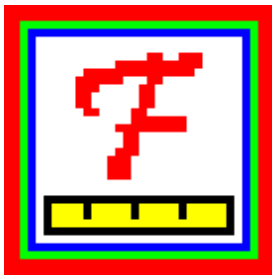


A User's Guide to

# FACETS

Rasch-Model
Computer Programs

Program Manual 3.87.0 by
John M. Linacre

www.winsteps.com

© Copyright 2023 John M. Linacre. All rights reserved.
Permission to copy Facets/Minifac output and documentation is granted.
ISBN 0-941938-03-4
Winsteps® is a registered trademark

# Table of Contents


1 Specification Index 10

2 Specification Index by Function 12

3 Output tables - files - plots - graphs 13

4 Introduction 14

4.1 Introduction ... 14
4.2 The theory behind Facets ... 16
4.3 References ... 19

5 Running Facets - Tutorial 31

5.1 Facets Tutorial PDFs ... 31
5.2 Launching Facets under Windows ... 33
5.3 Facets Window ... 33
5.4 File Box - Right-click ... 36
5.5 Extra Specifications ... 36
5.6 Report Output File ... 37
5.7 Estimation Progess ... 38
5.8 Report Output ... 39
5.9 Output Tables ... 41
5.10 Stopping Facets ... 42
5.11 Launching at the DOS prompt ... 42

6 Menus 44

6.1 Edit pull-down menu ... 44
6.2 Estimation pull-down menu ... 44
6.3 Files pull-down menu ... 46
6.4 Font pull-down menu ... 46
6.5 Graphs pull-down menu ... 47
6.6 Help pull-down menu ... 50
6.7 Output Files pull-down menu ... 51
6.8 Output Tables and Plots pull-down menu ... 51
6.9 Edit Initial Settings ... 52
6.10 Quick-action buttons ... 54
6.11 Short-cut keys ... 54

7 Examples of Specifications and Data 56

7.1 3mile.txt: 1 facet fixed effects analysis: Stress at Three Mile Island ... 56
7.2 Baseball.txt: 1 facet paired comparison (BTL): League Baseball ... 57

7.3 Kct.txt: 2 facet dichotomy: The Knox Cube Test ... 58
7.4 Sportcas.txt: 2 facet partial credit/rank order: Sportscasting ... 60
7.5 Lfs.txt: 2 facet rating scale: Liking for Science ... 62
7.6 Kcta.txt: 3 facet dichotomy with item bias: The Knox Cube Test ... 63
7.7 Woodcut.txt: 3 facet Poisson count: Woodcutting ... 64
7.8 Guilford.txt: 3 facet rating scale: Creativity (with Excel input data file) ... 65
7.9 Subsets.txt: 3 facet Subset reconnection ... 67
7.10 Measure.txt: 4 facet Arithmetic Test: Measuring, Anchoring and Describing ... 69
7.11 Dives.txt: 4 facet rating scale with missing data: Diving ... 71
7.12 Divesb.txt: 4 facet Interactions and Dummy Facets: Diving + Creativity ... 72
7.13 Kctinter.txt: 4 facet item-type interactions: Knox Cube Test ... 75
7.14 Pair.txt: 4 facet judge pairs: Language test ... 77
7.15 Essays.txt: 4 facet rating scale with bias analysis: Essays ... 79
7.16 Olympics.txt: 4 facet rating scale with unobserved categories: ice-skating ... 79
7.17 Essayday.txt: 5 facet rating scale with bias analysis: Essays+Days ... 80
7.18 Gel.txt: 7 facet Paired comparison with ties: Flavor Strength of Gels ... 81
7.19 Chinese.txt: example of UTF-8 multibyte character codes ... 83

8 Building Specification and Data files 84

8.1 Facets requires specifications and data ... 84
8.2 Template for specifications and data ... 86
8.3 Excel data file ... 87
8.4 R data file ... 91
8.5 SAS data file ... 92
8.6 SPSS data file ... 95
8.7 STATA data file (version 13 and earlier) ... 96
8.8 Text data file ... 98
8.9 Facform data formatter ... 99
8.10 Waiting for imported datafile ... 100

9 Specifications * 102

9.1 %include file (included specification and data files) ... 102
9.2 Anchor output file = ... 103
9.3 Arrangement by Number/Alpha/Measure/Fit/PBS (Asc./Desc.) = N ... 106
9.4 ASCII output table display format = Yes ... 107
9.5 Barcharts output = Yes ... 108
9.6 Batch file processing = Yes ... 109
9.7 Beep sound = No ... 111
9.8 Bias (direction: difficulty/ability/omit) = Ability ... 111
9.9 Box show around Tables = Yes ... 111
9.10 Center facet = ... 112
9.11 Convergence (criteria to end iteration) = .5, .01, 0, 0 ... 113

9.12 CSV output format = " " ..... 114
9.13 Data for analysis = ..... 114
9.14 Delements element type in data = ..... 118
9.15 Dvalues for data facets = ..... 121
9.16 Elements in Labels= ..... 126
9.17 Entry order of facets in data = 1,2,3,4,5, ..... 132
9.18 Facets in data = 2 ..... 133
9.19 Fair averages based on = Mean ..... 133
9.20 Glabel group labels = ..... 134
9.21 Graph plotting file name = " " ..... 136
9.22 Hardware = 80 ..... 137
9.23 Heading lines = Yes ..... 137
9.24 Inter-rater Agreement Coefficients = 0 ..... 138
9.25 Iterations (stop after) = 0 (unlimited) ..... 138
9.26 Juxtapose (column headings after every this many lines) = 0 ..... 138
9.27 Labels of facets and elements = (required) ..... 138
9.28 Left-hand placement of row labels = No ..... 140
9.29 Missing data codes = ..... 141
9.30 Model to be used in the analysis = ?,?,D,1 ..... 142
9.31 Negatively-oriented facet = ..... 148
9.32 Newton-Raphson step size = 0 ..... 148
9.33 Noncenter facet = 1 ..... 149
9.34 Null element = 0 ..... 149
9.35 Omit unobserved elements = Yes ..... 150
9.36 Output results/anchors/residuals = ..... 151
9.37 Positively-oriented facet = 1 ..... 152
9.38 Pt-biserial correlation = Measure ..... 154
9.39 QM quotation marks around labels = ..... 157
9.40 Query to monitor matching of data to models = No ..... 158
9.41 Rating scale (or Response model) = ..... 159
9.42 Replication character = R ..... 167
9.43 Residuals / Responses Output File = "" ..... 168
9.44 Reversed scoring = Yes ..... 170
9.45 Score/Measure Output Files = "" ..... 171
9.46 SE (standard error is Model/Real) = Model ..... 174
9.47 Show unobserved elements = No ..... 174
9.48 Simulated data file = "" ..... 174
9.49 Subset detection = Yes ..... 176
9.50 T3onscreen show only one line on screen iteration report = Y ..... 178
9.51 T4maximum number of residuals in Table 4 = 100 ..... 178
9.52 T7LGS Table 7 List-Group details-Group Summary ..... 179

9.53 T8NBC Table 8 Numbers-Barcharts-Curves = NBC ... 179
9.54 Tables output from main analysis = Yes ... 179
9.55 Title for each results table = ... 179
9.56 Total scores reported = Yes ... 180
9.57 Udecimals (user-chosen decimal places) = 2 ... 180
9.58 Umean (user mean and scale and decimals) = 0, 1, 2 ... 180
9.59 Unexpected (standardized residuals reported at least) = 3 ... 181
9.60 Uscale (user rescaling of measures) = 1 ... 182
9.61 Usort sort order of unexpected responses = u ... 183
9.62 UTF8enc encoding = ? ... 184
9.63 UTF8sub substitute character = . ... 184
9.64 Vertical rulers in Table 6 = 1A,2A,3A,4A, ... 184
9.65 Whexact Wilson-Hilferty standardization = Yes ... 187
9.66 Write iteration details in output file = No ... 188
9.67 Xtreme (extreme score adjustment) = 0.3, 0.5 ... 188
9.68 Yardstick (Table 6 columns lines low high extreme) = ... 189
9.69 Zscore minimum for reported bias terms (bias size/t-statistic) = 0, 0 ... 190

10 The Screen Output 191

10.1 Processing specifications ... 191
10.2 Reading data ... 191
10.3 Calculating estimates ... 192
10.4 Fit calculation and writing out results ... 194
10.5 Performing Bias/Interaction analysis ... 195
10.6 Completing analysis ... 196

11 Tables Output from Main Analysis 197

11.1 Table 1 is a summary of your specifications ... 197
11.2 Table 2 is the data summary report ... 197
11.3 Table 3 is the iteration report for the main analysis ... 199
11.4 Table 4 reports Unexpected Responses ... 200
11.5 Table 5 is the measurable data summary ... 201
11.6 Table 6.0 is the All facet summary - Wright Map "rulers" ... 205
11.7 Table 6.0.0 Subsets - disjoint element listing ... 208
11.8 Table 6.2 Graphical description of facet statistics ... 209
11.9 Table 7 Facet elements measurement report ... 210
11.10 Table 7 Reliability and Chi-square statistics ... 214
11.11 Table 7 Agreement Statistics ... 216
11.12 Table 7 Group Summary ... 218
11.13 Table 8.1 Dichotomy, binomial trial and Poisson statistics ... 218
11.14 Table 8.1 Rating scale statistics ... 222
11.15 Table 8 Scale structure bar-chart ... 226

11.16 Table 8 Scale structure probability curves ... 228
11.17 Table 9 is the iteration report for Bias estimation ... 230
11.18 Table 10 lists responses, still unexpected after allowing for bias ... 231
11.19 Table 11 shows how many responses were used to calculate bias estimates ... 231
11.20 Table 12 is the Bias Summary Report ... 232
11.21 Table 13 is the bias report ... 233
11.22 Table 14 is the pairwise bias report ... 235

12 Output Tables from menu 238

12.1 Table 4 from menu ... 238
12.2 Table 5 from menu ... 239
12.3 Table 6 from menu ... 239
12.4 Table 7 from menu ... 241
12.5 Table 8 from menu ... 241
12.6 Modify output specifications ... 242
12.7 Table 13-14 Excel bias plot ... 243

13 Output Files from menu 245

13.1 Specification settings file ... 245
13.2 Anchor output file ... 246
13.3 Graph output file ... 247
13.4 Output Dialog Box ... 249
13.5 Residuals / Responses Output File ... 249
13.6 Score and measure file output ... 252
13.7 Web output table ... 256
13.8 Simulated data file ... 257
13.9 Subset group-anchor file ... 258
13.10 Winsteps control, data and anchor file ... 258

14 Graphs 261

14.1 Graph window ... 261
14.2 Category information function ... 267
14.3 Category probability curves ... 267
14.4 Conditional probability curves ... 269
14.5 Cumulative probabilities ... 270
14.6 Empirical category curves - frequency polygons ... 271
14.7 Empirical ICC or IRF ... 272
14.8 Expected score ICC or IRF ... 273
14.9 Expected+Empirical ICC or IRF ... 273
14.10 Item information function ... 275
14.11 Points and lines ... 275
14.12 Prob.+Empirical Cat. Curves ... 276

14.13 Test information function ... 276

15 Program Messages 279

15.1 Check! Warning! and other messages and their explanation ... 279
15.2 Disk file problem ... 281
15.3 Disk space exhausted ... 281
15.4 Memory problem ... 281
15.5 No valid matching responses found! ... 281
15.6 Overflow problem ... 282
15.7 Specification or data problem ... 282
15.8 Unexplained program failure ... 283

16 Special Topics 283

16.1 Analysis - step-by-step ... 283
16.2 Anchoring rating scale structures ... 284
16.3 ANOVA-like Analysis ... 286
16.4 BASIC program to build Facets data and specification files ... 286
16.5 Bayesian Logistic Regression ... 287
16.6 Bias interaction DIF DPF DRF estimation ... 289
16.7 Convergence problems ... 291
16.8 Cut-off points ... 292
16.9 Decimal and continuous data ... 293
16.10 Decision Consistency ... 294
16.11 Deleting or Ignoring data values, observations or elements ... 295
16.12 Diagnosing Misfit ... 297
16.13 Dichotomous Misfit Statistics ... 299
16.14 Differential performance on items ... 299
16.15 Disordered Rasch-Andrich thresholds ... 300
16.16 Double ratings ... 301
16.17 Dummy facets and group-anchoring for interactions ... 302
16.18 Effect size ... 303
16.19 Equating tests ... 304
16.20 Estimates do not converge ... 305
16.21 Estimation considerations: JMLE estimation bias ... 306
16.22 Excel files do not display ... 307
16.23 Expected score ... 308
16.24 Export Facets Tables and other files to Excel, SPSS or Word ... 308
16.25 External measures as anchor values or ratings ... 309
16.26 Extreme scores ... 309
16.27 FACDOS ... 310
16.28 Fair average ... 310
16.29 Fractional dichotomous and polytomous models ... 314

16.30 Frame of reference ... 314
16.31 Generalizability Theory ... 314
16.32 Glossary ... 315
16.33 Group comparisons ... 323
16.34 Guessing elimination ... 324
16.35 Hierarchical Rater Models HRM ... 324
16.36 Inestimable elements or excessive logit ranges ... 324
16.37 Instability of Long Links ... 326
16.38 Inter-rater and intra-rater Reliability ... 326
16.39 Inter-rater Correlations ... 332
16.40 Investigating differences between rater groups ... 334
16.41 Iterations do not stop ... 335
16.42 Judging Plans and Experimental Designs ... 337
16.43 LLTM - Linear Logistic Test Model ... 341
16.44 Maximum size limitations on specifications and data ... 344
16.45 Measures and calibrations ... 345
16.46 Measuring the elements ... 345
16.47 Missing and sparse data considerations: Sample size and Category counts ... 346
16.48 Mixed rating scale forms and MCQ scoring keys ... 346
16.49 Model - matching data with measurement models ... 348
16.50 Model statement examples ... 350
16.51 Multiple choice test ... 352
16.52 Nested facet designs: remedying ambiguities ... 353
16.53 Omitting elements without editing your data ... 355
16.54 One observation for an element ... 355
16.55 Paired comparisons of objects ... 356
16.56 Pairwise Maximum Likelihood Estimation PMLE with Facets ... 358
16.57 Polytomous Misfit Statistics ... 358
16.58 Prettifying (beautifying) Facets output ... 360
16.59 R Statistics ... 360
16.60 Rasch models: implement in Facets ... 360
16.61 Rater misbehavior ... 362
16.62 Ratio scale of logit measures ... 364
16.63 Recoding non-numeric and other data values ... 365
16.64 Rectangular copying ... 365
16.65 Reliability - separation - strata ... 366
16.66 Scoring table construction ... 367
16.67 Selecting elements ... 369
16.68 Silent installation of Facets from the DOS prompt ... 370
16.69 Specification format for Facets ... 371
16.70 Speeding up slow estimation ... 371

16.71 Subdimension person measurement ..... 372
16.72 Subscale measures ..... 372
16.73 Subset connectedness ..... 374
16.74 Subset connection demonstration ..... 379
16.75 Teaching Rasch measurement, Rasch analysis ..... 385
16.76 Text Editor ..... 386
16.77 t-statistics ..... 387
16.78 Unmeasurable ..... 388
16.79 Unobserved or null categories ..... 389
16.80 Weighting the data ..... 390
16.81 Zeroes: Structural and Incidental: Ordinal or Keep ..... 393
Index of Specifications and Files 396

