-side of the plot) indicates that even a mean-square of 1.2 (and perhaps 1.1) would be reported as misfitting highly significantly. So your mean-squares tell us: "these data fit the Rasch model usefully", and the Z-values tell us: "but not exactly". This situation is often encountered in situations where we know, in advance, that the null hypothesis will be rejected. The Rasch model is a theoretical ideal. Empirical observations never fit the ideal of the Rasch model if we have enough of them. You have more than enough observations, so the null hypothesis of exact model-fit is rejected. It is the same situation with Pythagoras theorem. No empirical right-angled-triangle fits Pythagoras theorem if we measure it precisely enough. So we would reject the null hypothesis "this is a right-angled-triangle" for all triangles that have actually been drawn. But obviously billions of triangle are usefully right-angled.

Example of computation:

Imagine an item with categories $j=0$ to m . According to the Rasch model, every category has a probability of being observed, P_j .

Then the expected value of the observation is $E = \sum (j * P_j)$

The model variance (sum of squares) of the probable observations around the expectation is $V = \sum (P_j * (j - E)^2)$.

This is also the statistical information in the observation.

For dichotomies, these simplify to $E = P_1$ and $V = P_1 * P_0 = P_1*(1-P_1)$.

For each observation, there is an expectation and a model variance of the observation around that expectation.

residual = observation - expectation

Outfit mean-square = $\sum (\text{residual}^2 / \text{model variance}) / (\text{count of observations})$

Infit mean-square = $\sum (\text{residual}^2) / \sum (\text{modeled variance})$

Thus the outfit mean-square is the accumulation of squared-standardized-residuals divided by their count (their expectation). The infit mean-square is the accumulation of squared residuals divided by their expectation.

Outlying observations have smaller information (model variance) and so have less information than on-target observations. If all observations have the same amount of information, the information cancels out. Then Infit mean-square = Outfit mean-square.

For dichotomous data. Two observations: Model $p=0.5$, observed=1. Model $p=0.25$, observed =1.

Outfit mean-square = $\sum ((\text{obs-exp})^2 / \text{model variance}) / (\text{count of observations}) = ((1-0.5)^2 / (0.5*0.5) + (1-0.25)^2 / (0.25*0.75)) / 2 = (1 + 3) / 2 = 2$

Infit mean-square = $\sum ((\text{obs-exp})^2) / \sum (\text{model variance}) = ((1-0.5)^2 + (1-0.25)^2) / ((0.5*0.5) + (0.25*0.75)) = (0.25 + 0.56) / (0.25 + 0.19) = 1.84$. The off-target observation has less influence.

The [Wilson-Hilferty](#) cube root transformation converts the mean-square statistics to the normally-distributed z-standardized ones. For more information, please see Patel's "Handbook of the Normal Distribution" or www.rasch.org/rmt/rmt162g.htm.

Diagnosing Misfit: Noisy = Underfit. Muted = Overfit				
Classification	INFIT	OUTFIT	Explanation	Investigation
	Noisy	Noisy	Lack of convergence Loss of precision Anchoring	Final values in Table 0 large? Many categories? Large logit range? Displacements reported?
Hard Item	Noisy	Noisy	Bad item	Ambiguous or negative wording? Debatable or misleading options?
	Muted	Muted	Only answered by top people	At end of test?
Item	Noisy	Noisy	Qualitatively different item Incompatible anchor value	Different process or content? Anchor value incorrectly applied?
		?	Biased (DIF) item	Stratify residuals by person group?
	Muted	Muted	Curriculum interaction	Are there alternative curricula?
Rating scale	Noisy	Noisy	Extreme category overuse	Poor category wording?
	Muted	Muted	Middle category overuse	Combine or omit categories? Wrong model for scale?
Person	Noisy	?	Processing error Clerical error Idiosyncratic person	Scanner failure? Form markings misaligned? Qualitatively different person?
High Person	?	Noisy	Careless Sleeping Rushing	Unexpected wrong answers? Unexpected errors at start? Unexpected errors at end?
Low Person	?	Noisy	Guessing	Unexpected right answers?

			Response set "Special" knowledge	Systematic response pattern? Content of unexpected answers?
	Muted	?	Plodding Caution	Did not reach end of test? Only answered easy items?
Person/Judge Rating	Noisy	Noisy	Extreme category overuse	Extremism? Defiance? Misunderstanding the rating scale?
			Middle category overuse	Conservatism? Resistance?
Judge Rating	Muted	Muted	Apparent unanimity	Collusion? Hidden constraints?
INFIT : information-weighted mean-square, sensitive to irregular inlying patterns OUTFIT : usual unweighted mean-square, sensitive to unexpected rare extremes Muted : overfit, un-modeled dependence, redundancy, the data are too predictable Noisy : underfit, unexpected unrelated irregularities, the data are too unpredictable.				

Guessing and Carelessness

Both guessing and carelessness can cause high outfit mean-square statistics. Sometimes it is not difficult to identify which is the cause of the problem. Here is a procedure:

1) Analyze all the data: output [PFILE=pf.txt](#) [IFILE=if.txt](#) [SFILE=sf.txt](#)

2) Analyze all the data with anchor values: [PAFILE=pf.txt](#) [IAFILE=if.txt](#) [SAFILE=sf.txt](#)

2A) [CUTLO=-2](#) this eliminates responses to very hard items, so person misfit would be due to unexpected responses to easy items

2B) [CUTHI=2](#) this eliminates responses to very easy items, so person misfit would be due to unexpected responses to hard items.

What is your primary concern? Statistical fit or productive measurement?

Statistical fit is like "beauty". Productive measurement is like "utility".

Statistical fit is dominated by sample size. It is like looking at the data through a microscope. The more powerful the microscope (= the bigger the sample), the more flaws we can see in each item. For the purposes of beauty, we may well scrutinize our possessions with a microscope. Is that a flaw in the diamond? Is that a crack in the crystal? There is no upper limit to the magnification we might use, so there is no limit to the strictness of the statistical criteria we might employ. The nearer to 1.0 for mean-squares, the more "beautiful" the data.

In practical situations, we don't look at our possessions through a microscope. For the purposes of utility, we are only concerned about cracks, chips and flaws that will impact the usefulness of the items, and these must be reasonably obvious. In terms of mean-squares, the range 0.5 to 1.5 supports productive measurement.

However, life requires compromise between beauty and utility. We want our cups and saucers to be functional, but also to look reasonably nice. So a reasonable compromise for high-stakes data is mean-squares in the range 0.8 to 1.2.

Fit statistics for item banks are awkward. They depend on the manner in which the items in the item banks are used, and also the manner in which the item difficulties are to be verified and updated. Initial values for the items often originate in conventional paper-and-pencil tests, but if the item bank is used to support other styles of testing, these initial values will be superseded by more relevant values. Constructing fit statistics for other styles of testing has proved challenging for theoreticians. This has forced the relaxation of the strict rules of conventional statistical analysis. This is probably why you are having difficulty finding appropriate literature.

My person sample size is 2,000. Many items have significant misfit. What shall I do?

Don't despair! Your items may not be as bad as those statistics say.

If the noise in the data is homogeneous, then the noise-level is independent of sample size. The mean-square statistics will also be independent of sample size.

t-statistics are sensitive to the power of the statistical test (sample size). The relationship between mean-squares and *t*-statistics is shown in the Figures above which suggests that, for practical purposes, *t*-statistics are under-powered for sample sizes less than 100 and over-powered for sample sizes greater than 300.

So my recommendation (not accepted by conventional statisticians) is that mean-squares be used in preference to *t*-statistics. In my view, the standard *t*-tests are testing the wrong hypothesis. Wrong hypothesis = "The data fit the model (perfectly)". Right hypothesis = "The data fit the model (usefully)". Unfortunately conventional statisticians are not interested in usefulness and so have not formulated *t*-tests for it.

Alternatively, random-sample 300 from your 2,000 test-takers. Perform your *t*-tests with this reduced sample. Confirm your findings with another random-sample of 300.

19.46 Global fit statistics

Please use [Table 44](#). This uses simulations to estimate the degrees of freedom.

Here is an earlier computation of degrees of freedom which tended to exaggerate misfit of the data to the Rasch model:

Winsteps reports global fit statistics and approximate global log-likelihood chi-square (or chi-squared†) statistic in [Table 3.1](#). The usual significance levels are $p \leq .05$ and $p \leq .01$.

Example: To compare the fit of "Rating Scale Model" (RSM) and "Partial Credit Model" (PCM, [ISGROUPS=0](#)) analyses. The number of categories and items is in the [Table heading](#). The chi-square test is: (global chi-square for RSM analysis - global chi-square for PCM analysis) with d.f. ((PCM Categories - 2*Items) - (RSM categories - 2)) where the 2 is 1 for the bottom category and 1 for the centering of the rating/PCM thresholds

The [variance](#) tables report the relative sizes of explained and unexplained variances.

The chi-square value is approximate. It is based on the current reported estimates which may depart noticeably from the "true" maximum likelihood estimates for these data. The degrees of freedom are the number of datapoints used in the free estimation (i.e., excluding missing data, data in extreme scores, etc.) less the number of free parameters. For an unanchored analysis, free parameters = non-extreme items + non-extreme persons - 1 + (categories in estimated rating-scale structures - 2 * rating-scale structures).

Thus, for the "Liking for Science" data of 75 children administered 25 items. There are 74 non-extreme children and 25 non-extreme items. The data are complete so there are $74 \times 25 = 1850$ data points. The free parameters are $74 + 25 - 1 + (3\text{-category rating scale} - 2 \times 1 \text{ rating scale}) = 99$ parameters. So the degrees of freedom are $1850 - 99 = 1751$. The log-likelihood chi-square is 2657.91. So that the significance $p = .000$, i.e., the data exhibit highly significant misfit to the Rasch model, as is nearly always expected.

If you wish to compute your own global (or any other) fit test, the response-level probabilities, residuals etc. are reported in the [XFILE=](#). For instance, for a global fit test, you could add up all the log-probabilities. Then chi-square estimate = $-2 * \log\text{-probability}$. A different chi-square estimate is the sum of squared-standardized residuals. You can count up the number of free parameters. For complete dichotomous data, it is usually the minimum of (number of different person marginal raw scores, number of different item marginal scores) - 1.

For a more local fit test, the chi-square estimate is $-2 * \text{sum of log-probabilities of relevant data or sum of squared-standardized residuals for the relevant data}$. The degrees of freedom approximate the count of data points less L/L for each relevant person parameter and N/N for each relevant item parameter, where L is the number of responses by the person included in the local test and L is the total number of responses by the person. N is the number of responses to the item included in the local test and N is the total number of responses to the item.

Deviance statistics are more trustworthy. They are the difference between the chi-squares of two analyses, with d.f. of the difference between the number of free parameters estimated.

The Rasch model is an idealization, never achieved by real data. Accordingly, given enough data, we expect to see statistically significant misfit the model. If the current data do not misfit, we merely have to collect more data, and they will! In

essence, the null hypothesis of this significance test is the wrong one! We learn nothing from testing the hypothesis, "Do the data fit the model (perfectly)?" Or, as usually expressed in social science, "Does the model fit the data (perfectly)?" Perfection is never obtained in empirical data. What we really want to test is the hypothesis "Do the data fit the model usefully?" And, if not, where is the misfit, and what is it? Is it big enough in size (not "statistical significance") to cause trouble? This is the approach used in much of industrial quality-control, and also in Winsteps.

The general principle for degrees of freedom is "number of data points used for estimating the parameters - number of free parameters estimated". So, in most Rasch situations, when computing the d.f.:

1. omit extreme scores (persons and items).
2. data points = number of datapoints in non-extreme scores.
3. number of free person parameters = number of persons
4. number of free item parameters = number of items - 1, because the items are usually centered at 0 so one parameter is constrained to be the negative of the sum of the other parameters.
5. number of free parameters for each rating scale structure = number of categories in the structure - 2.
6. So dichotomies have no free parameters

"Rating scale model" = one structure for all the data

"Partial credit model" = one structure per item

Grouped-item model" = one structure per item-group.

However, this formula-based computation tends to exaggerate misfit of the data to the Rasch model. This, is superseded by the simulation method in [Table 44](#).

19.47 Glossary

Glossary - Dictionary - Lexicon of Rasch Measurement Terminology	
Glosario Español www.rasch.org/rmt/glosario.htm	
Ability	the level of successful performance of the objects of measurement (persons) on the latent variable. Each person's location on the unidimensional variable measured in "additive Rasch units", usually logits.
Additive scale	Scale of measurement in which the units have the properties of simple addition, so that "one more unit = the same amount extra regardless of the amount you already have". Typical measuring devices such as tape measures and thermometers have additive scales. Rasch additive scales are usually delineated in logits.
Agent of Measurement	the tool (items, questions, etc.) used to define a latent variable, and to position objects of measurement (persons etc.) along that variable.
Analytic rating	a rating of a specific aspect of a performance (cf. Holistic rating)
Anchor	the process of using anchor values to insure that different analyses produce directly comparable results.
Anchor Value	a pre-set logit value assigned to a particular object, agent or step to be used as a reference value for determining the measurements or calibrations of other objects, agents or steps.
Anchor Table	the table of Anchor Values used during Rasch analysis of an Input Grid and so included in the Results Table produced. The Anchor Table has the same format as the Results Table.
Anchoring	the process of using anchor values to insure that different analyses produce directly comparable results.
Best Test Design	Wright, B.D. & Stone, M.H., Best Test Design: Rasch Measurement. Chicago: Mesa Press, 1979
Bias	A change in logit values based on the particular agents or objects measured.
BOTTOM	The value shown in the Results Table for an agent on which all objects were successful, (so it was of bottom difficulty), or for an object which had no success on any agent (so it was of bottom ability)
Bottom Category	the response category at which no level of successful performance has been manifested.

Calibration	a difficulty measure in logits used to position the agents of measurement (usually test items) along the latent variable.
CAT Test	Computer-Adaptive Test. A test administered by computer in which the display of the next item depends on the response to the previous item.
Categories CATS	qualitative levels of performance on an observational or response format, e.g., a rating scale.
Cell	Location of data in the spreadsheet, given by a column letter designation and row number designation e.g. B7
Classical Test Theory	Item analysis in which the raw scores are treated as additive numbers.
Common Scale	a scale of measurement on which all agents and objects can be represented.
Column	Vertical line of data in the Spreadsheet data, usually representing in an Input Grid all responses to a particular item, or in a Results Table, all statistics measuring the same attribute of agents or objects.
Comment	A semi-colon ; followed by text. This is ignored by Winsteps and Facets.
Complete data	Data in which every persons responds to every item. It makes a completely-filled rectangular data matrix. There are no missing data.
Computer-Adaptive Test	CAT Test. A test administered by computer in which the display of the next item depends on the response to the previous item.
Construct validity	The correlation between the item difficulties and the latent trait as intended by the test constructor. "Is the test measuring what it is intended to measure?"
Content	the subject area evoked and defined by an agent.
Continuation line	A separate line of text which Winsteps analyses as appended to the end of the previous line. These are shown with "+".
Contrast component	In the principal components analysis of residuals, a principal component (factor) which is interpreted by contrasting the items (or persons) with opposite loadings (correlations) on the component.
Control file	A DOS-text file on your disk drive containing the Winsteps control variables.
Control variable	In Winsteps, "control variable = value", is an instruction for controlling the computer program, e.g., "ITEM1 = 4".
Convergence	the point at which further improvement of the item and person estimates makes no useful difference in the results. Rasch calculation ends at this point.
Correlation	the relationship between two variables
CTT	Classical Test Theory
Data file	Winsteps: file containing the person labels and the responses to the items. It is part of the Control file if DATA= or MFORMS= are not used.
Demographics	Information about the person included the person label, e.g., "F" for female or "M" for male
Deterministic	Exactly predictable without any uncertainty. This contrasts with Probabilistic.
Dichotomous Response	a response format of two categories such as correct-incorrect, yes-no, agree-disagree.
DIF Differential item functioning	Change of item difficulty depending on which person classification-group is responding to the item, also called "item bias"
Difficulty	the level of resistance to successful performance of the agents of measurement on the latent variable. An item with high difficulty has a low marginal score. The Rasch item difficulty is the location on the unidimensional latent variable, measured in additive Rasch units, usually logits. Item difficulty measures are the locations on the latent variable (Rasch dimension) where the highest and lowest categories of the item are equally probable, regardless of the number of categories the item has.
Dimension	a latent variable which is influencing the data values.
Discrepancy	one or more unexpected responses.

Distractor	Incorrect answer to a multiple-choice question, which is intended to distract the examinee away from the correct option. Sometimes all the options, correct and incorrect, are called "distractors".
Disturbance	one or more unexpected responses.
Diverging	the estimated calibrations at the end of an iteration are further from convergence than at the end of the previous iteration.
Easiness	the level of susceptibility to successful performance of the agents of measurement on the latent variable. An item with high easiness has a high marginal score.
Eigenvalue	The value of a characteristic root of a matrix, the numerical "size" of the matrix
Element	Individual in a facet, e.g., a person, an item, a judge, a task, which participates in producing an observation.
Empirical	Based on observation or experiment
Empirical data	data derived from observation or experimentation
END LABELS END NAMES	Winsteps: the end of the list of item identifying labels. This is usually followed by the data.
Entry number	Sequence number of the person or item in the dataset. Person: Entry number 1 is the top row of the response-level data. Item: Entry number 1 is the left-hand column of item-response data.
Equating	Putting the measures from two tests in the same frame of reference
Estimate	A value obtained from the data. It is intended to approximate the exactly true, but unknowable value.
EXP Expected value	Value predicted for this situation based on the measures
Expected Response	the predicted response by an object to an agent, according to the Rasch model analysis.
EXP() Exponential	Mathematical function used in estimating the Rasch measures
Exponential form	The Rasch model written in terms of exponentials, the form most convenient for computing response probabilities.
Extreme item	An item with an extreme score. Either everyone in the sample scored in the top category on the item, or everyone scored in the bottom category. An extreme measure is estimated for this item, and it fits the Rasch model perfectly, so it is omitted from fit reports.
Extreme person	A person with an extreme score. This person scored in the top category on the every item, or in the bottom category on every item. An extreme measure is estimated for this person, who fits the Rasch model perfectly, so is omitted from fit reports.
Facet	The components conceptualized to combine to produce the data, e.g., persons, items, judges, tasks.
Fit Statistic	a summary of the discrepancies between what is observed and what we expect to observe.
Focal group	The person classification-group which is the focus of a differential-item-functioning investigation
Frame of reference	The measurement system within which measures are directly comparable
Fundamental Measurement	1. Measurement which is not derived from other measurements. 2. Measurement which is produced by an additive (or equivalent) measurement operation.
Guttman	Louis Guttman (1916-1987) organized data into Scalograms intending that the observed response by any person to any items could be predicted deterministically from its position in the Scalogram.
Guttman pattern	Success on all the easier items. Failure on all the more difficulty items.
Heading	an identifier or title for use on tables, maps and plots.

Holistic rating	One rating which captures all aspects of the performance (cf. Analytic rating)
Hypothesis test	Fit statistics report on a hypothesis test. Usually the null hypothesis to be tested is something like "the data fit the model", "the means are the same", "there is no DIF". The null hypothesis is rejected if the results of the fit test are significant ($p \leq .05$) or highly significant ($p \leq .01$). The opposite of the null hypothesis is the alternate hypothesis.
Imputed data	Data generated by the analyst or assumed by the analytical process instead of being observed.
Independent	Not dependent on which particular agents and objects are included in the analysis. Rasch analysis is independent of agent or object population as long as the measures are used to compare objects or agents which are of a reasonably similar nature.
Infit	an information-weighted or inlier-sensitive fit statistic that focuses on the overall performance of an item or person, i.e., the information-weighted average of the squared standardized deviation of observed performance from expected performance. The statistic plotted and tabled by Rasch is this mean square normalized.
Interval scale	Scale of measurement on which equal intervals represent equal amounts of the variable being measured. Rasch analysis constructs interval scales with additive properties.
Item	agent of measurement (prompt, probe, "rating scale"), not necessarily a test question, e.g., a product rating. The items define the intended latent trait.
Item bank	Database of items including the item text, scoring key, difficulty measure and relevant statistics, used for test construction or CAT tests
Iteration	one run through the data by the Rasch calculation program, done to improve estimates by minimizing residuals.
Knox Cube Test	a tapping pattern test requiring the application of visual attention and short term memory.
Latent Trait	The idea of what we want to measure. A latent trait is defined by the items or agents of measurement used to elicit its manifestations or responses.
Link	Relating the measures derived from one test with those from another test, so that the measures can be directly compared.
LN() Logarithm	Natural or Napierian logarithm. A logarithm to the base e, where $e = 2.718...$ This contrasts with logarithms to the base 10.
Local origin	Zero point we have selected for measurement, such as sea-level for measuring mountains, or freezing-point for Celsius temperature. The zero point is chosen for convenience (similarly to a "setting-out point"). In Rasch measurement, it is often the average difficulty of the items.
Log-odds	The natural logarithm of the ratio of two probabilities (their odds).
Logit	"Log-odds unit": the unit of measure used by Rasch for calibrating items and measuring persons on the latent variable. A logarithmic transformation of the ratio of the probabilities of a correct and incorrect response, or of the probabilities of adjacent categories on a rating scale.
Logistic curve-fitting	an estimation method in which the improved value of an estimate is obtained by incrementing along a logistic ogive from its current value, based on the size of the current raw-score residual.
Logistic ogive	the relationship between additive measures and the probabilities of dichotomous outcomes.
Logit-linear	The Rasch model written in terms of log-odds, so that the measures are seen to form a linear, additive combination
Map	a bar chart showing the frequency and spread of agents and objects along the latent variable.
Matrix	a rectangle of responses with rows (or columns) defined by objects and columns (or rows) defined by agents.

MCQ Multiple-Choice Question.	This is an item format often used in educational testing where the examinee selects the letter corresponding to the answer.
Mean-square MnSq	Also called the relative chi-square and the normed chi-square. A mean-square fit statistic is a chi-square statistic divided by its degrees of freedom (d.f.). Its expectation is 1.0. Values below 1.0 indicate that the data are too predictable = overly predictable = overfit of the data to the model. Values above 1.0 indicate the data too unpredictable = underfit of the data to the model
Measure Measurement	the location (usually in logits) on the latent variable. The Rasch measure for persons is the person ability. The Rasch measure for items is the item difficulty.
Menu bar	This is at the top of a program's window, and shows a list of standard program operations
Misfit	Any difference between the data the model predictions. Misfit usually refers to "underfit". The data are too unpredictable.
Missing data	Data which are not responses to the items. They can be items which the examinees did not answer (usually score as "wrong") or items which were not administered to the examinee (usually ignored in the analysis).
Model	Mathematical conceptualization of a relationship
Muted	Overfit to the Rasch model. The data are too predictable. The opposite is underfit, excessive noise.
Newton-Raphson iteration	A general method for finding the solution of non-linear equations
Noise	1. Randomness in the data predicted by the Rasch model. 2. Underfit: excessive unpredictability in the data, perhaps due to excessive randomness or multidimensionality.
Normal	a random distribution, graphically represented as a "bell" curve which has a mean value of 0 and a standard deviation of 1.
Normalized	1. the transformation of the actual statistics obtained so that they are theoretically part of a unit-normal distribution. "Normalized" means "transformed into a unit-normal distribution". We do this so we can interpret the values as "unit-normal deviates", the x-values of the normal distribution. Important ones are ± 1.96 , the points on the x-axis for which 5% of the distribution is outside the points, and 95% of the distribution is between the points. 2. linearly adjusting the values so they sum to a predetermined amount. For instance, probabilities always sum to 1.0.
Not administered	an item which the person does not see. For instance, all the items in an item bank which are not part of a computer-adaptive test.
Object of Measurement	person, product, site, to be measured or positioned along the latent variable.
OBS Observed	Value derived from the data
Observation Observed Response	the actual response by an object to an agent.
Odds ratio	Ratio of two probabilities, e.g., "odds against" is the ratio of the probability of losing (or not happening) to the probability of winning (or happening).
Outfit	an outlier-sensitive fit statistic that picks up rare events that have occurred in an unexpected way. It is the average of the squared standardized deviations of the observed performance from the expected performance. Rasch plots and tables use the normalized unweighted mean squares so that the graphs are symmetrically centered on zero.
Outliers	unexpected responses usually produced by agents and objects far from one another in location along the latent variable.
Overfit	The data are too predictable. There is not enough randomness in the data. This may be caused by dependency or other constraints.

Perfect score	Every response "correct" or the maximum possible score. Every observed response in the highest category.
Person	the object of measurement, not necessarily human, e.g., a product.
Plot	an x-y graph used by Rasch to show the fit statistics for agents and objects.
Point Labels	the placing on plots of the identifier for each point next to the point as it is displayed.
Point-measure correlation PT-MEASURE, PTMEA	The correlation between the observations in the data and the measures of the items or persons producing them.
Poisson Counting	a method of scoring tests based on the number of occurrences or non-occurrences of an event, e.g. spelling mistakes in a piece of dictation.
Polarity	The direction of the responses on the latent variable. If higher responses correspond to more of the latent variable, then the polarity is positive. Otherwise the polarity is negative.
Polytomous response	responses in more than two ordered categories, such as Likert rating-scales.
Population	Every person (or every item) with the characteristics we are looking for. A sample of persons or items is usually assumed to be a random sample from the population.
Predictive validity	This is the amount of agreement between results obtained by the evaluated instrument and results obtained from more directly, e.g., the correlation between success level on a test of carpentry skill and success level making furniture for customers. "Do the person measures correspond to more and less of what we are looking for?"
Probabilistic	Predictable to some level of probability, not exactly. This contrasts with Deterministic.
Process	the psychological quality, i.e., the ability, skill, attitude, etc., being measured by an item.
PROX	the "Normal Approximation" estimation algorithm (Cohen, 1979). used to obtain initial estimates for the iterative estimation process.
Rack	Placing the responses to two tests in adjacent columns for each person, as though the items were being placed on a rack, c.f., stack.
Rasch, Georg	Danish Mathematician (1906-1980), who first propounded the application of the statistical approach used by Rasch.
Rasch measure	linear, additive value on an additive scale representing the latent variable
Rasch Model	a mathematical formula for the relationship between the probability of success (P) and the difference between an individual's ability (B) and an item's difficulty (D). $P = \exp(B-D) / (1 + \exp(B-D))$ or $\log [P / (1-P)] = B - D$
Rasch-Andrich Threshold	Step calibration. Location on the latent variable (relative to the center of the rating scale) where adjacent categories are equally probable.
Rating Scale	A format for observing responses wherein the categories increase in the level of the variable they define, and this increase is uniform for all agents of measurement.
Rating Scale Analysis	Wright, B.D. & Masters, G.N., Rating Scale Analysis: Rasch Measurement. Chicago: Mesa Press, 1982.
Raw score	the marginal score; the sum of the scored observations for a person, item or other element.
Reference group	The person classification-group which provides the baseline item difficulty in a differential-item-functioning investigation
Reliability	Reliability (reproducibility) = True Variance / Observed Variance (Spearman, 1904, etc.). It is the ratio of sample or test variance, corrected for estimation error, to the total variance observed.
Residuals	the difference between data observed and values expected.
Response	The value of an observation or data-point indicating the degree of success by an object (person) on an agent (item)
Response set	Choosing the same response on every item, such as always selecting option "C" on a multiple-choice test, or always selecting "Agree" on an attitude survey.

Results Table	a report of Rasch calculations.
Rigidity	when agents, objects and steps are all anchored, this is the logit inconsistency between the anchoring values, and is reported on the Iteration Screen and Results Table. 0 represents no inconsistency.
Row	a horizontal line of data on a Spreadsheet, usually used, in the Input Grid, to represent all responses by a particular object. The top row of each spreadsheet is reserved for Rasch control information.
Rule-of-thumb	A tentative suggestion that is not a requirement nor a scientific formula, but is based on experience and inference from similar situations. Originally, the use of the thumb as a unit of measurement.
Sample	the persons (or items) included in this analysis
Scale	the quantitative representation of a latent variable.
Scalogram	Picture of the data in which the persons (rows) and items (columns) are arranged by marginal raw scores.
Score points	the numerical values assigned to responses when summed to produce a score for an agent or object.
Scoring key	The list of correct responses to multiple-choice (MCQ) items.
Scree plot	Plot showing the fraction of total variance in the data in each variance component.
Separation	the ratio of sample or test standard deviation, corrected for estimation error, to the average estimation error. This is the number of statistically different levels of performance that can be distinguished in a normal distribution with the same "true" S.D. as the current sample. Separation = 2: high measures are statistically different from low measures.
Specification	A Winsteps control-variable and its value, e.g., "Name1=17"
Stack	Analyzing the responses of the same person to multiple administrations of the same test as though they were made by separate persons, by "stacking" the person records in one long data file, c.f., "rack"
Standard Deviation: P.SD, S.SD	The root mean square of the differences between the sample of values and their mean value. In Winsteps, all standard deviations are "population standard deviations" (the sample is the entire population) = P.SD. For the larger "sample standard deviation" (the sample is a random selection from the population) = S.SD, please multiply the Winsteps standard deviation by square-root (sample-size / (sample size - 1)).
Standard Error	An estimated quantity which, when added to and subtracted from a logit measure or calibration, gives the least distance required before a difference becomes meaningful.
Step calibration Step difficulty	Rasch-Andrich threshold. Location on the latent variable (relative to the center of the rating scale) where adjacent categories are equally probable.
Steps	the transitions between adjacent categories as ordered by the definition of the latent variable.
Strata	$= (4 * \text{Separation} + 1) / 3$ This is the number of statistically different levels of performance that can be distinguished in a normal distribution with the same "true" S.D. as the current sample, when the tails of the normal distribution are due to "true" measures, not measurement error. Strata=3: very high, middle, and very low measures can be statistically distinguished.
Sufficient statistic	A statistic (a number) which contains all the information in the data from which to estimate the value of a parameter.
Suffix	The letters added to a file name which specify the file format, e.g., ".txt" means "text file". If you do not see the suffix letters, instruct Windows to display them. See the Lesson 1 Appendix.
Table	Lists of words and numbers, arrange in columns, usually surrounded by " ".
Targeted	when the item difficulty is close to the person ability, so that the probability of success on a dichotomous item is near to 50%, or the expected rating is near to the center of

	the rating scale.
Targeting	Choosing items with difficulty equal to the person ability.
Task bar	This shows the Windows programs at the bottom of your computer screen
Template	a specially formatted input file.
Test length	The number of items in the test
Test reliability	The reliability (reproducibility) of the measure (or raw score) hierarchy of sample like this sample for this test. The reported reliability is an estimate of (true variance)/ (observed variance), as also are Cronbach Alpha and KR-20.
TOP	The value shown in the Results Table for an agent on which no objects were successful, (so it was of top difficulty), or for an object which succeeded on every agent (so it was of top ability)
Top Category	the response category at which maximum performance is manifested.
UCON	the unconditional (or "joint" JMLE) maximum likelihood estimation formula, used by some Rasch programs for the second part of the iteration process.
Underfit	The data are too unpredictable. The data underfit the model. This may be because of excessive guessing, or contradictory dimensions in the data.
UNSURE	Rasch was unable to calibrate this data and treated it as missing.
Unweighted	the situation in which all residuals are given equal significance in fit analysis, regardless of the amount of the information contained in them.
Variable	a quantity or quality which can change its value
Weighted	the adjustment of a residual for fit analysis, according to the amount of information contained in it.
Zero score	Every response "incorrect" or the minimum possible score. Every observed response in the lowest category.
ZSTD	Probability of a mean-square statistic expressed as a z-statistic, i.e., a unit-normal deviate. For $p \leq .05$ (double-sided), $ZSTD > 1.96 $.
&END	The end of the list of Winsteps control variables
&INST	The beginning of the list of Winsteps control variables. This is not necessary.

19.48 Guttman parameterization of rating scales

Rating scales with many categories often have unobserved categories or categories with very low frequencies. These make the standard estimation of Andrich thresholds (step difficulties), based on the relative frequencies of adjacent categories, subject to accidents in the data. Modeling the relationship between the thresholds with a continuous function smooths out the lumpiness in the data.

Louis Guttman (1941, etc.) proposes that responses to attitude-survey items can be summarized by the scale location and scale intensity. Guttman called these "principal components". Here named "Guttman components". The family of Guttman components was extended by Pender Pedler (1987), also Andrich and Luo (2003), as a set of orthogonal polynomials summarizing the Andrich thresholds.

The Orthogonal Polynomials

Pedler's (1987) orthogonal-polynomial Guttman components of the rating-scale thresholds are:

Polynomial where x is the observation	Formula: categories numbered 0..m	Minimum Categories / Thresholds	Interpretation
$T_0(x)$	0	1 / 0	rating-scale exists
$T_1(x)$	1	2 / 1	mean: central location of Andrich thresholds

$T_2(x)$	$2(x - (m+1)/2)$	$3/2$	dispersion, variance: linear relationship
$T_3(x)$	$3(x - (m+1)/2)^2 - (m^2 - 1)/4$	$4/3$	skewness: quadratic
$T_4(x)$	$4(x - (m+1)/2)^3 - (x - (m+1)/2)(3m^2 - 7)/5$	$5/4$	kurtosis: cubic
higher order: $T_{k+1}(x)$	$[(k+1)/k](x - (m+1)/2)T_k(x) - (((m^2 - (k-1)^2)(k^2 - 1))/[4(2k - 1)(2k-3)])T_{k-1}(x)$	$k+2/k+1$	higher moments

These are operationalized in Winsteps using [SFUNCTION=](#). For instance, SFUNCTION=3:
Andrich threshold between categories x-1 and x = $F(x) = p_1T_1(x) + p_2T_2(x) + p_3T_3(x)$ where p_1, p_2, p_3 are the data-dependent coefficients.

Estimating the Coefficients of the Polynomials

The initial values of the coefficients, p , are zero. In Winsteps, the central location of the Andrich thresholds of a rating scale is set at zero logits, so p_1 stays at zero. Initially all the $F(x)$ are also zero, so the thresholds are all at the central location.

The category probabilities are computed by applying the threshold values to the Andrich Rating-Scale Model, Partial Credit Model, etc. This produces:

1. an observed and expected frequency for each category
2. a modeled variance of the probability of each category

The likelihood of the data = $\Lambda = \text{Product}(\text{Probability}(x_{ni}))$ where x_{ni} are the observations.

Log-likelihood of the data = $\lambda = \sum (\text{Ln}(P_{x_{ni}})) = \sum (x_{ni}(B_n - D_i) - \sum_{j=1,x} F_j) - \sum (\text{Ln}(\text{probability normalizer}))$

$$\partial\lambda / \partial p_k = \sum \sum_{j=1,x} T_k(j) - \sum P_{ni} (\sum_{j=1,m} T_k(j))$$

$$\partial^2\lambda / \partial p_k^2 = \sum P_{ni} (\sum_{j=1,m} T_k(j))^2 - (\sum P_{ni} (\sum_{j=1,m} T_k(j)))^2$$

then, by Newton-Raphson,

$$\text{change in } P_k = \partial\lambda / \partial p_k / \partial^2\lambda / \partial p_k^2$$

Andrich, D. & Luo, G. (2003). Conditional Pairwise Estimation in the Rasch Model for Ordered Response Categories using Principal Components. *Journal of Applied Measurement*, 4(3), 205-221.

Guttman, L. (1950). The principal components of scale analysis. In S.A. Stouffer, L. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star and J.A. Clausen (Eds.), *Measurement and Prediction*, pp. 312-361. New York: Wiley.

Pedler, P.J. (1987) Accounting for psychometric dependence with a class of latent trait models. *Ph.D. dissertation*. University of Western Australia.

19.49 Half-rounding

Rounding occurs when a number, such as 5.032 is to be displayed with only two decimal places, e.g., as 5.03.

The general rule followed by Winsteps is to round to the nearest displayable value. Examples:

5.034 rounds to 5.03

5.036 rounds to 5.04

-5.034 rounds to -5.03

-5.036 rounds to -5.04

Rounding errors may arise with 5.035 and -5.035. Winsteps intends to round these away from zero to 5.04 and -5.04.

In practice, the computer arithmetic sometimes loses precision due to hardware limitations, so that 5.035 becomes an internal number like 5.034999997 - a number which is the same computationally as 5.035. But this value half-rounds to 5.03. This behavior is impossible to predict, because most displayed numbers are the result of a long chain of internal computations, .

Increasing the value of UDECIM= may display more decimal places, and so better display the numerical results.

In recent versions, Winsteps adds .0000005 to positive numbers, and subtracts that amount from negative numbers, in order that .005 will almost always round to .01.

19.50 How big an analysis can I do?

Winsteps will operate with a minimum of two observations per item or person, or even 1 observation with anchoring. This is useful for producing [score tables](#) with anchored items and dummy response strings. For statistically stable measures to be estimated, 30 observations per element (item or person) are needed: www.rasch.org/rmt/rmt74m.htm

The maximum number of reportable persons is 9,999,999 persons. The maximum number of items (NI=) is 65,000 items. The maximum length of an item or person label is 300 ASCII characters. The maximum printable line is 3,003 characters.

There can be up to 32,768 ordinally numbered categories per item (polytomies, rating scales, partial credit etc.).

Subset evaluation: up to 50,000 potential subsets, depending on available computer memory.

The widest logit range of measures that maintains computational precision is 90 logits, but useful results can be reported with a measure range up to 700 logits wide.

In order to maintain high computational precision, while maintaining computational speed, Winsteps implements West, D.H.D. (1979), "Updating Mean and Variance Estimates An Improved Method," Communications of the Association for Computing Machinery, 22, 532-535.

Example: a dataset of almost 500,000 persons and 20,000 items was analyzed with Winsteps. The data were 99% missing. This type of analysis can take several hours, depending on your computer hardware.

19.51 How long will an analysis take?

Most analyses have reached convergence within 30 iterations. A tentative guideline is:

$$\text{length of analysis in minutes} = (\text{number of persons}) * (\text{length of test}) / 1,000,000$$

An analysis of 7000 persons on 2700 items using a 2.5GHz PC with 2GB RAM required 10 minutes.

For a faster analysis,

at least 2GB of RAM memory. 4GB is better, see Memory.

specify [PRCOMP=NO](#) if you don't want to produce [ICORFILE=](#), [PCORFILE=](#), [Table 23](#) or [Table 24](#)

specify [DISTR=NO](#) if you don't want [distractor](#)/option sub-tables

specify [SUBSETS=NO](#) if you don't need to check for disjoint subsets in the data

loosen the convergence criteria: [CONVERGE=](#), [LCONV=](#), [RCONV=](#), or reduce the maximum number of iterations [MJMLE=](#)

put the input files and the work files on different disk drives, see [Edit Initial Settings](#).

19.52 Inestimable measures

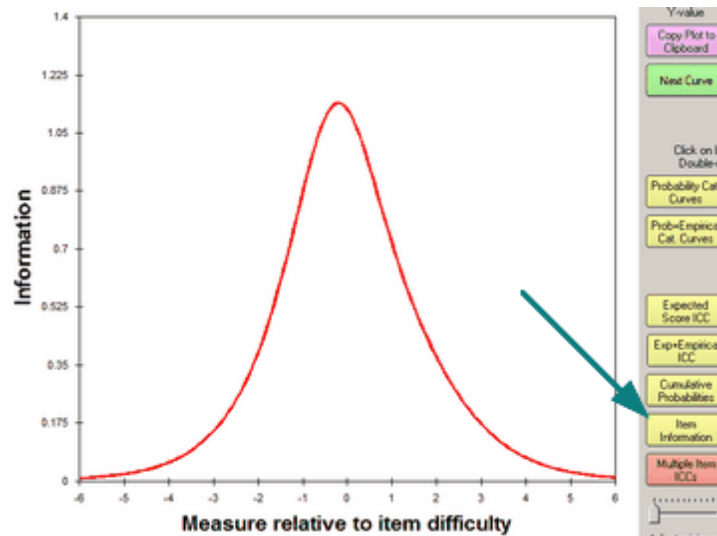
Winsteps may not be able to provide an estimate of a measure. INESTIMABLE is reported if all observations are eliminated as forming part of extreme response strings. To make such measures estimable, further data (real or artificial) or Winsteps specifications are required.

Example: [ISGROUPS=0](#) is specified (partial credit model) but all responses are in the same category. Winsteps cannot construct a rating scale structure. If possible, include this item in the same ISGROUPS= grouping as an estimable item.

19.53 Information - item and test

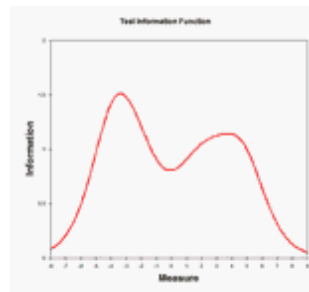
Fisher information is the amount of information data provide about a parameter. For item information, this is the amount of information the response to an item provides about a person parameter. For test information, this is the amount of information all of the items encountered by a person provide about a person parameter.

All dichotomous items have the same Rasch-derived item information. For a response of probability p , the item information is the binomial variance of the probability, $p(1-p)$. This function is:



For polytomous items, the information function depends on the rating scale structure. The better targeted the item is on the person the more Fisher information the item provides about the person parameter. This motivates some item selection techniques for computer-adaptive testing. An algorithm for the Fisher item information function for polytomous items is given at www.rasch.org/rmt/rmt191a.htm.

The test information function is the sum of the item information functions. The standard error of a measure is the inverse square-root of the test information at the location of the maximum-likelihood parameter estimate (or higher). Here is the test information function for the Knox Cube Test in [Example 1](#).



Test (sample) [reliability](#) is a summary of the interaction between the sample distribution and the test information function.

19.54 Installing multiple versions of Winsteps

If you need to install several versions of Winsteps on the same computer, install them in different folders. This is enough for them to work independently.

19.55 IRT to Rasch conversion

The 3-PL b values are roughly equivalent to the Rasch difficulties. The b difficulties are probably in probits, which could be converted to logits by multiplying them by 1.702. However, empirically, the conversion factor differs.

Do you have any of the 3-PL data? If so, use the 3-PL parameters to estimate the 3-PL thetas for those data. Then analyze the same data with Rasch and obtain the Rasch person measures (Rasch thetas). Cross-plot the two sets of thetas. The slope of the best-fit line gives the conversion between the empirical 3-PL scaling and the Rasch logit scaling.