# 1 Specification Index


| |
|---|
| % include files |
| Anchor output file = |
| Arrangement by Number/Alpha/Measure/Fit/PBS (Asc./Desc.) = mN, fN, N |
| ASCII output table display format = |
| Barcharts output = Yes |
| Batch processing = No |
| Beep sound = No |
| Bias (direction: difficulty/ability/omit) = Difficulty |
| Boxshow (show boxes around tables) = Yes |
| Center facet = |
| Convergence (criteria to end iteration) = .5, .01 |
| CSV output format = " " |
| Data file name = facets.dat |
| Delements element type = |
| Dvalues for data facet = |
| Entry order of facets in data = 1,2,3,4,5,.. |
| Facets in data = 2 |
| Fair score based on = Mean |
| Glabel group label= |
| Graph plotting file name = " "," " |
| Hardware = 80 |
| Heading lines = Y |
| Inter-rater agreement coefficients = 0 |
| Iterations (stop after) = 0 |
| Juxtapose (column headings after every this many lines) = 0 |
| Keep as null element = 0 |
| Labels of facets and elements = |
| Left-hand placement of row labels = No |
| Missing data codes = |
| Model to be used in the analysis = ?,?,D |
| Negatively-oriented facet = |
| Newton-Raphson = 0 |

| |
|---|
| Noncenter facet = 1 |
| Null element number = 0 |
| Omit unobserved elements = Yes |
| Output results/anchors/residuals = |
| Positively-oriented facet = 1 |
| Pt-biserial correlation =Yes |
| QM quotation marks around labels = |
| Query to monitor matching of data to models = No |
| Rating scale (or Response model) = |
| Replication character = R |
| Residuals output file= |
| Score output files name and CSV status = " "," " |
| SE (standard error is Model/Real) = Model |
| Show unobserved elements = No |
| Simulated data file= |
| Subset detection = Yes |
| T3onscreen show only one line on screen iteration report = Y |
| T4maximum number of residuals in Table 4 = 100 |
| T8NBC Table 8 Numbers-Barcharts-Curves = NBC |
| Tables are displayed after main analysis = Yes |
| Title for each results table = |
| Totalscore (includes extreme responses) = No |
| Udecimals (decimal places) = 2 |
| Umean (user mean, scale, decimal places) = 0, 1, 2 |
| Unexpected (standardized residuals reported, if not less than) = 3 |
| Uscale (measure scaling) = 1 |
| Usort sort order of unexpected responses = u |
| UTF8enc encoding = ? |
| UTF8sub substitute character = . |
| Vertical rulers in results file = 1A,2A,3A,4A,.. |
| Whexact exact fit standardization = Yes |
| Write iteration details in results file = No |
| Xtreme (score adjustment) = 0.3, 0.5 |

| |
|---|
| Yardstick (columns,lines,low, high) = 80,,, |
| Zscore minimum for reported bias terms (bias/zscore) = 0, 0 |

# 2 Specification Index by Function

**Facet conceptualization:**
Facets in data = 2
Center facet =
Inter-rater agreement coefficients = 0
Model to be used in the analysis = ?,?,D
Negatively-oriented facet =
Noncenter facet = 1
Positively-oriented facet = 1
Rating scale (or Response model) =

**Data layout:**
Data file name = facets.dat
Delements element type =
Dvalues for data facet =
Entry order of facets in data = 1,2,3,4,5,..
Glabel group label=
Labels of facets and elements =
Missing data codes =
Null element = 0
Replication character = R
% include files

**Estimation control:**
Batch processing = No
Beep sound = No
Convergence (criteria to end iteration) = .5, .01
Fair score based on = Mean
Hardware = 80
Iterations (stop after) = 0
Newton-Raphson = 0
Query to monitor matching of data to models = No
SE (standard error is Model/Real) = Model
Subset detection = Y
T3onscreen show only one line on screen iteration report = Y
Whexact exact fit standardization = Yes
Xtreme (score adjustment) = 0.3, 0.5

**Output options:**
Arrangement by Number/Alpha/Measure/Fit/PBS (Asc./Desc.) = mN, fN, N
ASCII output table display format =
Barcharts output = Yes
Bias (direction: difficulty/ability/omit) = Difficulty
Boxshow (show boxes around tables) = Yes
CSV output format = " "
Juxtapose (column headings after every this many lines) = 0
Left-hand placement of row labels = No
Pt-biserial correlation =Yes
QM quotation marks around labels =
Show unobserved elements = No
T4maximum number of residuals in Table 4 = 100

T8NBC Table 8 Numbers-Barcharts-Curves = NBC
Tables are displayed after main analysis = Yes
Title for each results table =
Totalscore (includes extreme responses) = No
Umean (user mean, scale, decimal places) = 0, 1, 2
Unexpected (standardized residuals reported, if not less than) = 3
Usort sort order of unexpected responses = u
Vertical rulers in results file = 1A,2A,3A,4A,..
Write iteration details in results file = No
Yardstick (columns,lines,low, high) = 80,,,
Zscore minimum for reported bias terms (bias/zscore) = 0, 0

**Output files:**
Anchor output file =
Graph plotting file name = " "," "
Heading lines = Yes
Output results/anchors/residuals =
Residuals output file=
Score output files name and CSV status = " "," "
Simulated data file=
Totalscore (includes extreme responses) = No

# 3 Output tables - files - plots - graphs

| Output Tables in Output Report or from the Output Tables pull-down menu<br>*If the Tables do not display correctly, see Text Editor* | |
|---|---|
| Table 1 | a summary of your specifications |
| Table 2 | the data summary report |
| Table 3 | the iteration report for the main analysis |
| Table 4 | reports Unexpected Responses |
| Table 5 | the measurable data summary |
| Table 6.0 | the All facet summary Wright map - "rulers" |
| Table 6.0.0 | the Disjoint element listing |
| Table 6.2 | Graphical description of facet statistics |
| Table 7 measurement | the facet measurement report |
| Table 7 reliability | reliability and chi-squared statistics |
| Table 7 agreement | agreement statistics |
| Table 8.1 dichotomous | Dichotomy, binomial trial and Poisson statistics |
| Table 8.1 polytomous | Rating scale (or partial credit) statistics |
| Table 8 barchart | Scale structure bar-chart |
| Table 8 curves | Scale structure probability curves |
| Table 9 | the iteration report for Bias estimation |
| Table 10 | lists responses, still unexpected after allowing for bias |

| | |
|---|---|
| Table 11 | shows how many responses were used to calculate bias estimates |
| Table 12 | the DIF/bias summary report |
| Table 13 | the DIF/bias report |
| Table 14 | the pairwise bias report |
| DIF/bias plot | Excel DIF/bias plot (only from Output Tables pull-down menu) |

| Output Files in specification file or from the Output Files pull-down menu: |
|---|
| Specification settings file= (only from Output Files pull-down menu) |
| Anchor output file= |
| Graph plotting file= |
| Output report file= |
| Residuals output file= |
| Score output file= |
| Simulated data file= |
| Subset group-anchor file= (only from Output Files pull-down menu) |
| Winsteps control & data file: output the Facets data in a format compatible with Winsteps® |

| Graphical output can be generated with the Graphs pull-down menu. |
|---|
| Category Probability Curves |
| Expected Score ICC/IRF |
| Cumulative Probability Curves |
| Test Information Function |
| Category Information Function |
| Conditional Probability Curves |

# 4 Introduction

## 4.1 Introduction

*Facets* is Windows-based software which assists in many applications of the Rasch model, particularly in the areas of performance assessment and paired comparisons. We recommend that you first become familiar with the conceptually and operationally simpler Rasch-measurement program WINSTEPS before embarking on Facets. We also offer Training Courses - please contact us for details. There is more information at: www.winsteps.com

The development of Facets is described in "Many-Facet Rasch Measurement" (2nd Ed., Linacre, 1994), available by free download from www.winsteps.com/manuals.htm

The algorithm implemented in Facets are fully explained in the book. Basically, Facets data are a set of non-linear simultaneous equations:

person + judge + item = f (observed rating)

f () indicates an inverse Rasch polytomous logistic function.

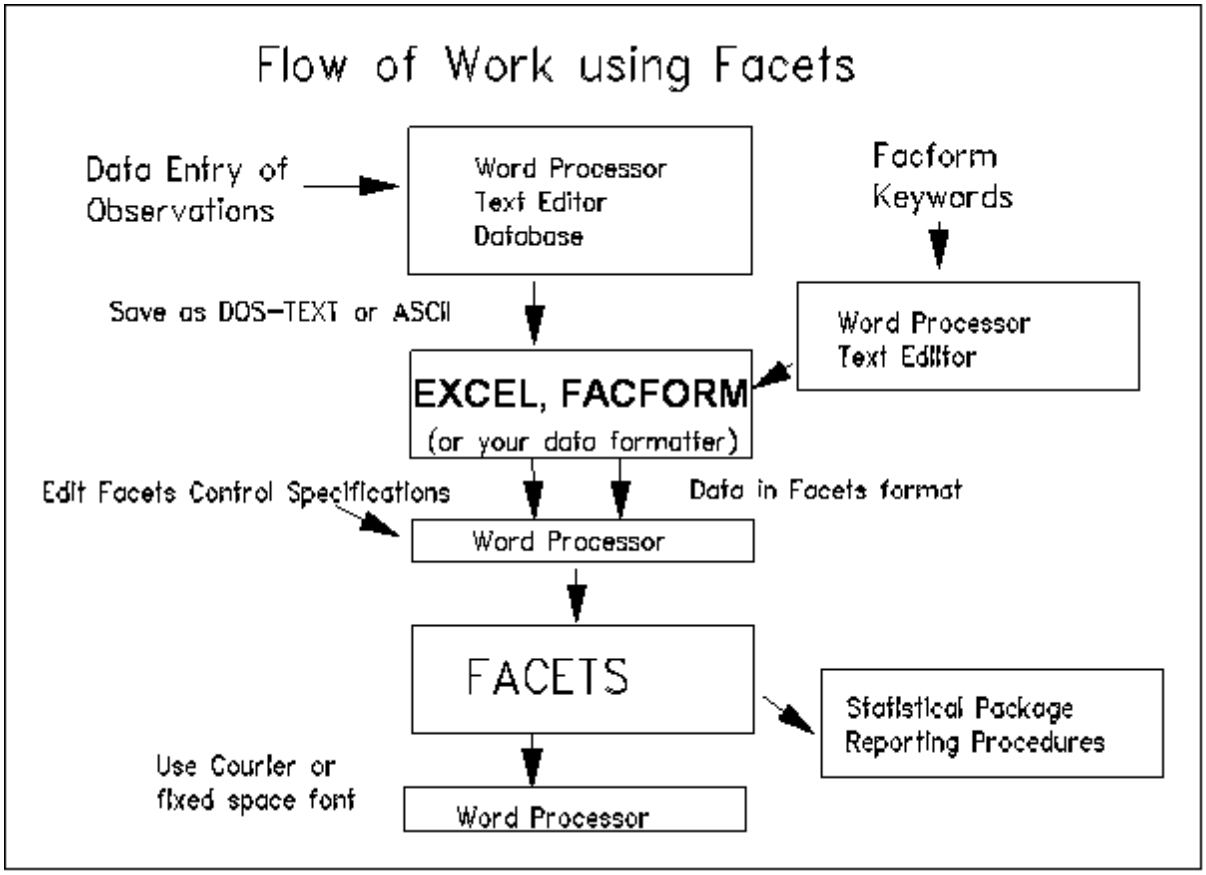
There is one equation for eachobservation, so there are more equations than unknowns (persons+judges+items). Facets uses the method of "Maximum Likelihood" to find the best set of person, judge and item measures for these data. The algorithm for this is documented in the book. It is heavy going!

Then we can discover the "Fair M Average" rating for each person, which transform the Rasch person measure back to the rating metric:

person + average judge + average item -> f ( "Fair M Average" rating )

Software validation: Facets has been operational for over 30 years and has been verified many times by users with their own data and with standard datasets. Since Facets can also analyze 2-facet data (the usual rectangular person-item datasets), its results can be compared with other Rasch software. For instance, Winsteps and Facets produce the same numbers for the same data when configured the same.

Facets is also self-checking. For the "best" set of Rasch estimates, the expected total raw score for each applicant, judge or item must equal the observed total raw score. When this does not happen, Facets reports a "displacement", indicating roughly how far away the reported estimate is from the best estimate. We can check this happens by stopping the iterative estimation process early (ctrl+f in the Facets analysis window) or by deliberately anchoring unknowns at their "wrong" values. We can also look at the Facets "Residual file" which shows the details of the estimation for each observation, i.e., the details of each of the equations as shown above.




This manual contains instructions for operating Facets, a computer program for the construction of linear measures from qualitatively-ordered counts by means of many-facet Rasch analysis. The theory underlying the program is described in reference MFRM. These instructions also apply to MINIFAC, the student/evaluation version of Facets, which has exactly the same functionality, but is limited in the number of observations it can analyze.

A basic many-facet Rasch model for observation Xnmij is:

```
log ( Pnmijk / Pnmij(k-1) ) = Bn - Am - Di - Cj - Fk
```

where

Bn is the ability of person n, e.g., examinee: Mary,
Am is the challenge of task m, e.g., an essay: "My day at the zoo",
Di is the difficulty of item i, e.g., punctuation,
Cj is the severity of judge j, e.g., the grader: Dr. Smith,
Fk is the barrier to being observed in category k relative to category k-1.
Pnmijk is the probability of category *k* being observed.
Pnmij(k-1) is the probability of category *k-1* being observed.

Persons, tasks, items and judges are facets. The elements include Mary, "My day at the zoo", punctuation, and Dr. Smith. For each element, *Facets* provides a measure (linear quantity), its standard error (precision) and five fit statistics (statistical validity). The fit statistics enable diagnosis of aberrant observations and idiosyncratic elements. *Facets* also provides calibrations of response format structures, such as rating scales, partial credit items, letter grades and ranks. Results are presented in tables and graphically. The graphical presentation of measures is in a "ruler" format especially useful to non-technical users.

Facets can also quantify discrepant interactions between elements of different facets. Once measures have been estimated from a data set, differential facet functioning, equivalent to differential item functioning or "item bias", can be investigated automatically. A judge's bias on one item, or an item's bias against a group of persons can be identified and its size and statistical significance estimated.

Facets is ideally suited for essay grading, portfolio assessment and other kinds of judged performances. Its use is not limited to educational and psychological testing. It is employed to convert qualitative observations to linear measures in many areas of research and practice, including product development, sports science, pollution control, public speaking and the arts.

See References.

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

## 4.2 The theory behind Facets

The computer program *"Facets"* implements the "many-facet Rasch measurement model" (Linacre, 1989). Each ordinal observation is conceptualized to be the outcome of an interaction between elements, e.g., a student, an item and a rater. These interacting elements are modeled to be operating independently and their measures to combine additively on the latent variable. For instance each rater is modeled to exhibit a specific amount of leniency or severity, and to act as an independent expert, not as a "scoring machine". The relationship between the ordinal observations and the linear measures of the elements is non-linear.

Danish Mathematician Georg Rasch (1960) constructed the necessary and sufficient mathematical model for the transformation of ordinal observations into linear measures. This model has the form of a logistic regression model, but each person and item is individually parameterized. In fact, it looks like a regression in which each item and each person is parameterized as a coefficient applied to a dummy variable. The dummy variable is "1" if the person or item participates in the observation, "0" otherwise. In principle this could be estimated with standard statistical software, but such software rarely allows for the estimation of the hundreds or thousands of parameters that can be encountered in just one Rasch analysis.

There are currently (2020) about 1,000 serious uses of the *Facets* software package. Many of these are in the medical field. This is probably because raters in that field behave like independent experts, the judging designs are irregular and pass-fail decisions for individuals (either for credentialing or patient treatment) are crucial. In contrast, in most educational testing situations, e.g., essay grading, raters are intended to behave like scoring machines, and the judging designs are regimented. Individual educational decisions are not of interest to educational administrators

(unless theirs are the relevant children!) Thus, provided the random behavior is small (verified using G-theory), administrators are not interested in Facets-style corrections to student ability estimates.

The standard Rasch model for dichotomous data with persons and items is:

$$\log ( P_{ni}/(1-P_{ni})) = B_n - D_i$$

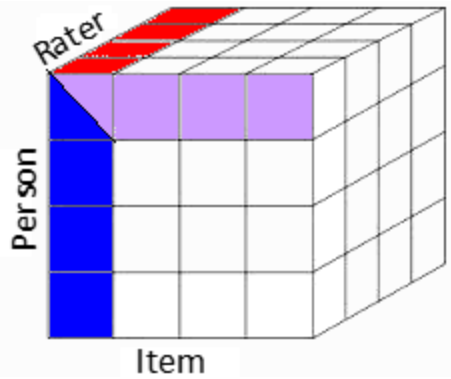
where Pni is the probability that person n will succeed on item i, where person n has ability Bn and item i has difficulty Di. It can be seen that the model is additive in the parameters (Bn) and (-Di). Thus it meets the first requirement for interval measurement. From the estimation standpoint, the maximum-likelihood of the parameter estimate for each parameter occurs when the expected raw score corresponding to the parameter estimate equals the observed raw score. This is Fisher's principle of statistical sufficiency. The model has other nice properties, such as conjoint ordering, stochastic Guttman transitivity, concatenation, and infinite divisibility. This model has been applied productively to educational tests for over 40 years.

Statisticians can find it difficult to adjust to Rasch methodology. They tend to believe that the data points tell the truth and that it is the task of statisticians to find models which explain them and to find the latent variables which underlie them. Rasch methodology takes an opposite position. It says that the latent variable is the truth, and when that latent variable is expressed in linear terms, it is the Rasch model that is necessary and sufficient to describe it. Consequently those data points which do not accord with the Rasch model are giving a distorted picture of the latent variable. They may be telling us very important things, e.g., "the students were disinterested", "the scoring key was wrong" - but those do not pertain to the central variable .

The Rasch model has been extended to rating scale and partial credit observations, while maintaining the same mathematical properties. This "rating scale" model has been used successfully for 40 years in the analysis of attitude surveys and other rated assessments. An extended version of this model (Andrich, 1978, Masters 1982) is the grouped-item rating scale model:

$$\log ( P_{nik}/P_{ni(k-1)}) = B_n - D_{gi} - F_{gk}$$

where Pnik is the probability of observing category k for person n encountering item i.
Pni(k-1) is the probability of observing category k-1
Fgk is the difficulty of being observed in category k relative to category k-1, for an item in group g.




**rows = person**
**columns = items**
**3rd dimension (slices) = raters**

Among many other extensions to the Rasch model is the Many-Facets Rasch Model. This extends the polytomous form of the model:

$$\log ( P_{nijk}/P_{nij(k-1)}) = B_n - D_{gi} - C_j - F_{gk}$$

Again, the mathematical properties of the model are maintained, but one (or more) extra components of the measurement situation are introduced. In this example, Cj, represents the severity (or leniency) of rater (judge) j, who awards the ratings {k} to person n on item i. As in the dichotomous model, the raw scores are the sufficient statistics for the Bn, Dgi and Cj. The counts of observations in each category are the sufficient statistics for

estimating the {Fk}. The model also supports powerful quality-control fit statistics for assessing the conformance of the data to the model. The model is robust against many forms of misfit, so that the typical perturbations in data tend to have little influence on the measure estimates. A further feature of the model is its robustness against missing data. Since the model is parameterized at the individual observation level, estimates are obtained only from the data that has been observed. There is no requirement to impute missing data, or to assume the overall form of the distribution of parameters.

In estimating the measures, the model acts as though the randomness in the data is well-behaved. This is not a blind assumption, however, because the quality-control fit statistics immediately report where, and to what extent, this requirement has not been exactly met.
After measures have been constructed from data, they exist in a strictly linear frame of reference. This means that plots of the measures do, in fact, have the geometric properties generally assumed by unsophisticated readers to exist in all numbers. Ordinal numbers, such as the original ordered observations, do not have these strict geometric properties.

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 all unanchored elements and rating-scale categories. For anchored values, this convergence criterion is never met, but the fit statistics etc. are computed and reported by *Facets* as though it has been met. Convergence is based on the unanchored estimates. For more about estimation including CMLE, MMLE PMLE, see Estimation Considerations.

---

**Historical Note on Inter-Rater Reliability:**

Inter-rater Reliability is really slippery. Reliable in what way?

If we need raters to agree about who is better and who is worse, then correlations can work fine.

If we need raters to agree with the official ratings, often the aim of rater training, then a root-mean-square of the difference between the rater's ratings and the official ratings can work fine.

If we need raters to agree with who passes and who fails, then direct comparison of the ratings with the cut-point ratings can work fine up to a point. The problem here is rater leniency/severity. This goes all the way back to 1890 and the first mathematical study of rating. F. Y. Edgeworth, in his paper "The Element of Chance in Competitive Examinations", discovered that the spread of rater leniencies was about half the spread of person abilities. This has been seen many times since. In 1931, the "Conference on Examinations" (Eastbourne, England) tried to solve this problem mathematically, but they got bogged down in arguments between proponents of different standard statistical models. These models were descriptive, lacking any clear foundation in theory. It was not until 1986 that a solid foundation based on Rasch theory was constructed and solved the problem. This was the development of MFRM. a statistical methodology designed to produce linear measures that adjust for rater leniency and missing data with the minimum load on the raters, while also producing useful diagnostic information about each rater's behavior.

The benchmark paper for standard statistical approaches is Saal, F.E., Downey, R.G. and Lahey, M.A (1980) Rating the Ratings: Assessing the Psychometric Quality of Rating Data, Psychological Bulletin, 88(2), 413-428. There have been many methodological tweaks since then, but their message remains solid.

There were many papers written comparing MFRM with CTT and Generalizability Theory around 1990. In the Language Testing arena, https://www.winsteps.com/facetman/references.htm suggests 60 or so papers. (I have not kept this up to date, so there would be many more now.) The most comprehensive current comparison of standard statistics and MFRM is the book, "Introduction to Many-Facet Rasch Measurement" by Thomas Eckes published by Peter Lang. Probably also relevant, the book "Fairness, Justice and Language Assessment" by Tim McNamara, Ute Knoch, and Jason Fan published by Oxford UP.

## 4.3 References

*Please cite the current version of Facets as:*
Linacre, J. M. (2023) Facets computer program for many-facet Rasch measurement, version 3.87.0. Beaverton, Oregon: Winsteps.com

For a webpage in Facets Help:
Linacre, J. M. (%YEAR%, %MONTH%) Web page title, Retrieved from www.winsteps.com/facetman/webpage.htm

*Books*
MFRM means Linacre J.M. (1994). Many-Facet Rasch Measurement, 2nd Ed. Chicago: MESA Press www.winsteps.com - free
BTD means Wright B.D. & Stone M.H. (1979) Best Test Design, Chicago: MESA Press www.rasch.org - free
RSA means Wright B.D. & Masters G.N. (1982) Rating Scale Analysis, Chicago: MESA Press www.rasch.org - free

"Introduction to Many-Facet Rasch Measurement" by Thomas Eckes (2015, 2nd Edn.), Frankfurt am Main: Peter Lang.
"Invariant Measurement with Raters and Rating Scales: Rasch Models for Rater-Mediated Assessments" (2017, 2018) George Engelhard Jr., Stefanie Wind. Routledge - Taylor & Francis
"Statistical Analyses for Language Testers" (2013) Rita Green, Palgrave Macmillan

"Measuring Second Language Performance" by T. F. McNamara, Addison-Wesley Longman, 1996.
"Applying the Rasch Model: Fundamental Measurement in the Human Sciences", by Trevor G. Bond & Christine M. Fox, 4th edn 2020. Routledge
"Introduction to Rasch Measurement", Everett V. Smith, Jr. & Richard M. Smith (Eds.) JAM Press, 2004: jampress.org
Rasch G. (1960, 1980, 1992) Probabilistic Models for Some Intelligence and Attainment Tests. Chicago: MESA Press. www.rasch.org/models.htm

*Other recommended sources:*
Rasch Measurement Transactions: www.rasch.org/rmt/
Journal of Applied Measurement: jampress.org

*References: (If you publish a paper using many-facet Rasch measurement, or know of one not on this list, please notify www.winsteps.com.)*
For quick access to online papers, Google Scholar "+Facets +Rasch +Linacre filetype:pdf" - as of January 2018, this produced about 600 hits.


Ahlström S., Bernspång B. (2003) Occupational Performance of Persons Who Have Suffered a Stroke: a Follow-up Study. Scandinavian Journal of Occupational Therapy, 10, 2, 88 - 94.
Allen J.M. & Schumacker R.E.(1998) Team Assessment Utilizing a Many-Facet Rasch Model. Journal of Outcome Measurement 2:2, 142-158.
Atchison B.T., Fisher A.G., Bryze K. (1998) Rater reliability and internal scale and person response validity of the School Assessment of Motor and Process Skills. American Journal of Occupational Therapy, 52, 843-850.
Bachman L. F., Davidson, F., Ryan, K., & Choi, I. (1995). An investigation into the comparability of two tests of English as a foreign language: The Cambridge-TOEFL comparability study. Cambridge University Press.
Bachman L.F., Lynch, B.K., and Mason M. (1995) Investigating variability in tasks and rater judgements in a performance test of foreign language speaking. Language Testing, 12, 238-257.
Banerji M. (2000) Construct Validity of Scores/Measures from a Developmental Assessment of Mathematics using Classical and Many-Facet Rasch Measurement. Journal of Applied Measurement, 1:2, 177-198.
Barnett R.V., Easton J., Israel G.D. (2002) Keeping Florida's Children Safe in School: How One State Designed a Model Safe School Climate Survey. School Business Affairs, 68, 6, 31-38.
Barrett, S. (2001) The impact of training on rater variability. International Education Journal, 2 (1), 49-58
Basturk, R. (2008) Applying the many-facet Rasch model to evaluate PowerPoint presentation performance in higher education. Assessment and Evaluation in Higher Education, 33, (4) 431-444.

Batty, A.O. (2015). A comparison of video- and audio-mediated listening tests with many-facet Rasch modeling and differential distractor functioning. Language Testing, 32: 3-20.
Beck C.T. & Gable R.K. (2000) Postpartum Depression Screening Scale: Development and Psychometric Testing. Nursing Research. 49(5):272-282 www.nursingresearchonline.com
Beck C.T. & Gable R.K. (2001) Comparative Analysis of the Performance of the Postpartum Depression Screening Scale With Two Other Depression Instruments. Nursing Research. 50(4):242-250. www.nursingresearchonline.com
Beck C.T. & Gable R.K. (2001) Further Validation of the Postpartum Depression Screening Scale. Nursing Research. 50(3):155-164. www.nursingresearchonline.com
Beck C.T. & Gable R.K. (2003) Postpartum Depression Screening Scale: Spanish Version. Nursing Research. 52(5):296-306. www.nursingresearchonline.com
Bernspång B. (1999) Rater Calibration Stability for the Assessment of Motor and Process Skills. Scandinavian Journal of Occupational Therapy, 6, 3, 101-109
Bernspång, B., & Fisher, A.G. (1995) Differences between persons with right or left CVA on the Assessment of Motor and Process Skills. Archives of Physical Medicine and Rehabilitation, 76, 1144-1151.
Bernspång, B., & Fisher, A.G. (1995) Validation of the Assessment of Motor and Process Skills for use in Sweden. Scandinavian Journal of Occupational Therapy, 2, 3-9.
Bode R.K., Klein-Gitelman M.S., Miller M.L., Lechman T.S., Pachman L.M. (2003) Disease activity score for children with juvenile dermatomyositis: Reliability and validity evidence . Arthritis Care & Research, 49, 1, 7-15.
Bonk W. (2000) KEPT ARGENTINA: 2000 Administration Final Report. Kanda University of International Studies.
Bonk W.J. & Ockey G.J. (2003) A many-facet Rasch analysis of the second language group oral discussion task. Language Testing, 20, 1, 89-110
Bray, K., Fisher, A.G., Duran, L. (2001) The validity of adding new tasks to the Assessment of Motor and Process Skills. American Journal of Occupational Therapy, 55, 409-415.
Breton, G., Lepage, S., & North, B. (2008). Cross-language benchmarking seminar to calibrate examples of spoken production in English, French, German, Italian and Spanish with regard to the six levels of the Common European Framework of Reference for Languages (CEFR). Strasbourg: Language Policy Division.
Brindley, G. (1998) Outcomes-based assessment and reporting in language learning programs: a review of the issues. Language Testing, 15, 1: 45-85.
Brown A. & Hill K. (1998) Interviewer style and candidate performance in the IELTS oral interview. IELTS Research 1, 1
Brown A. (2003) Interviewer variation and the co-construction of speaking proficiency. Language Testing, 20, 1, 1-25
Caban H.L. (2003) Rater group bias in the speaking assessment of four l1 Japanese ESL students. Second Language Studies, 21(2), 1-44.
Campbell S.K., Kolobe T.H.A., Osten E.T., Lenke M., Girolami G.L. (1995). Construct Validity of the Test of Infant Motor Performance. Physical Therapy 75:7 p.585-596.
Campbell S.K., Kolobe T.H.A., Wright B.D., Linacre J.M. (2002) Validity of the Test of Infant Motor Performance for prediction of 6-, 9- and 12-month scores on the Alberta Infant Motor Scale. Developmental Medicine & Child Neurology, 44: 263 - 272
Campbell, S.K., Osten, E.T., Kolobe, T.H. A., Fisher, A.G. (1993) Development of the Test of Infant Motor Performance. Physical Medicine and Rehabilitation Clinics of North America: New Developments in Functional Assessment, 4, 541-550.
Chatterji, M. (2002). Measuring leader perceptions of school readiness for reforms: Use of an iterative model combining classical and Rasch methods. Journal of Applied Measurement, 3, 455-485.
Chesnut R.M. et al. (1999) AHRQ Evidence Reports, Number 2. Rehabilitation for Traumatic Brain Injury. National Library of Medicine. HSTAT.
Chi, E. (2001) Comparing Holistic and Analytic Scoring for Performance Assessment with Many-facet Rasch Measurement. Journal of Applied Measurement 2:4, 379-388.
Choi, S.E. (1997) Rasch invents "Ounces". Rasch Measurement Transactions 11:2, 557.
College Board (2003) Monitoring Faculty Consultant Performance in the Advanced Placement English Literature and Composition Program with a Many-Faceted Rasch Model, RR No. 2003-1
Congdon, P. (1998). Unmodelled rater discrimination error. In M. Wilson & G. Engelhard (Eds.), Objective measurement: Theory into practice (Vol. 5). Stamford, CT: Ablex.
Coniam D. & Falvey P. (2002) Does Student Language Ability Affect the Assessment of Teacher Language Ability? Journal of Personnel Evaluation in Education, 16, 4, 269-285