19.56 Item difficulty: definition

As modeled in Winsteps, the difficulty (challenge, easiness, etc.) of an item (task, prompt, etc.) is the point on the latent variable (unidimensional continuum) at which the highest and lowest categories have equal probability of being observed. This is usually near the center of the middle category of an odd number of categories, or close to the transition between adjacent central categories of an even number of categories.

For a dichotomous item, this is the point at which each category has a 50% probability of being observed.

For a Rasch-Andrich rating-scale item, this definition implies that the sum of the rating-scale-structure measures sum to zero relative to the item difficulty, i.e., the sum of the Rasch-Andrich thresholds is zero, i.e., $\sum(F_j) = 0$.

For a Masters partial-credit item, this definition implies that the item difficulty is the average of the difficulties of the Rasch-Masters thresholds for the item, i.e., $D_i = \text{average of the delta } (D_{ij}) \text{ values}$, so that re parameterizing, $D_{ij} = D_i + F_j$, then $\sum(F_{ij}) = 0$ for each item i .

19.57 Item difficulties from score-to-measure table

```
' This is Visual Basic Code

' This is Visual Basic Code

Option Explicit
Public Sub main()

' solving non-linear equations by targeted trial-and-error

' Example using Knox Cube Test data
' in this example, arrays start at 0

Dim maxi&, maxn&
maxi& = 14 ' 14 items
maxn& = 14 ' scores are in range 0 to 14

ReDim r#(maxn&) ' scores
ReDim b#(maxn&) ' measures for scores
ReDim d#(maxi&) ' target items

' list of scores
r(0) = 0.25 ' this was the extreme score adjustment
r(1) = 1
r(2) = 2
r(3) = 3
r(4) = 4
r(5) = 5
r(6) = 6
r(7) = 7
r(8) = 8
r(9) = 9
r(10) = 10
r(11) = 11
r(12) = 12
r(13) = 13
r(14) = 13.75

' list of measures
b(0) = -6.66
b(1) = -5.3
b(2) = -4.35
b(3) = -3.64
b(4) = -2.97
```

```

b(5) = -2.26
b(6) = -1.39
b(7) = -0.26
b(8) = 0.94
b(9) = 1.96
b(10) = 2.88
b(11) = 3.76
b(12) = 4.65
b(13) = 5.73
b(14) = 7.15

' size of estimation adjustment
Dim diff#
diff# = 0.1 ' too big or too small and this process fails
          ' tweaking this value may give slightly better, or hugely worse, results

ReDim scoren#(maxn) ' expected scores based on current set of item difficulties

Dim oldsqrdiff# ' sum of squared differences from previous iteration
Dim newsqrdiff# ' sum of squared differences for this iteration

Dim sumn#(2) ' 0,1,2 = 3 squared differences for tweaking current item
Dim dhold# ' holding current item

Dim i&, n&, k&, j& ' indexes into arrays

oldsqrdiff# = 9999 ' big starting value
newsqrdiff = 9998 ' big, but slightly smaller starting value

While newsqrdiff < oldsqrdiff ' continue iterating while squared difference reduces
  oldsqrdiff = newsqrdiff ' save current difference
  newsqrdiff = 0 ' reset new difference
  For i = 1 To maxi& ' loop down the items
    dhold = d(i) ' hold the current item
    For k = 0 To 2 ' try 3 different item values
      d(i) = dhold + (k - 1) * diff#
      sumn(k) = 0 ' squared difference for this target item value
      For n = 0 To maxn& ' loop through all the scores
        scoren(n) = 0 ' expected score
        For j = 1 To maxi ' loop through all the items
          scoren(n) = scoren(n) + 1# / (1# + Exp(d(j) - b(n))) ' sum the expected scores
        Next j
        sumn(k) = sumn(k) + (scoren(n) - r(n)) ^ 2 ' accumulate squared differences between
          ' expected and observed scores
      Next n ' next score
    Next k ' next target item value
    If sumn(0) < sumn(1) And sumn(0) < sumn(2) Then ' is target value k=0 the lowest?
      d(i) = dhold - diff# ' yes: revise target item value
      newsqrdiff = sumn(0) ' save differences in case this is the last score
    ElseIf sumn(2) < sumn(1) And sumn(2) < sumn(0) Then ' is target value k=2 the lowest?
      d(i) = dhold + diff#
      newsqrdiff = sumn(2)
    Else ' no change - k=1, the original d(i) is good
      d(i) = dhold
      newsqrdiff = sumn(1)
    End If
  Next i ' next target item
  Debug.Print oldsqrdiff, newsqrdiff ' display squared differences for confirmation
Wend

For n = 0 To maxn
  Debug.Print n, r(n), scoren(n), r(n) - scoren(n) ' list of scores and differences
Next n

```

```

For i = 1 To maxi
  Debug.Print i, d(i) ' list of item difficulties
Next i

```

```
Stop
```

```
End Sub
```

```
Iteration report for KCT data:
```

```

Previous iteration  This iteration
9998                83.3772167769053
83.3772167769053   82.8586294323071
.....
7.18316390024933E-03  7.08576670997411E-03
7.08576670997411E-03  7.08576670997411E-03

```

```
List of observed scores, expected scores, differences
```

Observed	Expected	Differences
0	0.25	0.29 -0.04
1	1 0.99	0.01
2	2 2.00	0.00
3	3 3.02	-0.02
4	4 4.03	-0.03
5	5 5.00	0.00
6	6 5.97	0.03
7	7 6.99	0.01
8	8 8.01	-0.01
9	9 8.99	0.01
10	10 9.98	0.02
11	11 10.99	0.01
12	12 11.98	0.02
13	13 12.99	0.01
14	13.75	13.69 0.06

```
List of item difficulties (item order does not matter)
```

Item	Estimate	Actual
10	5.0	4.80
12	4.9	4.80
14	4.5	4.80
8	3.0	3.37
6	2.9	2.24
4	1.7	1.95
2	0.6	.79
9	-1.1	-1.57
3	-2.9	-2.35
7	-3.1	-3.38
13	-3.5	-3.38
1	-3.8	-3.83
5	-4.1	-3.83
11	-4.1	-4.40

19.58 Item discrimination or slope estimation

In Winsteps, item discrimination is not a parameter. It is merely a descriptive statistic.

Thoughts about item discrimination: Item discrimination and Rasch fit statistics are asymmetric. Noisy, unpredictable misfit (high mean-squares, large positive t-values, very low discrimination) degrade measurement. But too predictable overfit (low mean-squares, large negative, t-statistics, very high discrimination) do not degrade measurement. They merely cause standard errors to be estimated too low and reliabilities to be estimated too high - this is equivalent to the well-known "attenuation paradox" of CTT. So the useful range of discrimination is something like the range 0.5 to 2 or it could be 0.5 to infinity!! But when discrimination becomes very high (mean-squares very low), that usually means the items are malfunctioning in some way. Here is an informative paper on the topic of high item discrimination:

Journal of Educational Measurement, 25, 1, 15, 1988. "Item Discrimination: When More Is Worse" by Geofferey N. Masters.

High item discrimination can be a symptom of a special kind of measurement disturbance introduced by an item that gives persons of high ability a special advantage over and above their higher abilities. This type of disturbance, which can be interpreted as a form of item "bias," can be encouraged by methods that routinely interpret highly discriminating items as the "best" items on a test and may be compounded by procedures that weight items by their discrimination. The type of measurement disturbance described and illustrated in this paper occurs when an item is sensitive to individual differences on a second, undesired dimension that is positively correlated with the variable intended to be measured. Possible secondary influences of this type include opportunity to learn, opportunity to answer, and test-wiseness.

The Rasch model specifies that item discrimination, also called the item slope, be uniform across items. This supports additivity and construct stability. Winsteps estimates what the item discrimination parameter would have been if it had been parameterized. The Rasch slope is the average discrimination of all the items. It is not the mean of the individual slopes because discrimination parameters are non-linear. Mathematically, the average slope is set at 1.0 when the Rasch model is formulated in logits, or 1.70 when it is formulated in probits (as 2-PL and 3-PL usually are). 0.59 is the conversion from logits to probits.

The empirical discrimination is computed after first computing and anchoring the Rasch measures. In a post-hoc analysis, a discrimination parameter, a_i , is estimated for each item. The estimation model is of the form:

$$\log\left(\frac{P_{nij}}{P_{ni(j-1)}}\right) \equiv a_i(\hat{B}_n - \hat{D}_i - \hat{F}_j)$$

This has the appearance of a 2-PL IRT or "Generalized Partial Credit" model, but differs because the discrimination or slope parameter is not used in the estimation of the other parameters. The reported values of item discrimination, DISCR, are a first approximation to the precise value of a_i obtained from the Newton-Raphson estimation equation:

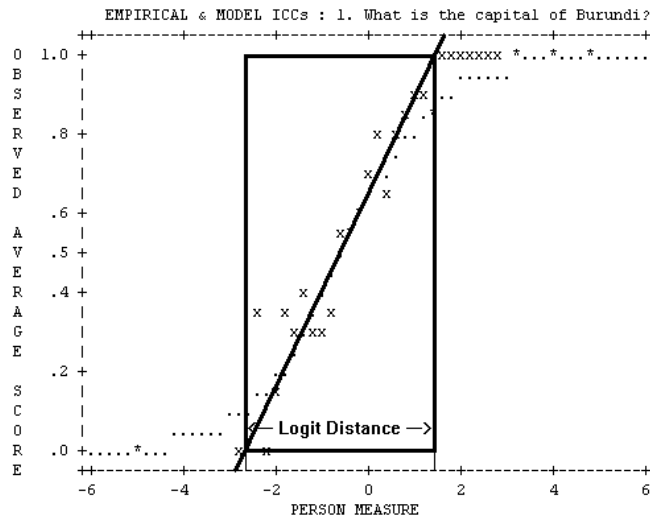
$$\hat{a}_i = 1 + \left[\frac{\sum_n (X_{ni} - P_{ni})(\theta_n - b_i)}{\sum_n P_{ni}(1 - P_{ni})(\theta_n - b_i)^2} \right]$$

The possible range of a_i is $-\infty$ to $+\infty$, where $+\infty$ corresponds to a Guttman data pattern (perfect discrimination) and $-\infty$ to a reversed Guttman pattern. Rasch estimation usually forces the average item discrimination to be near 1.0. Consequently an estimated discrimination of 1.0 accords with Rasch model expectations. Values greater than 1.0 indicate over-discrimination, and values less than 1.0 indicate under-discrimination. Over-discrimination is thought to be beneficial in many raw-score and IRT item analyses. High discrimination usually corresponds to low MNSQ values, and low discrimination with high MNSQ values.

From an informal simulation study, Edward Wolfe reports Winsteps discrimination to have a .88 correlation with the generating slope parameters for a 2-PL dataset. BILOG has a .95 correlation.

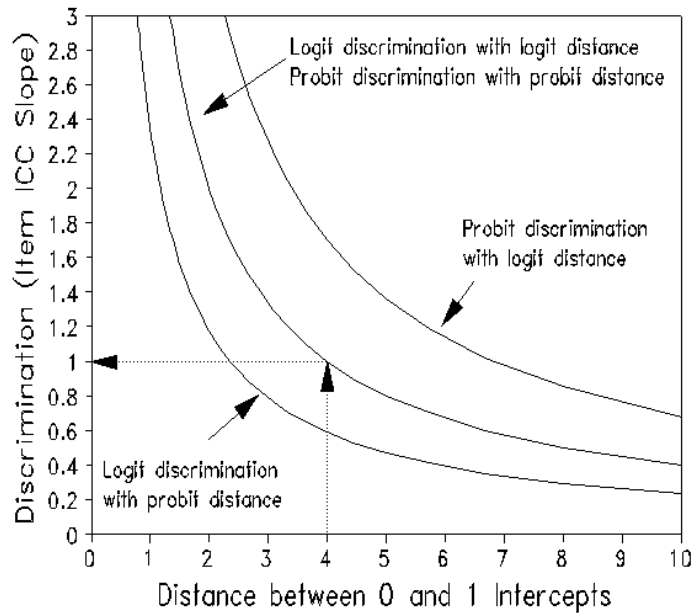
Table 29.1 allows you to estimate the empirical item discrimination, at least as well as a 2-PL IRT computer program. This is because 2-PL discrimination estimation is degraded by the imputation of a person distribution and constraints on discrimination values. It is also skewed by accidental outliers which your eye can disregard. When [Discrimination=Yes](#), exact computation is done in the measure tables.

In Table 29.1 draw in the line that, to your eye, matches the central slope of the empirical item characteristic curve (ICC).



Estimate the logit distance from where the line intercepts the .0 score value to where it intercepts the 1.0 score value (for dichotomies). The logit distance here is about 4.0 logits.

Use the central logit measure to logit discrimination line in this nomogram to estimate discrimination. In this nomogram, a logit distance of 4.0 logits, corresponds to a logit discrimination of 1.0, in accordance with model prediction. Steeper slopes, i.e., higher discriminations, correspond to shorter distances.



19.59 Item map in Excel or Word

See [Wright map](#)

19.60 Item Maps - Combining

Procedure for Combining Item maps

1. Run the separate analyses. You may need to adjust [UPMEAN=](#) to include extreme person measures.
2. Output an [IAFILE=](#) from each analysis to Excel.

3. Copy all the separate [IAFILE=](#) into one Excel worksheet with all the measures in the same column
4. Number the items sequentially in column A. The item measures are in column B
5. In a new text file


```
TITLE = Combined item map
ITEM1 = 1
NI = (total number of items)
NAME1=1 (it will not be used)
NAMELENGTH=1
CODES=01
IAFILE=*
(paste column A and B from the combined Excel worksheet)
*
&END
(list of all the item labels from the combined Excel worksheet)
END LABELS
1010101010 ... (for NI= columns)
0101010101 ... (for NI= columns)
```
6. Analyze this dummy data file and produce [Table 12](#)

19.61 Iterations - PROX & JMLE

The Rasch model formulates a non-linear relationship between non-linear raw scores and additive measures. So, estimating measures from scores requires a non-linear process. This is performed by means of iteration. Two [estimation methods](#) are used, PROX and JMLE.

The fundamental transformation in linearizing raw scores is:

$\log \left(\frac{\text{observed raw score} - \text{minimum possible raw score}}{\text{maximum possible raw score} - \text{observed raw score}} \right)$

In Winsteps, initially every person is estimated to have the same ability measure at the origin of the measurement scale. Each item is estimated to have the same difficulty measure, also at the origin of the measurement scale. Each rating scale structure parameter, Rasch-Andrich threshold, is also estimated to be 0.

In Winsteps, the first phase of estimation uses the PROX (normal approximation) estimation algorithm. This takes the initial set of estimates and produces revised estimates:

$$B_n = \mu_n + \sqrt{1 + \sigma_n^2 / 2.9} \log_e (R_n / (N_n - R_n))$$

where B_n is the revised ability estimate for person n , μ_n is the mean difficulty of the items encountered by person n , and σ_n is the standard deviation of those item difficulties. R_n is the observed raw score for person n and N_n is a perfect, maximum possible score on those same items. Similarly, for the items,

$$D_i = \mu_i - \sqrt{1 + \sigma_i^2 / 2.9} \log_e (R_i / (N_i - R_i))$$

where D_i is the revised difficulty estimate for item i , μ_i is the mean ability of the persons encountering by item i , and σ_i is the standard deviation of those person abilities. R_i is the observed raw score on item i and N_i is a perfect score by those same persons.

To update these PROX estimates, Winsteps traverses the data computing the values of all the terms on the right-side of the estimation equations. This traversal is called an "iteration". When the increase in the range of the person or item measures is smaller than 0.5 logits, or when [MPROX=](#) is reached, iteration ceases.

Initial estimates of the Rasch-Andrich threshold between category k and category $k-1$ are obtained from \log (observed frequency of category $k-1$ / observed frequency of category k) normalized to sum to zero across the thresholds of a rating scale.

The PROX estimates become the starting values for JMLE (Joint Maximum Likelihood Estimation). Using these person, item and rating scale structure estimates, Winsteps computes the expected value, according to the Rasch model, corresponding to each observation in term. After iterating through the entire data set, the marginal sums of these expected values, the person expected raw scores and the item expected raw scores, are compared with their observed (empirical) values. If a person's expected raw score is less than that person's observed raw score, then the ability estimate raised. If the person's expected raw score is greater than the observed score, then the ability estimate is lowered. For items, if the expected raw score is less than the observed score, then the difficulty estimate is lowered. If the item's expected raw score is greater than the observed score, then the difficulty estimate is raised.

The estimation equations for JMLE are derived in [RSA](#), where Newton-Raphson iteration is employed.

$$y' = y + (\text{observed score} - \text{Rasch expected score based on current estimates}) / (\text{modeled variance})$$

where y = a current estimated person measure and y' is the improved estimate.

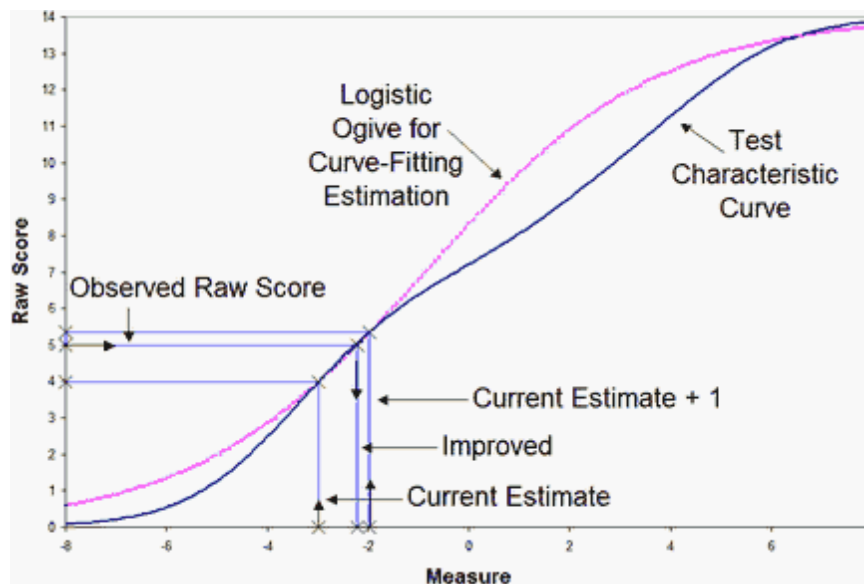
To see this process in an Excel spreadsheet, go to www.rasch.org/moulton.htm

Newton-Raphson estimation has proved unstable with sparse data sets and also with rating scales which have alternating very high and very low frequency categories. Accordingly, Winsteps implements a more robust proportional-curve-fitting algorithm to produce JMLE estimates. The relationship between raw scores and measures is always monotonic, so the characteristic curve for each person or item parameter is modeled to have the local form of a logistic ogive:

$$y = a * \log((x-l)/(h-x)) + c$$

where y = an estimated measure, a = slope of the ogive, x = a raw score, l = the known minimum possible raw score for the parameter, h = the known maximum possible raw score for the parameter, c = location of ogive relative to local origin.

Values of x are obtained from the current estimated measure y and a nearby measure ($y + d$). From these, a and c are estimated. The revised measure y' is obtained by evaluating the equation using the observed raw score as the value of x . In the plot below for [Example 1](#), the current estimate, y , is -3 logits, a nearby estimate, $y+d$, is -2 logits. These both estimate raw scores on the currently-estimated test characteristic curve (TCC, the remainder of which is not yet known). The violet line is the logistic ogive going through these two known points. It is close to the putative TCC. The observed score of "5" is then found on the logistic ogive and an improved estimate is obtained. After all the person and item estimates are improved, the estimated TCC changes and this estimation process is repeated by performing another iteration through the data. Winsteps sets d equal to the biggest logit change in any item or person estimate in the previous iteration. This rapidly becomes much less than 1.0.



For the rating scale structure, the estimate, y_k , for Rasch-Andrich threshold k is improved by $y_k' = y_k - \log(\text{observed count category } k / \text{observed count category } k-1) + \log(\text{estimated count category } k / \text{estimated count category } k-1)$

When the various [convergence](#) criteria are satisfied, iteration ceases and the final estimates are obtained. These are used in computing fit statistics.

Example: Here is the iteration Table for [example0.txt](#):

CONVERGENCE TABLE

-----+-----								
PROX		ACTIVE COUNT			EXTREME 5 RANGE		MAX LOGIT CHANGE	
ITERATION	KIDS	ACTS	CATS	KIDS	ACTS	MEASURES	STRUCTURE	
-----+-----								
1	75	25	3	3.78	3.20	3.8918	.0740	
2	74	25	3	4.59	3.71	.8258	-.6158	
3	74	25	3	4.83	3.92	.2511	-.1074	
-----+-----								
JMLE		MAX SCORE	MAX LOGIT	LEAST CONVERGED			CATEGORY STRUCTURE	
ITERATION	RESIDUAL*	CHANGE	KID	ACT	CAT	RESIDUAL	CHANGE	
-----+-----								
1	2.84	-.1955	60	22*	2	21.44	.0076	
2	.71	-.0335	53	15*	0	-5.89	.0144	
3	-.43	.0293	53	5*	1	3.48	.0101	
4	.32	.0235	18	11*	1	2.71	.0079	
5	.24	.0184	18	11*	0	-2.09	.0060	
6	.19	.0141	18	11*	0	-1.63	.0045	
7	.14	.0108	18	11*	0	-1.25	.0035	
8	.11	.0082	18	11*	0	-.96	.0026	
9	.08	.0062	18	11*	0	-.73	.0020	
10	.06	.0048	18	11*	0	-.56	.0015	
-----+-----								

In the top section of the Convergence Table are reported the number of active persons, items and categories. The range of item and person measures at the end of each PROX iteration is shown, also the biggest change in any person or item, and in any Rasch-Andrich threshold. PROX iteration ceases with iteration 3 because the "KIDS" (persons) and "ACTS" (items) range has increased by less than 0.5 logits.

In the lower section, for each JMLE iteration, the maximum score residual, the biggest difference between any observed and expected marginal score is shown. Also the biggest change in any measure. Iteration ceases when the values, in iteration 10, are less than the [convergence](#) criteria.

Newton-Raphson (quadratic estimation)	$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$
Generalized Newton-Raphson (cubic estimation)	$x_{k+1} = x_k - \frac{f'(x_k) - \sqrt{f'^2(x_k) - 2f(x_k)f''(x_k)}}{f''(x_k)}$
HouseHölder's method (cubic estimation)	$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} - \frac{f^2(x_k)f''(x_k)}{2f'^3(x_k)}$

W. Nazeer et al. (2016) "[Generalized Newton Raphson's method free from second derivative](#)." J. Nonlinear Sci. Appl. 9 (2016), 2823–2831.

19.62 JMLE is inconsistent!

Reviewers may reject Winsteps and its estimation method, JMLE, because they are "inconsistent", quoting perhaps Christensen (2012) "Estimation of the item parameters using the joint likelihood function leads to inconsistent item parameter estimates because the number of parameters increases with the number of persons", and recommending CMLE or MMLE

Response: It is true that JMLE (UMLE, UCON) is statistically inconsistent for infinite data. For finite data this is seen as "estimation bias". During many discussions/arguments in the 1970s and 1980s it was discovered that the JMLE estimation bias is inconsequential for most datasets and can easily be corrected where it is consequential. See, for instance, www.rasch.org/memo45.htm which appeared in Applied Psychological Measurement 12 (3) pp. 315-318, September 1988.

Of course, MMLE is estimation-biased if the person distribution mismatches the assumed person theta distribution (which it always does, empirical data never matches a theoretical distribution, - as the reviewer implicitly admits). CMLE, as usually implemented, is estimation-biased for the person measures (thetas) - see my note in Rasch Measurement Transactions - www.rasch.org/rmt/rmt331.pdf "CMLE - a Problem, its Solution and a Useful Approximation". Further depending on the nature of your data and analysis, CMLE and MMLE may be impossible to implement.

In my experience over 40 years JMLE estimation bias is only consequential for pairwise data, so Winsteps has a special command for this unique situation: "PAIREDdata=Yes". In other situations, I recommend against bias correction, STBIAScorrection=Yes in Winsteps, because of its side-effects. For instance, only for uncorrected JMLE can we complete this loop: *raw person and item scores with original data -> item estimates and person estimates -> original person and item raw scores*. This is important if we intend to predict person raw scores and item p-values for future data from the current estimates.

19.63 Local Dependence

In some data designs, data are collected from the same persons more than once, or observations are collected on equivalent items. Consequently there is reason to suspect that local dependence exists in the data. What is its impact on the Rasch measures?

Local dependence usually squeezes or stretches the logit measures, but does not usually change cut-points much when they are expressed in raw-score terms.

Procedure A. If a subset of items may have strong local dependence, then rescore the subset as one partial-credit item.

- 1) Analyze all the data with subset of items as separate items. Output the person measures to Excel. PFILE=
- 2) Combine the scores on the subset of items into one partial credit item. All the other items remain unchanged. Analyze all the data. Output the person measures to Excel. PFILE=
- 3) In Excel, cross-plot the person measures from (1) and (2). The curvature of the plot shows the influence of the local dependence on linearity.
- 4) Is the curvature big enough to be important to your audience? If yes, use the person measures from (3). If no, use the person measures from (2). My guess is usually "No".

Procedure B. To avoid local dependence in measuring change from pre-test to post-test:

1. Create a random, "stacked" dataset, in which each patient only appears once, either at pre-test or post-test. For instance, see [FORMAT=](#) example 7.
2. Run an initial Winsteps analysis and create item and step anchor files: [IFILE=if.txt](#), [SFILE=sf.txt](#)
3. Run Winsteps on the full pre-test data set using the [IAFILE=if.txt](#) and [SAFILE=sf.txt](#) from step #2
4. Run Winsteps on the full post-test data set using the [IAFILE=if.txt](#) and [SAFILE=sf.txt](#) from step #2
5. You can determine the effect of local dependence by cross-plotting the person measures from 3) and 4) against the measures from an unanchored stacked analysis of all the data.

Marais I. Response dependence and the measurement of change. J Appl Meas. 2009;10(1):17-29

Procedure C. Here is an experiment to determine whether local dependence is a problem. Assuming that data from the same persons may be a problem, select from your cases one of each different response string. This will make the data as heterogeneous as possible. Perform an analysis of this data set and see if that changes your conclusions markedly. If it does, then local dependence may be a concern. If it doesn't then local dependence is having no substantive impact.

Using Excel, a method of obtaining only one of each different response string:

0. Import the data into excel as a "character" column
1. from the Excel data pull down menu choose -> filter -> advanced filter
2. under "action" choose "copy to another location"
3. click "list range" and highlight the range of element numbers - if you want the whole column click on the letter at the top of the column
4. click "copy to" and choose an empty column, e.g., column J.
5. click "unique records only"
6. click "OK"
7. look at column J. The data are unique.

19.64 Logit and probit

When USCALE=1 (or USCALE= is omitted), measures are reported in logits. When USCALE=0.59, measures are reported in approximated probits.

Logit: A logit (log-odds unit, pronounced "low-jit") is a unit of additive measurement which is well-defined within the context of a single homogeneous test. When logit measures are compared between tests, their probabilistic meaning is maintained but their substantive meanings may differ. This is often the case when two tests of the same construct contain items of different types. Consequently, logit measures underlying different tests must be equated before the measures can be meaningfully compared. This situation is parallel to that in Physics when some temperatures are measured in degrees Fahrenheit, some in Celsius, and others in Kelvin.

As a first step in the equating process, plot the pairs of measures obtained for the same elements (e.g., persons) from the two tests. You can use this plot to make a quick estimate of the nature of the relationship between the two logit measurement frameworks. If the relationship is not close to linear, the two tests may not be measuring the same thing.

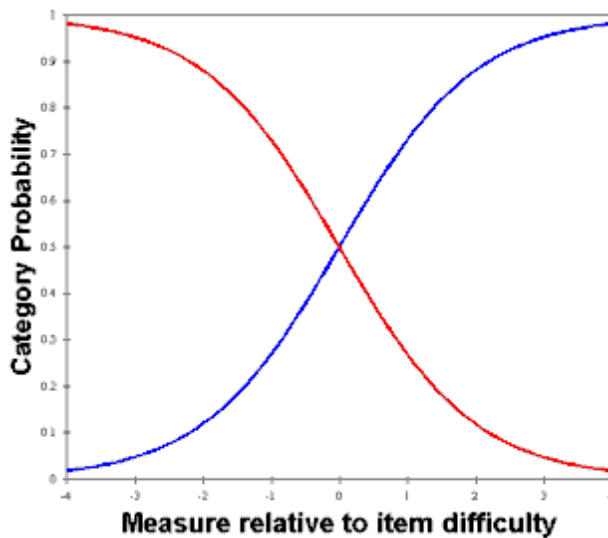
Logarithms: In Rasch measurement all logarithms, "log", are "natural" or "Napierian", sometime abbreviated elsewhere as "ln". "Logarithms to the base 10" are written log10. Logits to the base 10 are called "lods". 1 Lod = 0.4343 * 1 Logit. 1 Logit = 2.3026 * 1 Lod.

Logit-to-Probability Conversion Table

Logit difference between ability measure and item calibration and corresponding probability of success on a dichotomous item is shown in the table. A rough approximation between -2 and +2 logits is:

$$\text{Probability\%} = (\text{logit ability} - \text{logit difficulty}) * 20 + 50$$

Logit Difference	Probability of Success	Probit Difference		Logit Difference	Probability of Success	Probit Difference
5.0	99%	2.48		-5.0	1%	-2.48
4.6	99%	2.33		-4.6	1%	-2.33
4.0	98%	2.10		-4.0	2%	-2.10
3.0	95%	1.67		-3.0	5%	-1.67
2.2	90%	1.28		-2.2	10%	-1.28
2.0	88%	1.18		-2.0	12%	-1.18
1.4	80%	0.85		-1.4	20%	-0.85
1.1	75%	0.67		-1.1	25%	-0.67
1.0	73%	0.62		-1.0	27%	-0.62
0.8	70%	0.50		-0.8	30%	-0.50
0.69	67%	0.43		-0.69	33%	-0.43
0.5	62%	0.31		-0.5	38%	-0.31
0.4	60%	0.25		-0.4	40%	-0.25
0.2	55%	0.13		-0.2	45%	-0.13
0.1	52%	0.06		-0.1	48%	-0.06
0.0	50%	0.00		-0.0	50%	-0.00



Example with dichotomous data:

In Table 1, it is the distance between each person and each item which determines the probability.

<more>	MEASURE		MEASURE	<rare>
-----	PERSONS	+-	ITEMS	-----
5.0		+		5.0
			XXX (items difficult for persons) <- 4.5 logits	
4.0		+		4.0
		XX	<- 3.7 logits	
3.0		+		3.0

The two persons are at 3.7 logits. The three items are at 4.5 logits. The difference is 3.7 - 4.5 = -0.8 logits. From the logit table above, this is predicted as 30% probability of success for persons like these on items like these. 30% probability of success on a 0-1 item means that the expected score on the item is 30%/100% score-points = 0.3 score-points.

Inference with Logits

Logit distances such as 1.4 logits are applicable to individual dichotomous items. 1.4 logits is the distance between 50% success and 80% success on a dichotomous items.

Logit distances also describe the relative performance on adjacent categories of a rating scale, e.g, if in a Likert Scale, "Agree" and "Strongly Agree" are equally likely to be observed at a point on the latent variable, then 1.4 logits higher, "Strongly Agree" is likely to be observed 8 times, for every 2 times that "Agree" is observed.

For sets of dichotomous items, or performance on a rating scale item considered as a whole, the direct interpretation of logits no longer applies. The mathematics of a probabilistic interpretation under these circumstances is complex and rarely worth the effort to perform. Under these conditions, logits are usually only of mathematical value for the computation of fit statistics - if you wish to compute your own.

Different tests usually have different probabilistic structures, so that interpretation of logits across tests are not the same as interpretation of logits within tests. This is why test equating is necessary.

Logits and Probits

Logits are the "natural" unit for the logistic ogive. Probits are the "natural" units for the unit normal cumulative distribution function, the "normal" ogive. Many statisticians are more familiar with the normal ogive, and prefer to work in probits. The normal ogive and the logistic ogive are similar, and a conversion of 1.7 approximately aligns them.

When the measurement units are probits, the dichotomous Rasch model is written:
 $\log (P / (1-P)) = 1.7 * (B - D)$

To have the measures reported in probits, set USCALE = 0.59 = 1/1.7 = 0.588

But if the desired probit-unit is based on the empirical [standard deviation](#), then set USCALE = 1 / (logit S.D.)

Some History

Around 1940, researchers focused on the "normal ogive model". This was an IRT model, computed on the basis that the person sample has a unit normal distribution $N(0,1)$.

The "normal ogive" model is: Probit (P) = theta - Di
where theta is a distribution, not an individual person.

But the normal ogive is difficult to compute. So they approximated the normal ogive (in probit units) with the much simpler-to-compute logistic ogive (in logit units). The approximate relationship is: logit = 1.7 probit.

IRT philosophy is still based on the $N(0,1)$ sample distribution, and so a 1-PL IRT model is:

$$\log(P/(1-P)) = 1.7 (\text{theta} - D_i)$$

where theta represents a sample distribution. Di is the "one parameter".

The Rasch model takes a different approach. It does not assume any particular sample or item distribution. It uses the logistic ogive because of its mathematical properties, not because of its similarity to the cumulative normal ogive.

The Rasch model parameterizes each person individually, B_n . As a reference point it does not use the person mean (norm referencing). Instead it conventionally uses the item mean (criterion referencing). In the Rasch model there is no imputation of a normal distribution to the sample, so probits are not considered.

The Rasch model is: $\log(P/(1-P)) = B_n - D_i$

Much IRT literature asserts that "1-PL model = Rasch model". This is misleading. The mathematical equations can look similar, but their motivation is entirely different.

If you want to approximate the "normal ogive IRT model" with Rasch software, then

- (a) adjust the person measures so the person mean = 0: UPMEAN=0
- (b) adjust the user-scaling: probits = logits/1.7: USCALE=0.59

After this, the sample may come close to having an $N(0,1)$ sample distribution - but not usually! So you can force S.D. = 1 unit, by setting USCALE = 1 / person S.D.

19.65 Mantel and Mantel-Haenszel DIF statistics

Mantel-Haenszel is the industry-standard DIF statistic, but it expects complete data because it stratifies the data by raw scores. Please Google "Mantel-Haenszel". The Winsteps implementation is slightly different because it stratifies by person measure (same as raw scores for complete data), so it is robust against missing data.

If the data are incomplete (missing data), then there are alternative methods, such as the Rasch-Welch method. There is no industry-standard method in this situation. Most methods require deletion of persons with incomplete response strings, see epm.sagepub.com/content/69/1/18.short - Rasch-Welch uses all available data.

Differential item functioning (DIF) can be investigated using log-odds estimators, Mantel-Haenszel (1959) for dichotomies or Mantel (1963) for polytomies. The sample is divided into difference classification groups (also called reference groups and focal groups) which are shown in [Table 30](#) and specified with [DIF=](#). And then sliced into strata by ability measure (equivalent to raw score for complete data).

The usual M-H computation stratifies the sample by raw scores, so it works with case-wise deletion of cases with missing data. Winsteps stratifies cases by measure, so cases with missing data are stratified at their estimated measure. For complete data and thin-slicing, the conventional M-H computation and the Winsteps M-H computation produce the same numbers. With missing data or thick-slicing, the conventional M-H computations and the Winsteps M-H computations may differ.

M-H and the t-tests in Winsteps should produce the same results, because they are based on the same logit-linear theory. But, in practice, M-H will be more accurate if the data are complete and there are large numbers of subjects at every score level, so called "thin" matching. Under other circumstances, M-H may not be estimable, or must use grouped-score "thick" matching, in which case the t-test method will probably be more accurate. Similar conclusions can also be inferred from http://www.eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?_nfpb=true&_ERICExtSearch_SearchValue_0=ED334230&ERICExtSearch_SearchType_0=no&accno=ED334230

MHSLICE= controls the width of each slice, thin or thick. MHSLICE= specifies the width of the slice (in logits) of the latent variable be included in each cross-tab. The lower end of the lowest slice is always the lowest observed person measure.

MHSLICE = 0 bypasses Mantel-Haenszel or Mantel computation.

MHSLICE = .1 logits and smaller. The latent variable is stratified into thin slices. This corresponds to the slicing by raw scores with complete data

MHSLICE = 1 logit and larger. The latent variable is stratified into thick slices.

For each slice, a cross-tabulation is constructed for each pair of person classifications against each scored response level. An odds-ratio is computed from the cross-tab. Zero and infinite ratios are ignored. A homogeneity chi-square is also computed when possible.

Thin slices are more sensitive to small changes in item difficulty across person classifications, but more persons are ignored in inestimable cross-tabs. Thick slices are more robust because fewer persons are ignored. Use the [Specification](#) pull-down menu to set different values of MHSLICE= and then produce the corresponding [Table 30](#).

In principle, when the data fit the Rasch model, the Mantel and Mantel-Haenszel estimators should concur with the Rasch DIF contrast measures. The Rasch DIF contrast weights each person equally. Mantel weights each cross-tabulation equally. Thus when the DIF estimates disagree, it indicates that the DIF in the data is [non-uniform](#) with ability level.

Computation:

Person classification groups are A, B, ... They are compared pairwise. Starting from the lowest person measure, each slice is MHSLICE= logits wide. There are K slices up through the highest person measure. For the target item, in the kth slice and comparing classification groups A and B, with categories renumbered from 0 to simplify the computation,

Person Classification Groups, $g = 1, 2, \dots$ by $j = 0, J$ categories by $k = 1, K$ ability strata (slices)								
		Rating-Scale Category					Total:	
		$j = 0$	1	2	3	j		J
Classification Group		Score = Y_0	Y_1	Y_2	Y_3	Y_j	Y_J	
Reference	$g = 1$	nk_{10}	nk_{11}	nk_{12}	nk_{13}	nk_{1j}	nk_{1J}	Nk_1
Focal	2	nk_{20}	nk_{21}	nk_{22}	nk_{23}	nk_{2j}	nk_{2J}	Nk_2
Total:		Mk_0	Mk_1	Mk_2	Mk_3	Mk_j	Mk_J	Tk

Then the Mantel (1963) or Mantel-Haenszel (1959) DIF chi-square for the target item is:

$$\chi_1^2 = \frac{\left[\sum_{k=1}^K \sum_{j=0}^J (nk1jYj - Nk1MkjYj / Tk) - continuity \right]^2}{\sum_{k=1}^K \frac{Nk1Nk2}{Tk^2(Tk-1)} \left[Tk \sum_{j=0}^J MkjYj^2 - \left(\sum_{j=0}^J MkjYj \right)^2 \right]}$$

where continuity = 0.5 for dichotomies and = 0 for polytomies.

The Mantel-Haenszel (1959) log-odds estimator for dichotomies and the Liu-Agresti (1996) cumulative log-odds estimator for polytomies are:

$$S_{nkj} = \sum_{h=0}^{j-1} nkgh$$

$$LOR = \frac{\sum_{k=1}^K \sum_{j=1}^J (S_{nk1j}(Nk2 - S_{nk2j}) / Tk)}{\sum_{k=1}^K \sum_{j=1}^J (S_{nk2j}(Nk1 - S_{nk1j}) / Tk)}$$

Liu, I-M, & Agresti, A. (1996). Mantel-Haenszel-type inference for cumulative odds ratios with a stratified ordinal response. *Biometrics*, 52, 1223-1234.

Mantel N. (1963) Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. *J Amer Stat Assoc* 58, 690-700.

Mantel, N. and Haenszel, W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 22, 719-748.

ETS DIF Category	with DIF Contrast and DIF Statistical Significance	
C = moderate to large	 DIF ≥ 0.64 logits	prob(DIF ≤ 0.43 logits) ≤ .05 (2-sided) <i>approximately: DIF > 0.43 logits + 2 * DIF S.E.</i>
B = slight to moderate	 DIF ≥ 0.43 logits	prob(DIF = 0 logits) ≤ .05 (2-sided) <i>approximately: DIF > 2 * DIF S.E</i>
A = negligible	-	-
C-, B- = DIF against focal group; C+, B+ = DIF against reference group		
ETS (Educational Testing Service) use Delta δ units. 1 logit = 2.35 Delta δ units. 1 Delta δ unit = 0.426 logits.		
Zwick, R., Thayer, D.T., Lewis, C. (1999) An Empirical Bayes Approach to Mantel-Haenszel DIF Analysis. . <i>Journal of Educational Measurement</i> , 36, 1, 1-28 More explanation at www.ets.org/Media/Research/pdf/RR-12-08.pdf pp. 3,4		

Example:

```
-----
| PERSON  DIF  DIF  PERSON  DIF  DIF      DIF  JOINT  Welch  Mantel-Haenszel Size ITEM
| CLASS  MEASURE S.E. CLASS  MEASURE S.E. CONTRAST S.E.  t  d.f. Prob. Chi-squ Prob. CUMLOR Number Name |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| A      1.47  .28  P      2.75  .33   -1.28  .43 -2.96 102 .0038  7.198 .0073 -1.20  1 Response|
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
Size of Mantel-Haenszel slice = .100 logits
```

```
title="MH computation"
; d.f.=1 chi=7.198 p=0.073
; log-odds = -1.20
codes=01
cfile=*
1 Better
```

```

0 Same
*
item1=1
name1=1
NI=1
pweight=$s9w2 ; weighting substitutes for entering multiple records
PAFILE=$S6W1 ; anchoring forces stratification
DIF = $4W1 ; cross-tab by Gender, F or M
&end
Response
;234567890
END LABELS
1 FA 1 16
0 FA 1 11
1 FP 1 5
0 FP 1 20
1 MA 2 12
0 MA 2 16
1 MP 2 7
0 MP 2 19

```

19.66 Misfit diagnosis: infit outfit mean-square standardized

Remember that our purpose is to measure the persons, not to optimize the items and raters. A good approach is to compute the person measures based on all the different item selections that you think are reasonable. Start with all the items, and then reduce to smaller sets of items. Cross-plot the person measures. If the person measures are collinear, use the larger set of items. If the person measures are not collinear, use the set of items which produces the more meaningful set of person measures.