Connally J, Jorgensen K, Gillis S, Griffin P. (2003) A multi-source measurement approach to the assessment of higher order competencies. AVETRA.
Cooke, K.Z., Fisher, A.G., Mayberry, W., Oakley, F. (2000) Differences in activities of daily living process skills of persons with and without Alzheimer's disease. Occupational Therapy Journal of Research, 20, 87-104.
Dapueto J.J., Francolino C., Servente L., Chang C.H., Gotta I., Levin R., del Carmen Abreu M. (2003) Evaluation of the Functional Assessment of Cancer Therapy-General (FACT-G) Spanish Version 4 in South America: Classic Psychometric and Item Response Theory Analyses. Health and Quality of Life Outcomes, 1, 32.
Darragh A.R., Sample P.L., Fisher A.G. (1998) Environment effect of functional task performance in adults with acquired brain injuries: use of the assessment of motor and process skills. Arch Phys Med Rehabil. 79, 4, 418-23.
Daud N.M. & Kasim N.L.A. (2004) Class assessment: cans students be relied on? The Australian Association for Research in Education.
de Jong J.H.A.L. & Bernstein J. (2001) Relating PhonePassTM Overall Scores to the Council of Europe Framework Level Descriptors. Proceedings of Eurospeech 2001
DeCarlo L. T. (2005) A Model of Rater Behavior in Essay Grading Based on Signal Detection Theory. Journal of Educational Measurement, 42, 1, 53-76
Derrickson J.P., Fisher A.G., Anderson J.E.L. (2001) Lessons learned from an assessment of the Individual-Level Core Food Security Module. In Andrews M.S. & Prell M.A. (Eds.) Second Food Security Measurement and Research Conference, Volume II: Papers. USDA, 50-58.
Derrickson J.P., Fisher A.G., Anderson J.E.L., Brown A.C. (2001) An Assessment of Various Household Food Security Measures in Hawaiì Has Implications for National Food Security Research and Monitoring. Journal of Nutrition. 131:749-757.)
Derrickson, J.P., Anderson, J.E. L., Fisher, A.G. (2000) Concurrent validity of a face valid food security measure [Online]. Institute for Research on Poverty, University of Wisconsin (Discussion Paper DP-1206-00).
Derrickson, J.P., Fisher, A.G., Anderson, J.E. L. (2000) The Core Food Security Module scale measure is valid and reliable when used with Asians and Pacific Islanders. Journal of Nutrition, 130, 2666-2674.
DeShea L. (2003) A scenario-based scale of Willingness to Forgive. Individual Differences Research, 1, 3, 201-217.
Dickerson, A.E., & Fisher, A.G. (1993) Age differences in functional performance. American Journal of Occupational Therapy, 47, 686-692.
Dickerson, A.E., & Fisher, A.G. (1995) Culture-relevant functional performance assessment of the Hispanic elderly. Occupational Therapy Journal of Research, 15, 50-68.
Dickerson, A.E., & Fisher, A.G. (1997) The effects of familiarity of task and choice on the functional performance of young and old adults. Psychology and Aging, 12, 247-254.
Doble, S.E., & Fisher, A.G. (1998) The dimensionality and validity of the Older Americans Resources and Services (OARS) activities of daily living (ADL) scale. Journal of Outcome Measurement, 2, 2-23.
Doble, S.E., Fisk, J.D., Fisher, A.G., Ritvo, P.G., Murray, T.J. (1994) Functional competence of community-dwelling persons with multiple sclerosis using the Assessment of Motor and Process Skills. Archives of Physical Medicine and Rehabilitation, 75, 843-851.
Doble, S.E., Fisk, J.D., MacPherson, K.M., Fisher, A.G., Rockwood, K. (1997) Measuring functional competence in older persons with Alzheimer's disease. International Psychogeriatrics, 9, 25-38.
Doble, S.E., Fisk, J.D., Rockwood, K. (1999) Assessing the ADL Functioning of Persons With Alzheimer's Disease: Comparison of Family Informants' Ratings and Performance-Based Assessment Findings. International Psychogeriatrics (1999), 11: 399-409
Du, Y. and Brown, W.L. (2000) Raters and Single Prompt-to-Prompt Equating Using the FACETS Model in a Writing Performance Assessment. In Objective Measurement: Theory into Practice, Vol. 5 Westport Ct: Ablex.
Du, Y. and Wright, B.D. (1997) Effects of Item Characteristics in a Large-scale Direct Writing Assessment. In M. Wilson, G. Engelhard, Jr., & K. Draney (Eds.), Objective Measurement: Theory into Practice (Vol. 4, pp. 1-24). Norwood, NJ: Ablex.
Duran, L., & Fisher, A.G. (1996) Male and female performance on the Assessment of Motor and Process Skills. Archives of Physical Medicine and Rehabilitation, 77, 1019-1024.
Dávid, G.A. (2007) Investigating the Performance of Alternative Types of Grammar Items. Language Testing. 24/1, pp. 65-97.
Dávid, G.A. (2010). Linking the General English Suite of Euro Examinations to the CEFR: a case study report. In: Martyniuk, W. (Szerk.). Aligning Tests with the CEFR: Reflections on Using The Council of Europe's Draft Manual. Cambridge: Cambridge University Press, 177-203.

Dávid, G.A. (2011). Linking the Euroexams to the Common European Framework of Reference: A full report. Budapest: Euroexam Nyelvvizsga Kft.
Dávid, G.A. (2012). A szintleírások nyelvének szerepe a közös európai referenciakeret magyar, angol és német nyelvu kiadásában. Magyar Pedagógia.112(1), 19-39. (Hungarian). English summary.
Eckes T. (2004) Beurteilerübereinstimmung und Beurteilerstrenge: Eine Multifacetten-Rasch-Analyse von Leistungsbeurteilungen im "Test Deutsch als Fremdsprache" (TestDaF). Diagnostica, 50, 2, 65-77.
Eckes T. (2004) Facetten des Sprachtestens: Strenge und Konsistenz in der Beurteilung sprachlicher Leistungen. In A. Wolff, T. Ostermann & C. Chlosta (Hrsg.), Integration durch Sprache (Materialien Deutsch als Fremdsprache, Bd. 73, S. 485 - 518). Regensburg: FaDaF.
Eckes T. (2005) Analyse und Evaluation sprachproduktiver Prüfungen beim TestDaF. In I. Kühn, W. Timmermann & M. Lehker (Hrsg.), Sprachtests in der Diskussion (S. 60 - 93). Frankfurt: Lang.
Eckes T. (2005) Evaluation von Beurteilungen: Psychometrische Qualitätssicherung mit dem Multifacetten-Rasch-Modell. Zeitschrift für Psychologie, 213, 77 - 96.
Eckes T. (2005) Examining rater effects in TestDaF writing and speaking performance assessments: A many-facet Rasch analysis. Language Assessment Quarterly, 2, 3, 197-221.
Eckes T. (2006) Multifacetten-Rasch-Analyse von Personenbeurteilungen. [German] Zeitschrift fur Sozialpsychologie, 37, 3, 185-195.
Eckes T. (2008) Assuring the quality of TestDaF examinations: A psychometric modeling approach. In L. Taylor & C. J. Weir (Eds.), Multilingualism and assessment: Achieving transparency, assuring quality, sustaining diversity – Proceedings of the ALTE Berlin Conference May 2005 (pp. 157–178). Cambridge, UK: Cambridge University Press.
Eckes T. (2008) Rater types in writing performance assessments: A classification approach to rater variability. Language Testing, 25, 155–185.
Eckes T. (2009) Manual for relating Language Examinations to the Common European Framework of Reference for Languages (CEFR). Council of Europe. www.coe.int/t/dg4/linguistic/Source/CEF-refSupp-SectionH.pdf
Eckes T. (2016) Setting cut scores on an EFL placement test using the prototype group method: A receiver operating characteristic (ROC) analysis. Language Testing.
Eckes, T. & Grotjahn, R. (2006) A closer look at the construct validity of C-tests. Language Testing 2006 23 (3) 290 - 325
Elder, C., Barkhuizen, G., Knoch, U. and von Randow, J. (2007). Evaluating rater responses to an online training program for writing assessment. Language Testing 24, 1, 1-28.
Elder, C., Iwashita, N., & McNamara, T. (2002). Estimating the difficulty of oral proficiency tasks: What does the test-taker have to offer? Language Testing, 19, 347-368.
Elder, C., Knoch, U., Barkhuizen, G. and von Randow, J. (2005). Feedback to enhance rater training. Does it work? Language Assessment Quarterly 2, 3, 175-196
Ellison, S., Fisher, A.G., Duran, L. (2001) The alternate forms reliability of the new tasks added to the Assessment of Motor and Process Skills. Journal of Applied Measurement, 2, 120-133.
Engelhard G., Jr. & Anderson, D.W. (1998). A binomial trials model for examining the ratings of standard-setting judges. Applied Measurement in Education, 11(3), 209-230.
Engelhard G., Jr. & Myford, C.M. (2003) Monitoring Faculty Consultant Performance in the Advanced Placement English Literature and Composition Program with a Many-Faceted Rasch Model, RR No. 2003-1. College Board.
Engelhard G., Jr. & Stone, G.E. (1998) Evaluating the Quality of Ratings Obtained From Standard-Setting Judges, Educational and Psychological Measurement, 58(2), 179-196.
Engelhard G., Jr. (1992). The measurement of writing ability with a many-faceted Rasch model. Applied Measurement in Education, 5(3), 171-191.
Engelhard G., Jr. (1994). Examining rater errors in the assessment of written composition with a many-faceted Rasch model. Journal of Educational Measurement, 31(2), 93-112.
Engelhard G., Jr. (1996). Evaluating rater accuracy in performance assessments. Journal of Educational Measurement, 33(1), 56-70.
Engelhard G., Jr. (1997). Constructing rater and task banks for performance assessments. Journal of Outcome Measurement, 1(1), 19-33.
Engelhard G., Jr. (2001) Examining the Psychometric Quality of the National Board for Professional Teaching Standards Early ... Journal of Personnel Evaluation in Education 15:4, 253-285
Engelhard G., Jr. (2002). Monitoring raters in performance assessments. In G. Tindal & T. Haladyna (Eds.), Large-scale assessment programs for ALL students: Development, implementation, and analysis (p. 261-287). Mahway, NJ: Lawrence Erlbaum Associates, Pub.

Engelhard G., Jr., Cramer S. (1997) Using Rasch Measurement to Evaluate Ratings of Standard-Setting Judges. In M. Wilson, G. Engelhard, Jr., & K. Draney (Eds.), Objective Measurement: Theory into Practice (Vol. 4, pp. 97-112). Norwood, NJ: Ablex.
Engelhard G., Jr., David, M., & Hansche, L. (1999). Evaluating the accuracy of judgments obtained from item review committees. Applied Measurement in Education, 12, 199-210.
Engelhard G., Jr., Myford, C. M. (2003). Monitoring Faculty Consultant Performance in the Advanced Placement English Literature and Composition Program with a Many-Faceted Rasch Model. College Board Research Report No. 2003-1. ETS RR-03-01.
Engelhard G., Jr., Myford, C. M., & Cline, F. (2000). Investigating assessor effects in National Board for Professional Teaching Standards assessments for Early Childhood/Generalist and Middle Childhood/Generalist certification (ETS Research Report RR-00-13). Princeton, NJ: Educational Testing Service.
Englund, B., Bernspång, B., Fisher, A.G. (1995) Development of an instrument for assessment of social interaction skills in occupational therapy. Scandinavian Journal of Occupational Therapy, 2, 17-23.
Fisher A.G. (1992) Commentary [Haley, S.M., Ludlow, L.H. Applicability of the hierarchical scales of the Tufts Assessment of Motor Performance for school-aged children and adults with disabilities]. Physical Therapy, 72, 202-203.
Fisher A.G. (1993) The assessment of IADL motor skills: An application of many-faceted Rasch analysis. American Journal of Occupational Therapy, 47, 319-329.
Fisher A.G. (1993, April) The assessment of IADL motor skills: An application of many-faceted Rasch analysis. American Journal of Occupational Therapy, 47(4), 319-329.
Fisher A.G. (1994) Development of a functional assessment that adjusts ability measures for task simplicity and rater leniency. In M. Wilson (Ed.), Objective measurement: Theory into practice. Vol II (pp. 145-175) Norwood, New Jersey: Ablex Publishing Corporation.
Fisher A.G. (1997) Multifaceted measurement of daily life task performance : Conceptualizing a test of instrumental ADL and validating the addition of personal ADL tasks. Physical medicine and rehabilitation : State of the Art Reviews. 11(2) : 289-303.
Fisher A.G. (2003) Assessment of Motor and Process Skills (AMPS). Vol. 1: Development, Standardization, and Administration Manual (5th ed.) Fort Collins, CO: Three Star Press. www.ampsintl.com
Fisher A.G. (2004) Assessment of Motor and Process Skills (AMPS). Vol. 2: User Manual (revised 5th ed.) Fort Collins, CO: Three Star Press.
Fisher A.G., & Duran, G.A. (2004) Schoolwork Task Performance of Students At Risk for Delays. Scandinavian Journal of Occupational Therapy, 11, 1-8.
Fisher A.G., Bryze K.A., Granger C.V., Haley S.M., Hamilton B.B., Heinemann A.W., Puderbaugh J.K., Linacre J.M., Ludlow L.H., McCabe M.A. & Wright B.D. (1994) Applications of conjoint measurement to the development of functional assessments. International Journal of Educational Research, 21(6), 579-593.
Fisher A.G., Bryze, K., Hume. V. (2002) School AMPS: School Version of the Assessment of Motor and Process Skills. Ft. Collins, CO: Three Star Press.
Fisher A.G., Liu, Y., Velozo, C.V. Pan, A.W. (1992) Cross-cultural assessment of process skills. American Journal of Occupational Therapy, 46, 876-885.
Fisher P. (1991) Baseball Plays and Players, Rasch Measurement Transactions, 5:2 p.142
Fisher P. (1994) Measurement of Golf proficiency, Rasch Measurement Transactions, 1994, 7:4 p.332
Fisher W.P. (2003) Mathematics, Measurement, Metaphor and Metaphysics II. Theory & Psychology, 13, 6, 791-828
Fisher W.P., & Fisher, A.G. (1993) Applications of Rasch analysis to studies in occupational therapy. Physical Medicine and Rehabilitation Clinics of North America: New Developments in Functional Assessment, 4, 551-569.
Fitzpatrick A.R, Ercikan K., Yen W., Ferrara S. (1998) The consistency between raters scoring in different test years. Applied Measurement in Education, 11, 195-208.
Fujioka-Ito, N (2004) Development of Speech Contest appraisal. Central Association of Teachers of Japanese. tell.fll.purdue.edu/CATJ16/fujioka.html
Galbraith P. & Haines C. (1998) Disentangling the Nexus: Attitudes to Mathematics and Technology in a Computer Learning Environment. Educational Studies in Mathematics, 36, 3, 275-290
Gallegos P.J., Peeters M.J.(2011). A measure of teamwork perception for team-based learning. Currents in Pharmacy Teaching and Learning, 3(1), 30-35.
Gerdes R., Bauske E, Kaiser F.G.(2023) A general explanation for environmental policy support: An example using carbon taxation approval in Germany. Journal of Environmental Psychology, 90. https://doi.org/10.1016/j.jenvp.2023.102066

Girard C.R., Fisher A.G., Short M.A., Duran L. (1999) Occupational performance differences between psychiatric groups. Scandinavian Journal of Occupational Therapy, 6, 119 - 126
Goldman, S.L., & Fisher, A.G. (1997) Cross-cultural validation of the Assessment of Motor and Process Skills (AMPS) British Journal of Occupational Therapy, 60, 77-85.
Goto, S., Fisher, A.G., Mayberry, W.L. (1996) AMPS applied cross-culturally to the Japanese. American Journal of Occupational Therapy, 50, 798-806.
Griffin P. & Gillis S. (2000) A multi source measurement approach to assessment of higher order competencies. British Education Research Association Annual Conference.
Haans A., Kaiser F.G., Bouwhuis D.G., IJsselsteijn W.A. (2012) Individual differences in the rubber-hand illusion: Predicting self-reports of people's personal experiences. Acta Psychologica, 141, 169–177
Haines C. & Crouch R. (2001) Recognizing constructs within mathematical modelling. Teaching Mathematics and its Applications, 20, 3, 129-138
Haines C.R. & Houston S.K. (2004?) Assessing student project work. ICMI Study On the Teaching and Learning of Mathematics at University Level.
Haladyna T. & Hess R. (1999) An Evaluation of Conjunctive and Compensatory Standard-Setting Strategies for Test Decisions. Educational Assessment. 6, 2, 129-153
Harasym, P. H., Woloschuk, W., & Cunning, L. (2008). Undesired variance due to examiner stringency/leniency effect in communication skill scores assessed in OSCEs. Advances in Health Sciences Education, 13, 617–632.
Hartman M.L., Fisher A.G., Duran L. (1999) Assessment of functional ability of people with Alzheimer's disease, 6, 111 - 118
Hayase D., Mosenteen D., Thimmaiah D., Zemke S., Atler K., Fisher A.G. (2004) Age-related changes in activities of daily living ability . Australian Occupational Therapy Journal, 51, 4, 192
Heller, J. I., Sheingold, K., & Myford, C. M. (1999). Reasoning about evidence in portfolios: Cognitive foundations for valid and reliable assessment. Educational Assessment, 5 (1), 5-40.
Hermansson L.M., Fisher A.G., Bernspång B., Eliasson A.-C. (2004) Assessment of Capacity for Myoelectric Control: a new rasch-built measure of prosthetic hand control. Journal of Rehabilitation Medicine, 36, 1-6.
Hess R.J. & Becker M.S. (1997) Applied Assessment in the Glendale Union High School District: An Application of the Many-Faceted Rasch Model. National Association of Test Directors.
Hickey D.T., Wolfe E.W., Kindfield A.C.H. (2001) Assessing Learning in a Technology-Supported Genetics Environment: Evidential and Systemic Validity Issues. Educational Assessment, 6, 3, 155-196.
Hoskens, M., & Wilson, M. (2001). Real-time feedback on rater drift in constructed response items: An example from the Golden State Examination. Journal of Educational Measurement, 38, 121-146.
Houran J. (2004) The public's expectation of finding a soul mate. True Online Magazine.
Hsieh, Mingchuang (2013) An application of Multifaceted Rasch measurement in the Yes/No Angoff standard setting procedure. Language Testing Journal, 30, 491-512.
Hung, L-F, Wang, W-C (2012). The Generalized Multilevel Facets Model for Longitudinal Data. Journal of Educational and Behavioral Statistics, 37, 2, 231–255.
Iramaneerat, C., Myford, C., Yudkowsky, R., Lowenstein T. (2009). Evaluating the effectiveness of rating instruments for a communication skills assessment of medical residents. Adv in Health Sci Educ (2009) 14:575–594
Iramaneerat, C., Yudkowsky, R., Myford, C., & Downing, S. (2007). Quality control of an OSCE using generalizability theory and many-faceted Rasch measurement. Advances in Health Sciences Education.
Iwashita, N., McNamara, T., & Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information-processing approach to task design. Language Learning, 51, 401-436.
Johnson, J. S., and Lim, G. S. (2009). The influence of rater language background on writing performance assessment. Language Testing, 26(4), 485-505.
Jones N. & Shaw S.D. (2003) Task difficulty in the assessment of writing: Comparing performance across three levels of CELS. Cambridge Esol Research Notes, 11, 11-15.
Kaliski, P.K., Wind, S.A., Engelhdard, G. Jr., Morgan, D.L., Plake, B.S., Reshetar, R.A. (2013) Using the Many-Faceted Rasch Model to Evaluate Standard Setting Judgments: An Illustration With the Advanced Placement Environmental Science Exam. Educational and Psychological Measurement, 73(3) 386-411.
Kassim, Noor Lide Abu (2011) Judging behaviour and rater errors: an application of the many-facet Rasch model. GEMA: Online Journal of Language Studies, 11 (3). pp. 179-197.
Kinnman J., Andersson U., Wetterquist L., Kinnman Y., Andersson U. (2000) Colling suit for multiple sclerosis: functional improvement in daily living? Scandinavian Journal of Rehabilitation Medicine, 32, 1, 20 - 24