What do Infit Mean-square, Outfit Mean-square, Infit Zstd (z-standardized), Outfit Zstd (z-standardized) mean?

Every observation contributes to both infit and outfit. But the weighting of the observations differs. On-target observations contribute less to outfit than to infit.

Outfit: outlier-sensitive fit statistic. This is based on the conventional chi-square statistic. This is more sensitive to unexpected observations by persons on items that are relatively very easy or very hard for them (and vice-versa).

Infit: inlier-pattern-sensitive fit statistic. This is based on the chi-square statistic with each observation weighted by its statistical information (model variance). This is more sensitive to unexpected patterns of observations by persons on items that are roughly targeted on them (and vice-versa).

Outfit = $\text{sum}(\text{residual}^2 / \text{information}) / (\text{count of residuals}) = \text{average}((\text{standardized residuals})^2) = \text{chi-square/d.f.} = \text{mean-square}$

The standardized residual is also called the *Pearson* residual.

Infit = $\text{sum}((\text{residual}^2 / \text{information}) * \text{information}) / \text{sum}(\text{information}) = \text{average}((\text{standardized residuals})^2 * \text{information}) = \text{information-weighted mean-square}$

Mean-square: this is the chi-square statistic divided by its degrees of freedom. Consequently its expected value is close to 1.0. Values greater than 1.0 (underfit) indicate unmodeled noise or other source of variance in the data - these degrade measurement. Values less than 1.0 (overfit) indicate that the model predicts the data too well - causing summary statistics, such as reliability statistics, to report inflated statistics. See further [dichotomous](#) and [polytomous](#) mean-square statistics. The mean-square Outfit statistic is also called the *Reduced* chi-square statistic. For computations, see

If the mean-squares average much below 1.0, then the data may have an almost-Guttman pattern. Please use much tighter [convergence criteria](#).

Z-Standardized: these report the statistical significance (probability) of the chi-square (mean-square) statistics occurring by chance when the data fit the Rasch model. "Standardized" means "transformed to conform to a unit-normal distribution". The values reported are unit-normal deviates, in which .05% 2-sided significance corresponds to 1.96. Overfit is reported with negative values. These are also called [t-statistics](#) reported with infinite degrees of freedom.

ZSTD probabilities: two-sided unit-normal deviates	
1.00	p= .317
1.96	.050
2.00	.045
2.58	.01
3.00	.0027
4.00	.00006
5.00	.0000006

Infit was an innovation of Ben Wright's (G. Rasch, 1980, Afterword). Ben noticed that the standard statistical fit statistic (that we now call Outfit) was highly influenced by a few outliers (very unexpected observations). Ben need a fit statistic that was more sensitive to the overall pattern of responses, so he devised Infit. Infit weights the observations by their statistical information (model variance) which is higher in the center of the test and lower at the extremes. The effect is to make Infit less influenced by outliers, and more sensitive to patterns of inlying observations.

Ben Wright's Infit and Outfit statistics (e.g., [RSA](#), p. 100, www.rasch.org/rmt/rmt34e.htm) are initially computed as mean-square statistics (i.e., chi-square statistics divided by their degrees of freedom). For Outfit the d.f. is the count of observations. For Infit the d.f. is the sum of the information in the observations = 1 / item or person logit S.E.**2. Their likelihood (significance) is then computed. This could be done directly from chi-square tables, but the convention is to report them as unit normal deviates (i.e., [t-statistics](#) corrected for their degrees for freedom). I prefer to call them z-statistics, but the Rasch literature has come to call them [t-statistics](#), so now I do to. It is confusing because they are not strictly Student [t-statistics](#) (for which one needs to know the degrees of freedom) but are random normal deviates.

General guidelines:

First, investigate **negative point-measure or point-biserial correlations**. Look at the Distractor Tables, e.g., [10.3](#). Remedy miskeys, data entry errors, etc.

Then, the general principle is:

**Investigate outfit before infit,
mean-square before t standardized,
high values before low or negative values.**

There is an asymmetry in the implications of out-of-range high and low mean-squares (or positive and negative [t-statistics](#)). High mean-squares (or positive [t-statistics](#)) are a much greater threat to validity than low mean-squares (or negative fit statistics).

Poor fit does not mean that the Rasch measures (parameter estimates) aren't additive. The Rasch model forces its estimates to be additive. Misfit means that the reported estimates, though effectively additive, provide a distorted picture of the data.

The fit analysis is a report of how well the data accord with those additive measures. So a MnSq >1.5 suggests a deviation from unidimensionality in the data, not in the measures. So the unidimensional, additive measures present a distorted picture of the data.

High outfit mean-squares may be the result of **a few random responses** by low performers. If so, drop with [PDFILE=](#) these performers when doing item analysis, or use [EDFILE=](#) to change those response to missing.

High infit mean-squares indicate that the items are mis-performing for the people on whom the items are targeted. This is a bigger threat to validity, but more difficult to diagnose than high outfit.

Mean-squares show the size of the randomness, i.e., the amount of distortion of the measurement system. 1.0 are their expected values. Values less than 1.0 indicate observations are too predictable (redundancy, model overfit). Values greater than 1.0 indicate unpredictability (unmodeled noise, model underfit). Mean-squares usually average to 1.0, so if there are high values, there must also be low ones. Examine the high ones first, and temporarily remove them from the analysis if necessary, before investigating the low ones.

Zstd are t-tests of the hypotheses "do the data fit the model (perfectly)?" ZSTD (standardized as a z-score) is used of a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or

the Student's [t-statistic](#) value has been adjusted to a unit normal value. They show the improbability (significance). 0.0 are their expected values. Less than 0.0 indicate too predictable. More than 0.0 indicates lack of predictability. If mean-squares are acceptable, then Zstd can be ignored. They are truncated towards 0, so that 1.00 to 1.99 is reported as 1. So a value of 2 means 2.00 to 2.99, i.e., at least 2. For exact values, see Output Files. If the test involves less than 30 observations, it is probably too insensitive, i.e., "everything fits". If there are more than 300 observations, it is probably too sensitive, i.e., "everything misfits".

Interpretation of parameter-level mean-square fit statistics:	
>2.0	Distorts or degrades the measurement system.
1.5 - 2.0	Unproductive for construction of measurement, but not degrading.
0.5 - 1.5	Productive for measurement.
<0.5	Less productive for measurement, but not degrading. May produce misleadingly good reliabilities and separations.

In general, mean-squares near 1.0 indicate little distortion of the measurement system, regardless of the Zstd value. Evaluate high mean-squares before low ones, because the average mean-square is usually forced to be near 1.0. Mean-square fit statistics will average about 1.0, so, if you accept items (or persons) with large mean-squares (low discrimination), then you must also accept the counter-balancing items (or persons) with low mean-squares (high discrimination).

Outfit mean-squares: influenced by outliers. Usually easy to diagnose and remedy. Less threat to measurement.
 Infit mean-squares: influenced by response patterns. Usually hard to diagnose and remedy. Greater threat to measurement.

Extreme scores always fit the Rasch model exactly, so they are omitted from the computation of fit statistics. If an extreme score has an anchored measure, then that measure is included in the fit statistic computations.

Anchored runs: Anchor values may not exactly accord with the current data. To the extent that they don't, the fit statistics may be misleading. Anchor values that are too central for the current data tend to make the data appear to fit too well. Anchor values that are too extreme for the current data tend to make the data appear noisy.

Question: Are you contradicting the usual statistical advice about model-data fit?

Statisticians are usually concerned with "how likely are these data to be observed, assuming they accord with the model?" If it is too unlikely (i.e., significant misfit), then the verdict is "these data don't accord with the model." The practical concern is: "In the imperfect empirical world, data never exactly accord with the Rasch model, but do these data deviate seriously enough for the Rasch measures to be problematic?" The builder of my house followed the same approach (regarding Pythagoras theorem) when building my bathroom. It looked like the walls were square enough for his practical purposes. Some years later, I installed a full-length rectangular mirror - then I discovered that the walls were not quite square enough for my purposes (so I had to make some adjustments) - so there is always a judgment call. The table of mean-squares is my judgment call as a "builder of Rasch measures".

Question: My data contains misfitting items and persons, what should I do?

Let us clarify the objectives here.

A. www.rasch.org/rmt/rmt234g.htm is aimed at the usual situation where someone has administered a test from somewhere to a sample of people, and we, the analysts, are trying to rescue as much of that data as is meaningful. We conservatively remove misfitting items and persons until the data makes reasonable sense. We then anchor those persons and items to their good measures. After reinstating whatever misfitting items and persons we must report, we do the final analysis.

B. A pilot study is wonderfully different. We want to optimize the subset of items. The person sample and the data can be tailored to optimize item selection. Accordingly,

First, even before data analysis, we need to arrange the items into their approximately intended order along the latent variable. With 89 items, item can be conceptually grouped into clusters located at 5 or more levels of the latent variable, probably more than 5. This defines what we want to measure. If we don't know this order, then we will not know whether we have succeeded in measuring what we intended to measure. We may accidentally construct a test that measures a related

variable. This happened in one edition of the MMPI where the test constructors intended to measure depression, but produced a scale that measured "depression+lethargy".

Second, we analyze the data and inspect the item hierarchy. Omit any items that are locating in the wrong place on the latent variable. By "omit", I mean give a weight of zero with IWEIGHT=. Then the item stays in the analysis, but does not influence other numbers. This way we can easily reinstate items, if necessary, knowing where they would go if they had been given the usual weight of 1.

Third, reanalyze the data with the pruned item hierarchy. Omit all persons who severely underfit the items, these are contradicting the latent variable. Again, "omit" means PWEIGHT= 0. Also omit persons whose "cooperation" is because they have an overfitting response set, such as the middle category of every item.

Fourth, analyze again. The data should be coherent. Items in the correct order. Persons cooperating. So apply all the other selection criteria, such as content balancing, DIF detection, to this coherent dataset.

Question: Should I report Outfit or Infit?

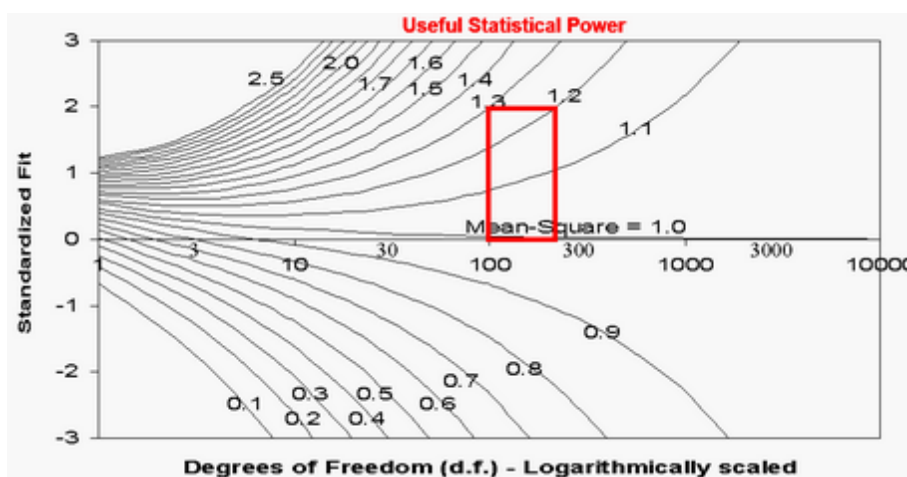
A chi-square statistic is the sum of squares of standard normal variables. Outfit is a chi-square statistic. It is the sum of squared standardized residuals (which are modeled to be standard normal variables). So it is a conventional chi-square, familiar to most statisticians. Chi-squares (including outfit) are sensitive to outliers. For ease of interpretation, this chi-square is divided by its degrees of freedom to have a mean-square form and reported as "Outfit". Consequently I recommend that the Outfit be reported unless there is a strong reason for reporting infit.

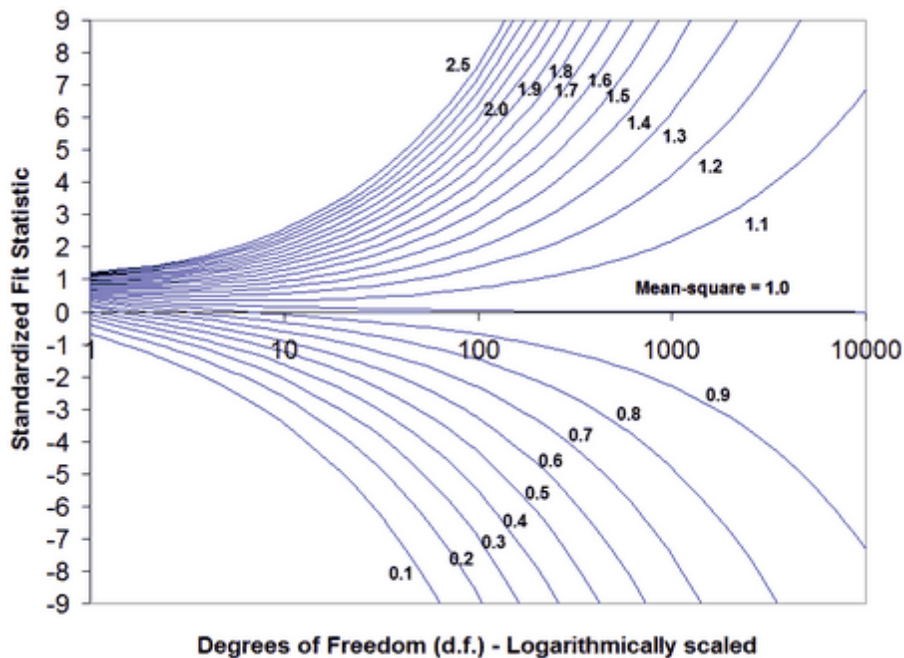
In the Rasch context, outliers are often lucky guesses and careless mistakes, so these outlying characteristics of respondent behavior can make a "good" item look "bad". Consequently, Infit was devised as a statistic that down-weights outliers and focuses more on the response string close to the item difficulty (or person ability). Infit is the sum of (squares of standard normal variables multiplied by their statistical information). For ease of interpretation, Infit is reported in mean-square form by dividing the weighted chi-square by the sum of the weights. This formulation is unfamiliar to most statisticians, so I recommend against reporting Infit unless the data are heavily contaminated with irrelevant outliers.

Question: Are mean-square values, >2 etc, sample-size dependent?

The mean-squares are corrected for sample size: they are the chi-squares divided by their degrees of freedom, i.e., sample size. The mean-squares answer "how big is the impact of the misfit". The [t-statistics](#) answer "how likely are data like these to be observed when the data fit the model (exactly)." In general, the bigger the sample the less likely, so that t-statistics are highly sample-size dependent. We eagerly await the theoretician who devises a statistical test for the hypothesis "the data fit the Rasch model usefully" (as opposed to the current tests for perfectly).

The relationship between mean-square and z-standardized [t-statistics](#) is shown in this plot. Basically, the standardized statistics are insensitive to misfit with less than 30 observations and overly sensitive to misfit when there are more than 300 observations.





Question: For my sample of 2400 people, the mean-square fit statistics are close to 1.0, but the Z-values associated with the INFIT/OUTFIT values are huge (over 4 to 9.9). What could be causing such high values?

Your results make sense. Here is what has happened. You have a sample of 2,400 people. This gives huge statistically power to your test of the null hypothesis: "These data fit the Rasch model (exactly)." In the nomographs above, a sample size of 2,400 (on the right-hand-side of the plot) indicates that even a mean-square of 1.2 (and perhaps 1.1) would be reported as misfitting highly significantly. So your mean-squares tell us: "these data fit the Rasch model usefully", and the Z-values tell us: "but not exactly". This situation is often encountered in situations where we know, in advance, that the null hypothesis will be rejected. The Rasch model is a theoretical ideal. Empirical observations never fit the ideal of the Rasch model if we have enough of them. You have more than enough observations, so the null hypothesis of exact model-fit is rejected. It is the same situation with Pythagoras theorem. No empirical right-angled-triangle fits Pythagoras theorem if we measure it precisely enough. So we would reject the null hypothesis "this is a right-angled-triangle" for all triangles that have actually been drawn. But obviously billions of triangle are usefully right-angled.

Example of computation:

Imagine an item with categories $j=0$ to m . According to the Rasch model, every category has a probability of being observed, P_j .

Then the expected value of the observation is $E = \sum (j * P_j)$

The model variance (sum of squares) of the probable observations around the expectation is $V = \sum (P_j * (j - E) **2)$.

This is also the statistical information in the observation.

For dichotomies, these simplify to $E = P_1$ and $V = P_1 * P_0 = P_1*(1-P_1)$.

For each observation, there is an expectation and a model variance of the observation around that expectation.

residual = observation - expectation

Outfit mean-square = $\sum (\text{residual}**2 / \text{model variance}) / (\text{count of observations})$

Infit mean-square = $\sum (\text{residual}**2) / \sum (\text{modeled variance})$

Thus the outfit mean-square is the accumulation of squared-standardized-residuals divided by their count (their expectation). The infit mean-square is the accumulation of squared residuals divided by their expectation.

Outlying observations have smaller information (model variance) and so have less information than on-target observations. If all observations have the same amount of information, the information cancels out. Then Infit mean-square = Outfit mean-square.

For dichotomous data. Two observations: Model $p=0.5$, observed=1. Model $p=0.25$, observed =1.

Outfit mean-square = $\sum ((\text{obs-exp})**2 / \text{model variance}) / (\text{count of observations}) = ((1-0.5)**2)/(0.5*0.5) + (1-0.25)**2/(0.25*0.75))/2 = (1 + 3)/2 = 2$

Infit mean-square = $\sum (obs - exp)^2 / \sum(model\ variance) = ((1-0.5)^2 + (1-0.25)^2) / ((0.5*0.5) + (0.25*0.75)) = (0.25 + 0.56) / (0.25 + 0.19) = 1.84$. The off-target observation has less influence.

The [Wilson-Hilferty](#) cube root transformation converts the mean-square statistics to the normally-distributed z-standardized ones. For more information, please see Patel's "Handbook of the Normal Distribution" or www.rasch.org/rmt/rmt162g.htm.

Diagnosing Misfit: Noisy = Underfit. Muted = Overfit				
Classification	INFIT	OUTFIT	Explanation	Investigation
	Noisy	Noisy	Lack of convergence Loss of precision Anchoring	Final values in Table 0 large? Many categories? Large logit range? Displacements reported?
Hard Item	Noisy	Noisy	Bad item	Ambiguous or negative wording? Debatable or misleading options?
	Muted	Muted	Only answered by top people	At end of test?
Item	Noisy	Noisy	Qualitatively different item Incompatible anchor value	Different process or content? Anchor value incorrectly applied?
		?	Biased (DIF) item	Stratify residuals by person group?
		Muted	Curriculum interaction	Are there alternative curricula?
	Muted	?	Redundant item	Similar items? One item answers another? Item correlated with other variable?
Rating scale	Noisy	Noisy	Extreme category overuse	Poor category wording? Combine or omit categories?
	Muted	Muted	Middle category overuse	Wrong model for scale?
Person	Noisy	?	Processing error Clerical error Idiosyncratic person	Scanner failure? Form markings misaligned? Qualitatively different person?
High Person	?	Noisy	Careless Sleeping Rushing	Unexpected wrong answers? Unexpected errors at start? Unexpected errors at end?
Low Person	?	Noisy	Guessing Response set "Special" knowledge	Unexpected right answers? Systematic response pattern? Content of unexpected answers?
	Muted	?	Plodding Caution	Did not reach end of test? Only answered easy items?
Person/Judge Rating	Noisy	Noisy	Extreme category overuse	Extremism? Defiance? Misunderstanding the rating scale?
	Judge Rating	Muted	Muted	Middle category overuse Conservatism? Resistance?
Apparent unanimity			Collusion? Hidden constraints?	
INFIT: information-weighted mean-square, sensitive to irregular inlying patterns OUTFIT: usual unweighted mean-square, sensitive to unexpected rare extremes Muted: overfit, un-modeled dependence, redundancy, the data are too predictable Noisy: underfit, unexpected unrelated irregularities, the data are too unpredictable.				

Guessing and Carelessness

Both guessing and carelessness can cause high outfit mean-square statistics. Sometimes it is not difficult to identify which is the cause of the problem. Here is a procedure:

- 1) Analyze all the data: output [PFILE=pf.txt](#) [IFILE=if.txt](#) [SFILE=sf.txt](#)
 - 2) Analyze all the data with anchor values: [PAFILE=pf.txt](#) [IAFILE=if.txt](#) [SAFILE=sf.txt](#)
- 2A) [CUTLO=-2](#) this eliminates responses to very hard items, so person misfit would be due to unexpected responses to easy items

2B) [CUTH](#) 2 this eliminates responses to very easy items, so person misfit would be due to unexpected responses to hard items.

What is your primary concern? Statistical fit or productive measurement?

Statistical fit is like "beauty". Productive measurement is like "utility".

Statistical fit is dominated by sample size. It is like looking at the data through a microscope. The more powerful the microscope (= the bigger the sample), the more flaws we can see in each item. For the purposes of beauty, we may well scrutinize our possessions with a microscope. Is that a flaw in the diamond? Is that a crack in the crystal? There is no upper limit to the magnification we might use, so there is no limit to the strictness of the statistical criteria we might employ. The nearer to 1.0 for mean-squares, the more "beautiful" the data.

In practical situations, we don't look at our possessions through a microscope. For the purposes of utility, we are only concerned about cracks, chips and flaws that will impact the usefulness of the items, and these must be reasonably obvious. In terms of mean-squares, the range 0.5 to 1.5 supports productive measurement.

However, life requires compromise between beauty and utility. We want our cups and saucers to be functional, but also to look reasonably nice. So a reasonable compromise for high-stakes data is mean-squares in the range 0.8 to 1.2.

Fit statistics for item banks are awkward. They depend on the manner in which the items in the item banks are used, and also the manner in which the item difficulties are to be verified and updated. Initial values for the items often originate in conventional paper-and-pencil tests, but if the item bank is used to support other styles of testing, these initial values will be superseded by more relevant values. Constructing fit statistics for other styles of testing has proved challenging for theoreticians. This has forced the relaxation of the strict rules of conventional statistical analysis. This is probably why you are having difficulty finding appropriate literature.

My person sample size is 2,000. Many items have significant misfit. What shall I do?

Don't despair! Your items may not be as bad as those statistics say.

If the noise in the data is homogeneous, then the noise-level is independent of sample size. The mean-square statistics will also be independent of sample size.

t -statistics are sensitive to the power of the statistical test (sample size). The relationship between mean-squares and t -statistics is shown in the Figures above which suggests that, for practical purposes, t -statistics are under-powered for sample sizes less than 100 and over-powered for sample sizes greater than 300.

So my recommendation (not accepted by conventional statisticians) is that mean-squares be used in preference to t -statistics. In my view, the standard t -tests are testing the wrong hypothesis. Wrong hypothesis = "The data fit the model (perfectly)". Right hypothesis = "The data fit the model (usefully)". Unfortunately conventional statisticians are not interested in usefulness and so have not formulated t -tests for it.

Alternatively, random-sample 300 from your 2,000 test-takers. Perform your t -tests with this reduced sample. Confirm your findings with another random-sample of 300.

19.67 Missing data

One of Ben Wright's requirements for valid measurement, derived from the work of L.L. Thurstone, is that "Missing data must not matter." Of course, missing data always matters in the sense that it lessens the amount of statistical information available for the construction and quality-control of measures. Further, if the missing data, intentionally or unintentionally, skew the measures (e.g., incorrect answers are coded as "missing responses"), then missing data definitely do matter. But generally, missing data are missing essentially at random (by design or accident) or in some way that will have minimal impact on the estimated measures (e.g., adaptive tests).

Winsteps does not require complete data in order to make estimates. One reason that Winsteps uses [JMLE](#) is that it is very flexible as regards estimable data structures. For each parameter (person, item or Rasch-Andrich threshold) there are sufficient statistics: the marginal raw scores and counts of the non-missing observations. During Winsteps estimation, the observed marginal counts and the observed and expected marginal scores are computed from the same set of non-missing observations. Missing data are skipped over in these additions. When required, Winsteps can compute an expected value for every observation (present or missing) for which the item and person estimates are known.

The basic estimation algorithm used by Winsteps is:

$$\text{Improved parameter estimate} = \text{current parameter estimate} + (\text{observed marginal score} - \text{expected marginal score}) / (\text{modeled variance of the expected marginal score})$$

The observed and expected marginal scores are obtained by summing across the non-missing data. The expected score and its variance are obtained by Rasch estimation using the current set of parameter estimates, see *RSA*.

If data are missing, or observations are made, in such a way that measures cannot be constructed unambiguously in one frame of reference, then the message

WARNING: DATA MAY BE AMBIGUOUSLY CONNECTED INTO nnn SUBSETS

is displayed on the Iteration screen to warn of [ambiguous connection](#).

Missing data in [Tables 23, 24](#): Principal Components Analysis.

For raw observations, missing data are treated as missing. Pairwise deletion is used during the correlation computations. For residuals, missing data are treated as 0, their expected values. This attenuates the contrasts, but makes them estimable.

You can try different methods for missing data by writing an [IPMATRIX=](#) of the raw data to a file, and then using your own statistical software to analyze.

Example 1: Missing observations are scored "1"

```
ni=4
codes=01
name1=1
item1=1
codes = 01A
ptbis=YES
misscore=1
&end
END LABELS
0110
1001
A001 ; A is in Codes= but scored "missing"
B101 ; B is not in Codes= but is scored 1 by MISSCORE=1
```

ENTRY	DATA	SCORE	DATA	AVERAGE	S.E.	OUTF	PTBSE			
NUMBER	CODE	VALUE	COUNT	%	MEASURE	MEAN	MNSQ	CORR.	ITEM	Correlation:
										Code with Score
1	A	***	1	25*	-.70			-.50	I0001	A {0,0,1,-}{2,1,1,2}
	0	0	1	33	-.01	.7	1.00			0 {1,0,-,-}{2,1,1,2}
	1	1	1	33	-.01*	1.4	-1.00			1 {0,1,-,-}{2,1,1,2}
	MISSING	1	1	33	1.26	.4	.50			B {0,0,-,1}{2,1,1,2}

Example 2: Missing data: two types: "skipped" and "not reached"

Missing responses in a dataset do not all have the same meaning. For instance, in a time multiple-choice test. Missing responses between observed responses may mean "skipped" - the respondent decided this question was too hard.

Missing responses at the end of the test can mean "not reached", because time ran out before the respondent could respond to these items.

Solution: enter two different missing data codes in the data file, for instance, "S" for skipped and "R" for not reached. Then,

when calibrating the items, we want to ignore "not reached" responses, but score skipped responses as wrong:

`CODES=01S`

`NEWSCORE=010` ; S is scored 0

`MISSING-SCORED=-1` ; data code R is not in CODES= so it will be scored -1 = "ignore", "not administered"

`IFILE=` item-calibrations.txt

when we measure the persons, skipped and not reached responses are wrong:

`IAFILE=` item-calibrations.txt ; anchor the item difficulties at their good calibrations

`CODES=01SR`

`NEWSCORE=0100` ; S and R are scored 0

19.68 Mixed-Mixture Models and Saltus models

Rasch models are grounded in the concept of the unidimensional latent variable, i.e., the items defining the latent variable operate in the same way for all members of the target population. Of course, this is a fiction. But reality can often be made to cooperate.

But there are occasions when a population is comprised of different classes of persons with the items comprising a different latent variable for each class. The classes are called "Latent Classes".

Standard Rasch "latent trait" models can be extended to allow for latent classes. These are called "Mixture Models" (Rost, 1990). The Saltus model (Mark Wilson, 1989) is a mixed model in which segments of items are modeled to shift their difficulties together, and by the same amount, for different latent classes. In these models, the different latent variables are defined by item difficulties, but individual respondents are not assigned to a particular class, but rather the probability that each respondent belongs to each class is reported.

Winsteps does not do a mixture or Saltus analysis directly, but it can provide much of the same information, and also can indicate whether a more rigorous latent class analysis is likely to be productive.

Here is an approach:

Step 1. Identify meaningful potential respondent classes, e.g., male/female, high/low performers. The Winsteps PCA analysis (e.g., [Table 24.4](#)) may help identify potential classes.

Step 2. Mark in the person label the class codes. The Microsoft Word "[rectangle copy](#)" function may be useful. High and low performers do not need to be flagged, instead the [MA2](#) function can be used.

Step 3. Perform [DIF](#) analysis based on the class codes. Items displaying strong DIF may be exhibiting class-related behavior.

Step 4. Flag the items by class in the item identification.

Step 5. Look for item-classification by person-classification interactions (differential classification-grouped functioning, DGF, [Table 33](#)). These would approximate the Saltus findings.

Rost, Jürgen. (1990) Rasch Models in Latent Classes: An Integration of Two Approaches to Item Analysis, *Applied Psychological Measurement*, 14, 271-282

Wilson, M. (1989). Saltus: A psychometric model for discontinuity in cognitive development. *Psychological Bulletin*, 105, 276-289.

19.69 Müller's continuous rating scale model CRSM

Müller (1987) suggests that continuous-looking data can be modeled with a conceptually infinite number of categories in a fixed interval. The infinite number of Andrich thresholds between the thresholds is modeled with a linear function. Linacre (2001) remarks that the linear function can be generalized to any convenient function and suggests other functions.

An approximate implementation of these continuous models in Winsteps can be obtained by using a high number of categories and a convenient polynomial function. This is done using [SFUNCTION=](#) to define the polynomial, and [ISRANGE=](#) to define the range of the categories. Müller's continuous rating scale model, CRSM, is approximately equivalent to SFUNCTION=2.

Linacre, J. M. (2001). Percentages with continuous Rasch models. *Rasch Measurement Transactions*, 14, 771-774.

Müller, H. (1987). A Rasch model for continuous ratings. *Psychometrika*, 52, 165-181.

19.70 Multidimensional Model - Multidimensionality

Winsteps estimates unidimensional Rasch models, but many of the benefits of multidimensional modeling can be obtained through a Winsteps analysis. The Winsteps results are often easier to understand and use than those of a multidimensional analysis.

Identifying dimensions among the items. This can be done by inspecting the content of the items or through Table 23.

1. Do a standard unidimensional Rasch analysis
2. Use Winsteps Table 23.1 and other information to identify each item's "dimension"
3. Enter a code for the item's dimension into column 1 of the item label
4. Do the standard Rasch analysis
5. Winsteps Table 31, Differential Person Functioning, will give a measure for each person on each "dimension", all in the same frame-of-reference.

Example 1: I always have a problem in unidimensionality. How can I solve this problem in this data and others?
Unexplained variance in 1st contrast = 4.6 so that the eigenvalue of first contrast is noticeably greater than 2.0

Reply:

(1) Is it truly a problem? There are many reasons for dependency in the data. For instance, if your items are groups using [ISGROUPS=](#), then items in the same group (and so with the same rating scale) are slightly more dependent with each other than with other items.

(2) Look at Table 23.1.

Look at the content (wording) of the items at the top of the plot, and compare them with the items at the bottom of the plot. What is in the top items that contrasts with what is in the bottom item? Are they on different dimensions (for instance, geography and history) or is the difference only superficial (for instance, response format or position on page)?

(3) Please look at the disattenuated correlations of the person measures at the bottom of Table 23.1:

Approximate relationships between the KID measures

PCA	ACT	Pearson	Disattenuated	Pearson+Extr	Disattenuated+Extr
Contrast	Clusters	Correlation	Correlation	Correlation	Correlation
1	1 - 3	0.1404	0.2175	0.1951	0.2923 < far from 1.0

If the correlations are near 1.0, then the person measures for the different "dimensions" are statistically the same. No action is needed.

(4) If the items really are on different dimensions and the person measures have a low correlation, then split the items into two subsets.

Please also see:

[Dimensionality contrasts & variances](#)

[Dimensionality investigation - an example](#)

[Dimensionality: when is a test multidimensional?](#)

Item [principal components/contrasts](#) in Table 23 Identifies structure and dimensionality in response residuals
[Table 23.0](#) Variance components scree plot for items
[Table 23.1, 23.11](#) Principal components plots of item loadings
[Table 23.2, 23.12](#) Item Principal components analysis/contrast of residuals
[Table 23.3, 23.13](#) Item contrast by persons
[Table 23.4, 23.14](#) Item contrast loadings sorted by measure
[Table 23.5, 23.15](#) Item contrast loadings sorted by entry number
[Table 23.6, 23.16](#) Person measures for item clusters in contrast. [Cluster Measure Plot](#) for Table 23.6.
[Table 23.99](#) Largest residual correlations for items
[Youtube video](#) explaining Table 23

Example 2: I want to do a multi-dimensional Rasch analysis with Winsteps.

Multi-dimensional Rasch analysis is implemented directly in ConQuest. In Winsteps, the data for all the domains could be in one control file. Item weighting would be assigned to each item for each domain in separate [IWEIGHT=](#) files, one for each domain. Each domain would be analyzed separately with its own IWEIGHT= file, and person measures produced. Finally, the set of person measures (one for each person for each domain) would be summarized into one measure for each person by using a summarizing analysis, e.g., Q-mode analysis.

19.71 Multiple copy-and-paste or other repetitive actions

Sometimes you want to do the same action repeatedly, such as copying and pasting item curves for all the items into a document. Winsteps users report that a Windows keyboard macro can do this successfully. [Autohotkey](#) is a free Windows keyboard macro generator.

19.72 Multiple Response items - MR questions

Multiple Response MR questions look like multiple-choice questions, but there is more than one correct option. Respondents must select all the correct options, and none of the incorrect options, to succeed on the item.

Example: Which prepositions are grammatically correct in this sentence?

"I am fighting you."

Options: a. against b. at c. for d. from e. to f. with

Correct responses: a., c., f.

Incorrect responses: b., d., e.

For more, see [Scoring Multiple Response Items](#)

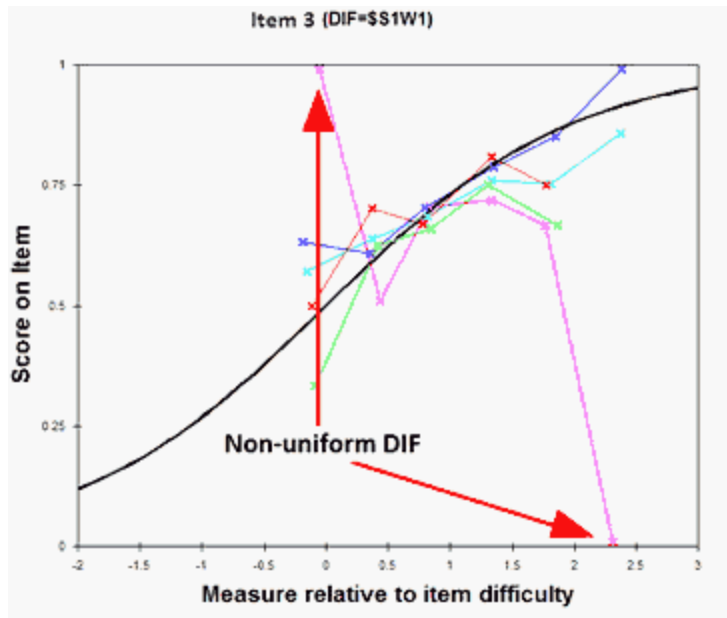
Winsteps cannot analyze these data directly. Please score each MR item into correct/incorrect (1-0) or correct/partially correct/incorrect (2-1-0). Then enter the scored responses into a standard rectangular data file. Winsteps can analyze this file. Alternatively, code each MR option as a separate 1-0 item.

19.73 Non-uniform DIF tables

When items exhibit different difficulties for different person groups this is called "item bias" or "differential item functioning", DIF. When the DIF is the same for all ability levels of the two groups this is termed Uniform DIF. [Table 30](#) reports this. When DIF differs with ability levels, this is termed Non-Uniform DIF, NUDIF. This is discussed here.

Non-uniform DIF is difficult to interpret. Please start by looking at Graphs Menu. [Non-uniform DIF ICCs](#).

This will show you which items have an interaction between item difficulty, group membership and ability.



If you see an item for which the item difficulty is noticeably different for high and low performers in a group, then [Table 30.1](#) will report the statistical significance of the difference.

DIF class specification is: DIF=\$S1W1+MA2

Examinee CLASS	Obs-Exp Average	DIF MEASURE	DIF S.E.	Examinee CLASS	Obs-Exp Average	DIF MEASURE	DIF S.E.	DIF CONTRAST	JOINT S.E.	Welch t	d.f.	Prob.	Mantel-Haenszel Chi-squ	Prob.	Size CUMLOR	Item Number	Item Name
21	.00	-.1950	.1151	22	-.13	.5092	.1836	-.7042	.2166	-3.25	300	.0013	.0000	1.000		3	1

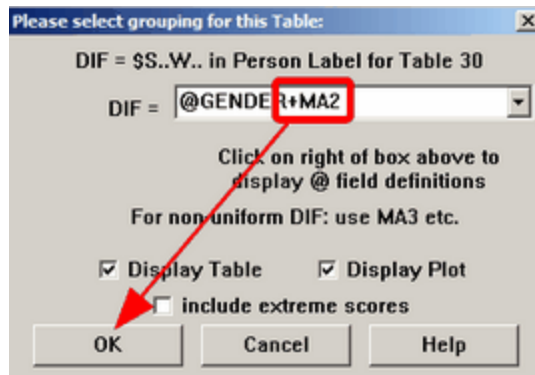
Conceptually, all DIF is non-uniform. "Uniform DIF" = "Average non-uniform DIF". For instance, in the [Mantel-Haenszel](#) DIF computation, every raw-score stratification reports a different DIF estimate. The Mantel-Haenszel statistic averages these estimates. Winsteps reports uniform DIF as a summary of a non-uniform DIF distribution in the same way that the mean is a summary of a sample distribution.

- 1) Look at the empirical ICC (Graphs menu, [Non-uniform DIF](#)). Is the empirical ICC clearly flatter or steeper than the model ICC for a classification group? Suppose that is group F
- 2) In the Winsteps DIF [table 30.1](#), we expect F1 vs. F2 to be both a large DIF contrast (>1.0 logits) and to be statistically significant ($p \leq .05$).
- 3) There should be a substantive reason for deciding that the NUDIF is not merely a statistical accident.

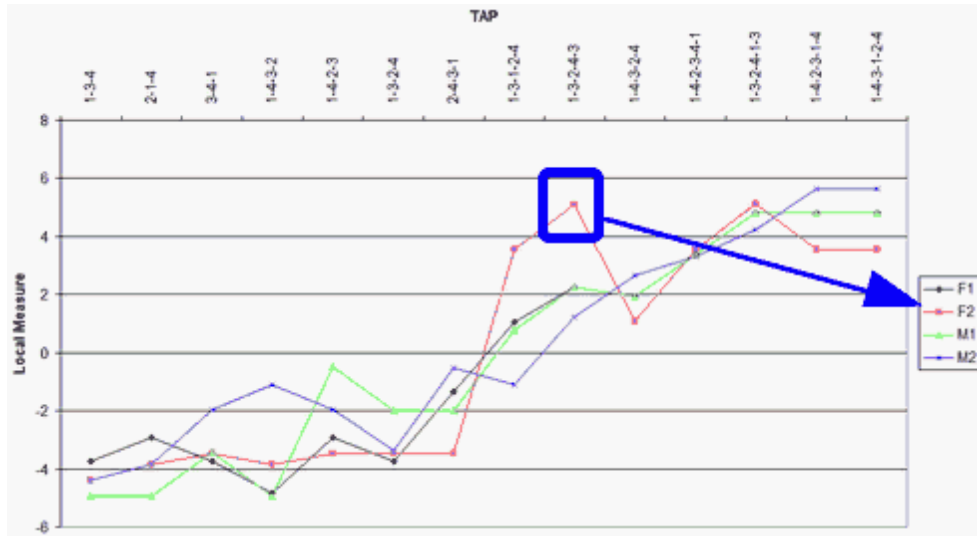
Example, the division sign, \div , is very difficult for those students who have not learned its meaning, and easy for those who have. So it exhibits NUDIF across grade-levels.

In detail:

1. Simple NUDIF reporting. In Table 30, specify both a person class column with [DIF=](#) and the number of ability strata to be reported, e.g., [\\$MA2](#), which reports 2 levels of ability, the top and bottom half of the ability range. Here is the [dialog box](#):



Here is the [plot](#):



The blue box flags an example of non-uniform DIF. It is "F2" where "F" (@GENDER) is the females, and "2" (MA2) the upper ability group. "F1" is "F" ability group "1", the lower ability group. The non-uniform DIF numbers are shown in [Table 30.2](#):

TABLE 30.2 KNOX CUBE TEST ZOU608W5.TXT Aug 21 6:04 2007
 INPUT: 35 KIDS 18 TAPS MEASURED: 35 KIDS 18 TAPS 2 CATS WINSTEPS 3.64.0

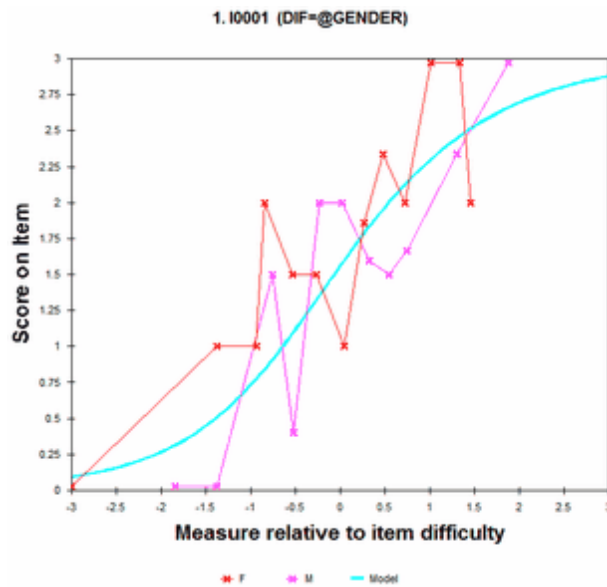
DIF class specification is: DIF=@GENDER+MA2

KID CLASS	OBSERVATIONS COUNT	AVERAGE	BASELINE EXPECT MEASURE	DIF SCORE	DIF MEASURE	DIF SIZE	DIF S.E.	DIF t	TAP Number	TAP Name
F1	7	.00	.01	2.24	-.01	2.24	.00	3.23	12	1-3-2-4-3
F2	11	.00	.20	2.24	-.20	5.10	2.80	1.94	12	1-3-2-4-3
M1	3	.00	.01	2.24	-.01	2.24	.00	6.68	12	1-3-2-4-3
M2	13	.46	.29	2.24	.17	1.21	-1.02	.65	12	1-3-2-4-3

Some person classes have extreme scores. These are shown in the green boxes, and are indicated by "<" for minimum extreme scores and ">" for maximum extreme scores. The plotted DIF measures are shown in the blue box, with the DIF size relative to the overall item difficulty level in the adjacent column.