Kirkley, K., & Fisher, A.G. (1999) Alternate forms reliability of the Assessment of Motor and Process Skills. Journal of Outcome Measurement, 3, 53-70.
Kjellberg A., Haglund L., Forsyth K., Kielhofner G. (2003) The measurement properties of the Swedish version of the assessment of communication and interaction skills. Scandinavian Journal of Caring Sciences, 17, 3, 271-277
Kline TL, Schmidt KM, Bowles RP. (2006). Using LinLog and FACETS to model item components in the LLTM. Journal of Applied Measurement, 7(1):74-91.
Knoch U. (2007). Do empirically developed rating scales function differently to conventional rating scales for academic writing? Spaan Fellow Working Papers in Second or Foreign Language Assessment, 5
Knoch U. (2007). 'Little coherence, considerable strain for reader': A comparison between two rating scales for the assessment of coherence. Assessing Writing 12, 108-128
Knoch U. (2008). Diagnostic writing ability: A rating scale for the assessment of accuracy, fluency and complexity. New Zealand Studies in Applied Linguistics 14, 2, 1-25
Knoch U. (2008). The assessment of academic style in EAP writing: The case of the rating scale. Melbourne Papers in Language Testing 13, 1, 34-67
Knoch U. (2009). The development and validation of a rating scale for diagnostic writing assessment. Language Testing 26, 2, 275-304
Knoch U., Read, J and von Randow, J. (2007). Re-training writing raters online: How does it compare with face-to-face training? Assessing Writing 12, 26-43
Kondo-Brown, K. (2002). A FACETS analysis of rater bias in measuring Japanese writing performance. Language Testing, 19(1), 3-31.
Kottorp A., Bernspång B., Fisher A.G. (2003) Validity of a performance assessment of activities of daily living for people with developmental disabilities. Journal of Intellectual Disability Research, 47, 8, 597 - November 2003
Kottorp A., Bernspång B., Fisher, A .G., Bryze, K. (1995) IADL ability measured with the AMPS: Relation to two classification systems of mental retardation. Scandinavian Journal of Occupational Therapy, 2, 121-128.
Kottorp A., Bernspång B., Fisher, A.G. (2003) Activities of daily living in persons with intellectual disability: Strengths and limitations in specific motor and process skills. Australian Occupational Therapy Journal, 50, 195-204.
Kottorp A., Bernspång, B., Fisher, A.G. (2003) Validity of a performance assessment of activities of daily living for persons with developmental disabilities. Journal of Intellectual Disability Research, 47, 597-605.
Kottorp A., Hällgren, M., Bernspång, B., Fisher, A.G. (2003) Client-centred occupational therapy for persons with mental retardation: Implemention of an intervention programme in activities of daily living tasks. Scandinavian Journal of Occupational Therapy, 10, 51-60.
Kozaki Y. (2004) Using GENOVA and FACETS to set multiple standards on performance assessment for certification in medical translation from Japanese into English. Language Testing, 21, 1, 1-27
Kubinger, K.D. (2012). Ein mathematisches Modell zur Fußball-Toto-Prognose (12er-Wette) (in German). In M. Voracek (Hrsg.), Of things past – Memorial Book for Anton K. Formann (1949-2010) (S. 71-78).
Lai J.S., Velozo C.A., Linacre J.M. (1997) Adjusting for Rater Severity in an Un-linked FIM national Data Base: An Application of the Many-Facets model. Physical Medicine and Rehabilitation 11:2 325-332.
Lai, J.S., Fisher, A.G., Magalhaes, L.C., Bundy, A.C. (1996) Construct validity of the praxis tests on the Sensory Integration and Praxis Tests. Occupational Therapy Journal of Research, 16, 75-97.
Lang S. (2002) Ranking and Measuring of Sailing Teams. Rasch Measurement Transactions. 15:4, 851.
Lange R.(2003) Scaling Methods: Model Sailplane Competition: From Awarding Points to Measuring Performance Skills. R/C Soaring Digest. 20:8, 4-11.
Lange R., Greyson B., Houran J. (2004) A Rasch scaling validation of a 'core' near-death experience. British Journal of Psychology, 95, 2, 161-177
Lange R., Jerabeck, I., Houran J. (2004) Building blocks for satisfaction in long-term romantic relationships: Evidence for the complementarity hypothesis for romantic compatibility. Adult Development Symposium Society for Research in Adult Development Preconference, AERA
Lange R.; Lange X. (2012) Quality Control in Crowdsourcing: An Objective Measurement Approach to Identifying and Correcting Rater Effects in the Social Evaluation of Products and Services. AAAI Spring Symposium Series, North America, Mar. 2012. Available at: www.aaai.org/ocs/index.php/SSS/SSS12/paper/view/4322/4691.
Lange, R., and Lange, X. (2012). Quality control in crowdsourcing: An Objective Measurement approach to identifying and correcting rater effects in the social evaluation of products and services. Paper presented at the AAAI Spring Symposium, Stanford University. March 26, 2012.
Lee M., Zhu W., Ulrich D.A. (2005) Many-Faceted Rasch Calibration of TGMD-2. AAHPERD aahperd.confex.com/aahperd/2005/preliminaryprogram/abstract_7094.htm

Lee Y.-W. & Kantor R. (2003) Investigating Differential Rater Functioning for Academic Writing Samples: An MFRM Approach. Educational Testing Service.
Li J. (2014) Examining genre effects on test takers' summary writing performance. Assessing Writing, 22,75-90.
Liao, P. M., Campbell, S.K. (2002) Comparison of Two Methods for Teaching Therapists to Score the Test of Infant Motor Performance. Pediatric Physical Therapy. 14, 4, 191-198.
Liao, P. M., Campbell, S.K. (2004) Examination of the Item Structure of the Alberta Infant Motor Scale. Pediatric Physical Therapy, 16, 1, 31-38.
Lim, G. S. (2010). Investigating prompt effects in writing performance assessment. Spaan Fellow Working Papers in Second or Foreign Language Assessment, 8, 95-116.
Linacre J.M. (1988) Rasch Analysis and Rank Ordering. Rasch Measurement Transactions, 2:4, 41.
Linacre J.M. (1991) Inter-rater reliability. Rasch Measurement Transactions, 5:3, 166.
Linacre J.M. (1991) Stress after Three Mile Island. Rasch Measurement Transactions, 5:4, 188.
Linacre J.M. (1992) Paired comparisons with ties. Rasch Measurement Transactions 9:2, 425.
Linacre J.M. (1993) Rasch-based generalizability theory. Rasch Measurement Transactions 7:1, 283-284.
Linacre J.M. (1994) Constructing measurement with a many-facet Rasch model. In M. Wilson (Ed.) Objective Measurement: Theory in Practice. Vol. II. Newark NJ: Ablex.
Linacre J.M. (1996) Generalizability Theory and Rasch Measurement, Chapter 5., In Engelhard G, & Wilson M. (Eds) Objective Measurement, Vol. 3. Newark NJ: Ablex.
Linacre J.M. (1997) Communicating Examinee Measures as Expected Ratings. Rasch Measurement Transactions, 11:1, 550-551.
Linacre J.M. (2001) Generalizability theory and Rasch measurement. Rasch Measurement Transactions 15:1, 806-7.
Linacre J.M. (2001) Paired Comparisons for Measuring Team Performance. Rasch Measurement Transactions 15:1, 812
Linacre J.M. (2002) Facets, Factors, Elements and Levels. Rasch Measurement Transactions, 16:2, 880.
Linacre J.M. (2002) Judging debacle in Pairs Figure Skating. Rasch Measurement Transactions 15:4, 839-40.
Linacre J.M., Engelhard, G., Jr., Tatum, D.S., & Myford, C.M. (1994). Measurement with judges: Many-faceted conjoint measurement. International Journal of Educational Research, 21(6), 569-577.
Linacre J.M., Wright B.D. (2002) Understanding Rasch Measurement: Construction of Measures from Many-facet Data. Journal of Applied Measurement, 3:4, 486-512.
Linacre J.M., Wright B.D. (2004) Construction of measures from many-facet data. Chapter 8 in E.V. Smith & R. M. Smith (Eds.) Introduction to Rasch Measurement. Maple Grove MN: JAM Press.
Linacre J.M., Wright B.D., Lunz, M.E. (1990) A Facets Model for Judgmental Scoring. Memo 61. MESA Psychometric Laboratory. University of Chicago. www.rasch.org/memo61.htm
Looney M.A. (1996) Figure skating fairness. Rasch Measurement Transactions, 10:2, 500.
Looney M.A. (2004) Evaluating Judge Performance in Sport. Journal of Applied Measurement 5:1, 31-47.
Ludlow L.H. & Haley S.M. (1996) Effect of context in rating of mobility activities in children with disabilities. Educational and Psychological Measurement, 56, 122-129.
Lumley T, Congdon P., Linacre J. (1999) Rater Variability. Rasch Measurement Transactions 12:4, 671.
Lumley T. (2002) Assessment criteria in a large-scale writing test: what do they really mean to the raters? Language Testing, 19, 3, 246-276
Lumley T., & McNamara, T.F. (1995). Reader characteristics and reader bias: Implications for training. Language Testing, 12, 54-71.
Lunz M.E. & Stahl J.A. (1993, April) The effect of rater severity on person ability measures: A Rasch model analysis. American Journal of Occupational Therapy, 47(4), 311-317.
Lunz M.E. & Stahl, J.A. (1990). Judge consistency and severity across grading periods. Evaluation and the Health Professions, 13, 425-444.
Lunz M.E. & Wright B.D. (1997) Latent Trait Models for Performance Examinations. In J. Rost & R. Langeheine (Hrsg.), Applications of latent trait and latent class models in the social sciences. Munster: Waxmann.
Lunz M.E. (2004?) Standardized oral examinations. Measurement Research Associates.
Lunz M.E., & Stahl, J.A. (1990). Judge consistency and severity across grading periods. Evaluation and the Health Professions, 13, 425-444.
Lunz M.E., Linacre J.M. (1998) Measurement designs using multi-facet Rasch modeling. Chapter 3 in Marcoulides G. (Ed.) Modern Methods for Business Research. New York: Lawrence Erlbaum.
Lunz M.E., Stahl, J.A., & Wright, B.D. (1996). The invariance of rater severity calibrations. In G. Engelhard, Jr., & M. Wilson (Eds.), Objective Measurement: Theory into Practice (Vol. 3, pp. 99-112). Norwood, NJ: Ablex.

Lunz M.E., Wright, B.D., & Linacre, J.M. (1990). Measuring the impact of judge severity on examination scores. Applied Measurement in Education, 3(4), 331-345.
Lynch B. & McNamara T.F. (1998) Using G-theory and many-facet Rasch measurement in the development of performance assessments of the ESL speaking skills of immigrants. Language Testing 15: 158-180.
MacMillan, P. (2000) Classical, generalizability and mutlifaceted Rasch detection of interrater variablity in large, sparse data sets. Journal of Experimental Education, 68, 167-190.
MacMillan, P. (2000) Simultaneous Measurement of Reading Growth, Gender, and Relative-Age Effects: Many-Faceted Rasch Approach to CBM Reading Scores. Journal of Applied Measurement 1:4, 393-408.
Magalhães, L., Fisher, A.G., Bernspång, B., Linacre, J.M. (1996) Cross-cultural assessment of functional ability. Occupational Therapy Journal of Research, 16, 45-63.
Malec. J.F. (2004) Comparability of Mayo-Portland Adaptability Inventory ratings by staff, significant others and people with acquired brain injury. Brain Injury, 18, 6, 563-575
Maris, G. & Bechger, T. (2003) Two methods for the practical analysis of rating data. CITO Measurement and Research Department Report 2003-1.
McCutcheon L.E., Lange R., Houran J. (2002) Conceptualization and measurement of celebrity worship. British Journal of Psychology, 93, 1, 67-87
McManus I.C, Thompson M, Mollon J. (2006) Assessment of examiner leniency and stringency ('hawk-dove effect') in the MRCP(UK) clinical examination (PACES) using multi-facet Rasch modelling. BMC Medical Education 2006, 6:42. www.biomedcentral.com/1472-6920/6/42/
McNulty, M.C., & Fisher, A.G. (2001) Validity of using the Assessment of Motor and Process Skills to estimate overall home safety in persons with psychiatric conditions. American Journal of Occupational Therapy, 55, 649-655.
Merritt B.K., Fisher A.G. (2003) Gender differences in the performance of activities of daily living. Arch Phys Med Rehabil. 84, 12, 1872-7
Micele R., Settanni M., Vidotto, G. (2008) Measuring change in training programs: An empirical illustration. Psychology Science Quarterly, Volume 50 (3), 433-447
Monsaas J, Engelhard G (1996) Examining Changes in the Home Environment with the Rasch Measurement Model. In G. Engelhard, Jr., & M. Wilson (Eds.), Objective Measurement: Theory into Practice (Vol. 3, pp. 127-142). Norwood, NJ: Ablex.
Mulqueen C., Baker D., Dismukes R.K. (2000) Using Multifacet Rasch Analysis to Examine the Effectiveness of Rater Training. SIOP.
Mulqueen C., Baker D., Dismukes R.K. (2002) Pilot Instructor Rater Training: The Utility of the Multifacet Item Response Theory Model. The International Journal of Aviation Psychology, 12(3), 87 - 303
Myford C.M. (2002) Investigating Design Features of Descriptive Graphic Rating Scales. Applied Measurement in Education, 15, 2, 187-215
Myford C.M. (2004). [Review of the book Automated essay scoring: A cross-disciplinary perspective]. Journal of Applied Measurement, 5 (1), 111-114
Myford C.M. & Engelhard, George, Jr. (2001) Examining the Psychometric Quality of the National Board for Professional Teaching Standards Early Childhood/Generalist Assessment System. Journal of Personnel Evaluation in Education, v15 n4 p253-85 Dec 2001
Myford C.M. & Mislevy, R.J. (1995). Monitoring and improving a portfolio assessment system. ETS Center for Performance Assessment Report No. MS 94-05. Princeton, NJ: Educational Testing Service.
Myford C.M. & Sims-Gunzenhauser, A. (2004). The evolution of large-scale assessment programs in the visual arts. In E. W. Eisner & M. D. Day (Eds.). Handbook of research and policy in art education (pp. 637-666). Mahwah, NJ: Lawrence Erlbaum Associates.
Myford C.M. & Wolfe, E.W. (2000). Monitoring sources of variability within the Test of Spoken English assessment system (TOEFL Research Report No. 65). Princeton, NJ: TOEFL Research Program, Educational Testing Service.
Myford C.M. & Wolfe, E.W. (2000). Strengthening the ties that bind: Improving the linking network in sparsely connected rating designs (TOEFL Technical Report No. 15). Princeton, NJ: TOEFL Research Program, Educational Testing Service. www.ets.org/Media/Research/pdf/RR-00-09.pdf
Myford C.M. & Wolfe, E.W. (2002). When raters disagree, then what: Examining a third-rater discrepancy resolution procedure and its utility for identifying unusual patterns of ratings. Journal of Applied Measurement, 3, 300-324.
Myford C.M. & Wolfe, E.W. (2003) Detecting and Measuring Rater Effects using Many-Facet Rasch Measurement: Part I. Journal of Applied Measurement, 4(4), 386-421 and in E. V. Smith, Jr. & R. M. Smith (Eds.), Introduction to Rasch measurement (pp. 460-517). Maple Grove, MN: JAM Press.

Myford C.M. & Wolfe, E.W. (2004) Detecting and Measuring Rater Effects using Many-Facet Rasch Measurement: Part II. Journal of Applied Measurement, 5(2), 189-227 and in E. V. Smith, Jr. & R. M. Smith (Eds.), Introduction to Rasch measurement (pp. 518-574). Maple Grove, MN: JAM Press.
Myford C.M., Marr D., Linacre J.M. (1996) Reader Calibrations and Its Potential Role in Equating for the Test of Written English. TOEFL Research Report RR-52. (ETS Center for Performance Assessment Report No. MS 95-02). Princeton, NJ: Educational Testing Service.
Nakamura Y, (2003) Oral proficiency assessment: Dialogue test and multilogue test. JALT Pan-SIG Conference Proceedings. www.jalt.org/pansig/2003/HTML/Nakamura.htm
Nilsson, I, Löfgren, B, Fisher, A., Bernspång, B. (2005) Focus on leisure repertoire in the oldest old, the Umeå 85+ study.
Nilsson, I., & Fisher, A.G. (2006) Evaluation of leisure activities in the oldest old. Manuscript submitted for publication. Scandinavian Journal of Occupational Therapy.
North B. & Schneider G. (1998). Scaling descriptors for language proficiency scales. Language Testing, 15, 217-263.
North B. (2000). The development of a common framework scale of language proficiency. New York: Peter Lang.
North, B. (2008). The CEFR levels and descriptive scales. In L. Taylor & C. J. Weir (Eds.), Multilingualism and assessment: Achieving transparency, assuring quality, sustaining diversity – Proceedings of the ALTE Berlin Conference May 2005 (pp. 21–66). Cambridge, UK: Cambridge University Press.
North, B., & Jones, N. (2009). Further material on maintaining standards across languages, contexts and administrations by exploiting teacher judgment and IRT scaling. Strasbourg: Language Policy Division.
Nygård, L., Bernspång, B., Fisher, A.G., Winblad, B. (1994) Comparing motor and process ability of persons with suspected dementia in home and clinic settings. American Journal of Occupational Therapy, 48, 689-696.
O'Loughlin K. (2000) The impact of gender in the IELTS oral interview. IELTS Research 3, 1
O'Neill, T.R., & Lunz, M.E. (2000). A method to study rater severity across several administrations. In M. Wilson & G. Engelhard, Jr. (Eds.), Objective Measurement: Theory into Practice (Vol. 5, pp. 135-146). Stamford, CT: Ablex.
O'Sullivan B. (2002) Investigating variability in a test of second language writing ability. Cambridge Esol Research Notes, 7, 14-17
Oakley F., Duran L., Fisher A.G., Merritt, B. (2003) Differences in activities of daily living motor skills of persons with and without Alzheimer's disease. Australian Occupational Therapy Journal, 50, 2, 72-78
Oakley F., Khin N.A., Parks R., Bauer L., Sunderland T. (2002) Improvement in activities of daily living in elderly following treatment for post-bereavement depression. Source: Acta Psychiatrica Scandinavica, 105, 3, 231-234
Oakley F., Sunderland T. (1997) Assessment of Motor and Process Skills as a Measure of IADL Functioning in Pharmacologic Studies of People With Alzheimer's Disease: A Pilot Study. International Psychogeriatrics, 9: 197-206
Pan, A.W., & Fisher, A.G. (1994) The Assessment of Motor and Process Skills of persons with psychiatric disorders. American Journal of Occupational Therapy, 48, 775-780.
Pape T. L.-B., Heinemann A.W., Kelly J.P., Hurder A.G. (2005) A measure of neurobehavioral functioning after coma. Part I: Theory, reliability, and validity of the Disorders of Consciousness Scale. Journal of Rehabilitation Research and Development, 42, 1, 1-18.
Park, S., Fisher, A.G., Velozo, C.A. (1994) Using the Assessment of Motor and Process Skills to compare occupational performance between clinic and home settings. American Journal of Occupational Therapy, 48, 697-709.
Park, T. (2004) An investigation of an ESL placement test using Many-Facet Rasch Measurement. Teachers College, Columbia University Working Papers in TESOL and Applied Linguistics, 4, 1.
Parra-López, E. & Oreja-Rodríguez J.R. (2014) Evaluation of the competitiveness of tourism zones of an island destination: An application of a Many-Facet Rasch Model (MFRM). Journal of Destination Marketing and Management, 3(2), 114-121
Patomella A.-H., Caneman G., Kottorp A., Tham K. (2004) Identifying scale and person response validity of a new assessment of driving ability.Scandinavian Journal of Occupational Therapy, 11, 2, 70-77.
Paulukonis, S.T., Myford, C.M., & Heller, J.I. (2000). Formative evaluation of a performance assessment scoring system. In M. Wilson & G. Engelhard, Jr. (Eds.) Objective measurement: Theory into practice (Vol. 5, pp. 15-40). Stamford, CT: Ablex Publishing.
Peeters M.J., Churchwell M.D., Mauro L.S., Cappelletty D.M., Stone G.E. (2010). A student-inclusive pharmacotherapeutic textbook selection process. Currents in Pharmacy Teaching and Learning, 2(1), 31-38.

Peeters M.J., Sahloff E.G., Stone G.E. (2010). A standardized rubric for student presentations. American Journal of Pharmaceutical Education, 74(9), article 171.
Peeters M.J., Serres M.L., Gundrum T.E. (2014). Improving reliability of a resident interview process. American Journal of Pharmaceutical Education, 77(8), article 168.
Pollitt A. & Elliott G. (2003) Monitoring and investigating comparability: a proper role for human judgement. UCLES.
Pollitt, A. (1997). Rasch measurement in latent trait models. In C. Clapham, & D. Corson (Eds.), Encyclopedia of language and education Vol. 7: Language testing and assessment (pp. 243-253). Netherlands: Kluwer Academic.
Pollitt, A., & Hutchinson, C. (1987). Calibrated graded assessment: Rasch partial credit analysis of performance in writing. Language Testing, 4, 72-92.
Pomplun, M. & Custer, M. (2004) The Equivalence of Three Data Collection Methods with Field Test Data: A FACETS Application, Journal of Applied Measurement 5:3, 319-327
Puderbaugh, J.K., Fisher, A.G. (1992) Assessment of motor and process skills in normal young children and children with dyspraxia. Occupational Therapy Journal of Research, 12, 195-216.
Ravaud, J.-F., Delcey M., Yelnik A. (1999) Construct validition of the Functional Independence Measure (FIM): questioning the unidimensionality of the scale and the "value" of FIL scores. Scandinavian Journal of Rehabilitation Medicine, 31, 1, 31-41
Rehfeldt, T.K. (1994) Ranks in sensory measurement. Rasch Measurement Transactions 8:2, 368.
Rexroth, P., Fisher, A.G., Merritt, B.K., Gliner, J. (2004) Ability differences in persons with unilateral hemispheric stroke. Manuscript submitted for publication.
Robinson, S.E., & Fisher, A.G. (1999) Functional and cognitive differences between cognitively-well people and people with dementia. British Journal of Occupational Therapy, 62, 466-471.
Révész, A. (2009). Task complexity, focus on form, and second language development. Studies in Second Language Acquisition, 31(3), 437-470.
Sampson S.O. & Bradley K.D. (2003) Rasch Analysis of Educator Supply and Demand Rating Scale Data. University of Kentucky.
Schatz, R., Belloto R.J.Jr., White D.B., Bachmann, K. (2003) Provision of Drug Information to Patients by Pharmacists: The Impact of the Omnibus Budget Reconciliation Act of 1990 a Decade Later. American Journal of Therapeutics. 10(2):93-103
Schulman J.A., Trujillo M.J., Karney B.R. (2001) Electronic Notes: Facets: Computer Software for Evaluating Assessment Tools. American Journal of Health Behavior, 25, 1, 25-77
Schumacker, R.E. (1999) Many-facet Rasch Analysis with Crossed, Nested, and Mixed Designs. Journal of Outcome Measurement 3:4. 323-338.
Schumacker, R.E. & Lunz M.E. (1997) Interpreting the Chi-Squared Statistics Reported in the Many-Faceted Rasch Model. Journal of Outcome Measurement 1:3. 239-257.
Sellers, S.W., Fisher, A.G., & Duran, L. (2001) Validity of the Assessment of Motor and Process Skills with students who are visually impaired. Journal of Visual Impairment and Blindness, 95, 164-167.
Sluijmans D. & Moerkerke G. (1999) Student involvement in performance assessment: A research project. European Journal of Open and Distance Learning.
Smith E.V.Jr. & Kulikowich J.M. (2004) An Application of Generalizability Theory and Many-Facet Rasch Measurement Using a Complex Problem-Solving Skills Assessment. Educational and Psychological Measurement, 64, 4, 617-639
Stahl J.A. & Lunz, M.E. (1991) Answering the "Call for a New Psychometrics". Rasch Measurement Transactions 5:1, 127
Stahl J.A. & Lunz, M.E. (1996) Judge Performance Reports: Media and Message. In G. Engelhard, Jr., & M. Wilson (Eds.), Objective Measurement: Theory into Practice (Vol. 3, pp. 113-126). Norwood, NJ: Ablex.
Stahl J.A. (1994) What Does Generalizability Theory offer that Many-Facet Rasch Measurement cannot duplicate? Rasch Measurement Transactions, 8:1, 342-3.
Stahl J.A., Shumway, R., Bergstrom, B., & Fisher, A.G. (1997) On-line performance assessment using rating scales. Journal of Outcome Measurement, 1, 173-191.
Stauffer, L.M., Fisher, A.G., & Duran, L. (2000) ADL performance of black and white Americans on the Assessment of Motor and Process Skills. American Journal Occupational Therapy, 54, 607-613.
Stemler, S. (2004) A Comparison of Consensus, Consistency, and Measurement Approaches to Estimating Interrater Reliability. Practical Assessment, Research & Evaluation, 9(4).
Tatum, D. (2003) Assessing E-Portfolios. National Communication Association. www.eport2passport.com/presentations/nca_2003/2003NCAPaperEportfolio.doc

Taylor, L. & Jones N. (2001) Revising the IELTS Speaking Test. Research Notes of the University of Cambridge Local Examination Syndicate English as a Second Language. 4, 9-11. Feb.
Tham K., Bernspaang B., Fisher A.G. (1999) Development of the assessment of awareness of disability. Scandinavian Journal of Occupational Therapy, 6, 184 - 190
Tham K., Ginsburg, E., Fisher, A.G., & Tegnér, R. (2001) Training to improve awareness of disabilities in clients with unilateral neglect. American Journal of Occupational Therapy, 55, 46-54.
Trace J., Janssen G., Meier V. (2105) Measuring the impact of rater negotiation in writing performance assessment. Language Testing, July 28, 2015
Twing J.S., Nichols P.D., Harrison I. (2003) The comparability of paper-based and image-based marking of a high-stakes, large-scale writing assessment. International Association for Educational Assessment.
Tyndall, B., & Kenyon, D. M. (1995). Validation of a new holistic rating scale using Rasch multifaceted analysis. In A. Cumming & Berwick (Eds), Validation in language testing (pp. 39-57). Clevedon, England: Multilingual Matters.
Upshur, J. A., & Turner, C. E. (1999). Systematic effects in the rating of second-language speaking ability: Test method and learner discourse. Language Testing, 16, 82-111.
Uto, M. (2021). A multidimensional generalized many-facet Rasch model for rubric-based performance assessment. Behaviormetrika, 48(2), 425-457. doi.org/10.1007/s41237-021-00144-w
Vianello, M., & Robusto, E. (2010). The Rasch Models in the Analysis of the Go/No Go Association Task. Behavior Research Methods, 42(4), 944-956. sites.google.com/site/michelangelovianello/rcv2008-pdf-1/VR.2010.pdf
Wang J.C., Chi C.C., Lian W.Y. (2005) The Application of Rasch Poisson counts model to construct norm-reference testing on the Badminton Skill Tests in Primary schools. Journal of Physical Education and Sports, 16, 1, 73-83.
Wang N. (2003) Examining reliability and validity of job analysis survey data. Journal of Applied Measurement, 4, 4, 358-369
Wang N., Schnipke D., Witt E.A. (2005) Use of Knowledge, Skill, and Ability Statements in Developing Licensure and Certification Examinations. Educational Measurement: Issues and Practice, 24, 1, 15
Wang W.-C., Cheng Y.-Y. (1999) A Multi-Facet Rasch Analysis of the College Teacher Evaluation Inventory. Paper presented at AARE.
Watson J., Callingham R. (2003) Statistical literacy: A complex hierarchical construct. Statistics Education Research Journal, 2, 2, 3-46.
Weigle S. (1998) Using FACETS to model rater training effects. Language Testing 15: 263-287.
Wigglesworth G. (1993) Exploring bias analysis as a tool for improving rater consistency in assessing oral interaction. Language Testing 10: 305-335.
Wigglesworth G. (1994) The investigation of rater and task variability using multi-faceted measurement. Report for the National Centre for English Language Teaching and Research, Macquarie University.
Williams E.J. (1999) Developmental Reading Assessment: Reliability Study. Pearson Learning Group.
Wilson M, & Wang W. (1995) Complex composites: Issues that arise in combining different modes of assessment. Applied Psychological Measurement, 19, 51-71.
Wilson M.R. & Case H. (2000) An examination of variation in rater severity over time: a study of rater drift. In M. Wilson and G. Engelhard, Jr. (Eds.) Objective Measurement: Theory into practice: Vol. 5 (pp. 113-134). Stamford CT: Ablex.
Winke P., Gass S., Myford C. (2013) Raters' L2 background as a potential source of bias in rating oral performance. Language Testing 30 231-252.
Wolfe E.W. (1997). The Relationship between essay reading style and scoring proficiency in a psychometric scoring system. Assessing Writing, 4, 83-106.
Wolfe E.W. (2004) Identifying rater effects using latent trait models. Psychology Science, 46, 1, 35-51.
Wolfe E.W., & Chiu C. W. T. (1997) Detecting rater effects with a multi-faceted rating scale model. East Lansing MI: National Center for Research on Teacher Learning. ERIC ED408 324.
Wolfe E.W., & Dobria, L. (2008). Applications of the multifaceted Rasch model. In J. W. Osborne (Ed.), Best practices in quantitative methods (pp. 71–85). Los Angeles: Sage.
Wolfe E.W., & Gitomer, D. (2001). The influence of changes in assessment design on the psychometric quality of scores. Applied Measurement in Education, 14, 91-107.
Wolfe E.W., Chiu, W. T., & Myford, C. M. (1999). The manifestation of common rater effects in multi-faceted Rasch analyses (RR-97-02). Princeton, NJ: Educational Testing Service, Center for Performance Assessment.
Wolfe E.W., Chiu, W. T., & Myford, C. M. (2000). Detecting rater effects in simulated data with a multi-faceted Rasch rating scale model. In M. Wilson & G. Engelhard, Jr. (Eds.), Objective measurement: Theory into practice (Vol. 5, pp. 147-164). Stamford, CT: Ablex Publishing Co.