[Table 30.1](#) reports if there is pairwise DIF between CLASSES/GROUPS.

2. For more insights, plot the [non-uniform DIF ICCs](#).



19.74 Null or unobserved categories: structural and incidental zeroes

There are two types of unobserved or null categories: structural zeroes and incidental/sampling zeroes.

Structural null categories occur when rating scale categories are number 10, 20, 30,... instead of 1,2,3. To force Winsteps to eliminate non-existent categories 11, 12, 13, either rescore the data [IVALUE=](#) or specify [STKEEP=NO](#).

For intermediate incidental null zeroes, imagine this scenario: The Wright & Masters "Liking for Science" data are rescored from 0,1,2 to 0,1,3 with a null category at 2. the categories now mean "disagree, neutral, agree-ish, agree". We can imagine that no child in this sample selected the half-smile of agree-ish.

The category frequencies of categories 0,1,2,3 are 378, 620, 0, 852

The three Rasch-Andrich threshold parameters are $-.89$, $+\infty$, $-\infty$.

The $+\infty$ is because the second parameter is of the order $\log(620/0)$. The $-\infty$ is because the third parameter is of the order $\log(0/852)$.

Mark Wilson's 1991 insight was that the leap from the 2nd to the 4th category is of the order $\log(620/852)$. This is all that is needed for immediate item and person estimation. But it is not satisfactory for anchoring rating scales. In practice however, a large value substitutes satisfactorily for infinity. So, a large value such as 40 logits is used for anchoring purposes. Thus the approximated parameters become $-.89$, 40.89 , -40.00 for [SAFILE=](#). With these anchored threshold values, the expected category frequencies become: 378.8, 619.4, .0, 851.8. None of these are more than 1 score point away from their observed values, and each represents a discrepancy of .2% or less of its category count.

Extreme incidental null categories (unobserved top or bottom categories) are essentially out of range of the sample and so the sample provides no direct information about their estimates. To estimate those estimates requires us to make an assertion about the form of the rating scale structure. The Rasch "Poisson" scale is a good example. All its infinitude of thresholds are estimable because they are asserted to have a specific form. But see [Example 12](#) for a different approach to this situation.

Our recommendation is that structural zeroes be rescored out of the data. If categories **may be** observed next time, then it is better to **include a dummy data record** in your data file which includes an observation of the missing category and reasonable values for all the other item responses that accord with that missing category. This one data record will have minimal impact on the rest of the analysis, especially if you give it a very small weight with [PWEIGHT=](#).

See also [Unobserved and dropped categories](#)

Another approach is to model the categories with a polynomial function: [SFUNCTION=](#).

19.75 One item test - One observation per respondent

One item test:

A dummy second item is needed. This can be given a very small weight, and its dummy data would be the reverse of the active item. Then every person will have a non-extreme score. If everybody is in the top (or bottom) category, then add a dummy person with a very small person and the opposite category response:

Example: Live data is:	Live data with Dummy data:
NI=1 CODES=01 &END Item 1 END LABELS 1 Person 1 1 Person 2 1 Person 3 1 Person 4	NI=2 CODES=01 IWEIGHT=* 2 .01 ; Adjust weighting up to make measures more central, down to make more extreme. * PWEIGHT=* 5 .01 * &END Item 1 Item 2 dummy END LABELS 10 Person 1 10 Person 2 10 Person 3 10 Person 4 01 Person 5 dummy

Some people (or items) have only one response:

An observation in an extreme category is treated the same as any other extreme score. An equivalent finite measure is reported. Intermediate categories generate finite measures. The measure will have very large standard errors (low precision).

This is like the first item of an adaptive test, or the first observation on a diagnostic instrument. It gives a rough idea of the measure. However, if we only have one observation, there is no opportunity for quality-control fit analysis. That is why carpenter's are encouraged to "measure twice!".

Every person (or item) has only one response:

Question: I'm trying to analyze a dataset where there are four test forms, and on each test form there is only one 4-point polytomous item. That is, each student took one and only one test question. Can this type of dataset be calibrated using Winsteps?

Reply: If there is only one response per person, there is not enough information to construct measures, but only enough to order the people by the raw score of that one response. But

If the people taking each of the 4 forms are supposed to be randomly equivalent, then we can [equate](#) the forms, and discover how a "3" on one form relates to a "3" on another form. To do this:

Enter the 4 forms as 4 items in Winsteps.
For each "item" enter the column of responses.
Anchor the rows at 0.
Set ISGROUPS=0
Run the analysis.

The measure corresponding to each score on each item is given in Table 3.2, "Score at Cat", and shown in Table 2.2. Use the measures in the "At Cat." column to correspond to the polytomous observations in summary analyses.

Example: The responses to the 4 forms, A, B, C, D, were:

```
A 1 3 2 4
B 2 4 3 1 1 3
```

```
C 3 2 2 3 1 4 1
D 4 4 3 2 1
```

Note that the order of the persons within form doesn't matter, and the number of respondents per form doesn't matter. Here is the *Winsteps* control file:

```
Title = "Measurement with 4 forms"
NI=4
Item1=1
Name1=1 ; there aren't any row names.
Codes=1234
ISGROUPS=0 ; allow each form its own rating (or partial credit) scale
Item=Form ; rename to remind ourselves
Person=Row ; Rows are anchored at zero, and so are all equivalent.
Pafile=*
1-7 0 ; anchor all rows at "0". 7 is the largest number of students who took any form.
*
CONVERGE=L ; only logit change is used for convergence
LCONV=0.005 ; logit change too small to appear on any report.
&end
A ; the 4 items are the 4 forms
B
C
D
END LABELS
1234 ; responses per form entered as columns with students in any order.
3424
2323
4132
.111
.34.
.1..
```

Resulting Table 2.2:

```
EXPECTED SCORE: MEAN (":" INDICATES HALF-POINT THRESHOLD)
-3      -2      -1      0      1      2      3
|-----+-----+-----+-----+-----+-----+-----| NUM FORM
1          1 :      2 :      3 :      4          4  2  B
|
|
|          1 :      2 :      3 :      4          4  1  A
1      1 :      2 :      3 :      4          4  3  C
|
|          1 :      2 :      3 :      4          4  4  D
|-----+-----+-----+-----+-----+-----+-----| NUM FORM
-3      -2      -1      0      1      2      3
                                7
                                M
                                ROWS
```

Table 3.2:

```
SUMMARY OF CATEGORY STRUCTURE. Model="R"
FOR GROUPING "0" FORM NUMBER: 1 A

FORM ITEM DIFFICULTY MEASURE OF .00 ADDED TO MEASURES
+-----+-----+-----+-----+-----+-----+-----+
|CATEGORY|OBSERVED|OBSVD|SAMPLE|INFIT|OUTFIT||ANDRICH|CATEGORY|
|LABEL|SCORE|COUNT|AVRGE|EXPECT|MNSQ|MNSQ||THRESHOLD|MEASURE|
+-----+-----+-----+-----+-----+-----+-----+
| 1 1 | 1 14| .00 .00| 1.00 1.00|| NONE | (-1.59)| 1
| 2 2 | 1 14| .00* .00| 1.00 1.00|| .00 | -.42 | 2
| 3 3 | 1 14| .00* .00| 1.00 1.00|| .00 | .42 | 3
| 4 4 | 1 14| .00* .00| 1.00 1.00|| .00 | ( 1.59)| 4
+-----+-----+-----+-----+-----+-----+
|MISSING| 3 43| .00 | | | | | |
+-----+-----+-----+-----+-----+-----+
AVERAGE MEASURE is mean of measures in category.

+-----+-----+-----+-----+-----+-----+
|CATEGORY|STRUCTURE|SCORE-TO-MEASURE|50% CUM.|COHERENCE|
| LABEL |MEASURE|S.E. | AT CAT. ----ZONE----|PROBABLY| M->C C->M|
+-----+-----+-----+-----+-----+-----+
| 1 | NONE | | (-1.59) -INF -1.01| | 0% 0%| 1
| 2 | .00 1.15 | -.42 -1.01 .00| -.61 | 50% 100%| 2
| 3 | .00 1.00 | .42 .00 1.01| .00 | 50% 100%| 3
| 4 | .00 1.15 | ( 1.59) 1.01 +INF | .61 | 0% 0%| 4
```

Form B:

```

+-----+
|CATEGORY  STRUCTURE | SCORE-TO-MEASURE | 50% CUM. | COHERENCE|
| LABEL    MEASURE  S.E. | AT CAT.  ---ZONE--- | PROBABILITY| M->C C->M|
+-----+
| 1        NONE      | (-1.09) -INF  -.63|          | 0%  0%| 1
| 2         1.10     .76 |  -.11  -.63   .28|  -.14 | 14% 100%| 2
| 3        -.69     .76 |   .70   .28  1.34|   .14 |  0%  0%| 3
| 4         .70     1.08 | ( 2.02)  1.34 +INF |  1.02 |  0%  0%| 4
+-----+

```

19.76 Order of elements in Control file

Element	Function	Status
&INST		optional, for backwards compatibility only
TITLE=	title of analysis	recommended
ITEM1=	starting column of items	Required
NI=	number of items	Required
ISGROUPS=	grouping information	optional, with GRPFRM=N (the standard)
MODELS=	model information	optional, with MODFRM=N (the standard)
RESCORE=	rescore information	optional, with RESFRM=N (the standard)
KEY1=	key information	optional, if KEYFRM= omitted (the standard)
KEY2= ..	(n=1 to 99, number of largest key)	optional, if KEYFRM= omitted (the standard)
KEYn= ..		optional, if KEYFRM= omitted (the standard)
other control variables		optional
; comments		optional
&END		Required
ISGROUPS=	(in data file format)	required if GRPFRM=Y
MODELS=	(in data file format)	required if MODFRM=Y
RESCORE=	(in data file format)	required if RESFRM=Y
KEY1=	(in data file format)	required if KEYFRM=1 or more
KEY2= ..	(in data file format)	required if KEYFRM=2 or more, and so on up to
KEYn=	(in data file format)	required if KEYFRM=n
Item Names	(expects NI= names)	expected if INUMB=N (the standard)
END NAMES		required if INUMB=N (the standard)
data records		required if DATA=" " (the standard)

19.77 Partial Credit model

The "Partial Credit" Rasch model was devised for multiple-choice questions in which credit is given for almost-correct distractors. But there is no reason to believe that the almost-correctness of distractors to different questions is the same. Consequently, each item is modeled to have its own response structure.

This model was extended to any questionnaire using ordered polytomies in which the response structure is modeled to be unique to each item.

Winsteps estimates response structures by item groupings, using ISGROUPS=. From this perspective, the Andrich Rating Scale Model includes all items in one grouping. The Masters Partial Credit Model allocates each item to its own grouping.

The conventional representation of the Partial Credit model is

$$\log (P_{nij} / P_{ni(j-1)}) = B_n - D_{ij} = \theta - \delta$$

Winsteps parameterizes D_{ij} as $D_i + F_{ij}$ where $\sum(F_{ij}) = 0$. And D_i is the average (D_{ij}).

$$\log (P_{nij} / P_{ni(j-1)}) = B_n - D_i - F_{ij}$$

Algebraically these two representations are identical.

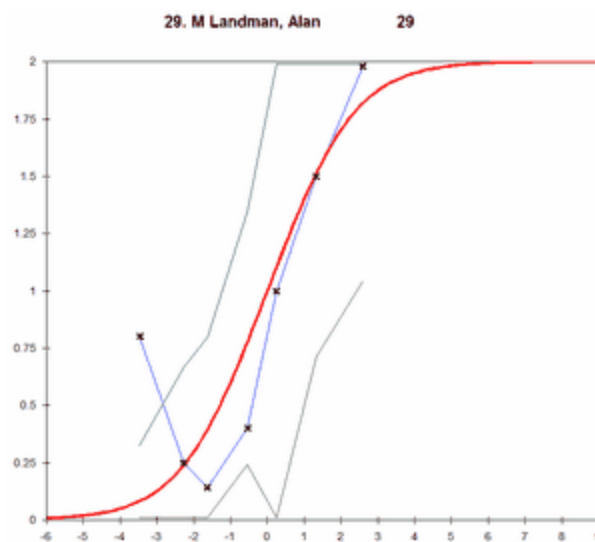
Thus every item has a mean difficulty, D_i . This simplifies communication, because the results of a Partial Credit analysis now have the same form as any other polytomous analysis supported by Winsteps.

19.78 Person Response Functions (PRF)

To produce [Person Response Functions](#) (PRF) or Person Characteristic Curves using Winsteps,

1. Transpose the scored data matrix: Winsteps Output File menu. [Transpose](#). So that the persons become the items.
2. [ISGROUPS=0](#) in the transposed file if you want to model each person to have a unique rating scale.
3. Analyze the transposed file.
4. The [Graphs menu](#) now produces person curves, not item curves, but with a reversed x-axis.
5. To smooth the PRF, which is called Hanning, adjust the [Smoothing](#) slider

Here is an example from [Example0.txt](#)



19.79 Person-free and Item-free measurement

A strength of Rasch methodology is that the estimates of Rasch measures ("Rasch scores", thetas and deltas) are person-free and item-free.

Person-free: as much as is statistically possible, the item-difficulty estimates are independent of the particular sample of persons from a homogeneous population that are used in the estimation. The actual distribution of the items and persons is irrelevant.

Item-free: as much as is statistically possible, the person-ability estimates are independent of the particular sample of items from a homogeneous population that are used in the estimation. The actual distribution of the items and persons is irrelevant.

Since each analysis has its own zero-point (frame-of-reference), we need to [equate or link](#) the tests to put all the analyses in the same frame-of-reference.

Ben Wright's (1967) demonstration of person-free and item-free estimation:

Item-free person-ability estimates:

1. Analyze a large dataset.
2. Split the items into easy items (low difficulties) and hard items (high difficulties)
3. Analyze each set of items separately.
4. Identify items and persons with extreme (zero, perfect) scores and remove from the analyses. Remove persons with extreme scores in one analysis also from the other analysis.
5. Reanalyze both sets of items.
6. Cross-plot the person estimates. They should form a fuzzy diagonal line. The Winsteps [scatterplot](#) includes confidence bands to show the statistical similarity of the two sets of person ability estimates.

Person-free item-difficulty estimates:

1. Analyze a large dataset.
2. Split the persons into high-ability persons and low-ability persons.
3. Analyze each set of persons separately.
4. Identify items and persons with extreme (zero, perfect) scores and remove from the analyses. Remove items with extreme scores in one analysis also from the other analysis.
5. Reanalyze both sets of persons.
6. Cross-plot the item estimates. They should form a fuzzy diagonal line. The Winsteps [scatterplot](#) includes confidence bands to show the statistical similarity of the two sets of item difficulty estimates.

Wright, B. D. (1967). Sample-Free Test Calibration and Person Measurement. In B. S. Bloom (Chair), Invitational Conference on Testing Problems (pp. 84-101). Princeton, NJ: Educational Testing Service. Available at www.rasch.org/memo1.htm.

19.80 Pivot anchoring

There is usually no problem defining the item difficulty of a standard dichotomous (right/wrong) item. It is the location on the latent variable where there is a 50% chance of success on the item.

Combining dichotomous items makes a polytomous super-item. But how do we define the difficulty of a super-item? Since the difficulty of a dichotomous item is the location on the latent variable where the top and bottom categories are equally probably (= 0.5), we apply the same logic to the super-item. Its difficulty is the location on the latent variable where the top and bottom categories are equally probably (= ?). But this definition does not make sense in every situation. So we may need to choose another definition.

For instance, if a super-item is a combination of 3 dichotomous items (so that its possible scores = 0,1,2,3), we might define its difficulty as the location on the latent variable where the expected score on the super-item is 1.5. Or the location on the latent variable where scores of 1 and 2 are equally probable. Or the location on the latent variable where the expected score is 1.0 or maybe 2.0. Or

For these alternative definitions, we need to compute the distance of the chosen location from the standard location and then apply that distance to the item difficulty using "pivot anchoring" implemented in Winsteps with `SAFILE=`. We can usually discover the distance we want by looking at the `GRFILE=` output.

The procedure is:

- (1) Analyze the data without pivot-anchoring
 - (2) Output `SFILE=sf.txt` which contains the standard Andrich thresholds
 - (3) Output `GRFILE=gr.txt` which contains the values connected with all the scores and probabilities on the item
 - (4) Identify the logit value corresponding to the desired location on the latent variable = M
 - (5) Subtract M from all the values for the super-item in `SFILE=sf.txt`
 - (6) The adjusted `SFILE=` is now specified as `SAFILE=sf.txt`, the pivot-anchor file
 - (7) Analyze the data with pivot-anchoring
- (The difficulty of the super-item should now have changed by the specified value, M

Pivots are the locations in the dichotomy, rating (or partial credit) scale at which the categories would be dichotomized, i.e., the place that indicates the transition from "bad" to "good", "unhealthy" to "healthy". Ordinarily the pivot is placed at the point

where the highest and lowest categories of the response structure are equally probable. Pivot anchoring redefines the item measures. The effect of pivot anchoring is to move the reported difficulty of an item relative to its rating scale structure. It makes no change to the fit of the data to the model or to the expected observation corresponding to each actual observation.

Dr. Rita Bode's procedure works well. The idea is to align the item difficulties so that the cut-point for each item (equivalent to the dichotomous item difficulty) is located at the reported item difficulty on the latent variable. So, we do an arithmetic sleight of hand. In the original analysis, we look at a Table such as [Table 2.2](#) and see where along the line (row) for each item is the substantive cut-point (pass-fail point, benchmark, etc.) We note down its measure value on the latent variable (x-axis of Table 2.2).

Then, for each item, we compare the measure value new item difficulty with its reported item difficulty. The difference is the amount we need to shift the Andrich thresholds for that item. Here is the computation:

$$\text{New average thresholds (excluding bottom "0")} = \text{Old item difficulty} + \text{Old average thresholds (excluding bottom "0")} - \text{Provisional New item difficulty}$$

If the analysis is unanchored, Winsteps will maintain the average difficulty of the items:

$$\text{New item difficulty} = \text{Provisional New item difficulty} - \text{Average(Provisional New item difficulty)} + \text{Average(Old item difficulty)}$$

Person measures with complete response strings will usually have very small or no changes.

Since this can become confusing. It is usually easiest to :

- 1) output the SFILE= from the original analysis to Excel.
- 2) add (item difficulty - cut-point) to the threshold values for each item. Example: we want to subtract 1 logit from the item difficulty to move the item difficulty to the cut-point on the latent variable. We add 1 logit to all the thresholds for an item, then Winsteps will subtract 1 logit from the item's difficulty.
- 3) Copy-and-paste the Excel SFILE= values into the Winsteps control file between SAFILE=* and *
- 4) Since each item now has different threshold values: [ISGROUPS=0](#)
- 5) This procedure should make no change to the person measures.

In Rita Bode's approach, we have a target item difficulty ordering. Usually most items are already in that order, but a few items are out of order. These out-of-order item need to be pivot-anchored to place them correctly in the item hierarchy.

- i) Output Table 13, the original ordering of the items.
- ii) Move the item rows up and down to give the desired ordering.
- iii) For items that are already in order (usually more than half the items), there is no change to SFILE= in the SAFILE=.
- iv) For the other items, change the SFILE= values enough to locate those items in the correct position. Example, to increase the item difficulty by one logit, decrease the thresholds from the SAFILE= to the SAFILE= by one logit.

See also [SAFILE=](#). PIVOT= was an earlier, unsuccessful attempt to automate this procedure.

For polytomies:

1) from your original analysis, with your GROUPS= (if any) and no SAFILE=, output an SFILE=

2) build an SAFILE=

a) for each item, use the SFILE= value for its group

b) add the pivot anchor value to the SFILE= value

c) include the new set of values for the item in the SAFILE=. There must be entries in SAFILE= for every item. Use the SFILE= values directly if there is no change

Example 1: SFILE= for the group with item 1:

```
1 0 .00
1 1 -.85
1 2 .85
```

We want to add one logit for item 1, two logits for item 2, no change for item 3

```

SAFILE=*
1 0 1.00 ; this is a placeholder, but is convenient to remind us of the pivot value
1 1 0.15
1 2 1.85
2 0 2.00
2 1 2.15
2 2 2.85
3 0 0.00
3 1 -.85
3 2 0.85
*

```

3) do the analysis again with ISGROUPS=0 and SAFILE=* ... *

4) The person measures will shift by the average of the pivot values.

Example 2: We can plot the IRF/ICC ogives for every PCM item on a graph. This gives a complete story of expected responses on all the items for each person on the x-axis, but it does not give us a useful depiction of the latent variable. When explaining the latent variable in terms of an item hierarchy, we need to take a horizontal slice across all the PCM ICCs/IRFs somewhere around the middle of the rating scale. Pivot anchoring lets us do this.

For instance, the rating scale may be a Likert scale from 1 to 5 with 4=Agree. We may decide that the "pivot" point on the IRF of each item is where the expected score = 4 (which is also the point where category 4=Agree has the highest probability of being observed). So we need to redefine the item difficulty of each PCM item away from its default value (the location where the highest and lowest categories are equal probable) to the location where the expected score on the item = 4. We do this using pivot-anchoring. Here is the procedure:

1. Do a standard PCM analysis ([ISGROUPS=0](#)). For each item, Table 3.2, etc., reports the item difficulty + Andrich thresholds: $D_i + \{F_{ij}\}$
2. Output [IFILE=](#) the item difficulty for each item = D_i . Example: $D=2.5$ logits
3. Output [ISFILE=](#) the location on the latent variable where the expected score on the item is 4 = E_i . Example $E_i = 3.0$ logits
4. Compute the desired shift of location for each item = $E_i - D_i = 3.0 - 2.5 = 0.5$ logits
5. Output [SFILE=](#) the Andrich thresholds (relative to the item difficulties) for each item = $\{F_{ij}\}$. Example: $\{-2, -1, 0, 3\}$
6. Subtract the shift from the Andrich thresholds $\{S_{ij}\} = \{F_{ij} - (E_i - D_i)\}$. Example: $\{S_{ij}\} = \{-2.5, -1.5, -0.5, -2.5\}$
7. Anchor (fix) the Andrich thresholds [SAFILE=](#) for each item at their $\{S_{ij}\}$ values.
8. Rerun the analysis. Item difficulties, D'_i , are now located at the points where the expected scores on the items are 4. Example: $D_i + F_{ij} = D'_i + S_{ij}$ for each threshold of each item. In our example:
 $D'_i = D_i + F_{ij} - S_{ij} = 2.5 + \{-2, -1, 0, 3\} - \{-2.5, -1.5, -0.5, -2.5\} = 2.5 + \{0.5, 0.5, 0.5, 0.5\} = 3.0$, the desired location.

19.81 Plausible values

Plausible values are estimates intended to represent the distribution of measures that could produce the observed scores. They were developed for large-scale educational assessments from which group-level measures are to be obtained, but with data too thin to support individual-level measurement.

Winsteps is designed for individual measurement. When this is possible, then group-level reporting can be done, e.g., with [PSUBTOT=](#). The Winsteps estimate approximates the mean of the plausible-value distribution.

Plausible values are values from the error distribution of the estimate. Winsteps reports each person's estimate (measure) and its standard error. Plausible values are values selected at random from a normal distribution with its mean at the estimated measure and with standard deviation equal to the standard error. You can generate these with Excel or other statistical software.

19.82 Poisson counts

Conceptually, the Poisson Counts model has an infinite number of categories, 0, 1, 2, onwards. In practice, due to time limits and other constraints, there is a finite maximum possible. If that maximum is much higher than the highest observed category, then the Poisson model continues to apply from a practical perspective. If the observed highest is close to the absolute maximum, then use a constrained rating scale. In Winsteps, do this with

[ISRANGE=](#)

to set the 0 to absolute maximum range of categories, and

[SFUNCTION=](#)

to set the function controlling the Andrich thresholds

SFUNCTION=2 is a good starting value.

The Winsteps program can analyze Poisson count data, with a little work. Poisson counts are a very long rating (or partial credit) scale with pre-set structure. The Andrich Thresholds are $\log_e(n)$, $n=1$ upwards. You can define a structure anchor file in this way:

```

XWIDE=2
STKEEP=YES
CODES = 00010203040506070809101112131415161718192021222324252627282930313233343536373839+
+40414243444546474849505152535455565758596061626364656667686970717273747576777879+
+8081828384858687888990919293949596979899
SAFILE=*
0 0 ; placeholder for the bottom count
1 0 ; the value corresponding to log(1) - the pivot point for the item measure
2 .693 ; the value corresponding to loge(2)
3 1.099 ; the value corresponding to loge(3)

```

4 1.386	28 3.332	52 3.951	76 4.331
5 1.609	29 3.367	53 3.970	77 4.344
6 1.792	30 3.401	54 3.989	78 4.357
7 1.946	31 3.434	55 4.007	79 4.369
8 2.079	32 3.466	56 4.025	80 4.382
9 2.197	33 3.497	57 4.043	81 4.394
10 2.303	34 3.526	58 4.060	82 4.407
11 2.398	35 3.555	59 4.078	83 4.419
12 2.485	36 3.584	60 4.094	84 4.431
13 2.565	37 3.611	61 4.111	85 4.443
14 2.639	38 3.638	62 4.127	86 4.454
15 2.708	39 3.664	63 4.143	87 4.466
16 2.773	40 3.689	64 4.159	88 4.477
17 2.833	41 3.714	65 4.174	89 4.489
18 2.890	42 3.738	66 4.190	90 4.500
19 2.944	43 3.761	67 4.205	91 4.511
20 2.996	44 3.784	68 4.220	92 4.522
21 3.045	45 3.807	69 4.234	93 4.533
22 3.091	46 3.829	70 4.248	94 4.543
23 3.135	47 3.850	71 4.263	95 4.554
24 3.178	48 3.871	72 4.277	96 4.564
25 3.219	49 3.892	73 4.290	97 4.575
26 3.258	50 3.912	74 4.304	98 4.585
27 3.296	51 3.932	75 4.317	99 4.595
			*

Arrange that the observations have an upper limit much less than 99, or extend the range of [CODES=](#) and [SAFILE=](#) to be considerably wider than the observations. Winsteps can go up to category 32767.

Use [UASCALE=](#) to multiply all Poisson Andrich Thresholds by a constant to adjust the "natural" form of the Poisson counts to the actual discrimination of your empirical Poisson process, or do the multiplication yourself, which can be different for different items. You need to adjust the constant so that the average overall mean-square of the analysis is about 1.0. See [RMT 14:2](#) about using mean-squares to adjust logit user-scaling. (The *Facets* program does this automatically, if so instructed.)

But my experience with running Poisson counts in the Facets program (which supports them directly) is that most "Poisson count" data do not match the Poisson process well, and are more usefully parameterized as a rating (or partial credit) scale. There is nearly always some other aspect of the situation that perturbs the pure Poisson process, especially by placing a ceiling value on the observations.

Theory: See [Wikipedia](#). The Poisson model is $P(k \text{ events in an interval}) = e^{-\lambda} \lambda^k / k!$ so that $P(k)/P(k-1) = \lambda/k$, which is implemented in Winsteps as $\ln(P(k)/P(k-1)) = B - D - \ln(k)/UASCALE$ where B is the person ability, D is the item difficulty, and $1/UASCALE$ is the Poisson-scale discrimination parameter.

Binomial trials: same as above, with $\log_e(n(m-n+1))$ where m is a fixed number of trials and n is the number of successes for $n=0$ to m, so that there are m thresholds. We can use [UASCALE=](#) to adjust for scale discrimination as above.

Negative (inverse) binomial trials: same as above where m is the number of trials and n is a fixed number of successes. For convenience, enter the data as x = (m-n) = number of failures, which will go from 0 to a large number (similar to Poisson counts above). Again we can pre-compute the Andrich thresholds for every observation = log (probability of observing x failures / probability of observing x-1 failures) when we are targeting n successes. If the number of successes is constant for each item, then we can use Winsteps. If it varies across observations within an item, we must use *Facets*. We can use UASCALE= to adjust for scale discrimination as above.

19.83 Polytomous mean-square fit statistics

For a general introduction, see [Diagnosing Misfit](#), also [Dichotomous mean-square fit statistics](#).

Response String Easy.....Hard	INFIT Mean-square	OUTFIT Mean-square	Point-measure correlation	Diagnosis
I. modeled:				
33333132210000001011	0.98	0.99	0.78	Stochastically
31332332321220000000	0.98	1.04	0.81	monotonic in form
33333331122300000000	1.06	0.97	0.87	strictly monotonic
33333331110010200001	1.03	1	0.81	in meaning
II. overfitting (muted):				
3322222221111111100	0.18	0.22	0.92	Guttman pattern
3333322221111100000	0.31	0.35	0.97	high discrimination
3222222221111111110	0.21	0.26	0.89	low discrimination
32323232121212101010	0.52	0.54	0.82	tight progression
III. limited categories:				
3333333332222222222	0.24	0.24	0.87	high (low) categories
2222222221111111111	0.24	0.34	0.87	central categories
3333322222222211111	0.16	0.2	0.93	only 3 categories
IV. informative-noisy:				
3222222201111111130	0.94	1.22	0.55	noisy outliers
33233332212333000000	1.25	1.09	0.77	erratic transitions
33133330232300101000	1.49	1.4	0.72	noisy progression
33333333330000000000	1.37	1.2	0.87	extreme categories
V. non-informative:				
2222222222222222222	0.85	1.21	0	one category
12121212121212121212	1.5	1.96	-0.09	central flip-flop
01230123012301230123	3.62	4.61	-0.19	rotate categories
03030303030303030303	5.14	6.07	-0.09	extreme flip-flop
03202002101113311002	2.99	3.59	-0.01	random responses
VI. contradictory:				
11111122233222111111	1.75	2.02	0	folded pattern *
11111111112222222222	2.56	3.2	-0.87	central reversal

22222222223333333333	2.11	4.13	-0.87	high reversal
00111111112222222233	4	5.58	-0.92	Guttman reversal
00000000003333333333	8.3	9.79	-0.87	extreme reversal

from [Smith R.M. \(1996\) Rasch Measurement Transactions 10:3 p. 516](#)

The z-score standardized statistics report, as unit normal deviates, how likely it is to observe the reported mean-square values, when the data fit the model. The term z-score is used of a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's [t-statistic](#) value has been adjusted to a unit normal value.

* "folded data" can often be rescued by imposing a theory of "not reached" and "already passed" on to the observations. For instance, in archaeological analysis, the absence of bronze implements can mean a "stone age" or an "iron age" society. A useful recoding would be "1" = "stone age", "2" = "early bronze", "3" = "bronze", "2=>4" = "late bronze", "1=>5" = "iron age". This can be done iteratively to obtain the most self-consistent set of 4's and 5's. (Folding is discussed in Clive Coombes' "A Theory of Data".)

Example: Identify persons who respond in the same category to every item:
Anchoring all the items at the same difficulty and using the Rating Scale Model. Then the distribution of the mean-square fit statistics will indicate the extent to which all items are rated the same. We expect strings like 333333333333333333 to have zero mean-square fit statistics. Set NAME1= and NAMLEN= so that the response strings are also the person labels.

If the responses start in column 1 and there are 9 items:

```
NAME1 = 1
NAMLEN = 9
ISGROUPS=
IAFILE=*
1-9 0
*
```

FITP=0 ; show everyone in Table 6.

Output PFILE= to Excel, then you can draw a histogram of the mean-square fit statistics.

For instance: www.ablebits.com/office-addins-blog/2016/05/11/make-histogram-excel/

19.84 Printing pretty tables and graphs

Please see [ASCII=](#) to print Tables in different format. See [Graph window-bottom](#) right for graphs.

19.85 Prior distributions

Question: We pre-tested some polytomous items under one timing condition and have now collected a very small sample of that same data under another timing condition. I would like to be able to use the item/step difficulties from the first condition as priors to estimate item/step difficulties for the second condition.

Answer: Winsteps does not accept prior distributions, so we need to simulate them as additional data.

For instance, we could use the pre-test dataset, down-weight it using [PWEIGHT=](#), and analyze it again together with the new data.

When we have a suitable set of item/step difficulties, then export them with [IFILE=if.txt](#), [SFILE=sf.txt](#)

Then analyze only the new data with anchored items and steps [IAFILE=if.txt](#), [SAFILE=sf.txt](#) to get a report on only the new data.

19.86 Probabilities from measures

Request: I'm trying to obtain the probabilities by working backwards from the parameters Winsteps estimated, to see if probabilities I calculate correspond to those used to estimate the expected value for each person-item encounter.

Reply:

Yes, if you have the reported person ability, item difficulty, and the Rasch-Andrich threshold values (rating scale structure), then you can calculate exactly the same probabilities as Winsteps.

For dichotomous responses,

Probability of 0 = $1 / (1 + \exp(\text{ability-difficulty}))$

Probability of 1 = $1 / (1 + \exp(\text{difficulty-ability}))$

These probabilities always sum to 1.

An easy way to do this for polytomous (rating-scale, partial-credit) responses is:

Value of bottom category $v_0 = 1$

Value of next category $v_1 = v_0 * \exp(\text{ability} - \text{difficulty} - \text{threshold}_1)$

Value of next category $v_2 = v_1 * \exp(\text{ability} - \text{difficulty} - \text{threshold}_2)$

....

Value of top category $v_m = v_{(m-1)} * \exp(\text{ability} - \text{difficulty} - \text{threshold}_m)$

then

Probability of category 0 = $p_0 = v_0 / \text{sum}(v_0+v_1+\dots+v_m)$

Probability of category 1 = $p_1 = v_1 / \text{sum}(v_0+v_1+\dots+v_m)$

....

Probability of category m = $p_m = v_m / \text{sum}(v_0+v_1+\dots+v_m)$.

These probabilities always sum to 1.

Please compare your probabilities with the [graphed](#) values.

19.87 Quality-control misfit selection criteria

Rasch measurement does not make any presumptions about the underlying distribution of the parameters. Maximum likelihood estimation expects "errors" in the observations to be more or less normally distributed around their expected values. Since all observations are integral values, this expectation can be met only asymptotically as the number of persons and items becomes infinite. The information-weighted fit statistic, "infit", and the outlier-sensitive fit statistic, "outfit", are described in *BTD* and *RSA*. Possible values, and hence interpretation, of these statistics is influenced by the observed distribution the person and item statistics. This is particularly true of their t standardized values which are designed to follow standard normal (0,1) distributions. The local significance of these statistics is best interpreted in terms of their means and [standard deviations](#) reported in Table 3.1. Start investigating the misfit causing the most extreme values of these statistics, and stop your investigation when the observed responses become coherent with your intentions.

The fit statistics reported will not exactly match those printed in *BTD* or *RSA*, or those produced by another program. This is because the reported values of these statistics are the result of a continuing process of development in statistical theory and practice. Neither "correct" fit statistics nor "correct" values exist, but see the Appendices for guidance.

Report measure in Tables 6 ([FITP=](#)) and Table 10 ([FITI=](#)) if any of:

Statistic	Less than	Greater than
t standardized INFIT	-(FITP or FITI)	FITP or FITI
t standardized OUTFIT	-(FITP or FITI)	FITP or FITI
mean-square INFIT	$1 - (\text{FITP or FITI})/10$	$1 + (\text{FITP or FITI})/10$
mean-square OUTFIT	$1 - (\text{FITP or FITI})/10$	$1 + (\text{FITP or FITI})/10$
point-biserial correlation	negative	

To include every person, specify FITP=0. For every item, FITI=0.

For Table 7, the diagnosis of misfitting persons, persons with a t standardized fit greater than FITP= are reported. Selection is based on the OUTFIT statistic, unless you set OUTFIT=N in which case the INFIT statistic is used.

For Table 11, the diagnosis of misfitting items, items with a t standardized fit greater than FITI= are reported. Selection is based on the OUTFIT statistic, unless you set OUTFIT=N in which case the INFIT statistic is used.

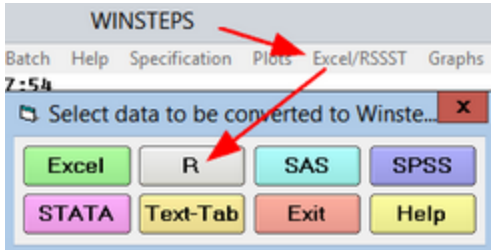
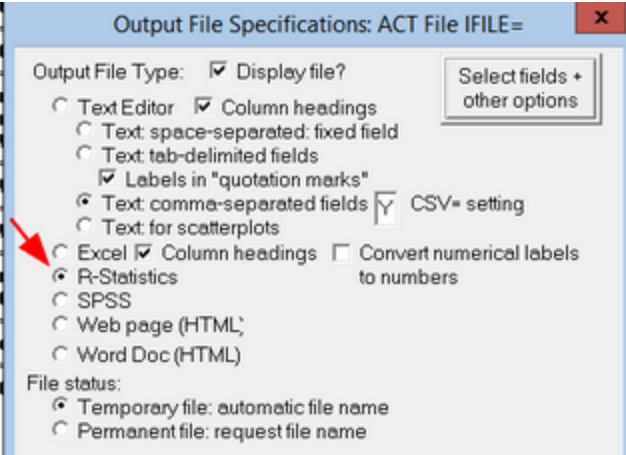
19.88 R Statistics

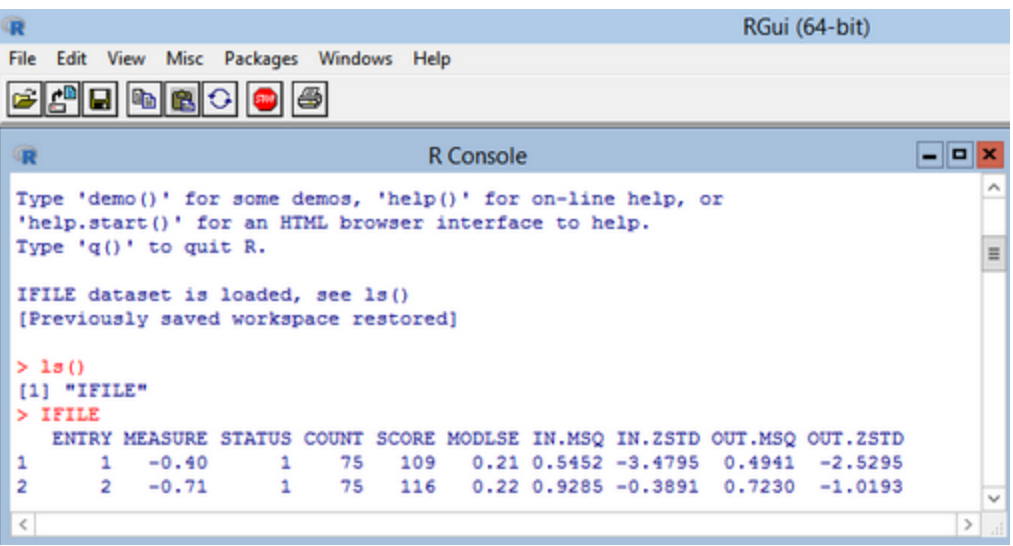
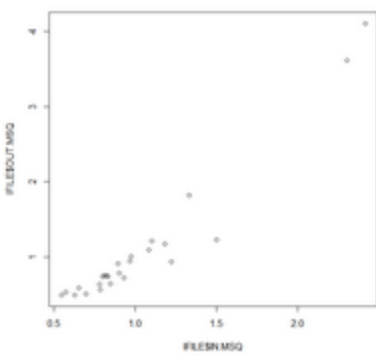
Winsteps can [import](#) and [export](#) files compatible with the R Statistics package.

The freeware R statistics package can be downloaded from www.r-project.org. Paul Murrell has posted useful instructional material online: <http://www.stat.auckland.ac.nz/~paul/> - especially good for graphing.

R command	R's response	Meaning
ls()	[1] "IFILE"	list of loaded datasets
help(ls)	Help windows opens	help for "ls" command
names(IFILE)	[1] "ENTRY" "NAME"	names of variables in the dataset "IFILE"
IFILE	(contents of IFILE)	displays the data in the dataset
attach(IFILE)	-	IFILE variables active without \$ reference
plot(x, y)	scatterplot displays	variable y is plotted against variable x
q()	(R terminates)	quit (exit from) R Statistics

An R package to interface with Winsteps is also available from CRAN: Rwinsteps (Albano and Babcock) - cran.r-project.org/web/packages/Rwinsteps/index.html

Action:	Screen:
<p>Import R statistics data in a .rdata or .rda file.</p> <p>Now follow the procedure at Excel/RSSST menu.</p>	
<p>Export Winsteps output files as R statistics files: .rdata or .rda</p> <p>from the Output Files menu</p>	

<p>R Statistics screen</p> <p>ls() at any time to see list of available objects, etc.</p>	 <pre> RGui (64-bit) File Edit View Misc Packages Windows Help R Console Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help. Type 'q()' to quit R. IFILE dataset is loaded, see ls() [Previously saved workspace restored] > ls() [1] "IFILE" > IFILE ENTRY MEASURE STATUS COUNT SCORE MODLSE IN.MSQ IN.ZSTD OUT.MSQ OUT.ZSTD 1 1 -0.40 1 75 109 0.21 0.5452 -3.4795 0.4941 -2.5295 2 2 -0.71 1 75 116 0.22 0.9285 -0.3891 0.7230 -1.0193 </pre>
<p>simple plot with R</p> <pre>attach(IFILE) plot(IN.MSQ, OUT.MSQ) or plot(IFILE\$IN.MSQ, IFILE\$OUT.MSQ)</pre>	

19.89 Random number generator

For generating pseudo-random numbers, used in [simulating data](#), Winsteps implements a FORTRAN 77 adaptation of this "Mersenne Twister" code:

- c A C-program for MT19937, with initialization improved 2002/1/26.
- c Coded by Takuji Nishimura and Makoto Matsumoto.
- c Copyright (C) 1997 - 2002, Makoto Matsumoto and Takuji Nishimura,
- c All rights reserved.
- c Copyright (C) 2005, Mutsuo Saito,
- c All rights reserved.