Wolfe E.W., Kao, C.W., & Ranney, M. (1998). Cognitive differences in proficient and nonproficient essay scorers. Written Communication, 15, 465-492.
Wolfe E.W., Moulder, B.C., Myford, C. (2001) Detecting Differential Rater Functioning over Time (DRIFT) Using the Rasch Multi-faceted Rating Scale Model. Journal of Applied Measurement 2:3, 256-280.
Wolfe E.W., Myford, C. M., Engelhard, G. E., & Manalo, J. R. (2007). Monitoring reader performance and DRIFT in the AP English Literature and Composition examination using benchmark essays. (College Board Research Report No. 2007-2). New York: The College Board.
Wright B.D. (1993) Discrete-time survival analysis. Rasch Measurement Transactions 7:3, 307
Yamauchi K. (1999) Comparing Many-facet Rasch Model and ANOVA model: Analysis of ratings of essays [in Japanese]. Japanese Journal of Educational Psychology. Vol 47(3), Sep., 383-392.
Zhang Y., Elder C. (2010) Judgments of oral proficiency by non-native and native English speaking teacher raters: Competing or complementary constructs? Language Testing 28(1) 31-50
Zhu W., Cole, E. L. (1996). Many-faceted Rasch calibration of a gross motor instrument. Research Quarterly for Exercise and Sport, 67(1), 24-34.
Zhu W., Ennis C.D., Chen A. (1998) Many-faceted Rasch modeling experts' judgement in test development. Measurement in Physical Education and Exercise Science, 221-39.
Zhu W., Updyke, W.F, & Lewandowski, C. (1997). Post-Hoc Rasch Analysis of Optimal Categorization of an Ordered-Response Scale. Journal of Outcome Measurement, 1(4), 286-304.

*Useful background:*

Andrich D. (1978) A rating scale formulation for ordered response categories. Psychometrika, 43, 561-573
Edgeworth F.Y. (1890) The element of chance in competitive examinations. Journal of the Royal Statistical Society 53, 460-75 and 644-63.
Lumley, T. (2005). Assessing second language writing: The rater's perspective. Frankfurt, Germany: Lang.
Masters G.N. (1982) A Rasch model for partial credit scoring. Psychometrika 47, 149-174.
Saal, F.E., Downey, R.G. and Lahey, M.A (1980) Rating the Ratings: Assessing the Psychometric Quality of Rating Data, Psychological Bulletin, 88(2), 413-428.

# 5 Running Facets - Tutorial

## 5.1 Facets Tutorial PDFs

Here are four free *Facets* tutorials in PDF format. They make it easier to learn about using *Facets*. Please download them.

ftutorial1.pdf
ftutorial2.pdf
ftutorial3.pdf
ftutorial4.pdf

***Example of a Tutorial page:***

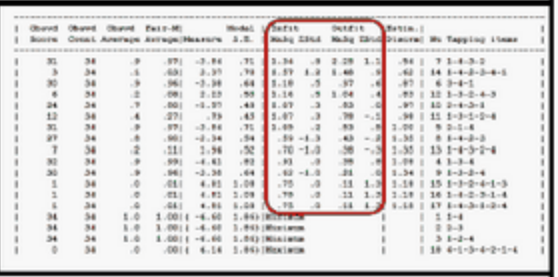
| | |
|---|---|
| **43.** | **Table 4: Unexpected Responses** |
| **44.** | We could go through this computation by hand for every observation, but it is easier to have *Facets* do it for us. Click on the *Facets* Report output file on your Windows Taskbar<br>(or click on *Facets* "Edit" menu, click on "Report output file")<br>Scroll down to Table 4. It is the last Table. It shows the unexpected responses (or unexpected observations) |
| **45.** | *Facets* has done the computation for Child 2 on Item 14 more precisely than I did. It reports the standardized residual ($Z_{ni}$, StRes) is 6.1. This is the most unexpected observation in these data. The observation is unexpectedly high (1) compared with its expected value (.0) |

'Knox Cube Test (Best Test Design p.31)' 04-15-2007 22:20
Table 4.1 Unexpected Responses (7 residuals sorted by u).

| Cat | Step | Exp. | Resd | StRes | Nu Chil | Nu Tapping items |
|---|---|---|---|---|---|---|
| 1 | 1 | .0 | 1.0 | 6.1 | 2 Boy 14 | 1-4-2-3-4-1 |
| 0 | 0 | 1.0 | -1.0 | -6.0 | 2 Boy 7 | 1-4-3-2 |
| 0 | 0 | 1.0 | -1.0 | -6.0 | 24 Girl 7 | 1-4-3-2 |
| 0 | 0 | 1.0 | -1.0 | -4.8 | 24 Girl 6 | 3-4-1 |
| 0 | 0 | .9 | -.9 | -3.5 | 20 Girl 5 | 2-1-4 |
| 1 | 1 | .1 | .9 | 3.5 | 24 Girl 12 | 1-3-2-4-3 |
| 1 | 1 | .1 | .9 | 3.5 | 33 Girl 12 | 1-3-2-4-3 |

| Cat | Step | Exp. | Resd | StRes | Nu Chil | Nu Tapping items |
|---|---|---|---|---|---|---|

| | |
|---|---|
| **46.** | Look at the next two observations listed. Both are on item 7 and they are equally unexpected StRes = -6.0. The minus - sign means "they did worse than we expected." Both children, 2 and 24, failed on the item when we expected them so succeed. We don't know why, but if we were serious about the children or the instrument we might inquire. The cubes are sequential 4-3-2. Perhaps the examiner sped up unintentionally, or perhaps he didn't clearly tap each cube so the children missed tapping "3".<br>The list of unexpected responses nearly always contains useful messages about the instrument, the sample, the judges, the dataset, or whatever ..... |
| **47.** | **Table 7: Quality-control fit statistics elements and observations** |
| **48.** | Looking down the list of unexpected responses is somewhat like looking at the needles in a haystack. You want to pay some attention to them (not too much, usually), but they don't tell you much about the haystack. For that we need to take a wider look. |
| **49.** | Scroll back up the Report Output file to Table 7.2.2. - the measure Table for Items in fit order, descending. You will see 4 columns: Infit and Outfit, MnSq and Zstd. These are quality-control fit statistics. They are central to the evaluation of the quality of the data for the construction of measures. |



| | |
|---|---|
| **50.** | How well do the observations of each element fit with the estimate of the parameter measure? If the fit is good, then we can have confidence that the measure means what it says. If the fit is bad, then the measure could mislead us. If the fit is too good, then perhaps something is constraining the data to be too coherent. |
| **51.** | So we have the process of fit evaluation. In *Facets*, most fit statistics are based on summarizing the residuals that we've already thought about one at a time.<br>The principles of fit are easier to explain with dichotomous that with polytomous data, so that is why we are starting here. |

Data → Parameter Estimates →
Expected Data → Residuals →
Fit statistics →
Validity of Parameter Estimates

## 5.2 Launching Facets under Windows

**To Run Facets:**

**On your Desktop, double-click on the Facets icon:**


**Or** Click on **Start** button
Move mouse pointer to **Programs**
Point to ***Facets*** folder
Click on ***Facets*** icon

Or **Drag** a Facets specification file onto the Facets icon.

or Add Facets to the Windows Sendto menu, and right-click on your control file.
To add Facets 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 Facets into the Send To folder.
3. To run an analysis, right click on the control file to open the Send To menu, select Facets.

---

**Super-Fast-Track:**
Click on Files menu
Click on Specification File Name?
Double-click on Essays.txt

Extra specifications?
Click on **OK**

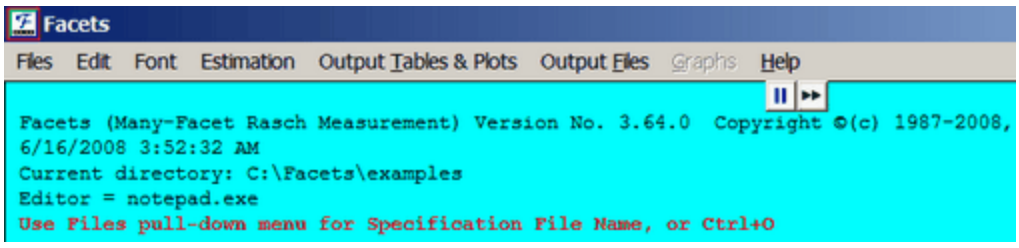
The program will ask you for the name of your analysis output file with
What is the Report Output file name?
Essays.out.txt displays.
**Click on Open**

The analysis will commence and run to its conclusion.
The Report Output file automatically displays.
Other output can be obtained with the Output Tables menu, Output Files menu and Graphs menu.

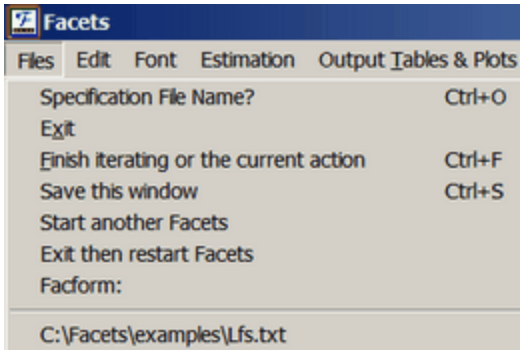
**Other icons in the Start Menu:**
The installation procedure installs icons for Facets, the Facform data formatter for Facets, the User Manuals and Help files.

## 5.3 Facets Window

Facets is a Windows-native program with many of the usual Windows menu functions, and plenty of added features.

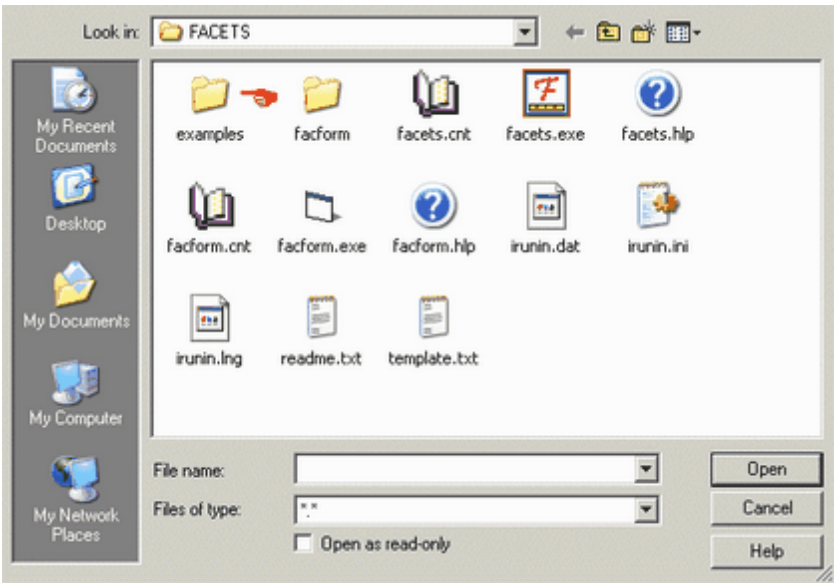


Click on the "Files" pull-down menu:

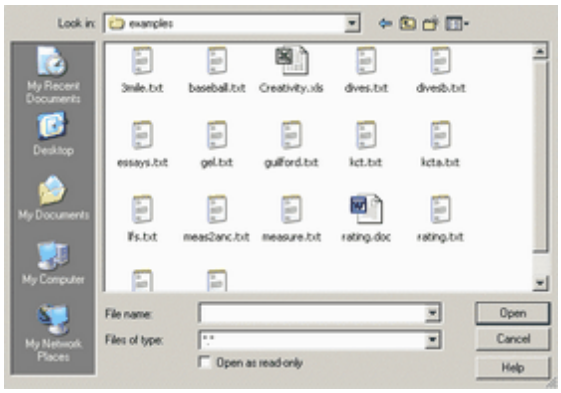



Select **Specification File Name?** or click on a file name at the bottom of the menu.

**The Opening Screen with Specification File Name?**




Double click on "Examples" (if not already there)
Clicking on "Cancel" takes you back to "Files" pull-down menu.




Click on the Facets specification file you prefer, e.g., essays.txt. Then click on "**Open**" or press the enter key.

Not sure what a Specification file looks like? click here

Want a quick way to look at file contents? click here

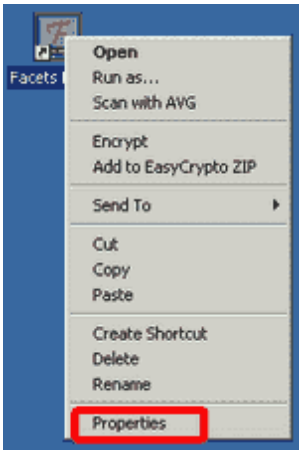
---

**The Start-In Folder**

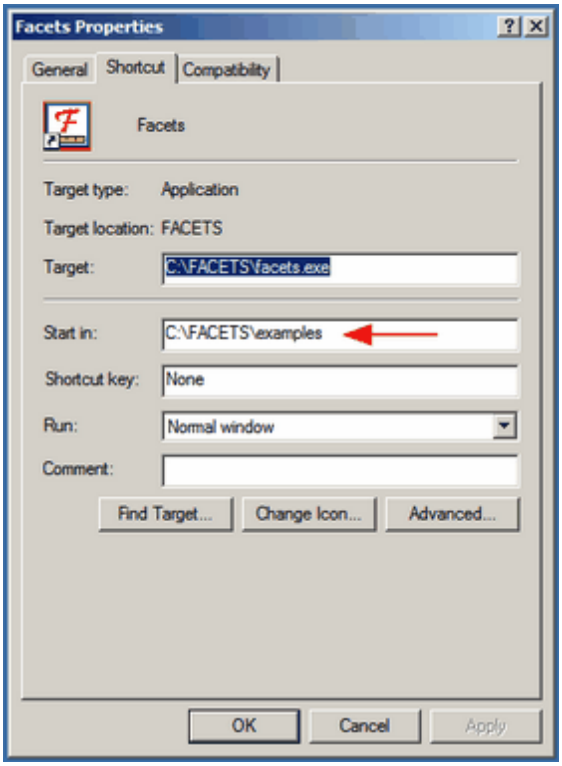
**When launched from the Desktop short-cut or Start menu, *Facets*** starts by displaying its standard folder, c:\Facets\examples. If you prefer Facets to display different folder, the easiest way is to change the "Start in" folder in the Facets short-cut on your desktop.

*Right-click* on

and you will see:

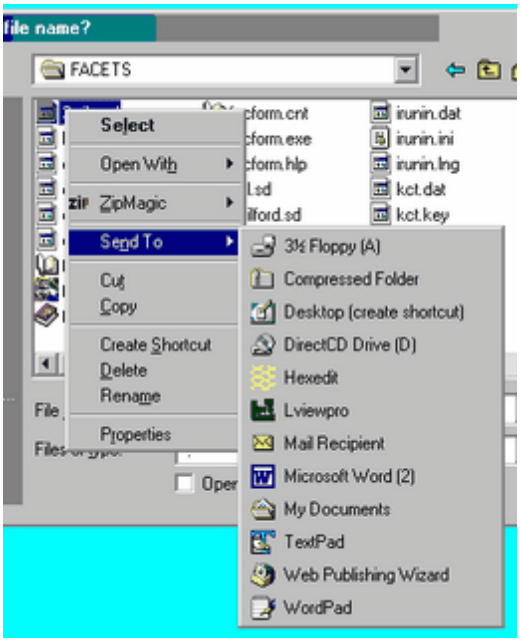



Click on *Properties,* and then change the "Start in" path to the one you prefer:



## 5.4 File Box - Right-click

**Mouse Right-Click in the File Dialog Box:**
Right-click a file name, and get the **Send To** menu. Click on NotePad to view the contents of a file.