Their website is www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/emt.html

The generator is [seeded](#) with the system time or a value chosen by the analyst.

Simulated data observations.

Winsteps simulates data using the estimated values of person ability, item difficulty and rating-scale thresholds. Winsteps implements www.rasch.org/rmt/rmt213a.htm for dichotomous and polytomous data.

If you use these simulated data to estimate the parameter values, then the estimated values will be somewhat wider than the generators. This is due to JMLE bias. It can usually be corrected with [STBIAS=](#)YES.

19.90 Rank order data

Rankings and partial rankings, with or without ties, can be conveniently analyzed using [ISGROUPS=0](#) and [STKEEP=No](#)

Each **row** is an element or object to be ranked.

Each **column** is a set of rankings by a respondent. In the item label, place any interesting demographics about the respondent doing the ranking.

ISGROUPS=0 - Each respondent (column) has their own "ranking scale". This is equivalent to the Partial Credit model.

STKEEP=No - tied ranks are OK, but if the ranking goes: 1, 2, 2, 4, 5, in the data then, from a Rasch perspective this should be: 1, 2, 2, 3, 4, which is what STKEEP=NO does for us.

The elements (rows) will have measures and fit statistics indicating respondent preference. The fit statistics for the respondents (columns) will indicate the extent of agreement of each respondent with the consensus. The measures for the respondents are usually meaningless (or the same) and can be ignored.

Fit statistics, [DIF](#) and [DPF](#) analysis, and [contrast analysis](#) of residuals are all highly informative.

Elements rated by only one respondent will be correctly measured (with large standard errors), except if that element is ranked top or bottom by that respondent. When that happens, construct a dummy ranking column in which the extreme elements are not-extreme. Give that ranking a low weight with [IWEIGHT=](#).

If every ranking set includes every element, and ties are not allowed, then the elements can be columns and the respondents can be rows. ISGROUPS=0 is not required.

Example:

In the data, as collect, the ranked objects are columns, the respondents are rows, and the rankings as numbers.

Set up a standard rating-scale analysis of these data.

Run an analysis in Winsteps. Ignore the results except to verify that the data have been input correctly.

Now we want the respondents as columns and the objects as columns:

Winsteps "output files", "transpose". [TRPROFILE=](#)

Edit the transposed control file with ISGROUPS=0 and STKEEP=NO

Now run the Winsteps analysis. The results should make sense. The objects (rows) should have measures and fit statistics indicating respondent preference. The fit statistics for the respondents (columns) will indicate the extent of agreement of each respondent with the consensus.

19.91 Rasch - Why use Rasch methodology?

William Boone discusses "Rasch Analysis for Instrument Development: Why, When, and How?" in a Journal article: www.ncbi.nlm.nih.gov/pmc/articles/PMC5132390/

In general, Rasch analysis constructs a unidimensional latent variable from the data. It verifies that each person and item can be placed on it, reporting the persons and items locations by means of linear (additive, interval) measures - the type of measures taken for granted in most scientific and day-to-day activities. Rasch identifies departures in the data for persons, items and even data points from the ideal of unidimensionality. These are reported with fit statistics that guide the improvement of the instrument and point out possible flaws in the data. Rasch also has techniques for identifying multi-dimensionality in the data (PCA of residuals).

Raw scores? Rasch confirms that raw scores mean what we think they mean. For instance, that a higher score indicates more of the latent variable than a lower score. Raw scores are non-linear. They have strong ceiling and floor effects, so Rasch also tells us how much more of the latent variable 1 more score-point indicates at different places along the latent

variable. This can have important consequences when change-scores influence decision-making. A 10-point gain near the ends of the raw score range can be 4 times more change along the latent variable than a 10-point gain near the center of the raw score range - www.rasch.org/memo62.htm

Factor analysis? Factor analysis is a useful technique, but is not as exact as Rasch for our data. Our intention is to construct a unidimensional latent variable from the data, not to describe the data with all its variety and intricacies. We need to identify specific areas in the data that do not aid in this construction. Then remedy them, eliminate them or decide that flaws in the data are inconsequential. For instance, in an arithmetic test, factor analysis may point out that "addition" and "subtraction" items are different factors, but provides little help in the decision as to whether "arithmetic" can be treated as one construct. Factor analysis also tends to assign items in different difficulty strata (nodes) to different factors. This has produced misleading findings, see www.rasch.org/rmt/rmt81p.htm - When we have a clear picture of the data and the latent variable, factor analytic techniques, such as PCA of residuals, can be helpful to identify subtle departures in the data from the ideal. Factor analysis may assist in identifying clusters of items which threaten the invariance of the measurement system, but this is indirect and inexact compared with Rasch-based identification of anomalies in the data.

Reliability? This applies equally to Rasch or Classical analysis: With 53 dichotomous items and a reasonable range of person abilities, we would hope the reliability is around 0.9. This gives us about 4 statistically-different measurement strata - www.rasch.org/rmt/rmt63i.htm - So, if this set of items is to be used for decision-making, and this person sample is representative, how many levels of performance need to be identified? If 4 levels, then around 53 items. If only high-low, then 25 items are probably enough.

Invariance? Rasch is rigorous in applying the ordinary definition of invariance: "The property of remaining unchanged regardless of changes in the conditions of measurement." (Dictionary.com). Rasch examines every observation to discover the extent to which it is independent of which particular person and which particular item participates, apart from the person's ability and the item's difficulty. In general, Rasch measures are independent of which items and which persons participate in the measurement process. This is the expected situation for invariant measurement in the physical sciences. In principle, it does not matter which thermometer is used to measure heat, or what the source of the heat is. Rasch applies this to social science data, operationalizing L.L. Thurstone's ideals that (1) "A measuring instrument must not be seriously affected in its measuring function by the object of measurement." (Thurstone, 1928, p.547) and "It should be possible to omit several test questions at different levels of the scale without affecting the individual score (measure)." (Thurstone, 1926, p.446) - cited in www.rasch.org/memo62.htm

19.92 Rasch Model

The Rasch Model is a mathematical formula for converting raw scores (test scores, ratings) into linear, additive, interval measures on a line (unidimensional latent variable). Items and persons are positioned on the same line (conjoint measurement). The Rasch model was invented by Georg Rasch (1900-1980), a Danish mathematician around 1950.

19.93 Rasch Software

There are many software packages available for estimating the parameters of Rasch Models. Winsteps and Facets are two packages, each has thousands of users. Winsteps (and its predecessors) and Facets have both been in use for over 30 years and are under constant development.

19.94 Rating scale conceptualization: Andrich, Thurstone, half-point thresholds

(See [Table 1](#), [Table 3.2](#), [Table 12](#), [Table 21](#), [Graphs](#))

There are several ways of conceptualizing a rating scale item. They all contain exactly the same measurement information, but communicated in different ways. Usually, one of these alternatives will be most meaningful for your audience. The plots corresponding to these approaches are shown in Table 21, and also on the Graphs screen.

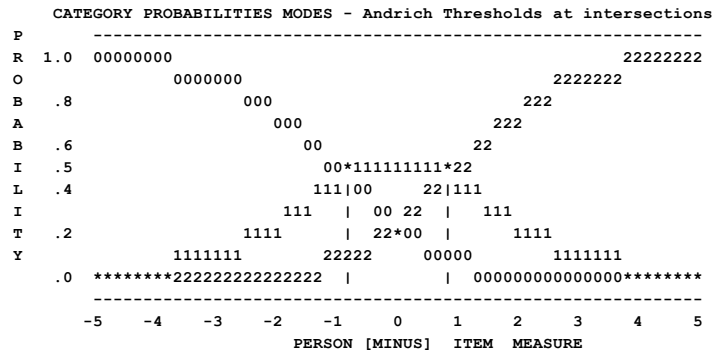
```

-----
|CATEGORY  OBSERVED|OBSVD  SAMPLE|INFIT  OUTFIT|| ANDRICH |CATEGORY|
|LABEL SCORE COUNT %|AVRGE  EXPECT|  MNSQ  MNSQ||THRESHOLD| MEASURE|
-----+-----+-----+-----+-----+-----+-----+-----+-----+
|  0   0   378  20|  -.87 -1.03|  1.08  1.19||  NONE  |(-2.07)| 0 Dislike
|  1   1   620  34|   .13  .33|   .85  .69||  -.86  | .00 | 1 Neutral
|  2   2   852  46|  2.24  2.16|  1.00  1.47||   .86  |( 2.07)| 2 Like
-----

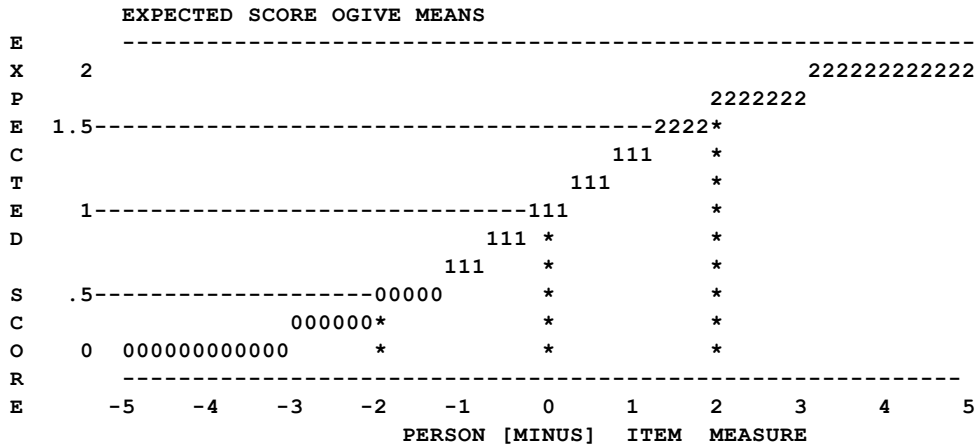
```


0. Distribution of frequencies of observation in each category. For a rating-scale for which inferences will be made at the category level, we like to see a uniform process at work. When applied to a sample, this would produce a smooth distribution of category frequencies. Unimodal without sharp peaks or troughs. No statistical test is intended, but merely Berkson's "inter-ocular traumatic test" (= what hits you between the eyes).

1. If you conceptualize the rating scale in terms of the probability of individual categories (Andrich's approach), then the Andrich Thresholds are of interest. The Andrich thresholds are the points at which adjacent categories are equally probable.



2. If you conceptualize the rating scale in terms of average ratings on the model (predicted) item characteristic curve (ICC), then "Category measures" are of interest. The "Category Measures" are the points on the latent variable at which the expected score on the item equals the category number. The Rasch-half-point thresholds define the ends of each category interval. These are shown in [Table 12.5](#).



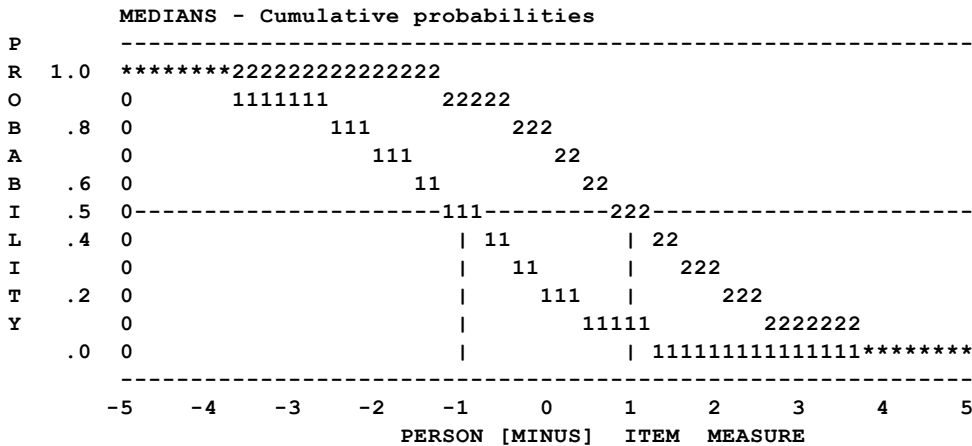
CATEGORY LABEL	STRUCTURE MEASURE	S.E.	SCORE-TO-MEASURE AT CAT.	50% CUM. PROBABILITY	COHERENCE	ESTIM
0	NONE		(-2.07)	-INF -1.19	62%	42%
1		.86 .07	.00	-1.19 1.19	54%	71% .73
2		.86 .06	(2.07)	1.19 +INF	1.00	85% 78% 1.19

3. If you conceptualize the rating scale in terms of the probability of accumulated categories (Thurstone's approach), then Rasch-Thurstone thresholds = "50% Cumulative Probabilities" are of interest. 50% Cum Probability is the point at which the probability of being observed in the categories below = the probability of being observed in this category or above. These are shown in [Table 12.6](#).

Thurstone thresholds are category boundaries on the latent variable, where we define "boundary" to mean "if someone has an ability measure exactly on a category boundary, then that person has a 50% chance of being observed in a category below the boundary (including categories down to the bottom of the rating scale) and a 50% chance of being observed in a category above the boundary (including categories up to the top of the rating scale)". This is the same definition that we

apply to the only category boundary for a dichotomous item, between 0 and 1. Someone exactly at the 0-1 boundary (the item difficulty) has this same 50-50 chance.

The Rasch-Thurstone thresholds also approximate the item difficulties when the rating scales are dichotomized between the categories below the target category and those at and above the target category. This is useful when the probability of "success" on a polytomous item must be computed. www.rasch.org/rmt/rmt233e.htm



19.95 Ratio scale of measures: twice as much

Question: I want to say that person A has twice the ability of person B. How can I do that with logit (or user-scaled) measures?

To say that "one person is twice as good as another" requires a reference point. Think of the same thing with mountains. Mountain A is twice as high as Mountain B. The reference point is "sea level".


Constructing a reference point: A useful approach is to add to the data a dummy person whose performance is at your hypothesized reference point. Include in the data ratings for this dummy person against real items. Weight the dummy person at 0 with [PWEIGHT=](#)

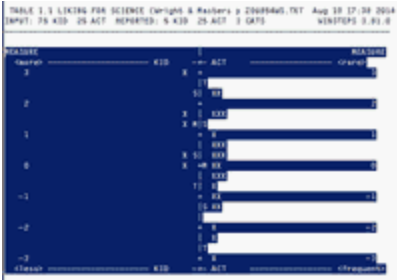



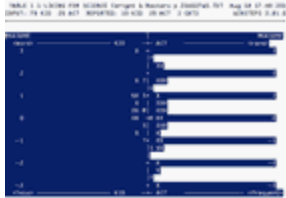


Similarly a dummy item at the item's reference point. Include in the data ratings for this dummy item against real persons. Weight the dummy person at 0 with [IWEIGHT=](#)

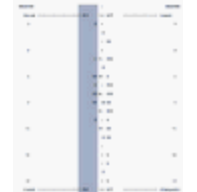


You can now use the person or item at the reference point as the basis for your "twice as much" statements.

For further convenience, you can anchor the dummy person at 0 with [PAFILE=](#) or the dummy item at 0 with [IAFILE=](#), but not both at the same time.

19.96 Rectangular copying

To copy a rectangle into a map or table with Microsoft Word:	
1.	Open a Word document 

2.	Select and copy the primary Winsteps Table	
3.	Paste into the Word document	
4.	Select all (ctrl+A) Change font to Courier New Font size: 8 (if smaller needed, type in the number)	
5.	Open another Word document	
6.	Select and copy the secondary Winsteps Table	
7.	Paste into the Word document	
8.	Select all (ctrl+A) Change font to Courier New Font size: 8 (if smaller needed, type in the number)	

9.	Rectangular-highlight (alt+mouse drag) the columns to be copied	
10.	Paste into the primary Word document. This may take several attempts to get right. ctrl+z to undo.	
11.	Change column headings Select all: Paragraph format: Single line: No spacing The extra columns should now be correctly inserted into the primary Table.	

To copy a rectangle of numbers:

1. Select the lines of text that include the rectangle of numbers.
 2. Copy the lines to the clipboard
 3. Paste the lines into a word-processing document or an Excel spreadsheet cell.
 4. Set the font of the lines to Courier.
- 5A. In Word, select rectangles with Alt+Mouse (see below)
5B. In TextPad or NotePad++, select rectangles with Alt+Mouse
5C. In WordPerfect, select "Edit > Select > Rectangle"
5D. In Excel, use "Data > Text to Columns" to select the column of numbers into a column.

You could also display the column of numbers on your computer screen and do a graphical copy. PrintScreen saves the screen to the clipboard, then paste into Paint and do a rectangle selection of what you want. Paste the selection into your document as a Figure.

Rectangular copy-and-paste with Microsoft Word

In Word, ctrl-A the whole document.
Select a "Courier New" font. Now everything lines up neatly in columns.
Click the left mouse button to un-highlight everything.
Move the mouse pointer to the top-left corner of the rectangle you want to copy.
Press and hold down the Alt-Key. Left-click and hold down the mouse. Release the Alt-Key
Drag down to the right-hand corner of what you want. The rectangle should high-light.
Release all keys and mouse-buttons.
Ctrl+C to copy the high-lighted section.
Move the mouse pointer to where you want the rectangle to go.
Ctrl+V to paste.
Add or delete blank spaces to line things up neatly

19.97 Reinstalling Winsteps

The best way to reinstall Winsteps is to run the original installation software, WinstepsPasswordInstall.exe

If this is no longer available:

1. Please download and install www.winsteps.com/ministep.htm
2. Delete ministep.exe and its shortcuts.
3. Copy and replace the folders c\Winsteps from your old computer or from backup to your new computer.
4. On your laptop, right-click on Winsteps.exe. Short-cut to desktop
6. Double click on the short-cut. Winsteps should run correctly.
7. Right-click on the short-cut. Properties. Change the "Start in" folder to your data folder.

19.98 Reliability and separation of measures

Winsteps report reliability and separation statistics treating the sample of measures as the population. If it is not the entire population, then the reliability and separation are slightly higher than the reported values.

Usually person and item reliability and separation have different applications and implications.

Person separation is used to classify people. Low person separation (< 2 , person reliability < 0.8) with a relevant person sample implies that the instrument may not be sensitive enough to distinguish between high and low performers. More items may be needed.

Item separation is used to verify the item hierarchy. Low item separation (< 3 = high, medium, low item difficulties, item reliability < 0.9) implies that the person sample is not large enough to confirm the item difficulty hierarchy (= construct validity) of the instrument.

Reliability (separation index) means "reproducibility of relative measure location". It does not report on the quality of the data. So "high reliability" (of persons or items) means that there is a high probability that persons (or items) estimated with high measures actually do have higher measures than persons (or items) estimated with low measures. If you want high reliability, you need a wide sample and/or low measurement error. So, if you want high person (test) reliability, you need a person sample with a large ability (or whatever) range and/or an instrument with many items (or long rating scales). If you want high item reliability, you need a test with a large item difficulty range and/or a large sample of persons. Usually low item reliability is because the person sample size is too small to establish a reproducible item difficulty hierarchy.

Missing data: if some persons have missing observations, these can considerably reduce precision, and so lower reliability estimates. Suggestion: omit person-records with missing data when estimating reliabilities.

Person (sample, test) reliability depends chiefly on

- 1) Sample ability variance. Wider ability range = higher person reliability.
- 2) Length of test (and rating scale length). Longer test = higher person reliability
- 3) Number of categories per item. More categories = higher person reliability
- 4) Sample-item targeting. Better targeting = higher person reliability

It is independent of sample size. It is largely uninfluenced by model fit.

In general, Test Reliability reported by Classical Test Theory (Cronbach Alpha, KR-20) is higher than Rasch Reliability. Rasch Reliability is higher than 3-PL IRT Reliability.

Rasch Person "Test" Reliability is given by

OV = observed variance of person ability measures

EV = mean of squared standard errors of person ability measures

Person "Test" Reliability = $(OV - EV) / EV$

Item reliability depends chiefly on

- 1) Item difficulty variance. Wide difficulty range = high item reliability
- 2) Person sample size. Large sample = high item reliability

It is independent of test length. It is largely uninfluenced by model fit.

Rasch Item" Reliability is given by
OV = observed variance of item difficulty measures
EV = mean of squared standard errors of item difficulty measures
Item Reliability = (OV-EV)/EV

Tentative guidelines:

Person reliability: Does your test discriminate the sample into enough levels for your purpose? 0.9 = 3 or 4 levels. 0.8 = 2 or 3 levels. 0.5 = 1 or 2 levels.

Item reliability: Low reliability means that your sample is not big enough to precisely locate the items on the latent variable.

Rater reliability: Low "separation" reliability is better, because we want raters to be reliably the same, not reliably different.

The Winsteps "person reliability" is equivalent to the traditional "test" reliability. Low values indicate a narrow range of person measures, or a small number of items. To increase person reliability, test persons with more extreme abilities (high and low), lengthen the test. Improving the test targeting may help slightly.

The Winsteps "item reliability" has no traditional equivalent. Low values indicate a narrow range of item measures, or a small sample. To increase "item reliability", test more people. In general, low item reliability means that your sample size is too small for stable item estimates based on the current data. If you have anchored values, then it is the item reliability of the source from which the anchor values emanate which is crucial, not the current sample.

The **"model"** person reliability (including measures for extreme scores) is an upper bound to this value, when persons are ordered by measures.

The **"real"** person reliability (including measures for extreme scores) is a lower bound to this value, when persons are ordered by measures

The traditional "test reliability", as defined by Charles Spearman in 1904, etc., is the "true person variance / observed person variance" for this sample on these test items. So it is really a "person sample reliability" rather than a "test reliability", where reliability = reproducibility of person ordering. The "true person variance" cannot be known, but it can be approximated. KR-20 approximates it by summarizing item point-biserials. Cronbach Alpha approximates it with an analysis of variance. Winsteps approximates it using the measure standard errors.

The separation coefficient and reliability computations are computed with and without any elements with extreme measures. Since the measures for extreme scores are imprecise, reliability statistics which include extreme scores are often lower than their non-extreme equivalents. Conventional computation of a reliability coefficient (KR-20, Cronbach Alpha) includes persons with extreme scores. The classical reliability computation includes extreme scores (if any) is the conventional reliability, and usually produces an estimate between the MODEL and REAL values, closer to the MODEL or even above it.

KR-20 value is an estimate of the value when persons are ordered by raw scores. CRONBACH ALPHA (KR-20) KID RAW SCORE RELIABILITY is the conventional "test" reliability index. It reports an approximate test reliability based on the raw scores of this sample. It is only reported for complete data. An apparent paradox is that extreme scores have perfect precision, but extreme measures have perfect imprecision.

Winsteps computes upper and lower boundary values for the True Reliability. The lower boundary is the Real Reliability. The upper boundary is the Model Reliability. The unknowable True Reliability lies somewhere between these two. As contradictory sources of noise are removed from the data, the True Reliability approaches the Model Reliability

Cronbach Alpha and KR-20 Reliability

Here is a check on the computations. Guilford reports 0.81. Winsteps reports 0.82. The difference is probably computational precision and rounding error.

Dichotomous:

```
Title = "Guilford Table 17.2. His Cronbach Alpha = KR-20 = 0.81"  
ni=8  
item1=1  
name1=1  
&END  
END LABELS  
00000000  
10000000
```

```
10100000
11001000
01010010
11101010
11111100
11111100
11110101
11111111
```

Polytomous (Partial Credit) with missing data:

```
ni=4
codes=01234
groups=0
name1=1
item1=1
$END
END LABELS
1213
2.04
3323
4433
```

This has Cronbach Alpha: $(4/3) * (1 - 3.35/10.25) = 0.90$

Winsteps uses the population variances (as used by Lee J. Cronbach). SPSS uses the sample variances. For a discussion, see www.pbarrett.net/techpapers/kr20.pdf

Conventionally, only a Person ("Test") Reliability is reported. The relationship between raw-score-based reliability (i.e., KR-20, Cronbach Alpha) and measure-based reliability is complex, see www.rasch.org/rmt/rmt1131.htm - in general, Cronbach Alpha overestimates reliability, Rasch underestimates it. So, when it is likely that the Rasch reliability will be compared with conventional KR-20 or Cronbach Alpha reliabilities (which are always computed assuming the data match their assumptions), then include extreme persons and report the higher Rasch reliability, the "Model" reliability, computed on the assumption that all unexpectedness in the data is in accord with Rasch model predictions.

The big differences between Score and Measure reliabilities occur when

- (a) there are extreme scores. These increase score reliability, but decrease measure reliability.
- (b) missing data. Missing data always decreases measure reliability. If the missing data are imputed at their expected values (in order to make conventional reliability formulas computable), they increase score reliability. Winsteps attempts to adjust the raw-score reliability for this inflation in the raw-score reliability, but can only do the adjustment in an approximate way.

Winsteps also reports an item reliability, "true item variance / observed item variance". When this value is low, it indicates that the sample size may be too small for stable comparisons between items.

Anchored values are treated as though they are the "true values" of the MLE estimates. Their local standard errors are estimated using the current data in the same way as unanchored MLE standard error estimates. It is the measures (anchored or unanchored) and local standard errors that are used in the reliability computations. If you wish to compute reliabilities using different standard error estimates (e.g., the ones when the anchor values were generated), then please perform a separate reliability computation (using Excel).

You can easily check the Winsteps reliability estimate computation yourself.

Read the Winsteps [PFILE=](#) into an Excel spreadsheet.

Compute the STDEVP standard deviation of the person measures. Square it. This is the "Observed variance".

"Model" Reliability: Take the [standard ERROR](#) column. Square each entry. Sum the squared entries. Divide that sum by the count of entries. This is the "Model Error variance" estimate. Then,
Model Reliability = True Variance / Observed Variance = (Observed Variance - Model Error Variance) / Observed Variance.

"Real" Reliability: Take the [standard ERROR](#) column. Square each entry, SE². In another column, put SE²*Maximum [1.0, INFIT mean-square). Divide that sum by the count of entries. This is the "Real Error variance" estimate. Then,
Real Reliability = True Variance / Observed Variance = (Observed Variance - Real Error Variance) / Observed Variance.

Separation, Strata and Reliability

The crucial elements in the computation of reliability are the "True" variance and the Error variance. These are squared distances and so difficult to conceptualize directly. It is easier to think of their square-roots, the "True" standard deviation (TSD) and the root-mean-square standard error (RMSE).

SEPARATION coefficient is the ratio of the PERSON (or ITEM) TRUE S.D., the "true" standard deviation, to RMSE, the error standard deviation. It provides a ratio measure of separation in RMSE units, which is easier to interpret than the reliability correlation. This is analogous to the Fisher Discriminant Ratio. SEPARATION coefficient² is the signal-to-noise ratio, the ratio of "true" variance to error variance.

RELIABILITY (separation index) is a separation reliability. The PERSON (or ITEM) reliability is equivalent to KR-20, Cronbach Alpha, and the Generalizability Coefficient. The relationship between SEPARATION coefficient and RELIABILITY (separation index) is

RELIABILITY = SEPARATION coefficient² / (1 + SEPARATION coefficient²)
 or SEPARATION coefficient = square-root(RELIABILITY / (1 - RELIABILITY)).

Separation (if the outlying measures are accidental) or **Strata** (if the outlying measures represent true performances). These numbers are statistical abstractions, but their empirical meaning is indicated by locating the Separation or Strata levels in the observed distribution at (3 * "Observed S.D." / Separation) units apart, centered on the sample mean.

Error RMSE	True SD	True Variance	Observed Variance	Signal-to-Noise Ratio	Separation = True SD / RMSE	Strata = (4*Sep.+1) / 3	Reliability = True Variance / Observed Variance
1	0	0	1	0	0	0	0
1	1	1	2	1	1	1.67	0.5
1	2	4	5	2	2	3	0.8
1	3	9	10	3	3	4.33	0.9
1	4	16	17	4	4	5.67	0.94

There is more at www.rasch.org/rmt/rmt94n.htm and www.rasch.org/rmt/rmt163f.htm

Spearman-Brown Prophecy Formula for person "Test" reliability with different numbers of items (test lengths)

How many items (or persons) are required to produce the reliability I want with the sample of persons and the same type of items (or this test and the same type of persons)?

T = target number of items, R_T = target person reliability

C = current number of items, R_C = current person reliability

$$1. \text{ Predict number of items} = T = C * R_T * (1 - R_C) / ((1 - R_T) * R_C)$$

Example: the current test length is C = 10 items, and the current person reliability is R_C = 0.3. We want a person reliability of R_T = 0.8.

Target number of items is T = 10 * 0.8 * (1 - 0.3) / ((1 - 0.8) * 0.3) = 94 items.

$$2. \text{ Predicted person reliability} = R_T = T * R_C / (C * (1 - R_C) + T * R_C)$$

Example: we have a test of C = 11 items of person reliability R_C = 0.5, what is the predicted reliability of a test of T = 17 items?

Predicted person reliability R_T = 17 * 0.5 / (11 * (1 - 0.5) + 17 * 0.5) = 0.61

Prophecy Formula for person "Test" reliability with different observed standard deviations of the person measures/scores

assuming the average error variance is unchanged.

T = target observed standard deviation, R_T = target person reliability

C = current observed standard deviation, R_C = current person reliability

1. Predict person standard deviation = $T = C * \sqrt{(1-R_C) / (1-R_T)}$

2. Predict person reliability = $R_T = 1 - ((1-R_C)*C^2 / T^2)$

Test-Retest Reliability

is the correlation between the person measures obtained from two administrations of the same test to the same persons. The expected value of the test-retest reliability is the person reliability of the first administration.

The "Smallest Detectable Difference" = "Smallest statistically significant difference in a person's measures" when a test is administered twice to a person under normal conditions = $1.96 * (\text{person standard deviation}) * (\text{test-retest reliability})$.

Population and Sample Standard Deviation, Reliability, Separation and Strata

Winsteps assumes that the N statistics being summarized are the entire population. It reports the population observed Standard Deviation = P.SD, "True" Standard Deviation = T.P.SD, Reliability = P.Rel, Separation = P.Sep and Strata =

P.Strata, average error S.D. = RMSE, where

$P.Sep = \sqrt{P.Rel / (P.Rel + 1)}$

$RMSE = P.SD * \sqrt{1 - P.Rel}$

If the N statistics are a sample of the entire population, then the summary statistics are the sample observed Standard Deviation = S.SD, "True" Standard Deviation = T.S.SD, Reliability = S.Rel, Separation = S.Sep and Strata = S.Strata, where $S.SD = P.SD * \sqrt{N / (N-1)}$ - the sample S.D. is bigger than the population S.D.

$T.S.SD = \sqrt{S.SD^2 - RMSE^2}$

$S.Sep = T.S.SD / RMSE$

$S.Strata = (4 * S.Sep + 1) / 3$

$S.Rel = (1 + (N-1)*P.Rel) / N$

Confidence Intervals of a Reliability Coefficient from <http://www.psych.org/cgi-bin/R.cgi/Feldt1.R>

19.99 Response-pattern frequency

Here is how to report the frequency of response patterns:

1. In your Winsteps control file,

a) give `NAME1=` the same value as `ITEM1=`

b) give `NAMELENGTH=` the length of the response string = the value of `NI=` * `XWIDE=`

c) `PSUBTOT=1Wn` where n is the same as `NAMELENGTH=`

2. Analyze the control file

3. Output [Table 28](#) person sub-totals: each response pattern will have a count in the first column.

19.100 Rules for assigning values to control variables (key-words)

Do not worry about these unless Winsteps does not respond to your control file the way you expected. If possible, compare your control file with what is shown in Table 0 of your report output file in order to isolate the problem.

1. Values are assigned to control variables by typing the name of the control variable (or enough of it to disambiguate it), an equal sign, "=", and the value, e.g.

```
TABLES=11011011100
```

or

```
TAB=1101101110 this is enough of TABLES to make clear what you mean.
```

2. You must use one line for each assignment, but continuation lines are permitted.

To continue a line, put a + at the end of the line. Then put a + at the start of the text in the next line. The two lines will be joined together so that the + signs are squeezed out, e.g.,

```
TITLE = "Analysis of medical+
```

```
+ research data"
```

is interpreted as

```
TITLE = "Analysis of medical research data"
```

Continuation lines "+" are helpful to make control files fit on your screen.

```
CODES = 01020304+
```

```
+0506+
```

```
+0708
```

is interpreted as

```
CODES = 0102030405060708
```

To comment out a continuation line:

```
; CODES = 01020304+
```

```
; +05060708
```

3. The control variables may be listed in any order.

4. Character strings must be enclosed in 'single quotes' or "double quotes" when they contain blanks, e.g.,

```
TITLE="Rasch Analysis of Test Responses"
```

or

```
TITLE='Rasch Analysis of Test Responses'
```

Quotes are not required for single words containing no blanks, e.g. PFILE=kctpf.txt

5. The control variables may be in upper or lower case or mixed,

e.g., Pfile = Person.txt

6. Blanks before or after control variables, and before or after equal signs are ignored, e.g.

```
TITLE="Test Report"
```

and

```
TITLE = "Test Report"
```

are equally correct.

7. Commas at the end of lines are ignored, so equally correct are:

```
NAME1 = 33,
```

and

```
NAME1 = 33
```

8. Control variables can be made into comments, and so be ignored, by entering a semi-colon in column 1, e.g.

```
; FITP=3 is ignored
```

9. When all control variables (required or optional) have been assigned values,

type [&END](#) (in upper or lower case) on the next line, e.g.,

```
Title ="A 30 Item test"
```

```
NI = 30
; this is a comment: person names in columns 1-20.
ITEM1= 21
&END
```

10. Values may be expressions, with no internal spaces
USCALE = 2*10 ; 20

19.101 Scoring items with multiple responses

Question: We have 7 items on the test and 3 items are standard 4 option multiple choice items with one correct option as correct. Also there are 3 items that have 5 answer options and have 3 or 4 answer options as correct to achieve one score point. And one item with 6 answer options with 4 options correct to count as 1 correct.

Answer: These creative item types are beyond the standard capability of Winsteps :-)

One solution is to pre-process the responses, for instance with Excel.

Build an Excel worksheet with one row for each person. The standard items are one column each. These can be scored later by Winsteps. For the other items, each item option is one column in the worksheet. So the 5 answer options to one item are 5 columns for each person. In a 6th column, these 5 columns are compared to the scoring key by Excel and the count of right answers shown. In a 7th column, the count in 6th column is compared with the "correct" count and the final 0/1 shown. This is done for all three special items. Then Winsteps converts the Excel file into a Winsteps file ([Excel/RSSST menu](#)), selecting only the final score column for each special item.

19.102 Shortcut Keys

Some very frequent operations are quicker to perform using shortcut keystrokes than by menu selection. Here are the shortcut keys implemented in Winsteps:

Alt+ hold down the Alt key and press the letter key.
Ctrl+ hold down the Ctrl key and press the letter key.

Alt+A start **A**nother copy of Winsteps
Alt+E **E**dit the control file
Alt+H display **H**elp file
Alt+R **R**estart Winsteps with this control file
Alt+S **S**pecification entry
Alt+X **eX**it this Winsteps, and restart with this control file

Ctrl+E **E**nd epochs
Ctrl+F **F**inish iterating or **F**inish factoring
Ctrl+O **O**pen a control or output file
Ctrl+Q **Q**uit Winsteps
Ctrl+S **S**ave on-screen activity log
Ctrl+P **P**rint on-screen activity log

Esc **E**scape from action

19.103 Simulating data

See [Simulated data file](#)

19.104 Smartphone - running Winsteps on your PC

Winsteps and all windows programs can be run from your smartphone using Google Chrome Remote Desktop. The remote desktop has mouse and keyboard items to access all Winsteps functions. Areas of the smartphone screen can be enlarged by using your smartphone's magnification capabilities so that all Winsteps output can be seen clearly.

19.105 Specifying how data are to be recoded

You will need to choose how this is done.

First, use [CODES=](#) to specify the response codes in your data file.

If there is only one type of recoding to be done, use [NEWSCORE=](#)

If this one type of rescoring only applies to some of the items, also use [RESCORE=](#)

If the rescoring is more complex, use [IREFER=](#) and [IVALUE=](#)

If the items are multiple-choice, use [KEYn=](#)

If missing values are not to be ignored, i.e., treated as not-administered, you will need [MISSCORE=](#)

If alphabetical codes are used to express two-digit numbers in one column, use [ALPHANUM=](#)

19.106 Speeded tests

Speeded tests are a challenge for measurement. Even Georg Rasch had trouble with them.

If the problem is that people run out of time before completing all the items, then a two-phase approach to Rasch analysis has been used:

In phase 1, item calibration, all not-reached and skipped responses are coded as "not administered". Clearly aberrant responses are ignored ([CUTLO=](#), [CUTHI=](#)). This analysis is done to get a good estimate of the true difficulty of the items. Output the item file: [IFILE=if.txt](#)

In phase 2, person measurement, all the data are used. Every response is scored right or wrong, including skips and not-reached. The items are anchored [IAFILE=if.txt](#). The person measures are produced.

19.107 Split items - procedure

There are some methods to split an item into two or more columns based on person classification group.

1. From Winsteps, output your data to Excel using Output Files menu, [RFILE=](#) or [IPMATRIX=](#)

In Excel:

- a. Sort the rows (cases) by the target demographic (or whatever) variable
- b. Cut the responses to the target items by one demographic group into new columns
- c. Label (row 1) the new columns with their item labels and demographic information
- d. Save the Excel file under a new name
- e. Use the [Excel/RSSST](#) menu to convert the new Excel file into a Winsteps control file
- f. Analyze the new control file which now has the split items.

2. Use [MFORMS=](#), then [EDFILE=](#)

3. Use [FORMAT=](#), then [EDFILE=](#)

These examples use [FORMAT=](#). Here are the changes to a control file to split an item:

Example 1: We want to split item 3 "Enjoy cooking", of our 12 item test into two items, based on the Gender (M,F) indicator in column 2 of the person label.

1. Increase number of items:

```
NI= 13 (add one NI=12)
NAME1= (add one if to the right of the items)
```

The new item will be the last item

2. Add coding, grouping for new item 13 in IREFER=, ISGROUPS=, KEY1=

3. Add a new item name before END LABELS for item 13:

Enjoy cooking (male)

Edit the current item name for item 3:

Enjoy cooking (female)

4. Use FORMAT= duplicate the old item into the new item's position, e.g.,

```
FORMAT = (12A,T3,1A,T13,99A)
```

adds the item in column 3 again in column 13. Then follows it with the old column 13 onwards, now starting in column 14..

5. Delete responses in the two versions of the item, based on the person demographics (Gender: M, F) in column 2 of the person label:

```
EDFILE=*
"?M" 3 x ; all Male responses to item 3 are changed to "x" (scored as missing - not administered)
"?(~M)" 13 x ; all not Male responses to item 13 are changed to "x"
*
```

6. The Control file is:

```
Title = "Survey of Personal Preferences"
NI = 13 ; was NI=12
ITEM1 = 1 ; responses start the line
NAME1 = 15 ; was NAME1=14
CODES=12345
FORMAT = (12A,T3,1A,T13,99A) ; repeat column 3 at column 13
EDFILE=*
"?M" 3 x ; all Male responses to item 3 are changed to "x" (scored as missing - not administered)
"?(~M)" 13 x ; all not Male responses to item 13 are changed to "x"
*
RFILE = rfile.txt ; File which shows the reformatted data
&END
Item 1
Item 2
Enjoy cooking (female) ; Item 3
Item 4
....
Item 12
Enjoy cooking (female) ; Item 12
END NAMES
321453324213 PM George ; M is Male in column 2 of person label
532134324522 RF Mary ; F is Female in column 2 of person label
.....

The data become:
32x4533242131 PM George ; M is Male in column 2 of person label
532134324522x RF Mary ; F is Female in column 2 of person label
.....
```

Example 2: Splitting item 19 in Liking for Science, example0.txt, data by gender, and moving gender to first column of person label:

```
Title= "Liking for Science"
ITEM1= 1 ; Starting column of item responses
NI= 26 ; Number of items ; one item added to the standard 25
XWIDE = 1 ; this matches the biggest data value observed
CODES= 012 ; matches the data
NAME1 = 28 ; Starting column for person label in formatted data record (gender)
; show column 19 again in column 26, and also column 48 (gender) again in column 28
FORMAT = "(25A1, T19, 1A1, T26, 1A1, T48, 2A1, T27, 50A1)"
EDITFILE=*
"{~F}" 26 m ; m = deliberately omitted response
```


220002000222000222122010	MULLER, JEFF M 46
1110220112212111022010000	VAN DAM, ANDY M 47
2221022222222221222122022	CHAZELLE, BERNIE M 48
110011010111011211111210	BAUDET, GERARD M 49
22221212222222222122122	DOEPPNER, TOM M 50
121002102222221122021022	REISS, STEVE M 51
1111111111111111122111111	STULTZ, NEWELL M 52
0100101010110011021010000	SABOL, ANDREW M 53
1111021121122112122021011	BABBOO, BABOO 54
1010010011122112222111110	DRISKEL, MICHAEL M 55
222101201222222222122022	EASTWOOD, CLINT M 56
22200220222222222022022	CLAPP, CHARLIE M 57
122101012222221222021021	CLAPP, DOCENT 58
222102112222221222022222	CLAPP, LB 59
122111112222121122121021	SQUILLY, MICHAEL M 60
2111010022122111122011022	SQUILLY, BAY OF 61
22200000222222022020000	BEISER, ED M 62
12110101222212222122010	BECKER, SELWIN M 63
2210022112222121222122011	CORLEONE, VITO M 64
121121100222012122121200	CORLEONE, MICHAEL M 65
12010100122212022022021	PINHEAD, ZIPPY 66
1211001001212121222121012	MALAPROP, MRS. F 67
1200010102120120022001010	BOND, JAMES M 68
12100101222212222021021	BLOFELD, VILLAIN M 69
22220220122212222022021	KENT, CLARK M 70
220122200122002222022201	STOLLER, DAVE M 71
1001000100011010021200201	JACKSON, SOLOMON M 72
2012010122210201102202022	SANDBERG, RYNE M 73
22200220022222220220212	PATRIARCA, RAY M 74
1200110112122221022020010	PAULING, LINUS M 75

Example 3: We have non-uniform DIF on an item and want to split the persons high-low.

We need to split the persons on the item by raw score or Rasch measure. To do this conveniently, use RFILE= to output data to Excel. Also output the PFILE= to Excel. Then you can copy the score/measure column from the PFILE to the RFILE. Sort the RFILE by the new score/measure column. then split the desired item column into two columns high/low or whatever.

19.108 Standard errors: model and real

A standard error quantifies the **precision** of a measure or an estimate. It is the standard deviation of an imagined error distribution representing the possible distribution of observed values around their "true" theoretical value. This precision is based on information within the data. The quality-control fit statistics report on **accuracy**, i.e., how closely the measures or estimates correspond to a reference standard outside the data, in this case, the Rasch model.

S.E.s are produced by models of the data and are estimates of precision. For the S.E. of the mean, the model of the data is a normal distribution of the values summarized by the mean. The mean is an estimate, because we never know the true mean of a distribution. The S.E. shows the precision of the mean estimate. For the Rasch "model S.E.", the model is the Rasch model of ordinal data summarized by a parameter estimate. It is an estimate because we never know the true value of the parameter. The S.E. shows the precision of the Rasch estimate.

Standard errors of Rasch estimates reported by Winsteps do not include the imprecision in the estimates of all the other persons or items. When estimating the standard error for a person or item, the other persons and items are treated as though their distributions exactly match their populations and their estimated values are their true values. The imprecision in the estimates due to sampling errors and basing person estimates on item estimates, and *vice-versa*, is usually an order of magnitude less than the reported standard errors.

Note: Survey-style "sample" standard errors and confidence intervals are equivalent to Rasch item-calibration standard errors. So

Survey sample 95% confidence interval on a dichotomous (binary) item reported with a proportion-correct-value as a % = $1.96 * 100\% / (\text{item logit standard error} * \text{sample size})$

Example: survey report gives: $p = 90\%$, sample size=100, confidence interval (95%) = $90 \pm 6\%$

Winsteps: logit S.E. of item calibration = $1/\sqrt{100 * .9 * .1} = \pm .33$ logits.

So survey C.I. % = $\pm 1.96 * 100 / (.33 * 100) = \pm 6\%$

Standard Errors of Items

The size of a standard error of an estimate is most strongly influenced by the number of observations used to make the estimate. We need measurement precision (standard error size) adequate for the purpose for which we are using the measures.

Probably the only time we need to be concerned about item standard errors within a test is when we want to say "Item A is definitely more difficult than Item B". For this to be true, their measures need to be more than 3 S.E.s different.

When comparing item difficulties estimated from different datasets, we use the item standard errors to identify when differences between the item difficulties of the same item are probably due to chance, and when they may be due to a substantive change, such as item drift.

Model "Ideal" Standard Error

The highest possible precision for any measure is that obtained when every other measure is known, and the data fit the Rasch model. The model standard error is $1/\sqrt{\text{Fisher information}}$. For well-constructed tests with clean data (as confirmed by the fit statistics), the model standard error is usefully close to, but slightly smaller than, the actual standard error. The "model" standard error is the "best case" error. It is the asymptotic value for [JMLE](#). For dichotomous data this is, summed over items $i=1,L$ for person n , or over person $n=1,N$ for item i :

$$S.E.(B_n, D_i) = 1/\sqrt{\sum (P_{ni}(1-P_{ni}))}$$

For polytomies (rating scales, partial credit, etc.), with categories $j=0,m$:

$$S.E. = 1/\sqrt{\sum_{n \text{ or } i} \left(\sum_{j=0}^m (jP_{nij} - \sum_{j=0}^m jP_{nij})^2 \right)}$$

and, for the Rasch-Andrich thresholds,

$$S.E.(F_j) = 1/\sqrt{\left(\sum_{n=1}^N \sum_{i=1}^L \left(\sum_{k=0}^j P_{nik} \sum_{k=j+1}^m P_{nik} \right) \right)}$$

where P_{nik} is the probability of observing category k for person n on item i .

Misfit-Inflated "Real" Standard Error

Wright and Panchapakesan (1969) www.rasch.org/memo46.htm discovered an important result for tests in which each examinee takes more than a handful of items, and each item is taken by more than a handful of examinees: the imprecision introduced into the target measure by using estimated measures for the non-target items and examinees is negligibly small. Consequently, in almost all data sets except those based on very short tests, it is only misfit of the data to the model that increases the standard errors noticeably above their model "ideal" errors. Misfit to the model is quantified by fit statistics. But, according to the model, these fit statistics also have a stochastic component, i.e., some amount of misfit is expected in the data. Discovering "perfect" data immediately raises suspicions! Consequently, to consider that every departure of a fit statistic from its ideal value indicates failure of the data to fit the model is to take a pessimistic position. What is useful, however, is to estimate "real" standard errors by enlarging the model "ideal" standard errors by the model misfit encountered in the data.

Recent work by Jack Stenner shows that the most useful misfit inflation formula is

$$\text{Real S.E. of an estimated measure} = \text{Model S.E.} * \text{Maximum} [1.0, \sqrt{\text{INFIT mean-square}}]$$

In practice, this "Real" S.E. sets an upper bound on measure imprecision. It is the "worst case" error. The actual S.E. lies between the "model" and "real" values. But since we generally try to minimize or eliminate the most aberrant features of a measurement system, we will probably begin by focusing attention on the "Real" S.E. as we establish that measurement system. Once we become convinced that the departures in the data from the model are primarily due to modeled stochasticity, then we may base our decision-making on the usually only slightly smaller "Model" S.E. values.

What about Infit mean-squares less than 1.0? These indicate overfit of the data to the Rasch model, but do not reduce the standard errors. Instead they flag data that is lacking in randomness, i.e., is too deterministic. Guttman data are like this. Their effect is to push the measures further apart. With perfect Guttman data, the mean-squares are zero, and the measures are infinitely far apart. It would seem that inflating the S.E.s would adjust for this measure expansion, but Jack Stenner's work indicates that this is not so. In practice, some items overfit and some underfit the model, so that the overall impact of low infit on the measurement system is diluted.

Standard Errors with Anchor Values

Anchored measures are shown in the Winsteps output Tables with "A". These are set with [IAFILE=](#), [PAFILE=](#) and [SAFILE=](#). Anchor values are exactly precise with zero standard error. But each anchor value is reported with a standard error. This is the standard error that the anchor value would have if it were the freely estimated maximum-likelihood value of the parameter.

Plausible Values

"Plausible values" are random draws from a parameter's posterior distribution. Here the posterior distribution is a normal distribution of $N(\text{mean}=\text{estimated measure}, \text{S.D.}=\text{standard error})$ for each parameter. Plausible values would be random draws from this distribution. The Excel formula to do this is $=\text{(Measure + S.E.*NORMSINV(RAND()))}$ which can be input into an extra column in a [PFILE=](#) or [IFILE=](#) written to Excel.

19.109 Standard setting

The Bookmark standard-setting procedure can be implemented using Winsteps. See <http://pareonline.net/getvn.asp?v=11&n=2>

MacCann, Robert G. & Gordon Stanley (2006). The use of Rasch modeling to improve standard setting. *Practical Assessment Research & Evaluation*, 11(2).

19.110 Starting Winsteps from the DOS prompt

Winsteps can also be invoked from the DOS prompt in a DOS window.

At the prompt enter

```
C:>Winsteps(Enter)
```

Winsteps proceeds with its standard operations.

You can enter control and output files directly on the prompt line.

```
C:>Winsteps SF.txt SF.OUT(Enter)
```

Winsteps starts analysis immediately. You will not be prompted for "Extra Specifications"

You can also enter extra specifications here:

```
C:>Winsteps SF.txt SF.OUT chart=yes distractors=no(Enter)
```

Leave no spaces within specifications, or place them in quotes, e.g.,

```
C:>Winsteps SF.txt SF.OUT "chart = yes" "distractors = no"(Enter)
```

To perform the previous analysis again, with a temporary report output file:

```
C:>Winsteps @(Enter)
```

@ is replaced by the top control file on the Files= menu. If no output file is specified, then a temporary one is used.

For Batch file operation, see [Batch=](#)

19.111 Subsets and connection ambiguities

You see: **Warning:** Data are ambiguously connected into 6 subsets. Measures may not be comparable across subsets.

Quick (but arbitrary) solution: add to the data file two dummy person records so that all persons and items become directly comparable.

Dichotomous data:

```
Dummy person 1: responses: 010101010...
```

```
Dummy person 2: responses: 101010101...
```

This says: "the middle level of performance for all subsets of persons is the same."

Rating scale data, where "1" is the lowest category, and "5" is the highest category.

```
Dummy person 1: responses: 121212121...
```

```
Dummy person 2: responses: 2121212121...
```

This says: "the bottom level of performance for all subsets of persons is the same."

Explanation: Connectivity (or subsetting) is a concern in any data analysis involving missing data. In general, nested data are not connected. fully-crossed data (also called "complete data") are connected. partially-crossed data may or may not be connected.

Winsteps examines the responses strings for all the persons. It verifies that every non-extreme response string is linked into one network of success and failure on the items. Similarly, the strings of responses to the items are linked into one network of success and failure by the persons.

If person response string A has a success on item 1 and a failure on item 2, and response string B has a failure on item 1 and a success on item 2, then A and B are connected. This examination is repeated for all pairs of response strings and all pairs of items. Gradually all the persons are connected with all the other persons, and all the items are connected with all the other items. But if some persons or some items cannot be connected in this way, then Winsteps reports a "connectivity" problem, and reports which subsets of items and persons are connected.

Mathematics: connectivity is part of [Graph Theory](#). The person/item/judge/... parameters of the Rasch model are the vertices and the observations are the edges. In an undirected graph, we need every vertex to be connected directly or indirectly to every other vertex. A connection is established between two vertices when one vertex is observed to have both a higher observation and a lower observation than another vertex in the same context, or when both both vertices have the same intermediate category of a rating scale in the same context.

Thus there are two situation for failure to connect:

1) there is no direct or indirect link between two vertices, e.g., two different datasets analyzed together with no common parameters. This is detected by the Winsteps/Facets subset routine.

2) the vertices are connected by observations, but the observations do not meet the requirements, e.g., all the person respond to all the items, but half the persons score in the upper half of the rating scale on every item, and the other half of the persons score in the lower half of the rating scale on every item. This is called a "Guttman split" in the data. This is usually obvious in the reported estimates as a big gap on the Wright maps between the two halves of the person distribution.

Example 1: Connection problems and subsets in the data are shown in this dataset. It is **Examsubs.txt**.

```

Title = "Example of subset reporting"
Name1 = 1
Namelength = 24 ; include response string in person label
Item1 = 13
NI = 12
CODES = 0123 ; x is missing data
ISGROUPS = DDDDDDDDDRRR ; items 1-10 are dichotomies; items 11-12 share a rating scale
MUCON = 3 ; Subsetting can cause very slow convergence
TFILE=*
18.1
14.1
0.4
*
&End
01 Subset 1
02 Subset 1
03 Subset 2
04 Subset 2
05 Subset 7
06 Subset 4
07 Subset 4
08 Subset 5
09 Subset 5
10 Subset 5
11 Subset 6
12 Subset 6
END LABELS
01 Extreme 11111
02 Subset 1 01111
03 Subset 1 10111
04 Subset 2 00101
05 Subset 2 00011
06 Subset 3 011
07 Subset 3 011
08 Subset 4 001
09 Subset 4 010
10 Subset 5 0x1
11 Subset 5 10x
12 Subset 5 x10
13 Subset 6 01
14 Subset 6 10
15 Subset 6 23
16 Subset 6 32

```

The Iteration Screen reports:

```

CONVERGENCE TABLE
-Control: \HOLDW95\examples\examsubs.txt Output: \examples\ZOU571WS.TXT
| PROX ACTIVE COUNT EXTREME 5 RANGE MAX LOGIT CHANGE |
| ITERATION PERSON ITEM CATS PERSON ITEM MEASURES STRUCTURE |
>=====<
| 1 15 12 8 2.00 1.06 -2.0794 |
>=====<
| 2 15 12 6 2.38 1.84 2.6539 -1.6094 |
>=====<

```

```

|      3      14      12      6      2.67      1.60      2.7231      .0000 |
>=====<
|      4      14      12      6      2.68      2.33      -2.3912      .0000 |
>=====<
|      5      14      12      6      2.97      1.77      2.3246      |
>=====<
|      6      14      12      6      2.97      2.54      -2.2191      |
>=====<
|      7      14      12      6      3.22      2.10      2.0372      |
Probing data connection: to skip out: Ctrl+F - to bypass: subset=no
Processing unanchored persons ...
>=====<
Consolidating 9 potential subsets pairwise ...
>=====
Consolidating 9 potential subsets indirectly pairwise ...
>=====<
Consolidating 8 potential subsets pairwise ...
>=====
Consolidating 7 potential subsets pairwise ...
>=====
Consolidating 7 potential subsets indirectly pairwise ...
>=====<
Warning: Data are ambiguously connected into 7 subsets. Measures may not be comparable
        across subsets.
        Subsets details are in Table 0.4

```

[Table 18.1](#)

PERSON STATISTICS: ENTRY ORDER

```

-----
|ENTRY  | |
|NUMBER | | PERSON
|-----+-----|
|   1   | | 01 Extreme 11111 | MAXIMUM MEASURE
|       | |                   | < Guttman split here >
|   2   | | 02 Subset 1 01111 | SUBSET 1
|   3   | | 03 Subset 1 10111 | SUBSET 1
|       | |                   | < Guttman split here >
|   4   | | 04 Subset 2 00101 | SUBSET 2
|   5   | | 05 Subset 2 00011 | SUBSET 2
|       | |                   | < Guttman split here >
|   6   | | 06 Subset 3   011 | SUBSET 3
|   7   | | 07 Subset 3   011 | SUBSET 3
|       | |                   | < Guttman split here >
|   8   | | 07 Subset 4   001 | SUBSET 4
|   9   | | 09 Subset 4   010 | SUBSET 4
|       | |                   | < Subset split here >
|  10   | | 10 Subset 5     0x1 | SUBSET 5
|  11   | | 11 Subset 5     10x | SUBSET 5 < Indirect connection >
|  12   | | 12 Subset 5     x10 | SUBSET 5
|       | |                   | < Subset split here>
|  13   | | 13 Subset 6     01 | SUBSET 6
|  14   | | 14 Subset 6     10 | SUBSET 6
|       | |                   | < undetected Guttman split here:
|       | | Winsteps failed! >
|  15   | | 15 Subset 6     23 | SUBSET 6
|  16   | | 16 Subset 6     32 | SUBSET 6
|-----+-----|

```

<i>In Tables and Notes:</i>	<i>Explanation:</i>
< Guttman split here >	The persons above the split performed an unknowable amount different from the persons below the split. There is no item on which this subset succeeded

	and another subset failed, and also this subset failed and the other subset succeeded. The data are not "well-conditioned" (Fischer G.H., Molenaar, I.W. (eds.) (1995) Rasch models: foundations, recent developments, and applications. New York: Springer-Verlag. p. 41-43).
< Subset split here >	The persons in this subset responded to different items than persons in other subsets. We don't know if these items are easier or harder than items in other subsets.
< Indirect connection >	The persons responded to different items, but they are connected by a loop of successes and failures.
< undetected Guttman split here >	Winsteps subset-detection did not report that persons 13 and 14 always score lower than persons 15 and 16, causing a Guttman split. We do not know how much better persons 15 and 16 are than persons 14 and 15. Winsteps subset-detection may fail to report subsets. Unreported subsets usually cause big jumps in the reported measures.
Data are ambiguously connected	Measures for persons in different subsets are not comparable. Winsteps always reports measures, but these are only valid within subsets. We do not know how the measures for persons in one subset compare with the measures for persons in another subset. Reliability coefficients are accidental and so is Table 20 , the score-to-measure Table. Fit statistics and standard errors are approximately correct.
Measures may not be comparable across subsets	Please always investigate when Winsteps reports subsets, even if you think that all your measures are comparable.
MAXIMUM MEASURE, MINIMUM MEASURE, DROPPED, INESTIMABLE	Persons and items with special features are not included in subsets. Extreme scores (zero, minimum possible and perfect, maximum possible scores) imply measures that are beyond the current frame of reference. Winsteps uses Bayesian logic to provide measures corresponding to those scores.
SUBSET 1, 2, 4	These are directly connected subsets. Within each subset, a person has succeeded on an item and failed on an item, and <i>vice-versa</i> . The person performances are directly pairwise comparable within the subset. The persons in this subset have either succeeded on items in other subsets, or failed on items in other subsets, or have missing data on items in other subsets.
SUBSET 3	These two persons have the same responses, so they are in the same subset. No one succeeded on their failed items item, and also failed on their successful item.
SUBSET 5	This is an indirectly connected subset. There is a loop of successes and failures so that the performances of all three persons are connected indirectly pairwise.
SUBSET 6	Persons 13 and 14 are directly comparable using categories 0 and 1 of the rating scale. Persons 15 and 16 are directly comparable using categories 2 and 3 of the rating scale. Winsteps has not detected that persons 13 and 14 always rate lower than persons 15 and 16, causing a Guttman split.
SUBSET 7 (Table 14.1)	No person is in the same subset as this item. There is no subset in which persons both succeeded and failed on this item.
Connecting SUBSETS	Here are approaches: 1. Collect more data that links items across subsets. Please start Winsteps analysis as soon as you start data collection. Then subset problems can be remedied before data collection ends. 2. Dummy data. Include data for imaginary people in the data file that connects the subsets. 3. Anchor persons or items. Anchor equivalent items (or equivalent persons) in the different subsets to the same values - or juggle the anchor values to make the mean of each subset the same (or whatever)

	4. Analyze each subset of persons and items separately. In Table 0.4, Winsteps reports entry numbers for each person and each item in each subset, so that you can compare their response strings. To analyze only the items and persons in a particular subset, such as subset 4 above, specify the items and persons in the subset: IDELETE= +9-10 PDELETE= +10-11
Memory was not allocatable to probe connectivity	If the data are complete, ignore this message. If the data are sparse, add dummy data records. They will have little influence on connected data, but will connected up data with subsets. See also Memory

Table 14.1

```

-----
|ENTRY      |          |
|NUMBER     | ITEM      | G |
|-----+-----|
|    1      | 01 Subset 1 D | SUBSET 1
|    2      | 02 Subset 1 D | SUBSET 1
|           |           |
|    3      | 03 Subset 2 D | SUBSET 2
|    4      | 04 Subset 2 D | SUBSET 2
|           |           |
|    5      | 05 Subset 7 D | SUBSET 7
|           |           |
|    6      | 06 Subset 4 D | SUBSET 4
|    7      | 07 Subset 4 D | SUBSET 4
|           |           |
|    8      | 08 Subset 5 D | SUBSET 5
|    9      | 09 Subset 5 D | SUBSET 5
|   10      | 10 Subset 5 D | SUBSET 5
|           |           |
|   11      | 11 Subset 6 R | SUBSET 6
|   12      | 12 Subset 6 R | SUBSET 6
|-----+-----|

```

< Guttman split here >

< Guttman split here >

< Guttman split here >

< Guttman split here >

< Guttman split here >

Table 0.4 reports

```

SUBSET DETAILS

Subset 1 of 2 ITEM and 2 PERSON
ITEM: 1-2
PERSON: 2-3
Subset 2 of 2 ITEM and 2 PERSON
ITEM: 3-4
PERSON: 4-5
Subset 3 of 2 PERSON
PERSON: 6-7
Subset 4 of 2 ITEM and 2 PERSON
ITEM: 6-7
PERSON: 8-9
Subset 5 of 3 ITEM and 3 PERSON
ITEM: 8-10
PERSON: 10-12
Subset 6 of 2 ITEM and 4 PERSON
ITEM: 11-12
PERSON: 13-16
Subset 7 of 1 ITEM
ITEM: 5

```

Example 2: Analyzing two separate datasets together.

Dataset 1. The Russian students take the Russian items. This is connected. All the data are in one subset.

Dataset 2. The American students take the American items. This is connected. All the data are in one subset.

Dataset 3. Datasets 1 and 2 are put into one analysis. This is not connected. The data form two subsets: the Russian one and the American one. The raw scores or Rasch measures of the Russian students cannot be compared to those of the American students. For instance, if the Russian students score higher than the American students, are the Russian students more able or are the Russian items easier? The data cannot tell us which is true.

Winsteps attempts to estimate an individual measure for each person and item within one frame of reference. Usually this happens. But there are exceptions.

19.112 Subtest scoring and weighting

A test or protocol may consist of a series of subtests. You want each person to have a separate measure on each subtest, but in the same frame of reference as the overall test measures.

First, code each item in its item label with what subtest it belongs to.

A. Using DPF.

1. Analyze all the data together.

2. Reporting of subtests.

Reporting subtest items: In an overall analysis, the items of individual subtests can be reported after applying [ISELECT=](#) from the [Specification](#) pull-down box.

Reporting subtest person measures: a measure for each person on each subtest can be obtained by specifying the subtest identification column in the item label in [DPF=](#) and producing [Table 31](#) or the [DPF plot](#).

B. Using separate analyses.

1. Analyze all the data together. Produce an IFILE=if.txt. Produce a PFILE= for global measures.

2. Anchor all items IAFILE=if.txt.

Select one subtest: ISELECT=subtest code

Run Winsteps

Produce IFILE= and PFILE for the subtest.

3. Repeat for the other subtests.

Subtest Weighting

The different subsections have different maximum scores, but you want each subsection to be equally influential in estimating the final measures.

1) Do a standard Rasch analysis to investigate coding problems, misfit, reliability, etc.

2) Weight the items in the different sections so that the overall weights for each section are the same.

Let's suppose that the items are all dichotomies scored 0,1, then

IWEIGHT=*

1-40 1 ; the 40 items in section 1 are given a weight of 1, so weight of section 1 = $40 \times 1 = 40$

41-70 1.333 ; the 30 items in section 2 are given a weight of 1.333, so weight of section 2 = $30 \times 1.333 = 39.99$

71-120 .8 ; the 50 items in section 3 are given a weight of 0.8, so weight of section 3 = $50 \times 0.8 = 40$

*

We want to maintain close to the original overall weight, so that the final reliabilities and standard error have reasonable values.

3) Perform the weighted analysis to report the final person measures.

There are other effective approaches. For instance, perform a separate Rasch analysis for each section, and then combine the subtest measures using www.rasch.org/rmt/rmt83f.htm:

combined measure = $\text{sum}(\text{section measure}/\text{S.E.}(\text{section measure}))/\text{sum}(1/\text{S.E.}(\text{section measure}))$

19.113 TAM R Statistics and Winsteps

TAM is an R Statistics package with some of the capabilities of Winsteps.

Look at the item estimates in <https://cran.r-project.org/web/packages/TAM/TAM.pdf> section: tam.jml Example 1.

	TAM	TAM	Winsteps
	item	Zero	Zero
1	-2.076	-2.02	-2.04
2	-1.743	-1.69	-1.70
3	-1.217	-1.16	-1.17
4	-0.733	-0.68	-0.68
5	-0.338	-0.28	-0.28
6	0.147	0.20	0.20
7	0.593	0.65	0.65
8	1.158	1.21	1.22
9	1.570	1.62	1.64
10	2.091	2.15	2.17

TAM item estimates are not zero-centered. TAM Zero column adjusts the TAM item estimates to be zero centered. These are almost the same as Winsteps item estimates, which are zero-centered. The difference is probably due to slightly different implementation of JMLE and the convergence criteria.

The person "measure" in Winsteps is the TAM theta value. The range of theta estimates in TAM & WINSTEPS are

	TAM original	zero	Winsteps
score 0:	-4.404	-4.35	-4.35
score 5:	-0.070	-0.02	-0.01
score 10:	4.315	4.37	-4.38

The TAM Zero column applies the TAM Zero item adjustment to the TAM thetas. The TAM Zero thetas and Winsteps person measures are essentially the same.

So, TAM and Winsteps estimates differ mainly because:

(1) Estimation bias correction: in Winsteps set

[STBIAS=Yes](#)

(2) Winsteps zero-centers the items: to zero-center the person thetas:

[UPMEAN=0](#)

[UEXTREME=Yes](#)

19.114 Teaching Rasch measurement, Rasch analysis_2

Congratulations! Teaching Rasch methodology can be rewarding, but also has its challenges.

You will probably find that your class is imbued with "Classical" ideas. CTT usually starts with data and statistics. Avoid arguing the wrongs and rights of CTT for as long as possible

So avoid CTT pitfalls by starting from Rasch first principles with lots of examples and classroom interactions.

1. What do we want to measure? Can we express it along a line (latent variable) from less (easy, beginner) to more (hard, expert)?
2. Can we convert this progression into dichotomous items? (relevant items = content validity, item hierarchy = construct validity)
3. How do we expect people of different abilities to respond to these items? (ability ordering = predictive validity)
4. What about people with abilities exactly at, then close to, an item's difficulty? This introduces the idea of probability.
5. We can tell when a person is centered on the dichotomous items when the person's raw score is about 50%
6. Similarly we can tell when an item is close to the mean of the person abilities (thetas) when the frequency (p-value, probability) of success is close to 50%
7. So, based on these probabilities, we can put items and persons on the same "map" of the latent variable based on probabilities: "conjoint measurement". (In my experience, people are amazed at this. They have been taught item analysis and person reporting as separate topics.)
8. Draw a picture of probability (% success) vs. latent variable: [logistic ogive](#) -> logits unit of measurement
9. So with these first principles firmly in mind, introduce the Rasch model. Perhaps starting with one of the "Rasch model derived from" in RMT - such as <https://www.rasch.org/rmt/rmt62c.htm>
10. And now you are off and running ... time to introduce software to do the hard work!
11. If people need to see what is going on "under the hood", there are the Excel worksheets at www.rasch.org/moulton.htm

19.115 Testlets

Testlets are groups of locally dependent items, such as the questions about the same topic. Testlets are obvious in theory, but obscure in practice. If the local dependency of all the items is about the same across testlets, then the testlet effect disappears. So, "testlets" are not the problem, the problem is unevenness of local dependency across the testlets. Table 23.99 - inter-item dependency - tends to be too local. It rarely gives us a good picture of dependency within testlets or across testlets.

Suggestion: in each item label, put a testlet code. Then from the individual-item analysis, output Winsteps Table 28 - item subtotals by testlet. Testlets with average item mean-squares much below 1.0 have noticeably greater dependency (of some type) than the other testlets. These testlets would be the obvious candidates to become PCM items. However, there are usually many reasons for local dependency across items within a dataset, so we would want to see the Table 28 findings confirmed by Table 23.1, etc. Each more dependent testlet should also be the dominant feature in a Contrast.

How are you conceptualizing the testlets?

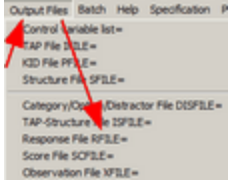
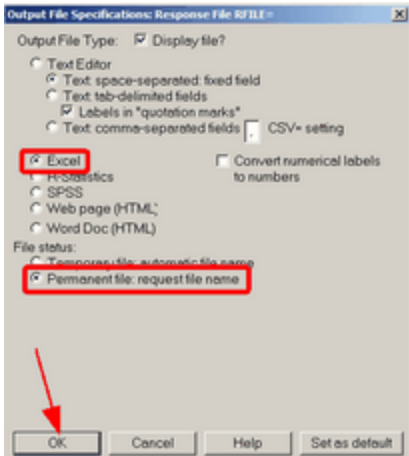
A. If the testlets are conceptualized as groups of separate items, then please put a testlet-code in column 1 of the item label. You can then obtain item-level summary statistics in Winsteps using ISUBTOTAL=1 and Table 27. You can also obtain person-by-testlet measures using DPF=1 and Table 31, and person-group-by-testlet using DPF=1, DIF=person group code column and Table 33.

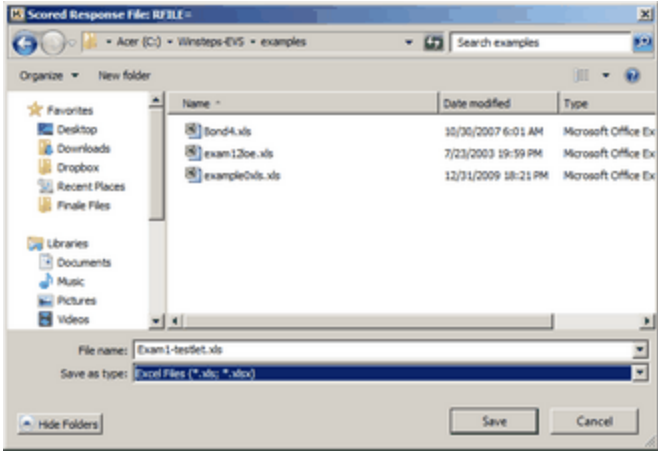
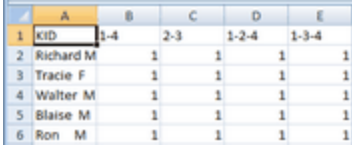
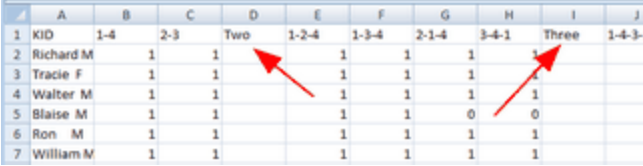
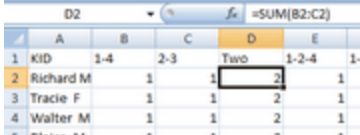
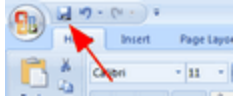

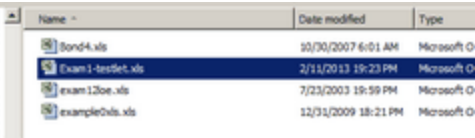
B. If two or more items are highly dependent (so that their residuals are highly correlated in [Table 23.99](#)), then it may be better combine them into one "testlet" item. Testlet Response Theory is technically complex and requires special software, but we can approximate it by combining the dependent items into partial-credit items.

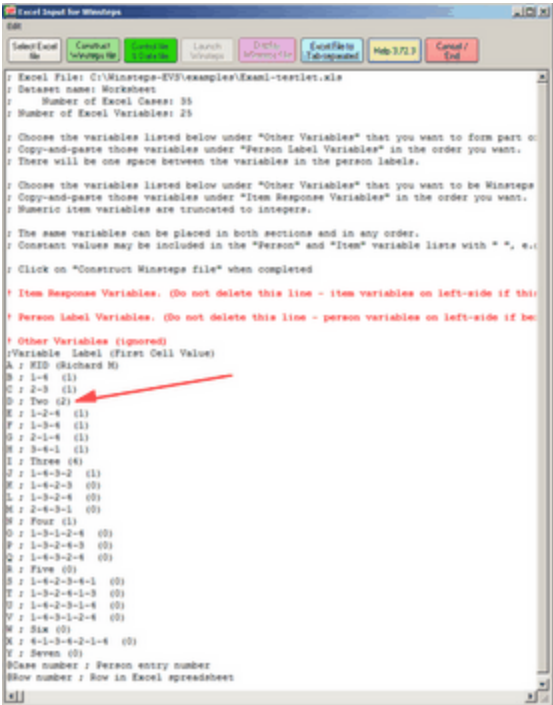
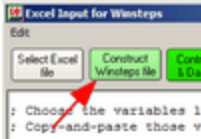
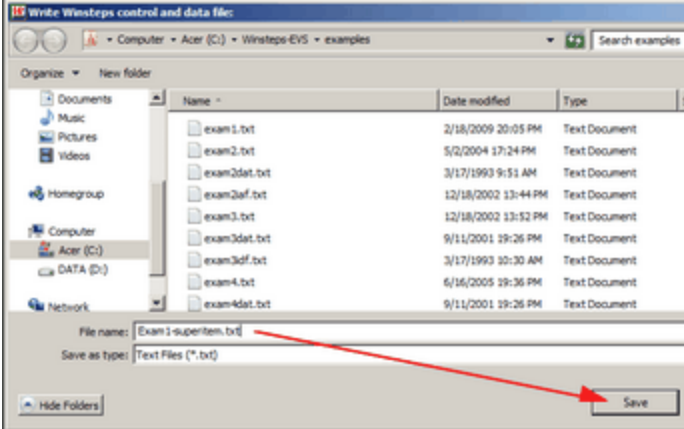
From a Winsteps analysis of the separate items, [Output files menu](#), [RFILE=](#) to Excel (with headings). In Excel, sum the separate items into new combined-item columns. Save the Excel file In Winsteps, [Excel/RSSST menu](#), [Excel option](#). Click on the saved Excel file. Select the combined items from the list of Excel columns. Format a Winsteps control and data file. (You can also select the original uncombined items, but will need to weight them zero with [IWEIGHT=](#) in the new analysis.) Analyze the new Winsteps control and data file with [ISGROUPS=0](#) (partial-credit model)

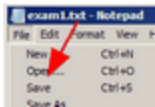

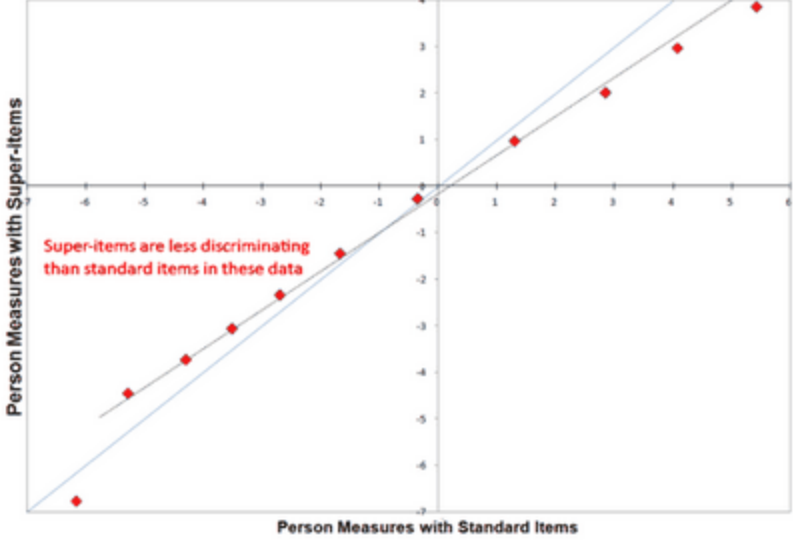
[Cross-plot](#) the person measures from the original analysis and the partial credit analysis. Are there any differences? Which analysis produces more meaningful results?

How to combine items into super-items or testlets

1.	If your data are already in columns in an Excel worksheet, jump to 8.	
2.	Place your data in a Winsteps control and data file. Our example is the Knox Cube Test data, exam1.txt	<pre> ; This file is EXAM1.TXT - (";" starts a comment) &INST : shows this is a control file (optional) TITLE="KNOX CUBE TEST" : Report title NAME1=1 : First column of person label in data file ITEM1=11 : First column of responses in data file NI=18 : Number of items CODES=01 : Valid response codes in the data file CLFILE="" : Labels the observations 0 Wrong : 0 in data is "wrong" 1 Right : 1 in data is "right" "++" : "++" is the end of a list PERSON=KID : Person title: KID means "child" ITEM=TAP : Item title: TAP means "tapping pattern" GENDER=SC9M1 : Gender indicator in column 9 of data record &END : Item labels for 18 items follow : tapping pattern of first item: cube 1 then cube 4 are tapped. 1-4 2-3 1-2-4 1-3-4 2-1-4 3-4-1 1-4-3-2 1-4-2-3 1-3-2-4 2-4-3-1 1-3-1-2-4 1-3-2-4-3 1-4-3-2-4 1-4-2-3-4-1 1-3-2-4-1-3 1-4-2-3-1-4 1-4-3-1-2-4 4-1-3-4-2-1-4 : last tapping pattern: 7 actions to remember! END NAMES : END NAMES or END LABELS must come at end of list Richard M : Here are the 35 person response strings Tracie F : Walter M : </pre>
3.	Launch Winsteps with your control and data file. Analyze your data Verify that all is correct	<pre> Calculating Fit Statistics >===== Standardized Residuals N(0,1) Mean: -.01 S.D.: .85 Time for estimation: 0:0:0.281 Processing Table 0 KNOX CUBE TEST ----- KID 35 INPUT 35 MEASURED INFIT OUTFIT TOTAL COUNT MEASURE REALSE IHNSQ ZSTD OHNSQ ZSTD MEAN 9.7 18.0 -.38 1.19 1.82 -.1 .72 -.1 S.D. 2.4 .0 2.29 .37 1.88 1.2 1.41 .71 REAL RHSE 1.24 TRUE SD 1.93 SEPARATION 1.55 KID RELIABILITY .97 ----- TAP 18 INPUT 18 MEASURED INFIT OUTFIT TOTAL COUNT MEASURE REALSE IHNSQ ZSTD OHNSQ ZSTD MEAN 94.7 175.0 -.95 .67 .99 -.1 .72 -.7 S.D. 69.8 .0 4.92 .63 .29 1.6 .64 1.2 REAL RHSE .92 TRUE SD 4.83 SEPARATION 5.27 TAP RELIABILITY .97 ----- Output written to C:\ANALYSIS\NEW03750\200960US.TXT CODES= 01 Measures constructed: use "Diagnosis" and "Output Tables" menus </pre>
4.	For the scored responses, Output Files menu RFILE= or for the original responses, Output Files menu IPMATRIX=	
5.	Output File dialog box: Excel Permanent file OK	

<p>6. File name dialog box: Enter name for testlet Excel file</p> <p>Our example: Exam1-testlet.xls</p> <p>Save</p>																																																																																									
<p>7. Excel displays with the data.</p> <p>Check that the item labels are in Row 1.</p>	 <table border="1"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>KID</td> <td>1-4</td> <td>2-3</td> <td>1-2-4</td> <td>1-3-4</td> </tr> <tr> <td>2</td> <td>Richard M</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> </tr> <tr> <td>3</td> <td>Tracie F</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> </tr> <tr> <td>4</td> <td>Walter M</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> </tr> <tr> <td>5</td> <td>Blaise M</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> </tr> <tr> <td>6</td> <td>Ron M</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> </tr> </tbody> </table>		A	B	C	D	E	1	KID	1-4	2-3	1-2-4	1-3-4	2	Richard M	1	1	1	1	3	Tracie F	1	1	1	1	4	Walter M	1	1	1	1	5	Blaise M	1	1	1	1	6	Ron M	1	1	1	1																																														
	A	B	C	D	E																																																																																				
1	KID	1-4	2-3	1-2-4	1-3-4																																																																																				
2	Richard M	1	1	1	1																																																																																				
3	Tracie F	1	1	1	1																																																																																				
4	Walter M	1	1	1	1																																																																																				
5	Blaise M	1	1	1	1																																																																																				
6	Ron M	1	1	1	1																																																																																				
<p>8. Insert columns for the testlets, super-items or combined items</p> <p>In our example, we will have a super-item item for each length of tapping pattern: Two, Three, ..., Seven</p>	 <table border="1"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> <th>F</th> <th>G</th> <th>H</th> <th>I</th> <th>J</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>KID</td> <td>1-4</td> <td>2-3</td> <td>Two</td> <td>1-2-4</td> <td>1-3-4</td> <td>2-1-4</td> <td>3-4-1</td> <td>Three</td> <td>1-4-3-</td> </tr> <tr> <td>2</td> <td>Richard M</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> </tr> <tr> <td>3</td> <td>Tracie F</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> </tr> <tr> <td>4</td> <td>Walter M</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> </tr> <tr> <td>5</td> <td>Blaise M</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>0</td> <td>0</td> <td>1</td> <td>1</td> </tr> <tr> <td>6</td> <td>Ron M</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> </tr> <tr> <td>7</td> <td>William M</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> <td>1</td> </tr> </tbody> </table>		A	B	C	D	E	F	G	H	I	J	1	KID	1-4	2-3	Two	1-2-4	1-3-4	2-1-4	3-4-1	Three	1-4-3-	2	Richard M	1	1	1	1	1	1	1	1	1	3	Tracie F	1	1	1	1	1	1	1	1	1	4	Walter M	1	1	1	1	1	1	1	1	1	5	Blaise M	1	1	1	1	1	0	0	1	1	6	Ron M	1	1	1	1	1	1	1	1	1	7	William M	1	1	1	1	1	1	1	1	1
	A	B	C	D	E	F	G	H	I	J																																																																															
1	KID	1-4	2-3	Two	1-2-4	1-3-4	2-1-4	3-4-1	Three	1-4-3-																																																																															
2	Richard M	1	1	1	1	1	1	1	1	1																																																																															
3	Tracie F	1	1	1	1	1	1	1	1	1																																																																															
4	Walter M	1	1	1	1	1	1	1	1	1																																																																															
5	Blaise M	1	1	1	1	1	0	0	1	1																																																																															
6	Ron M	1	1	1	1	1	1	1	1	1																																																																															
7	William M	1	1	1	1	1	1	1	1	1																																																																															
<p>9. Sum the responses to the relevant items into the super-item columns</p>	 <table border="1"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>KID</td> <td>1-4</td> <td>2-3</td> <td>Two</td> <td>1-2-4</td> </tr> <tr> <td>2</td> <td>Richard M</td> <td>1</td> <td>1</td> <td>2</td> <td>1</td> </tr> <tr> <td>3</td> <td>Tracie F</td> <td>1</td> <td>1</td> <td>2</td> <td>1</td> </tr> <tr> <td>4</td> <td>Walter M</td> <td>1</td> <td>1</td> <td>2</td> <td>1</td> </tr> </tbody> </table>		A	B	C	D	E	1	KID	1-4	2-3	Two	1-2-4	2	Richard M	1	1	2	1	3	Tracie F	1	1	2	1	4	Walter M	1	1	2	1																																																										
	A	B	C	D	E																																																																																				
1	KID	1-4	2-3	Two	1-2-4																																																																																				
2	Richard M	1	1	2	1																																																																																				
3	Tracie F	1	1	2	1																																																																																				
4	Walter M	1	1	2	1																																																																																				
<p>10. Save or "Save As" the Excel worksheet</p>																																																																																									
<p>11. Winsteps menu bar Excel/RSSST Selection box: Excel</p>																																																																																									
<p>12. Select the saved Excel worksheet</p>	 <table border="1"> <thead> <tr> <th>Name</th> <th>Date modified</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td>Bond4.xls</td> <td>10/30/2007 6:01 AM</td> <td>Microsoft O</td> </tr> <tr> <td>Exam1-testlet.xls</td> <td>2/11/2013 19:23 PM</td> <td>Microsoft O</td> </tr> <tr> <td>exam12oe.xls</td> <td>7/23/2003 19:59 PM</td> <td>Microsoft O</td> </tr> <tr> <td>examplexls.xls</td> <td>12/31/2009 18:21 PM</td> <td>Microsoft O</td> </tr> </tbody> </table>	Name	Date modified	Type	Bond4.xls	10/30/2007 6:01 AM	Microsoft O	Exam1-testlet.xls	2/11/2013 19:23 PM	Microsoft O	exam12oe.xls	7/23/2003 19:59 PM	Microsoft O	examplexls.xls	12/31/2009 18:21 PM	Microsoft O																																																																									
Name	Date modified	Type																																																																																							
Bond4.xls	10/30/2007 6:01 AM	Microsoft O																																																																																							
Exam1-testlet.xls	2/11/2013 19:23 PM	Microsoft O																																																																																							
exam12oe.xls	7/23/2003 19:59 PM	Microsoft O																																																																																							
examplexls.xls	12/31/2009 18:21 PM	Microsoft O																																																																																							

<p>13. The Excel column labels display with the items and super-items</p>	
<p>14. Copy-and-paste the super-item column labels under “Item Response Variables” Other items can be copied, especially if you plan to weight them zero with IWEIGHT= Also copy the person column labels under “Person Label Variables”</p>	<p>! Item Response Variables. D ; Two (2) I ; Three (4) N ; Four (1) R ; Five (0) W ; Six (0) Y ; Seven (0)</p> <p>! Person Label Variables. A ; KID (Richard M)</p>
<p>15. Click on “Construct Winsteps File”</p>	
<p>16. Save the Winsteps control and data file with a suitable name.</p>	

<p>17. The new Winsteps control and data file displays. It contains the superitems. CODES= contains the scores in the superitems.</p>	<pre> XWTFE = 1 : Matches the widest data value obser ; GROUPS = 0 ; Partial credit model: in case it CODES = 01234 ; matches the data TOTALSCORE = Yes ; Include extreme responses in ; Person Label variables: columns in label: col @KID = 1E9 ; \$C8W9 &END : Item labels follow: columns in label Two ; Item 1 : 1-1 Three ; Item 2 : 2-2 Four ; Item 3 : 3-3 Five ; Item 4 : 4-4 Six ; Item 5 : 5-5 Seven ; Item 6 : 6-6 END NAMES 241000 Richard M 244000 Tracie F </pre>																																																																										
<p>18. Remove the ; before GROUPS = 0 because we expect each super-item to have its own structure (partial credit model).</p>	<pre> XWTFE = 1 : Matches the widest data GROUPS = 0 ; Partial credit model: CODES = 01234 ; matches the data TOTALSCORE = Yes : Include extreme </pre>																																																																										
<p>19. Save the Winsteps control and data file</p>																																																																											
<p>20. Click on "Launch Winsteps"</p>																																																																											
<p>21. The analysis of the super-items is performed</p>	<table border="1"> <thead> <tr> <th>PERSON</th> <th>35 INPUT</th> <th>35 MEASURED</th> <th>INFIT</th> <th>OUTFIT</th> </tr> </thead> <tbody> <tr> <td>TOTAL</td> <td>COUNT</td> <td>MEASURE</td> <td>REALSE</td> <td>IMNSQ</td> <td>ZSTD</td> <td>OMNSQ</td> <td>ZSTD</td> </tr> <tr> <td>MEAN</td> <td>9.7</td> <td>6.8</td> <td>-.26</td> <td>1.38</td> <td>.94</td> <td>-.1</td> <td>.81</td> </tr> <tr> <td>S.D.</td> <td>2.4</td> <td>.8</td> <td>2.76</td> <td>.42</td> <td>1.22</td> <td>1.8</td> <td>1.79</td> </tr> <tr> <td>REAL RMSE</td> <td>1.37</td> <td>TRUE SD</td> <td>2.39</td> <td>SEPARATION</td> <td>1.75</td> <td>PERSON RELIABILITY</td> <td>.75</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th>ITEM</th> <th>6 INPUT</th> <th>4 MEASURED</th> <th>INFIT</th> <th>OUTFIT</th> </tr> </thead> <tbody> <tr> <td>TOTAL</td> <td>COUNT</td> <td>MEASURE</td> <td>REALSE</td> <td>IMNSQ</td> <td>ZSTD</td> <td>OMNSQ</td> <td>ZSTD</td> </tr> <tr> <td>MEAN</td> <td>67.8</td> <td>35.8</td> <td>.88</td> <td>.44</td> <td>1.88</td> <td>.8</td> <td>.89</td> </tr> <tr> <td>S.D.</td> <td>53.8</td> <td>.8</td> <td>3.55</td> <td>.12</td> <td>.11</td> <td>.4</td> <td>.31</td> </tr> <tr> <td>REAL RMSE</td> <td>.45</td> <td>TRUE SD</td> <td>3.52</td> <td>SEPARATION</td> <td>7.88</td> <td>ITEM RELIABILITY</td> <td>.98</td> </tr> </tbody> </table>	PERSON	35 INPUT	35 MEASURED	INFIT	OUTFIT	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD	MEAN	9.7	6.8	-.26	1.38	.94	-.1	.81	S.D.	2.4	.8	2.76	.42	1.22	1.8	1.79	REAL RMSE	1.37	TRUE SD	2.39	SEPARATION	1.75	PERSON RELIABILITY	.75	ITEM	6 INPUT	4 MEASURED	INFIT	OUTFIT	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD	MEAN	67.8	35.8	.88	.44	1.88	.8	.89	S.D.	53.8	.8	3.55	.12	.11	.4	.31	REAL RMSE	.45	TRUE SD	3.52	SEPARATION	7.88	ITEM RELIABILITY	.98
PERSON	35 INPUT	35 MEASURED	INFIT	OUTFIT																																																																							
TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD																																																																				
MEAN	9.7	6.8	-.26	1.38	.94	-.1	.81																																																																				
S.D.	2.4	.8	2.76	.42	1.22	1.8	1.79																																																																				
REAL RMSE	1.37	TRUE SD	2.39	SEPARATION	1.75	PERSON RELIABILITY	.75																																																																				
ITEM	6 INPUT	4 MEASURED	INFIT	OUTFIT																																																																							
TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD																																																																				
MEAN	67.8	35.8	.88	.44	1.88	.8	.89																																																																				
S.D.	53.8	.8	3.55	.12	.11	.4	.31																																																																				
REAL RMSE	.45	TRUE SD	3.52	SEPARATION	7.88	ITEM RELIABILITY	.98																																																																				
<p>22. Appendix: Impact of the super-items on measurement.</p> <p>Output PFILE=pf.txt from the super-item analysis, then in the standard analysis, Scatterplot the person measures against pf.txt</p> <p>In this example, the logit spread of the person measures is less for the super-items than for the standard items. This indicates that there is local dependency among the standard items that comprise each super-item.</p>																																																																											
<p>23. In the plot in 22. all the measures have shifted. The super-items are automatically correctly weighted. However, the mean difficulty of the items changes when separate items are collapsed into super-items. In this situation, either (i) set the mean ability of the persons to be the same in both analyses, UPMEAN=0 ; anchor the mean person ability at 0 logits or (ii) anchor an unchanged item at the same logit difficulty in both analyses,</p>																																																																											

[IAFILE=*](#)

1 0 ; anchor item 1 at 0 logits
*

19.116 Theta or B? parameterizing ability

Item Response Theory (IRT) generally uses "theta", θ , to symbolize a person ability distribution. This accords with the prevalent usage of theta in statistics: "In statistics, θ , the lowercase Greek letter 'theta', is the usual name for a (vector of) parameter(s) of some general probability distribution." (Wikipedia).

In principle, Rasch does not parameterize ability as a distribution, but as the property of an individual. The actual distribution of the individual abilities is usually inconsequential. So, in Winsteps documentation, ability is shown as B, usually with an individual's subscript, n, as B_n .

19.117 t-statistics

Prob. is the two-sided probability of the absolute value of the reported t with the reported d.f., so that a statistically significant finding for a single two-sided t-test is $\text{Prob.} < .05$, and a highly significant finding is $\text{Prob.} < .01$. Please interpret this conservatively, i.e., "barely significant is probably not significant". If you wish to make a [Bonferroni](#) multiple-comparison correction, compare this Prob. with your chosen significance level, e.g., $p < .05$, divided by the number of DIF tests in this Table. This is approximate because of dependencies between the statistics underlying the computation.

Many statistical tests are reported as Student's *t* statistics. This table shows the significance-level values for different degrees of freedom (d.f.). Often the reported *t*-statistics have effectively infinite degrees of freedom and so approximate a unit normal distribution. *t*-statistics with infinite degrees of freedom are also called z-statistics, paralleling the use of "z" in z-scores.

d.f.	p=.05	p=.01	d.f.	p=.05	p=.01	d.f.	p=.05	p=.01
1	12.71	63.66	11	2.20	3.11	21	2.08	2.83
2	4.30	9.93	12	2.18	3.06	22	2.07	2.82
3	3.18	5.84	13	2.16	3.01	23	2.07	2.81
4	2.78	4.60	14	2.15	2.98	24	2.06	2.80
5	2.57	4.03	15	2.13	2.95	25	2.06	2.79
6	2.45	3.71	16	2.12	2.92	30	2.04	2.75
7	2.37	3.50	17	2.11	2.90	100	1.98	2.63
8	2.31	3.36	18	2.10	2.88	1000	1.96	2.58
9	2.26	3.25	19	2.09	2.86	Infinite	1.96	2.58
10	2.23	3.17	20	2.09	2.85	(z-statistic)		

A calculator for the probability of any *t* value and d.f. is at <http://www.danielsoper.com/statcalc3/calc.aspx?id=8>

Welch's refinement of Student's *t*-test for possibly unequal variances:

For sample 1,

M_1 = mean of the sample

SS_1 = sum of squares of observations from the individual sample means

N_1 = sample size (or number of observations)

$SS_1 / (N_1 - 1)$ = sample variance around the mean (or the measure)

$SS_1 / ((N_1 - 1)(N_1))$ = standard error variance = $EV_1 = SE_1^2$

$SE_1 = \text{Sqrt}(EV_1)$ = standard error of the mean (or the measure)

Similarly for sample 2, then

$$t = (M_1 - M_2) / \sqrt{EV_1 + EV_2} = (M_1 - M_2) / \sqrt{SE_1^2 + SE_2^2}$$

with Welch-Satterthwaite d.f. = $(EV_1 + EV_2)^2 / (EV_1^2 / (N_1 - 1) + EV_2^2 / (N_2 - 1))$

which is the same as d.f. = $(SE_1^2 + SE_2^2)^2 / ((SE_1^4 / (N_1 - 1)) + (SE_2^4 / (N_2 - 1)))$

A calculator for this is at <http://www.graphpad.com/quickcalcs/ttest1.cfm?Format=SEM>

Satterthwaite, F. E. (1946), "An Approximate Distribution of Estimates of Variance Components.", Biometrics Bulletin 2: 110-114

Welch, B. L. (1947), "The generalization of "Student's" problem when several different population variances are involved.", Biometrika 34: 28-35

Example: Gender subtotals for [Example0.txt Table 28](#):

$M_1 = 1.62, M_2 = .76, SE_1 = .38, SE_2 = .16, N_1 = 18, N_2 = 57$

Welch: $t = 2.08, d.f. = 23, p = .049$

19.118 Unidimensionality

See [Multidimensionality](#)

19.119 Unobserved and dropped categories

If you have data in which a category is not observed, then you must make an assertion about the unobserved category. There are several options:

For intermediate categories: either

(a) this category will never be observed (this is called a "structural zero"). Generally, these categories are collapsed or recoded out of the rating scale hierarchy. This happens automatically with [STKEEP=No](#).

or (b) this category didn't happen to be observed this time (an "incidental" or "sampling" zero). These categories can be maintained in the rating scale hierarchy (using [STKEEP=Yes](#)), but are estimated to be observed with a probability of zero.

1. Dummy data

For extreme categories:

(a) if this category will never be observed, the rating scale is analyzed as a shorter scale. This is the Winsteps standard.

(b) if this category may be observed, then introduce a dummy record into the data set which includes the unobserved extreme category, and also extreme categories for all other items except the easiest (or hardest) item. This forces the rare category into the category hierarchy.

(c) If an extreme (top or bottom) category is only observed for persons with extreme scores, then that category will be dropped from the rating (or partial credit) scales. This can lead to apparently paradoxical or incomplete results. This is particularly noticeable with [ISGROUPS=0](#). Again, dummy data solves this.

In order to account for unobserved extreme categories, a dummy data record needs to be introduced. If there is a dropped bottom category, then append to the data file a person data record which has bottom categories for all items except the easiest, or if the easiest item is in question, except for the second easiest.

If there is a dropped top category, then append to the data file a person data record which has top categories for all items except the most difficult, or if the most difficult item is in question, except for the second most difficult.

This extra person record will have very little impact on the relative measures of the non-extreme persons, especially if you give it a very small weight with [PWEIGHT=](#), but will make all categories of all items active in the measurement process.

If it is required to produce person statistics omitting the dummy record, then at the [Specification Menu](#) use [PDELETE=](#) or [PSELECT=](#) to omit it, and regenerate Table 3.

See also [Null or unobserved categories: structural and incidental zeroes](#)

Example: when the "Liking for Science" data, [example0.txt](#), are analyzed with the Partial Credit Model, [ISGROUPS=0](#), item 18, "Go on a Picnic", does not have the bottom 0 category of the 3-category 0-1-2 rating scale. The next easiest item is item 19, "Go to the zoo", so add a dummy data record looking like, to which you can give a very small influence by using [PWEIGHT=](#) with a small value.

*****01***** Dummy data (0 for item 18, 1 for item 19)

Example 2: Item 1 has no category 0 in this dataset, but the category should exist.

```
Title="test"
Codes=012
Isgroups=0 ; each item has its own rating-scale
NI=3
Item1=1
Name1=1
Pweight=*
4 0.01 ; small weight for dummy person 4 - adjust this weight if 0 for item 4 is too far away
*
&END
END LABELS
211 Person 1
122 Person 2
100 Person 3
011 Dummy person 4 because there was no 0 for item 1
```

2. Forced category range

Another approach is to specify the unobserved categories with [ISRANGE=](#), and then model all the categories with a polynomial function: [SFUNCTION=](#).

3. Anchored thresholds

Using [SAFILE=](#), reasonable threshold values can be applied to the item so that thresholds for unobserved categories do not need to be estimated.

19.120 User-friendly rescaling: zero point and unit

Winsteps sets the zero point (local origin) at the mean item difficulty (including extreme items if [UEXTREME=Yes](#)). The unit of measurement is the [logit](#). Changing the zero point and transforming logits into other units can be done with [UIMEAN=](#), [UPMEAN=](#), [USCALE=](#). These transformed measurements can be more meaningful for particular applications, see Chapter 8 of *BTD*. Anchor values are treated according to [UASCALE=](#)

Example 0: we observe the (current mean) and (current S.D.) of a set of measures. We want (desired mean) and (desired S.D.) then

new USCALE = (current USCALE=) *(desired S.D.) / (current S.D.)

new UIMEAN = desired mean - (current mean - current UIMEAN=)*(new USCALE=) / (current USCALE=)

we observe the (current high) and (current low). We want (desired high) and (desired low) then

new USCALE = (current USCALE=) *(desired high-desired low) / (current high-current low)

new UIMEAN = desired mean - (current mean - current UIMEAN=)*(new USCALE=) / (current USCALE=)

Example 1: CHIPs are a useful transformation, in which 1 logit = 4.55 CHIPs. In this user-scaling system, standard errors tend to be about 1 CHIP in size. The recommended control variable settings are:

USCALE = 4.55

UIMEAN = 50

UDECIM = 1

MRANGE = 50

The probability structure of this relationship is:

Probability of Success

Difference between Person Ability Measure and Item Difficulty Measure in CHIPs

.10 -10

.25 -5

.50 0

.75 5

.90 10

Example 2: WITs are one tenth the size of CHIPs, enabling the elimination of decimals from your output tables.

USCALE = 45.5

UIMEAN = 500

UDECIM = 0

MRANGE = 500

Example 3: You want the lowest reportable person measure to be 0 and the highest to be 100. Looking at Table 20, you see the extreme values are -4.53 and +5.72. You have not used USCALE= and UIMEAN=, so the current values are USCALE=1, UIMEAN=0.

Therefore,

"wanted low" person measure = 0

"wanted high" person measure = 100

"wanted range" = "wanted high" - "wanted low" = 100 - 0 = 100

"current low" person measure = -4.53

"current high" person measure = 5.72

"current range" = "current high" - "current low" = 5.72 - -4.53 = 10.25

USCALE= (wanted range) / (current range) = 100 / 10.25 = 9.76

UIMEAN= (wanted low) - (current low * USCALE) = 0 - (-4.53 * 9.76) = 44.20

Required values are:

USCALE = 9.76

UIMEAN = 44.20

UDECIM = 0 to show no decimal places in report

Double checking, when previous values are UIMEAN=0, USCALE=1:

low value = (current low)*(USCALE=) + (UIMEAN=) = (-4.53 * 9.76) + 44.20 = -0.01 = 0

high value = (current high)*(USCALE=) + (UIMEAN=) = (5.72 * 9.76) + 44.20 = 100.02 - 100

Example 4: You want the lowest reportable person measure to be 0 and the highest to be 100. Looking at Table 20, you see the extreme values are -4.53 and +5.72. The current values are USCALE=2, UIMEAN=1.5

Therefore,

"wanted low" person measure = 0

"wanted high" person measure = 100

"wanted range" = "wanted high" - "wanted low" = 100 - 0 = 100

"current low" person measure = -4.53

"current high" person measure = 5.72

"current range" = "current high" - "current low" = 5.72 - -4.53 = 10.25

"current UIMEAN" = 1.5

"current USCALE" = 2

USCALE= [(wanted range) / (current range)]/(current USCALE) = [100 / 10.25]/2 = 9.76/2 = 4.88

$$UIMEAN = (\text{wanted low}) - [(\text{current low} - \text{current UIMEAN}) * USCALE] = 0 - [(-4.53 - 1.5) * 9.76] = 58.85$$

Required values are:

$$USCALE = 4.88$$

$$UIMEAN = 58.85$$

UDECIM = 0 to show no decimal places in report

Double checking is most easily done by looking at the new Table 20.

Example 5: You want the lowest reportable person measure to be 100 and the highest to be 900. Looking at Table 20, you see the extreme values are -4.53 and +5.72. Looking at the second page of output, you see the current values are USCALE=1 and UMEAN=0.

$$USCALE = (\text{previous USCALE}) * (\text{wanted range: } 900 - 100) / (\text{reported range: } 5.72 - -4.53) = 1 * 800 / 10.25 = 78.05$$

$$UMEAN = (\text{wanted low}) - (\text{reported low} - \text{previous UMEAN}) * (\text{wanted range}) / (\text{reported range}) = 100 - (-4.53 - 0) * 800 / 10.25 = 453.56$$

UDECIM = 0 to show no decimal places in report

Example 6: You want norm-referenced user-scaling, such that the person mean is 0.0, and the person [standard deviation](#) is 1.0.

In a standard analysis, set:

UDECIM=4

USCALE=1

UMEAN=0

Look at Table 18

```

+-----+
|ENTRY   RAW
|NUMBER  SCORE  COUNT  MEASURE
+-----+
| MEAN    6.7   14.0   -.3728
| P.SD    2.4    .0     2.2202
+-----+

```

Set (either in a new analysis, or using the "Specification" pull-down menu

$$USCALE = 1/\text{person S.D.} = 1/2.2202 = 0.4504$$

$$UMEAN = -\text{person mean}/\text{person S.D.} = -(-.3728)/2.2202 = 0.1679$$

Look at Table 18

```

+-----+
|ENTRY   RAW
|NUMBER  SCORE  COUNT  MEASURE
+-----+
| MEAN    6.7   14.0   .0000
| P.SD    2.4    .0     1.0000
+-----+

```

Example 7: You want to give your pass-fail point the value 400 and 100 to the lowest reported measure. Inspecting your output you see that the pass-fail point is at 1.5 logits, and that -3.3 logits is the lowest reported measure.

Then $400 - 100$ new units = $1.5 - (-3.3)$ logits, so

$$USCALE = 300 / 4.8 = 62.5$$

$$UMEAN = 400 - (1.5) * 62.5 = 306.25$$

$$\text{Then: } 1.5 \text{ logits} = 306.25 + 1.5 * 62.5 = 400$$

$$-3.3 \text{ logits} = 306.25 - 3.3 * 62.5 = 100$$

Example 8: Form 1 and Form 2 are obtained from an item bank. We rescale Form 1 and want to apply the same rescaling to Form 2.

Do the Form 1 analysis:

UASCALE= is the scaling of the item bank scale (1 if logits)

IAFILE= is the list of bank item difficulties for Form 1
No UIMEAN=, UPMEAN=, UMEAN= or USCALE= command.

We rescale Form 1 to have its own scale setting UIMEAN= and USCALE= with the [scaling calculator](#).
Note down the values or list them with the [Control Variable list](#) from the Output Files menu.

In the Form 2 control file:
UASCALE= is the scaling of the item bank scale (1 if logits)
IAFILE= is the list of bank item difficulties for Form 2
No UIMEAN, UPMEAN or UMEAN command.

In the Form 2 control file, add these commands:
TFILE=*
; these are actioned after the analysis, at the beginning of the reporting
14.1 ; this will report the bank item anchor values as a check
3.1 ; summary statistics
UIMEAN= value from Form 1 ; the values used for the analysis are rescaled
USCALE= value from Form 1
14.1 ; rescaled items
3.1 ; summary statistics
18.1 ; rescaled persons
*

You can add these to an analysis of the Form 1 data to check that it works.

Do the analysis of Form 2.

19.121 Using a word processor or text editor

If you don't like NotePad, then [change word processor](#)

a) Input files: all lines in your control and data files follow DOS text conventions. This means that files created with a Word Processor, such as "Word Perfect", must be saved as "DOS-text with line breaks" or "ASCII" files.

1. Lines must not contain tabs or word processor codes.
2. Lines cannot overflow onto the next line, except for data records which are processed using the [FORMAT=](#) or [MFORMS=](#) control variables.
3. Lines must end with DOS or ASCII Carriage Return and Line Feed codes.

Be particularly careful to instruct your Word Processor to allow more characters within each line than are present in the longest line in your control or data files. Then your Word Processor will not break long data or control lines into two or more text lines with "Soft Return" codes. These cause Winsteps to malfunction. Space for a large number of characters per line is obtained by specifying a very wide paper size and/or a very small type size to your Word Processor.

When using "Word Perfect" to edit control or data files, select the smallest available type size (often 20 cpi or 5 pt). Define and use a very wide (50 inch) page style. It does not matter whether your printer can actually print it. Always save control and data files as "DOS-text with line breaks" or ASCII files.

With WordStar, use "Non-Document" mode to avoid these difficulties.

b) Output files: when importing Winsteps output into a document file, the following options have proved useful:

Base Font - 17 cpi (or more) or 8 point (or smaller) or 132 characters per line (or more)
Left Justify
Page numbering
Margins: top = 1", bottom = 0.5", left = 1", right = 0"

19.122 UTF-8 Chinese character set

Multibyte UTF-8 codes (Chinese) are allowed in person and item labels..Winsteps control variables are ASCII characters.

Excel/RSSST menu:

Your data in an Excel worksheet:

È	Item 2	Item3	Name
1	1	0	管理和社会发展
1	0	1	中英文均可
0	1	0	及时通过邮箱投递

Copy your data to a text-tab file if Excel fails, then use text-tab on the Excel/RSSST menu.

```
È Item 2 Item3 Name
1 1 0 管理和社会发展
1 0 1 中英文均可
0 1 0 及时通过邮箱投递
```

In conversion window:

```
! Other Variables (ignored)
;Variable Label (First Active Cell Value)
1 Ã^ ; 1
2 Item 2 ; 1
3 Item3 ; 0
4 Name ; ç@;çtâ'@çç#â¼šâ`â±• ignore this
@Case number ; Person entry number
@Row number ; Row in Text file
```

Winsteps control file:

```
&INST
.....
&END ; Item labels follow
Item 1 ; Item 1 : 1-1
Item 2 ; Item 2 : 2-2
Item 3 ; Item 3 : 3-3
END NAMES
110 管理和社会发展
101 中英文均可
010 及时通过邮箱投递
```

Winsteps output table:

```
-----
|ENTRY  TOTAL|
|NUMBER SCORE| PERSON
|-----+-----|
| 1      2| 管理和社会发展
| 2      2| 中英文均可
| 3      1| 及时通过邮箱投递
```

UTF8CODE= substitutes for multibyte characters to maintain the spacing. Here, UTF8CODE= ~

```
-----
| DATA | ITEM          | PERSON
|-----+-----|
| 2     | ~~~~~~       | 馬時撰
| 0     | Item 1 ~~~~~ | 馬時撰
| 1     | Item 2 ~~~~~ | 及时通过邮箱投递
```

19.123 Validity investigation

Question: I really want you to help me in a simple explanation understand how in practice can I go about collecting validity evidences via Rasch Analysis with Winsteps to support use and inference of my test?

Answer: There are many types of validity described in the literature, but they summarize to two main topics:

1. Construct/Content validity: does the test measure what it is intended to measure?

2. Predictive validity: does the test produce measures which correspond to what we know about the persons?

Investigation of these validities is performed directly by inspection of the results of the analysis (Rasch or Classical or ...), or indirectly through correlations of the Rasch measures (or raw scores, etc.) with other numbers which are thought to be good indicators of what we want.

Question: That is what exactly type of validity questions should I ask and how can I answer them using Rasch analysis?

Answer: 1. Construct validity: we need a "construct theory" (i.e., some idea about our latent variable) - we need to state explicitly, before we do our analysis, what will be a more-difficult item, and what will be a less-difficult item.

Certainly we can all do that with arithmetic items: $2+2=?$ is easy. $567856+97765=?$ is hard.

If the Table 1 item map agrees with your statement. Then the test has "construct validity". It is measuring what you intended to measure.

2. Predictive validity: we need to we need to state explicitly, before we do our analysis, what will be the characteristics of a person with a higher measure, and what will be the characteristics of a person with a lower measure. And preferably code these into the person labels in our Winsteps control file.

For arithmetic, we expect older children, children in higher grades, children with better nutrition, children with fewer developmental or discipline problems, etc. to have higher measures. And the reverse for lower measures.

If the Table 1 person map agrees with your statement. Then the test has "predictive validity". It is "predicting" what we expected it to predict. (In statistics, "predict" doesn't mean "predict the future", "predict" means predict some numbers obtained by other means.

Question: More specifically, in using, for example, DIF analysis via Winsteps what type of validity question I am trying to answer?

Answer: 1. Construct validity: DIF implies that the item difficulty is different for different groups. The meaning of the construct has changed! Perhaps the differences are too small to matter. Perhaps omitting the DIF item will solve the problem.

Perhaps making the DIF item into two items will solve the problem.

For instance, questions about "snow" change their difficulty. In polar countries they are easy. In tropical countries they are difficult. When we discover this DIF, we would define this as two different items, and so maintain the integrity of the "weather-knowledge" construct.

2. Predictive validity: DIF implies that the predictions made for one group of persons, based on their measures, differs from the predictions made for another group. Do the differences matter? Do we need separate measurement systems? ...

Question: Similarly, in using Fit statistics, dimensionality, and order of item difficulty what type of validity questions I am attempting to answer via Winsteps?

Answer: They are the same questions every time. Construct Validity and Predictive Validity. Is there a threat to validity? Is it big enough to matter in a practical way? What is the most effective way of lessening or eliminating the threat?

Question: I used the numbers in my Winsteps output to prove the Validity of my instrument, but a reviewer says that is not enough.

Answer: Your Validity seems to be relating only to statistical validity. Generally speaking this is of lower concern than:

1. Construct/Content validity - is the instrument measuring what it is intended to measure: e.g., Are these arithmetic items? Does their difficulty order agree with the construct theory about which arithmetic items are easier (one digit addition) and which are harder (long division)? You may need a content expert to assist with this.

2. Predictive validity - do the measures make sense with our experience of people whom we perceive to have more and less of what we intend to measure? For instance, with increasing elementary-school grade-levels do person (children) measures (thetas) increase on average? We may tie this to the results of another accepted instrument = concurrent validity.

3. If we satisfy (1) and (2), we can then proceed to the type of fine-tuning that you are discussing: Statistical validity

Are there off-dimensional, ambiguous, duplicative, etc., items that should be dropped or rewritten?

Are there items that have DIF, e.g., gender DIF: an arithmetic item that references cooking or carpentry?

Then we need to decide how many levels of competence the instrument is intended to detect. This ties in with the Test (=Person Sample) reliability. If we only need to separate high performers from low performers, then a reliability of 0.8 is enough. High-middle-low we need 0.9. More levels we need to go further toward 1.0.

19.124 Videos and Tutorials

There are Winsteps tutorials at www.winsteps.com/tutorials.htm

Youtube has over 70 Rasch-related videos at www.youtube.com

19.125 Weighting items and persons

There are circumstances in which certain items are to be given more influence in constructing the measures than others. For instance, certain items may be considered critical to the demonstration of competence. Though Winsteps supports several methods, **IWEIGHT=** is simplest for items, and **PWEIGHT=** for persons. Another approach is to replicate the data for particular items. This can be done with **FORMAT=** without changing the data file. Items can also be **rescored** from say, 0-1 to 0-2, but this makes variable maps difficult to interpret.

Unweighted and Weighted analysis: unweighted data is preferable for calibrating the Rasch items. This is because each observation is modeled to contribute one unit of independent statistical information. The effect of weighting is to distort the distribution of independent statistical information in the data. A practical approach is:

Step 1. Analyze the data without weighting. Investigate misfit, construct validity, etc.

Step 2. Weight the items. Compare the item calibrations with weighted and unweighted data to identify where there are discrepancies.

The true reliability of the measures is from the unweighted analysis. Weighting introduces an arbitrariness into the analysis. One solution is to adjust the weights to maintain the unweighted reliability = R_u . The reliability of the weighted analysis, using an initial set of weights, = R_w . We can then scale the weights using the Spearman-Brown Prophecy Formula: $S = R_u * (1 - R_w) / ((1 - R_u) * R_w)$. Multiply the initial set of weights by S . Then the weighted and unweighted reliabilities should be the same.

Standard errors and fit statistics: The weights applied to items or persons are used in computing the measure estimates, **standard errors** and fit statistics. When using significance tests with weighting, normalize the weights so that the total amount of independent statistical information in the data is not over- or under-inflated, i.e., when using **PWEIGHT=** with an observed sample size of N , multiply all **PWEIGHT=** values by $N / (\text{sum of all weights})$.

The standard is weights = 1.

When an item or person is weighted as 2, then the estimation acts as though that item or person appears twice in the data file.

When an item or person is weighted as 0, then that person does not influence the estimates, standard errors or fit statistics of other persons and items, but does have measure, standard error and fit statistics computed on all observations for itself. This is useful for evaluating pilot or off-dimensional items, or measuring idiosyncratic persons.

Estimation with weighting

Observation = X_{ni}

Expected value (computed using the Rasch model) = E_{ni}

Accumulated raw score = Accumulated raw score = $X_{ni} * IWEIGHT * PWEIGHT$

Accumulated expected score = Accumulated expected score = $E_{ni} * IWEIGHT * PWEIGHT$

Accumulated marginal count for item = Accumulated marginal count for item + $IWEIGHT * PWEIGHT$

Accumulated marginal count for person = Accumulated marginal count for person + $IWEIGHT * PWEIGHT$

Special rules apply when $IWEIGHT=0$ or $PWEIGHT=0$.

$IWEIGHT=0$ the item totals are incremented by $PWEIGHT$. The person totals are not incremented.

$PWEIGHT=0$ the person totals are incremented by $IWEIGHT$. The item totals are not incremented.

JMLE Estimation Accumulated expected score (for each person and each item) = Accumulated raw score (for each person and each item).

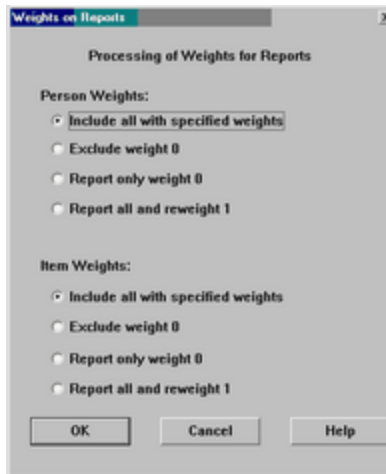
Weight Selection for Tables 23 and 24: On the [output tables menu](#), these are the options for persons and/or items. When [IWEIGHT=](#) or [PWEIGHT=](#) are used in estimation, reports can be adjusted to reflect those weights or not. Weights of zero are useful for pilot items, variant items or persons with unusual characteristics. These can be reported exclusively or excluded from reports.

(1) all items or persons are reported, with their weights (the standard). Tables 23 and 24 are computed as though all weights are 1.

(2) items or persons with a weight of 0 are excluded from the reporting. Tables 23 and 24 are computed as though all weights are 1, but zero weights are omitted.

(3) only items or persons with a weight of 0 are reported. Tables 23 and 24 are computed only from items or persons with a weight of 0.

(4) all items or persons are reported as though they have a weight of 1.



Example 1: MCQ items are scored 0-1. CR items are scored 0-0.5-1. How can we combine them in one Winsteps analysis.

Approach A (recommended). Double the scores of the CR items to 0-1-2, and then IWEIGHT= them by 0.5.

Approach B (not recommended). Double the scores of the MCQ items to 0-(1)-2, and double the scores of the CR items to 0-1-2. IWEIGHT= them all by 0.5. This gives the MCQ items a rating scale in which the middle category is not observed, making their ICCs steeper.

Example 2: What is the cut-score in a weighted analysis corresponding to a cut-score in an unweighted analysis?

Here is an approach:

1. In the unweighted analysis, identify the logit value of the cut-score. Save the person measures to Excel.
2. In the weighted analysis, save the person measures to Excel.
3. Cross-plot the weighted person measures (y-axis) against the unweighted person measures (x-axis)
4. Use the Excel "trend line" function to obtain a reasonable curve through the person-measure points.
5. Identify the value on the y-axis (weighted cut-score measure) corresponding to the value on the x-axis of the unweighted cut-score measure.

19.126 Winsteps: history and steps

What is the origin of Winsteps and to what does "steps" refer? Winsteps is an outcome of this process of development:

In 1983, Benjamin D. "Ben" Wright of the University of Chicago and John "Mike" Linacre released the first Rasch analysis program for personal computers (IBM PCs). It was also the first Rasch program to allow missing data. It was called Microscale, "Rasch scaling by microcomputer". Since MS-DOS was limited to 8-character program names, the actual execution name was "MSCALE".

1987: Mscale (dichotomies and Andrich rating scales) + Msteps (for partial credit "steps"). Ben implemented the Microscale algorithm on a Unix minicomputer, but kept the PC execution name, "Mscale".

1989: Bigscale (back to PCs). Again under MS-DOS but with much larger datasets. Mike takes over development again.

1991: Bigsteps (the functionality of Msteps was included in Bigscale). Ben interpreted this to mean "big steps forward in social science measurement"

1998: Winsteps (Windows-native Bigsteps). Ben interpreted this to mean "winning steps forward in social science measurement".

When talking about Rasch measurement, Ben Wright used "step" to mean:

(a) the category number counting up from 0 at the bottom. The bottom step, for dichotomies or polytomies, was the lowest category, always numbered 0. Ben would talk about going up and down the steps as one moved up and down the latent variable.

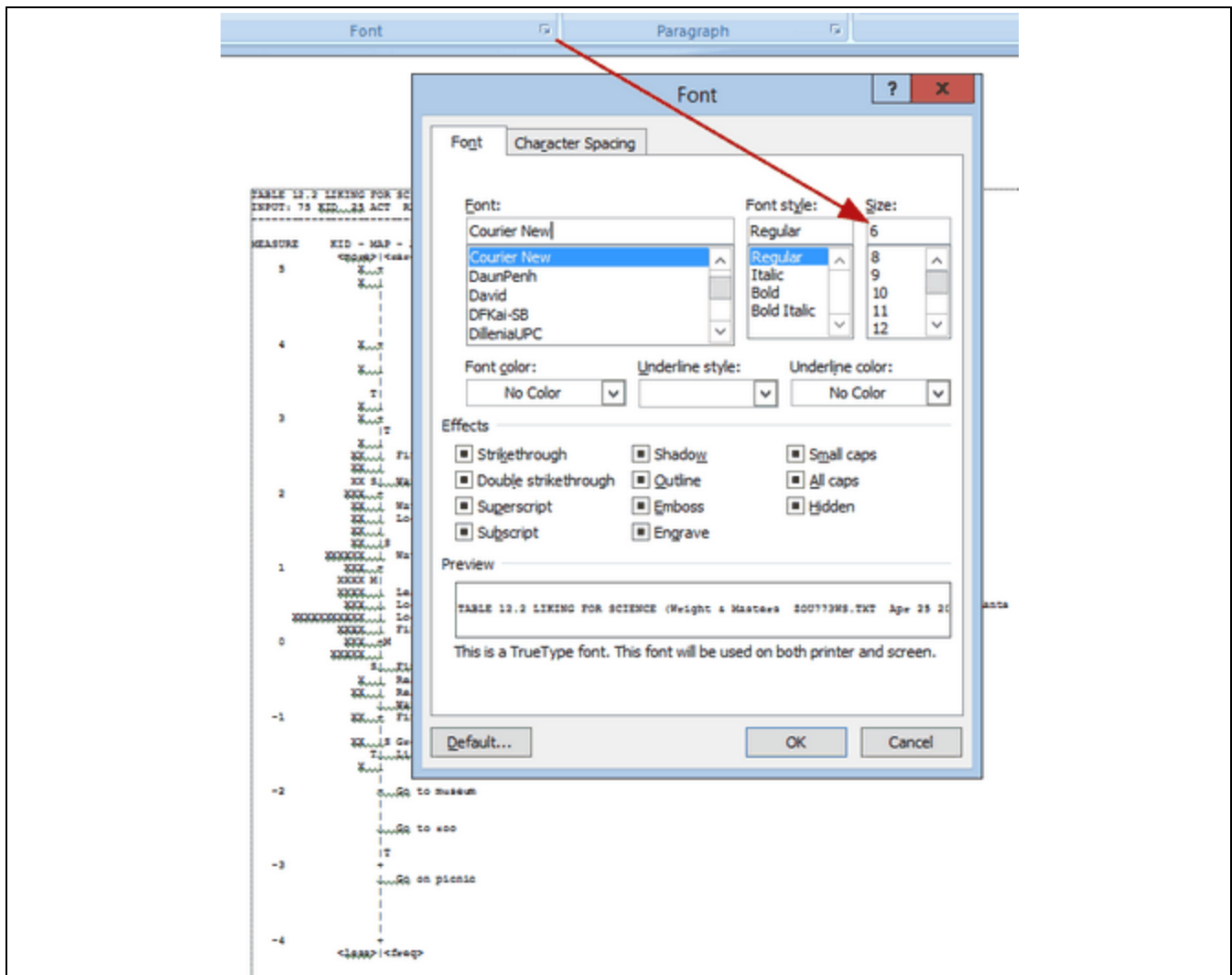
(b) the location of the transition from one category to the next higher category on the latent variable. Now called the Rasch-Andrich threshold for polytomies and the item difficulty for dichotomies.

(c) the process of moving from one category to the next as one's amount of the latent variable changes. A low negative threshold below a category indicates that the category is easy to step into as one moves up the latent variable. A high positive threshold below a category indicates a category that is hard to step into. So "[disordered](#)" thresholds around a category (high below, low above) indicate a category that is "hard to step into and easy to step out of" as one moves up the latent variable, i.e., a narrow category. The extreme of this is an infinitely-narrow, i.e., unobserved, category. It is infinitely hard to step into and infinitely easy to step out of.

19.127 Wright maps in Word, Excel, R, Notepad++

Item map in R: wrightmap.org

Transferring the item map to a Microsoft Word document:
<i>As a Word document:</i> Output the item map as a Word document using Winsteps specification ASCII=D .
<i>As text, copied into Word (also works for Notepad++):</i> Adjust the size of the map using Winsteps specifications LINELENGTH= , MRANGE= , MAXPAGE= As text: Copy-and-paste the map into the .doc file. In the .doc file, 1. copy-and-paste from Notepad to Word. 2. Select all (ctrl+a) in Word 3. Change font to courier New or another fixed-space font, such as Consol as, Luci da Consol e, Andal e Mono. 4. If the lines wrap, then 4a. change page orientation to "Landscape" 4b. smaller point size (you can type in point-sizes smaller than on the pull-down list): "6" works well.



5. Item maps can be combined in Word (or Notepad++) using rectangular-copy, alt+mouse.

6. Item maps can be rotated in Microsoft Word with:

Alt-F11 to launch VBA

In Immediate Window:

Selection.Orientation = wdTextOrientationVerticalFarEast

As a picture:

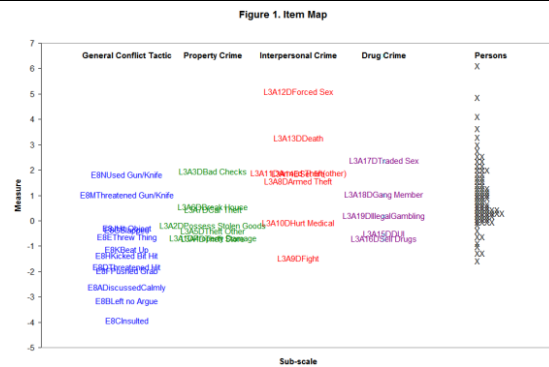
Display the .txt file on your screen. Press your PrintScreen key to copy a picture of the .txt file on your screen into the Windows clipboard. Paste into your Word document. Format the picture.

As a webpage:

Output the item map as a webpage using Winsteps specification [ASCII=W](#). Copy-and-paste the item map into your Word document.

Item Map in Excel

1. We want to produce this Excel Chart (or close approximation)



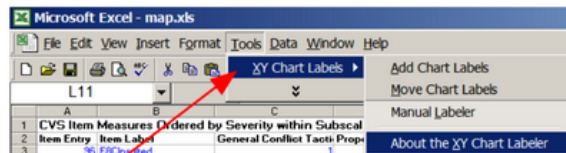
2. from this Excel worksheet.

Constructed with:

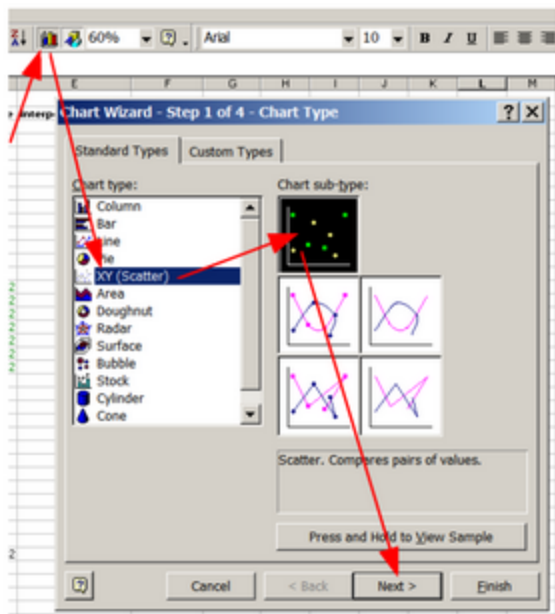
1. x-axis column is the sub-scale number for each item
2. y-axis column is the item measure from [IFILE=](#) output to an Excel file.
3. put in headings with x and y coordinates to place at the top or bottom of the columns
4. for the person measure distribution, Use the [Scorefile=](#) from the Output File menu. Write the file to Excel, and then use the "Measure" and "Frequency" columns. (If you want different "bins", use [PFILE=](#) and the Excel =frequency() formula.)
5. Then use the =REPT() Excel function to format a person sample bar

Item Entry	Item Label	General Conflict Tactic	Property Crime	Interpersonal Crime	Drug Crime	Measure
1	CVS Item Measures Ordered by Severity within Subscale					
2	Item Entry					
3	Item Label					
4	36 EBCInsulted	1				-3.97
5	35 EBLLeft No Argue	1				-2.79
6	34 EBADDiscussedCalmly	1				-2.65
7	33 EBPPushedClub	1				-2.01
8	32 EBTThreatenedHb	1				-1.86
9	31 EBPInsultedBtHb	1				-1.4
10	30 EBPBeardUp	1				-1.17
11	29 EBCThree Things	1				-0.67
12	28 EBCDiscussed	1				-0.4
13	27 EBPThreatenedGun/Knife	1				-0.32
14	26 EBPUsed Gun/Knife	1				-0.3
15	25 EBPThreatenedGun/Knife	1				-0.23
16	24 EBPUsed Gun/Knife	1				-0.18
17	23 EBPThreatenedGun/Knife	1				-0.13
18	22 EBPThreatenedGun/Knife	1				-0.11
19	21 EBPThreatenedGun/Knife	1				-0.1
20	20 EBPThreatenedGun/Knife	1				-0.09
21	19 EBPThreatenedGun/Knife	1				-0.08
22	18 EBPThreatenedGun/Knife	1				-0.07
23	17 EBPThreatenedGun/Knife	1				-0.06
24	16 EBPThreatenedGun/Knife	1				-0.05
25	15 EBPThreatenedGun/Knife	1				-0.04
26	14 EBPThreatenedGun/Knife	1				-0.03
27	13 EBPThreatenedGun/Knife	1				-0.02
28	12 EBPThreatenedGun/Knife	1				-0.01
29	11 EBPThreatenedGun/Knife	1				0
30	10 EBPThreatenedGun/Knife	1				0
31	9 EBPThreatenedGun/Knife	1				0
32	8 EBPThreatenedGun/Knife	1				0
33	7 EBPThreatenedGun/Knife	1				0
34	6 EBPThreatenedGun/Knife	1				0
35	5 EBPThreatenedGun/Knife	1				0
36	4 EBPThreatenedGun/Knife	1				0
37	3 EBPThreatenedGun/Knife	1				0
38	2 EBPThreatenedGun/Knife	1				0
39	1 EBPThreatenedGun/Knife	1				0
40	Measure					6.5
41	Frequency					46.26
42	Bar using =REPT					Persons
43	sub-scale					5
44						
45						
46						
47						
48						
49						
50						
51						
52						
53						
54						
55						
56						
57						
58						
59						
60						
61						
62						
63						
64						
65						
66						
67						
68						
69						
70						
71						
72						
73						
74						
75						
76						
77						
78						
79						
80						
81						
82						
83						
84						
85						
86						
87						
88						
89						
90						
91						
92						

3. On the Excel worksheet toolbar, verify that you have XY Chart Labels. If not, download and install from: www.appspro.com/Utilities/ChartLabeler.htm



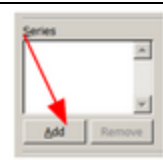
4. On the Excel Worksheet page,
 Click on Chart icon or "Insert" ribbon
 Click on XY (Scatter)
 Click on non-lines sub-type
 Click on Next



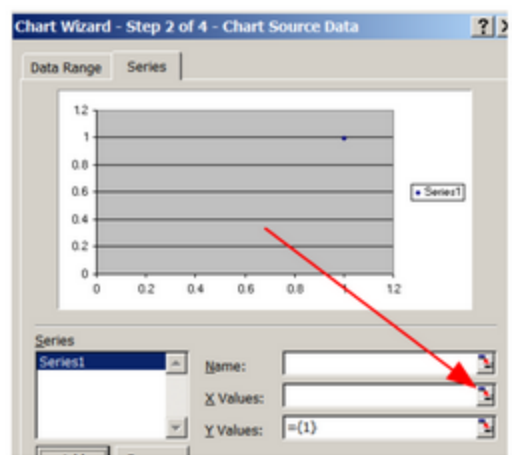
5. Click on Series



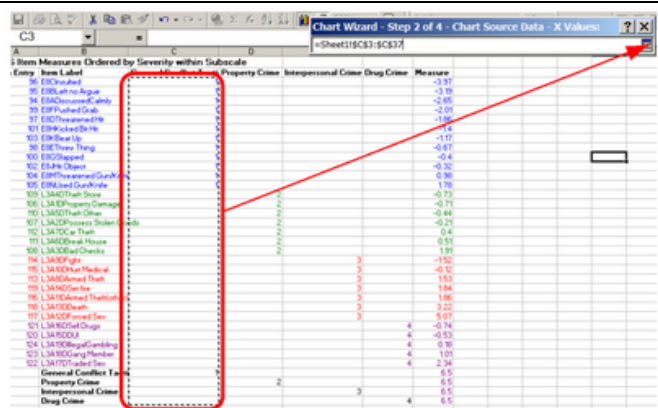
6. Click on Add



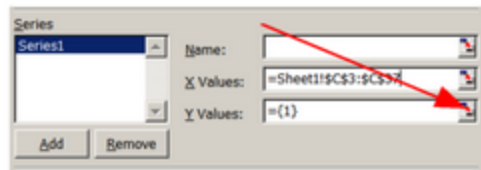
7. Click on X values



8. On Worksheet, drag mouse to highlight first sub-scale column.
Then click on series-selector icon



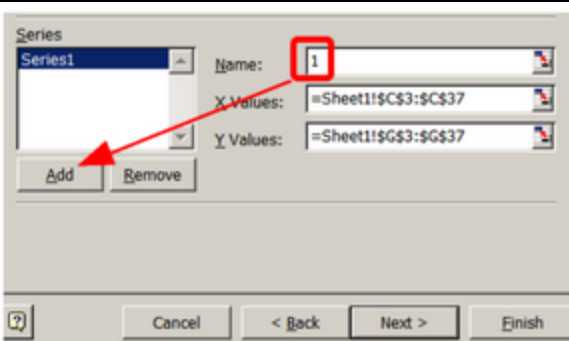
9. In Chart dialog
Click on Y Values selector icon



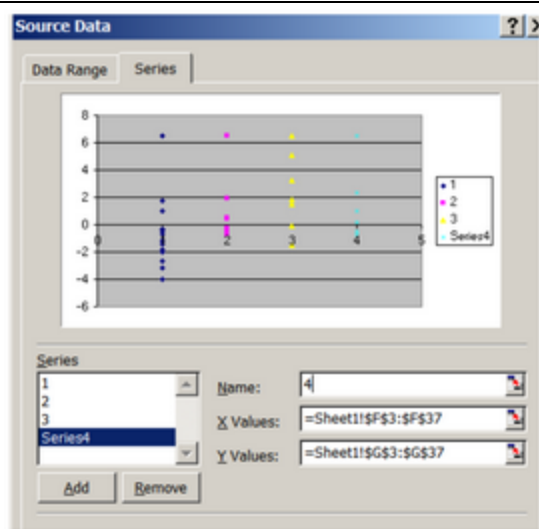
10. On Worksheet, drag mouse to highlight measure column.
Then click on series-selector icon



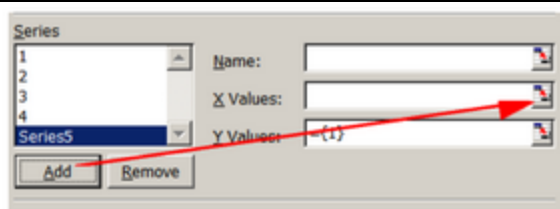
11. Type in the sub-scale number as the Name
For the next sub-scale:
Click Add



12. Repeat 7 through 11 for the other item sub-scales



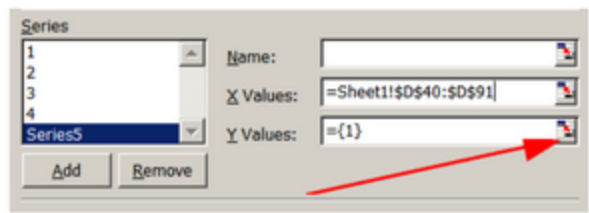
13. Now for the Person distribution
Click Add
Click on X Values icon



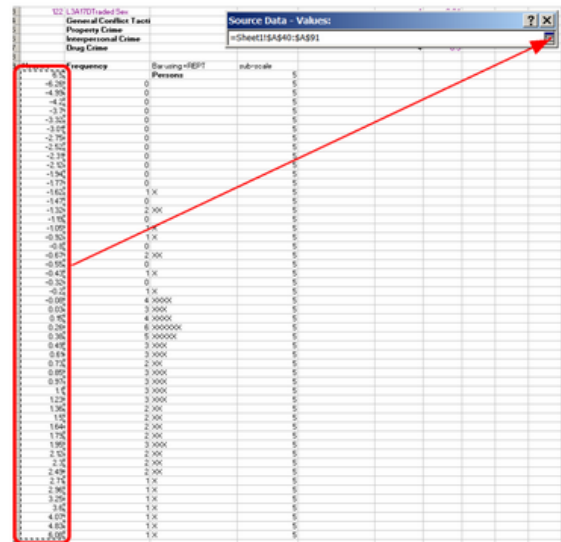
14. On worksheet, drag mouse to highlight the person sub-scale number
Click on range-selector icon



15. In Chart dialog, click on Y-values icon



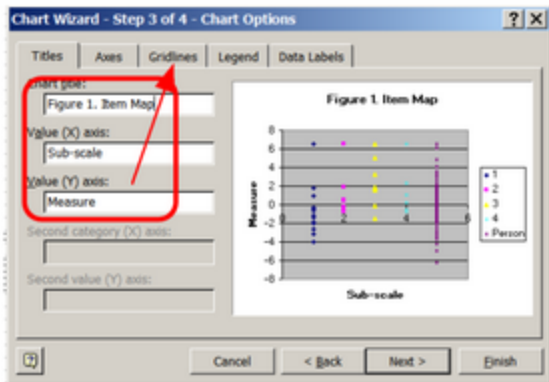
16. Drag mouse to select person measures
Click on range-selector icon



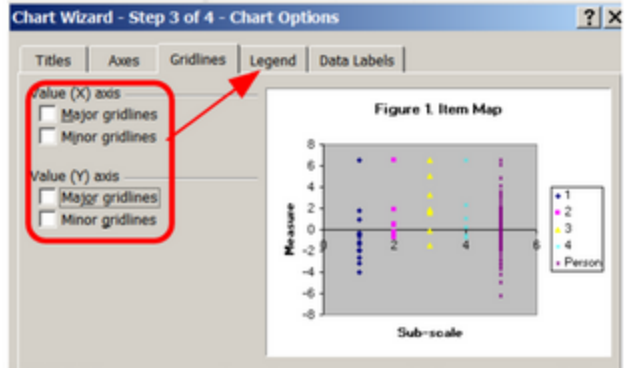
17. Type int Name: Person
Click Next



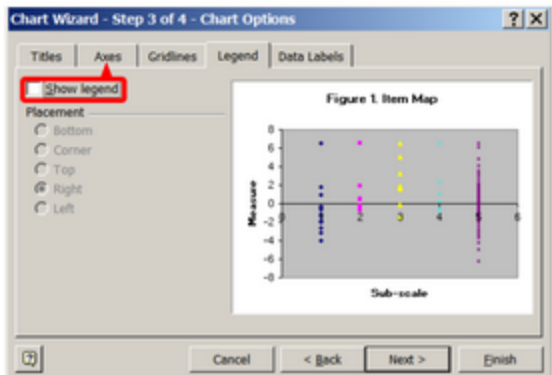
18. In Chart Wizard:
 On Titles Tab:
 Type in Chart Title: Figure 1. Item Map
 Value (X) axis: Sub-scale
 Value (Y) axis: Measure
 Click on "Grid lines" Tab



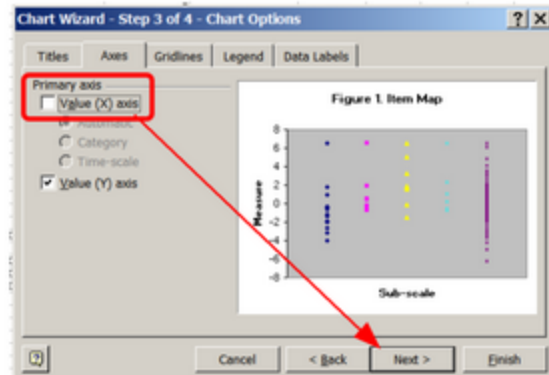
19. On Grid lines Tab:
 Uncheck all boxes
 Click on Legend Tab



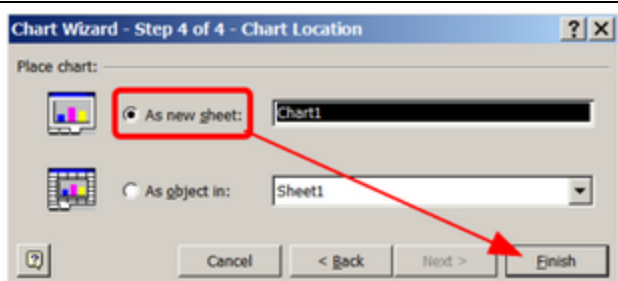
20. On Legend Tab:
 Uncheck "Show Legend" Box
 Click on "Axes" Tab



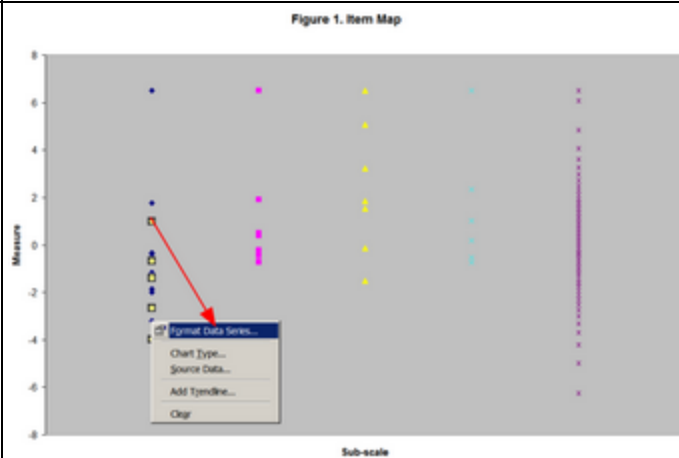
21. On Axes Tab
 Uncheck "Value (X) axis"
 Click on Next



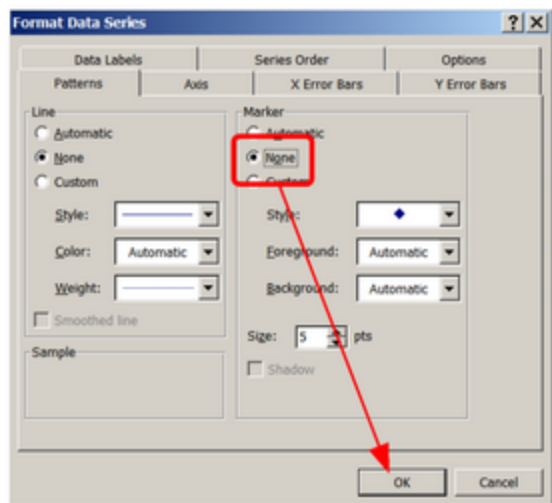
22. Click on "As new sheet"
Click on "Finish"



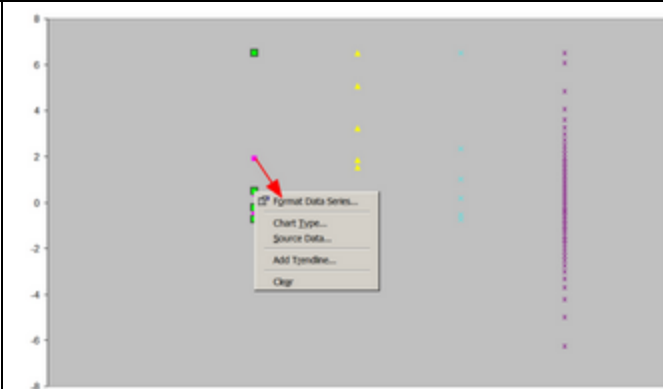
23. Right-click on first sub-scale
Click on "Format Data Series"



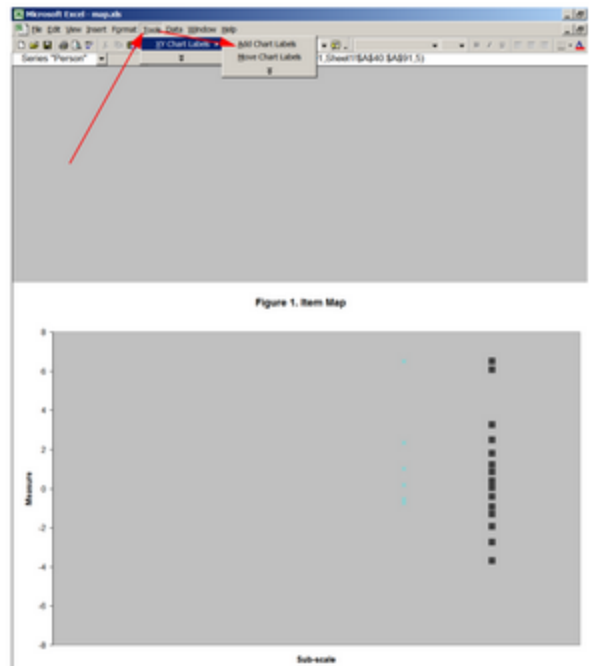
24. Click on "Marker: None"
Click on "OK"



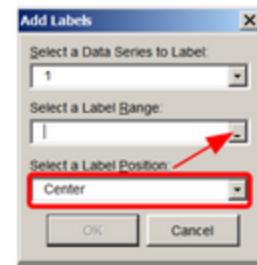
25. Same again for the other sub-scales:
Right-click on sub-scale
Click on "Format Data Series"
Click on "Marker: None"
Click on "OK"



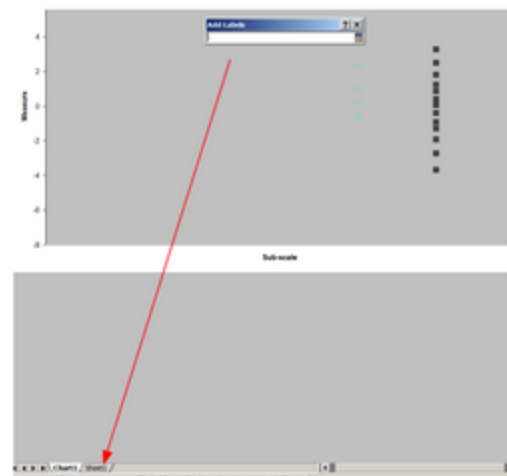
26. Click on "Tools"
Click on "XY Chart Labels"
Click on "Add Chart Labels"



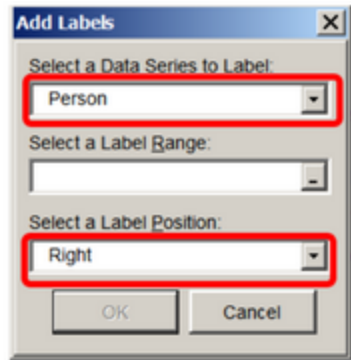
27. Select "Center" using pull-down menu
Click on "Label Range" selector



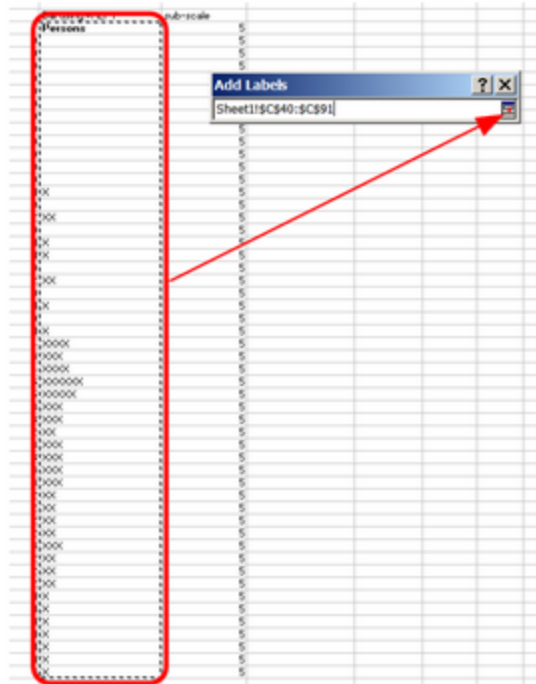
28. Click on Worksheet Tab



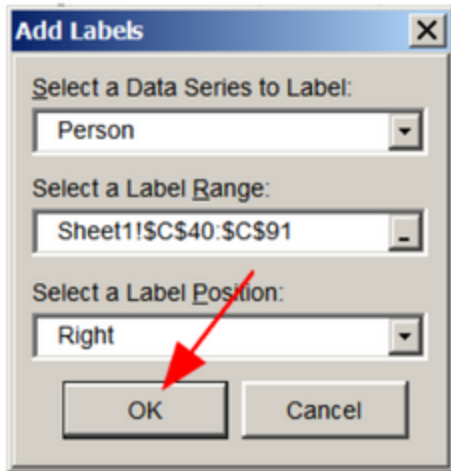
32. The once more for the Person sub-scale, but with Label Position: Right



33. On the Worksheet, Drag the mouse to select the XXXX column
Click on the range-selector



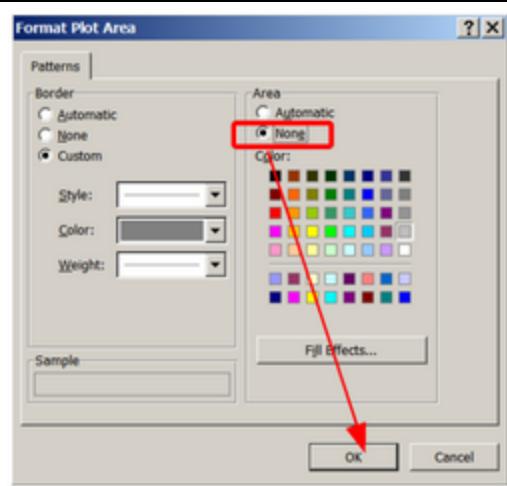
34. Click on OK



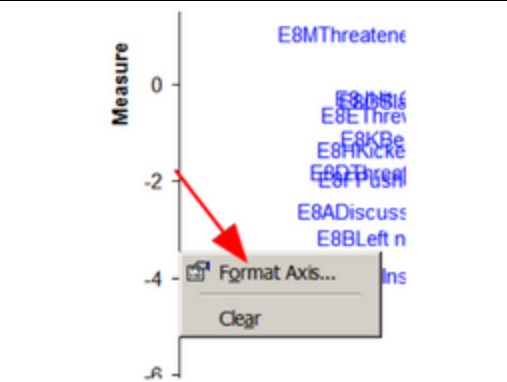
35. Right-click on plot background
Click on Format Plot Area



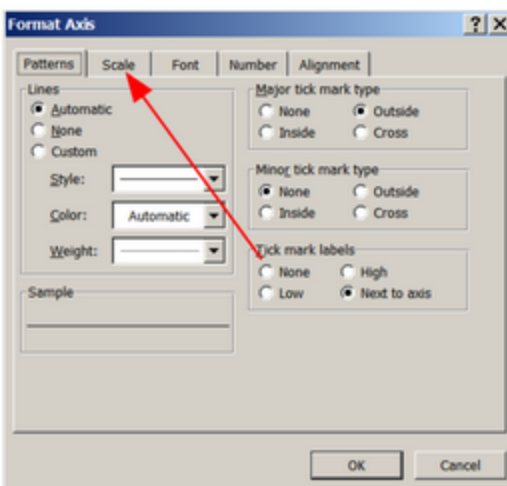
36. Click on "Area: None"
Click on OK



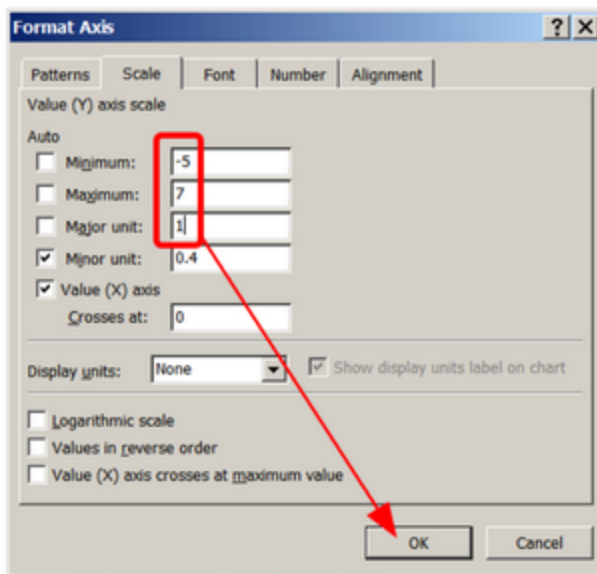
37. Right-click on the Y-axis
Click on "Format Axis"



38. Click on the Scale Tab



39. Choose suitable ranges and units,
Click OK

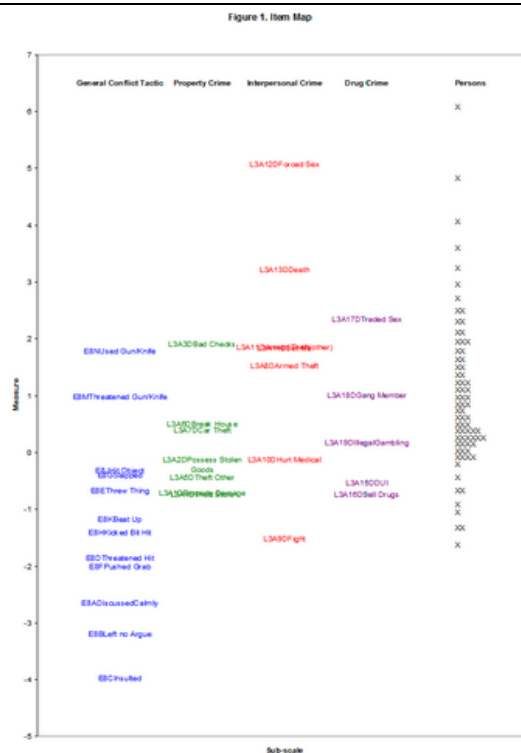


40. You can now right-click on the labels to change the font size.
You can also drag them so they don't overprint



41. To copy the plot,
click on it to select all of it
Copy
then
Paste Special - Picture into your Word document.

This is what I have done here:



Index of Specifications and Files

#1#= user-defined replaceable token.....	273
\$C= starting column in data record.....	557
\$S= starting column in label.....	557
&END end of control variables (required).....	273
&INST start of control variables (ignored).....	273
; starts a comment.....	273
@Fieldname = name for location in person or item label.....	274
~ File: Bond4.xls.....	72
~ File: Exam1.txt.....	277
~ File: Exam10a.txt.....	284
~ File: Exam11.txt.....	287
~ File: Exam12hi.txt.....	287
~ File: Exam13.txt.....	290
~ File: Exam14.txt.....	291
~ File: Exam15.txt.....	295
~ File: Exam16.txt.....	297
~ File: Exam17r.txt.....	300
~ File: Exam2.txt.....	278
~ File: Exam3.txt.....	278
~ File: Exam4.txt.....	279
~ File: Exam5.txt.....	280
~ File: Exam6.txt.....	280
~ File: Exam7.txt.....	282
~ File: Exam8.txt.....	282
~ File: Exam9.txt.....	283
~ File: Example0.txt.....	274
~ File: Examsubs.txt.....	302, 693
~ File: kct.txt.....	39
~ File: labels.txt.....	142
~ File: sf.txt.....	39
~ File: template.txt.....	70
~ File: winbatchbat.bat.....	83
AGREEFILE= paired-person agreement counts.....	78
AGREEGROUP= compare with other groups in Table 35 = Yes.....	80
ALPHANUM= alphabetic numbering.....	80
ANCESTIM= alternative method for anchored analyses.....	80
ASCII characters = Yes.....	80
ASYMPTOTE= item upper and lower asymptotes = No.....	82
BATCH= Running in batch mode = No.....	83
BOXSHOW= draw boxes around Output Tables = Yes.....	88
BYITEM= display graphs for items = Yes.....	89
CATREF= reference category for Table 2 = (item measure).....	89
CFILE= scored category label file.....	90

CHART= output graphical plots in Tables 10 & 13-15 = Yes.....	91
CHISQUARE= in IFILE= and PFILE= No.....	92
CLFILE= codes label file.....	92
CMATRIX= category matrix = Yes.....	94
CODERANGE = numerical range of data codes.....	95
CODES= valid data codes.....	96
CONVERGE= select convergence criteria = Either.....	98
CSV= comma-separated values in output files = No.....	100
CURVES= curves for Tables 2 & 21 = 111 all.....	100
CUTHI= cut off responses with high probability of success = 0 no.....	100
CUTLO= cut off responses with low probability of success = 0 no.....	101
DATA= name of data file = (data at end of control file).....	102
DATESHOW= show date in Table headings = Yes.....	103
DELIMITER= data field delimiters = " " fixed fields.....	103
DIF= columns within person label for Table 30 = \$S1W1.....	105
DISCRIMINATION= report item discrimination = No.....	107
DISFILE= category/distractor/option count file.....	108
DISOPTION= distractor file option.....	109
DISTR= output option counts in Tables 10 & 13-15 = Yes.....	109
DPF= columns within item label for Table 31 = \$S1W1.....	110
EDFILE= edit data file.....	110
EFILE= expected scores on items.....	113
END LABELS end of control variables.....	114
END NAMES end of control variables.....	114
EQFILE= code equivalences.....	114
ETYPE= option for EFILE=.....	115
EXCAT= extreme category adjustment = 0.....	115
EXCELNUMBER= Excel converts numeric labels into numbers = No.....	115
EXTRSC= extreme score correction for extreme measures = 0.3.....	116
FITHIGH= higher bar in charts = 0 none.....	117
FITI= item misfit criterion = 2.0.....	117
FITLOW= lower bar in charts = 0 none.....	117
FITP= person misfit criterion = 2.0.....	118
FORMAT= reformat data.....	118
FORMFD= the form feed character = ^ MS-DOS standard.....	124
FRANGE= half-range of fit statistics on plots = 0 auto-size.....	124
FREQ= heading in Tables 1, 12, 16 = <freq>.....	124
FSHOW= show files created from control file = No.....	124
G0ZONE= One-to-zero zone = 50 %.....	126
G1ZONE = Zero-to-one zone = 50 %.....	126
GRFILE= probability curve coordinate output file.....	124
GROUPS= assigns items to groups = " " all in one grouping.....	157
GRPFROM= location of GROUPS= = No before &END.....	126
GUFIL= Guttmanized response file.....	126
HEADER= display or suppress subtable headings = Yes.....	127
HIADJ= correction for top rating scale categories = 0.25.....	127
HLINES= heading lines in output files = Yes.....	128
IAFILE= item anchor file.....	128

IANCHQU= anchor items interactively = No.....	131
ICOLORFILE= colors for item labels in PKMAPs.....	133
ICORFILE= item residual correlation file.....	132
IDELETE= item one-line item deletion.....	134
IDELQU= delete items interactively = No.....	135
IDFILE= item deletion file.....	136
IDROPEXTREME= drop items with extreme scores = No.....	138
IFILE= item output file.....	139
ILFILE= item label file = (after &END).....	142
IMAP= item label on item maps Tables 1 & 12 = (whole label).....	143
INUMB= label items by sequence numbers = No.....	143
IOFSFIELDS= field selection for IFILE=.....	144
IPEXTREME= placement of extreme scores on maps = Yes.....	145
IPMATRIX= response-level matrix (from Output Files menu only).....	146
IPRFILE= change codes for blocks of persons and items.....	147
IREFER= identifying items for recoding (with IVALUE=).....	149
ISELECT= item selection criterion = * all items.....	151
ISFILE= item structure output file.....	153
ISGROUPS= assigns items to groups = " " all in one grouping.....	157
ISORT= column within item label for alphabetical sort in Table 15 = 1.....	161
ISRANGE= category range of an item group.....	162
ISUBTOTAL= columns within item label for subtotals in Table 27.....	162
ITEM= title for item labels = ITEM.....	163
ITEM1= column number of first response = (required).....	163
ITLEN= maximum length of item label = 30.....	164
IVALUEx= recoding of data (with IREFER=).....	165
IWEIGHT= item (variable) weighting.....	166
KEYFORM= skeleton for Excel keyform.....	506
KEYFROM= location of KEYn= = 0 before &END.....	168
KEYn= scoring key.....	169
KEYSCR= reassign scoring keys = 123.....	170
LCONV= logit change at convergence = .001 logits.....	172
LESS= heading in Tables 1, 12, 16 = <less>.....	172
LINELENGTH= length of printed lines in Tables 7 & 10-16 & 22 = 80.....	172
LOCAL= locally restandardize fit statistics = No.....	172
LOGFILE= accumulates control files.....	174
LOWADJ= correction for bottom rating scale categories = 0.25.....	174
MAKEKEY= construct MCQ key = No.....	174
MATRIX= correlation ooutput file layout = No.....	174
MAXPAGE= the maximum number of lines per page = 0 no limit.....	175
MFORMS= reformat input data and multiple data forms.....	175
MHSLICE= Mantel-Haenszel slice width = .1.....	181
MHZERO= adjustment to zero cells = 0.....	181
MISSCORE= scoring of missing data codes = -1(ignore).....	182
MJMLE= maximum number of JMLE (JCON) iterations = 0 no limit.....	184
MNSQ= show mean-square instead of t-standardized fit statistics = Yes.....	184
MODELS= assigns model types to items = R dichotomy or rating scale or partial credit.....	185
MODFROM= location of MODELS= = N before &END.....	186
MORE= heading in Tables 1, 12, 16 = <more>.....	187

MPROX= maximum number of PROX iterations = 10.....	189
MRANGE= half-range of measures on plots = 0 auto-size.....	190
MTICK= measure distance between tick marks on plots = 0, auto-size.....	190
MUCON= maximum number of JMLE (UCON) iterations = 0 no limit.....	184
NAME1= first column of person label = 1.....	189
NAMLEN= length of person label = (calculated).....	190
NAMLMP= name length on map for Tables 1 & 12 & 16 = (calculated).....	191
NEWSCORE= recoding values (with RESCORE=).....	191
NI= number of items (required).....	193
NORMAL= normal distribution for standardizing fit = No chi-square.....	193
OFFILE= report output file.....	193
OSORT= category/option/distractor sort = S score value order.....	194
OUTFIT= sort misfits on greater of infit or outfit = Yes.....	194
PAFILE= person anchor file.....	195
PAIRED= correction for paired comparison data = No.....	197
PANCHQU= anchor persons interactively = No.....	198
PCORFIL= person residual correlation file.....	198
PDELETE= person one-line item deletion.....	199
PDELQU= delete persons interactively = No.....	199
PDFILE= person deletion file.....	200
PDROPEXTREME= drop persons with extreme scores = No.....	201
PERSON= title for person labels = PERSON.....	201
PFILE= person output file.....	201
PKMAP= customize Table 36 diagnostic maps.....	204
PLFILE= person label file.....	211
PMAP= person label on person map: Tables 1 & 16 = (whole name).....	211
POFSFIELDS= field selection for PFILE=.....	212
PRCOMP= residual type for Tables 23-24 = S standardized.....	212
PSELECT= person selection criterion = * all persons.....	213
PSORT= column within person label for alphabetical sort in Table 19 = 1.....	215
PSUBTOTAL= columns within person label for subtotals in Table 28.....	215
PTBIS= point-biserial-type correlation coefficient = No.....	217
PVALUE= proportion correct or average rating = No.....	219
PWEIGHT= person (case) weighting.....	219
QCMLE = Quasi-CMLE estimates = No.....	220
QUOTED= quote marks around labels = Yes.....	220
RARE= heading in Tables 1, 12, 16 = <rare>.....	221
RCONV= score residual at convergence = 0.1.....	221
REALSE= inflate S.E. of measures for misfit = No.....	221
RESCORE= response recoding (with NEWSCORE= or KEYn=).....	222
RESFROM= location of RESCORE= No (before &END).....	223
RFILE= scored response file.....	224
RMSR= report root-mean-square residual = No.....	224
ROW1HEADING= heading in first row of output file.....	225
SAFILE= item structure anchor file.....	225
SAITEM= item numbers in SFILE= and SAFILE= (with one ISGROUPS=) = No.....	235
SANCHQ= anchor structures interactively = No.....	235
SCOREFILE= person score file.....	236

SCREEN= screen log file (command line only).....	238
SDELQU= delete structure interactively = No.....	238
SDFILE= item structure deletion file.....	238
SEPARATOR= data field delimiters.....	238
SEPARATOR= data field delimiters = " " fixed fields.....	103
SFILE= structure output file.....	238
SFUNCTION= function to model Andrich thresholds with polynomials.....	239
SICOMPLETE= simulate complete data = Yes.....	243
SIEXTREME= simulate extreme scores = Yes.....	244
SIFILE= simulated data file.....	244
SIMEASURE= measure or data = Yes.....	244
SINUMBER= number of simulated data files = 1.....	244
SIRESAMPLE= number of persons resampled.....	245
SISEED= simulated data seed = 0 (clock).....	245
SPFILE= supplementary control file.....	245
STBIAS= correct for JMLE statistical estimation bias = No.....	246
STEPT3= include structure summary in Table 3 (instead of Table 21) = Yes.....	248
STKEEP= keep non-observed intermediate categories in structure = Yes.....	248
SUBSETS= perform subset detection = Yes.....	250
SVDEPOCHS= maximum epochs for SVD estimation.....	251
SVDFACTORS= singular-value decomposition factors.....	251
SVDFILE= singular-values decomposition file.....	251
SVDMIN= singular-value decomposition minimum improvement.....	252
SVDTYPE= singular-value decomposition residual type = R.....	252
T1CAT= show category number in Table 1, 12 = (default).....	252
T1#= number of items summarized by "#" symbol in Table 1 = (auto-size).....	252
T1P#= number of persons summarized by "#" symbol in Table 1 = (auto-size).....	253
T1WEIGHT= show by weight in Table 1 = Yes.....	253
T2SELECT= selects Table 2 subtables.....	253
T45OPTIONS= options for Table 45.....	253
T7OPTIONS = selects Table 7.1 and 11.1 detail lines = OZ.....	254
TABLES= output tables.....	254
TARGET= estimate using information-weighting = No.....	255
TCCFILE= Test Characteristic Curve and Test Information Function.....	256
TFILE= list of Tables to be output.....	258
TITLE= title for output listing = (control file name).....	261
TOTALSCORE= show total scores with extreme observations = No.....	261
TRPOFILE= the transposed control and data file.....	261
TRPOTYPE= controls the format of the transposed file.....	262
UAMOVE= move all anchor values a pre-set amount.....	262
UANCHOR= anchor values supplied in user-scaled units = Yes.....	263
UASCALE= the anchor scale value of 1 logit = 1.....	263
UCOUNT = number of unexpected responses: Tables 6 & 10 = 50.....	264
UDECIMALS= number of decimal places reported = 2.....	264
UEXTREME= include extreme scores for UIMEAN= or UPMEAN= = No.....	264
UIMEAN= the mean or center of the item difficulties = 0.....	265
UMEAN= the mean or center of the item difficulties = 0.....	265
UPMEAN = the mean or center of the person abilities = (computed).....	266
USCALE= the user-scaled value of 1 logit = 1.....	267

UTF8ENC encoding = "?".....	268
UTF8SUB= substitute character for multibyte UTF-8 = ".".....	268
W300= Output files in Winsteps 3.00 format = No.....	269
WEBFONT = web-page fonts.....	269
WHEXACT= Wilson-Hilferty exact normalization = Yes.....	269
WMLE = Warm's Mean Likelihood Estimation = No.....	270
XFILE= analyzed response file.....	270
XWIDE= columns per response = 1.....	272