---

**If NotePad is not in your SendTo list:**

Add functionality to the **Send To** menu by copying short-cuts into **c:\windows\sendto** (or the equivalent in your version of Windows) - a useful program to add is **NotePad.**

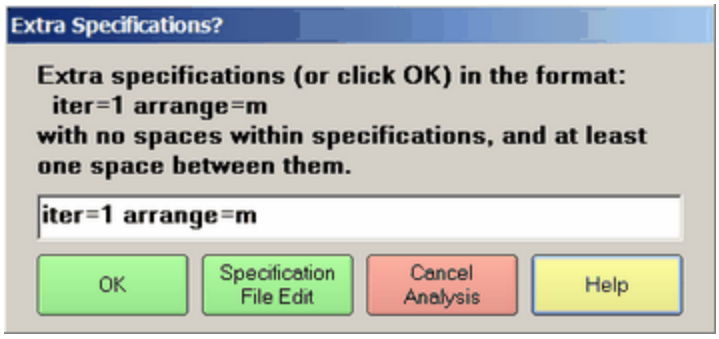
To do this:

Start
Find
Files or Folders
Named: **NotePad** in **C:**
when NotePad.exe appears, Right-click on it.
Send To: Desktop: Create shortcut
Exit from Find
On Desktop:
My Computer
Click on words "My Computer" to highlight them, and type:
C:\windows\sendto
The "SendTo" folder displays
Drag **NotePad shortcut** from Desktop into SendTo folder.
NotePad is now in the Send To menu.

## 5.5 Extra Specifications

After accepting the specification file name, *Facets* asks if you want to include "Extra Specifications"?

For most analyses, type nothing. Merely click "OK".




This figure shows an example of two other specifications "iter=1" and "arrange=m". These will be processed along with the specification file, e.g., essays.txt, and take precedence over any similar specifications in it.

No spaces within specifications. If the text is red, then spaces have been found.

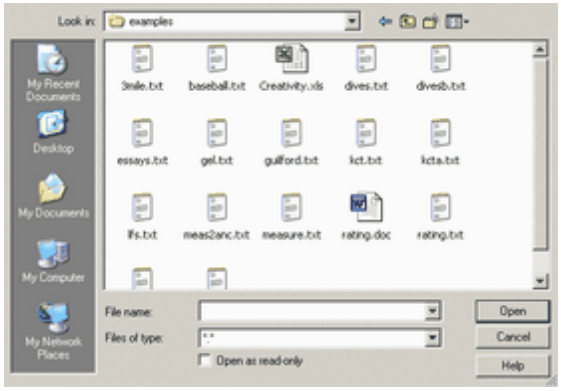
| | |
|---|---|
| **(Specification box)** | type in Specification=value with no internal blanks.<br>Multiple *Specification=value Specification=value* must be separated by a blank. |
| **OK** | accepts the extra specifications for processing and repeats them in the Facets window. |
| **Specification File Edit** | opens the specification file in a NotePad Edit window so it can be viewed or edited |
| **Cancel Analysis** | this Facets analysis is canceled |
| **Help** | displays this Help page |

## 5.6 Report Output File

*Facets* always produces a Report Output File into which are written at least the 8 standard Tables.

This name of this file can be specified in the Specification File using Output file=, or in "Extra Specifications?" in the same way.

If no Output File has been specified, then Facets asks:

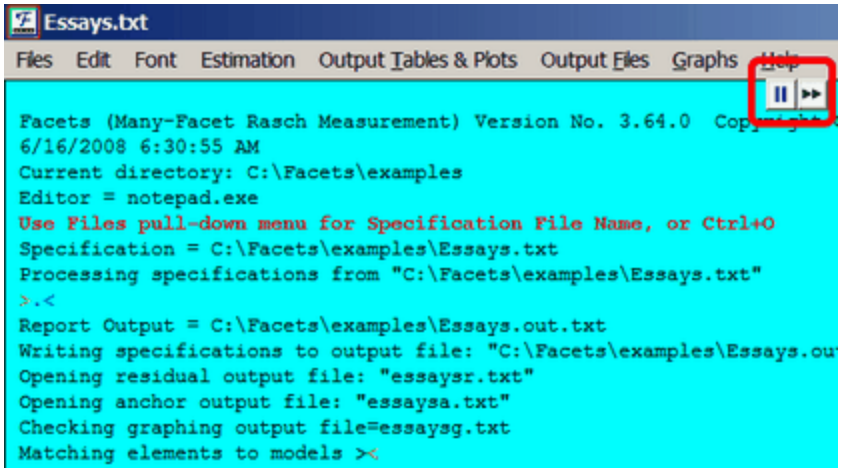



Select a file, then click on "open" or press Enter to accept the standard name, such as "essays.out.txt", or type in your own preference and click on "Open" or press Enter.

You can bypass this screen, and have Facets automatically accept the standard file name, see Edit Initial Settings - "Prompt for output file name?"

## 5.7 Estimation Progess

Once Facets has the needed information, it proceeds with constructing measures:




The first part reports which files are being read and written.

are the quick-action buttons.

>.....< is the **iteration bar** which indicates progress in processing the specification file.

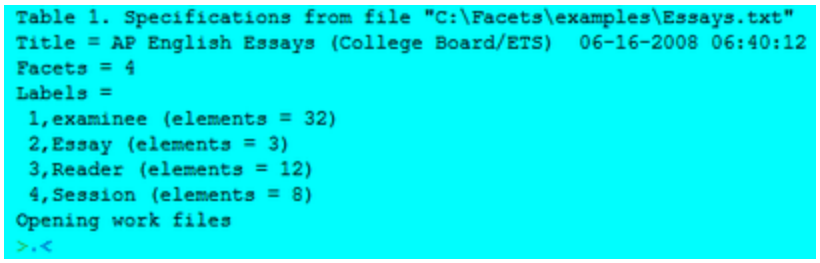



Table 1 summarizes key aspects of the specifications.

How many facets? - The number of different facets that combine together to produce the observations.

How many elements? - The number of different elements found in each facet.

**Compare these numbers with what you expect in order to verify that the specifications are correct.**

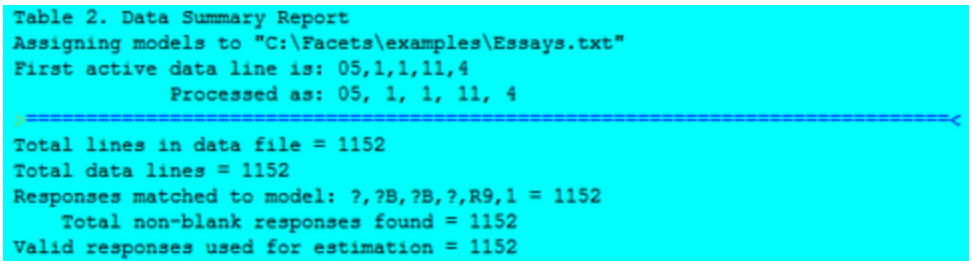



Table 2 reports on the data. Here there are 1152 observations, and all have been matched to the specified labels and the specified measurement model.

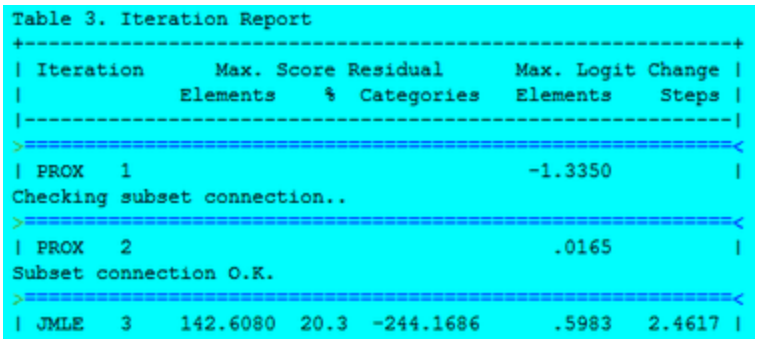
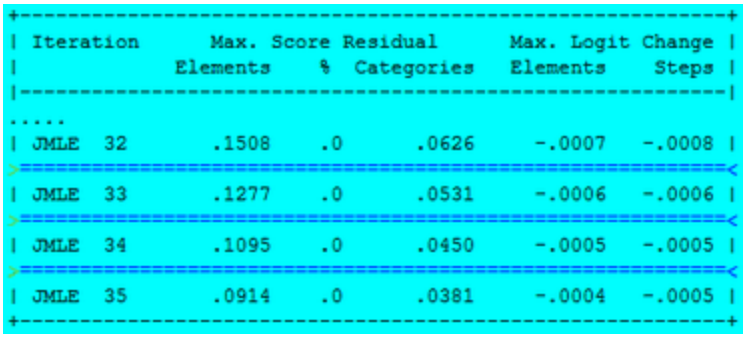


Table 3. Iteration Report

| Iteration | Max. Score Residual Elements | % | Categories | Max. Logit Change Elements | Steps |
|---|---|---|---|---|---|
| PROX 1 | | | | -1.3350 | |
| Checking subset connection.. | | | | | |
| PROX 2 | | | | .0165 | |
| Subset connection O.K. | | | | | |
| JMLE 3 | 142.6080 | 20.3 | -244.1686 | .5983 | 2.4617 |

Table 3 reports on the estimation of measures. The first estimation method used is PROX, the Normal Approximation algorithm, see "Rating Scale Analysis". The important consideration is that the "max.=" amount gets closer to zero.

At the same time, the data is checked for connectedness. Here "Subset connection O.K." is reported, so there is complete connectedness. If there is not, see Connectedness.



| Iteration | Max. Score Residual Elements | % | Categories | Max. Logit Change Elements | Steps |
|---|---|---|---|---|---|
| ..... | | | | | |
| JMLE 32 | .1508 | .0 | .0626 | -.0007 | -.0008 |
| JMLE 33 | .1277 | .0 | .0531 | -.0006 | -.0006 |
| JMLE 34 | .1095 | .0 | .0450 | -.0005 | -.0005 |
| JMLE 35 | .0914 | .0 | .0381 | -.0004 | -.0005 |

Estimation continues iteratively with JMLE, "joint maximum likelihood estimation", also known as UCON , "unconditional maximum likelihood", algorithm. This includes the *Facets* implementation of PMLE. "max. = .0914" at the bottom of this Figure means that the worst estimated measure predicts a raw score only .09 score points away from that observed. "max = -.0004" means that the biggest change in an estimate during this iteration is only .0004 logits. Since logits are only reported to 2 decimal places, this change is really meaningless. In this analysis, the convergence criteria appear to have been set more tightly than necessary, (but this is typical with high stakes examinations.) See also "My analysis does not converge."

When the convergence criteria are satisfied, estimation ceases. Or, to stop estimation more quickly, press Ctrl+F or select "Finish Iterating" on the Files pull-down menu.

## 5.8 Report Output

After measure estimation is completed, the quality-control fit statistics are computed, and the measures are reported. Table 4 is deliberately reported out of order because it is often lengthy.

```
Calculating fit statistics
>============================================================<
>.<
Computing inter-rater agreement:
>.<
Table 4. Unexpected Responses - appears after Table 8
Table 5. Measurable Data Summary
Writing anchor file: essaysa.txt
Computing fit summary statistics..
Writing score file for facet 1 to C:\Facets\examples\essayss1.txt
Writing score file for facet 2 to C:\Facets\examples\essayss2.txt
Writing score file for facet 3 to C:\Facets\examples\essayss3.txt
Writing score file for facet 4 to C:\Facets\examples\essayss4.txt
Writing graph output file ...
Writing residual output file ...
>============================================================<
Table 6.0  All Facet Vertical "Rulers"
>.<
Table 6.1  examinee Facet Summary
Table 6.2  Essay Facet Summary
Table 6.3  Reader Facet Summary
Table 6.4  Session Facet Summary
Table 7.1.1  examinee Measurement Report  (arranged by MN)
>.<
Table 7.2.1  Essay Measurement Report  (arranged by MN)
>.<
Table 7.3.1  Reader Measurement Report  (arranged by MN)
>.<
Table 7.4.1  Session Measurement Report  (arranged by MN)
Table 8. Category Statistics
>============================================================<
Table 4.1 Unexpected Responses (2 residuals sorted by 2,3,1)
>.<
```

The final automatic step is the reporting is of bias or interaction between elements, if specified in the specification file.

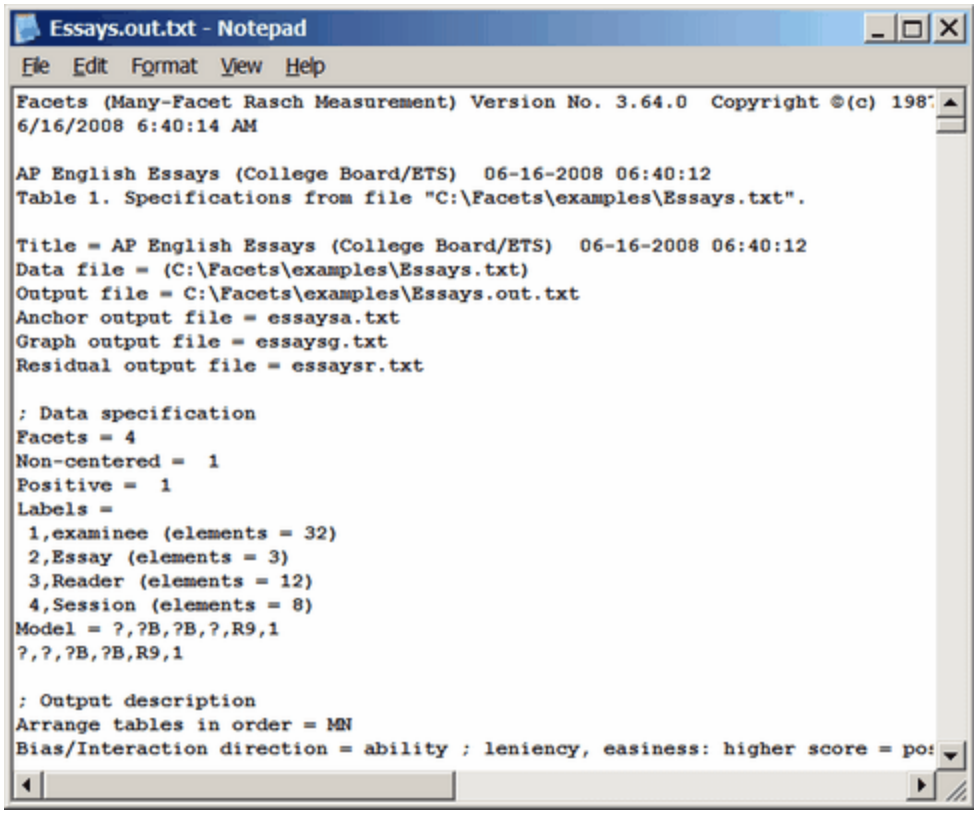
```
Table 9.1  Bias Iteration Report
Bias/Interaction: 2. Essay, 3. Reader
>============================================================<
There are empirically 36 Bias terms
+------------------------------------------------------------+
| Iteration      Max. Score Residual      Max. Logit Change |
|             Elements    %  Categories   Elements    Steps |
|------------------------------------------------------------|
>============================================================<
| BIAS   1     26.4675 -17.9                .5011           |
>============================================================<
| BIAS   2      -.4949   -.5                -.0100           |
>============================================================<
| BIAS   3      -.0006    .0                 .0000           |
+------------------------------------------------------------+
Calculating Bias fit statistics
>============================================================<
Table 11.1  Bias/Interaction Measurement Summary
Table 12.1  Bias/Interaction Summary Report
Table 13.1.1  Bias/Interaction Calibration Report (arranged by MN)
>.<
Press Ctrl+F to stop Table 14 bias/interaction reporting.
Table 14: There are 36 pairwise terms for 2. Essay
Table 14.1.1.2  Bias/Interaction Pairwise Report (arranged by MN)
>============================================================<
>.<
Table 14: There are 234 pairwise terms for 3. Reader
Table 14.1.1.3  Bias/Interaction Pairwise Report (arranged by MN)
>============================================================<
>.<
```

Then the automatic analysis ceases, and Facets waits for your instructions.

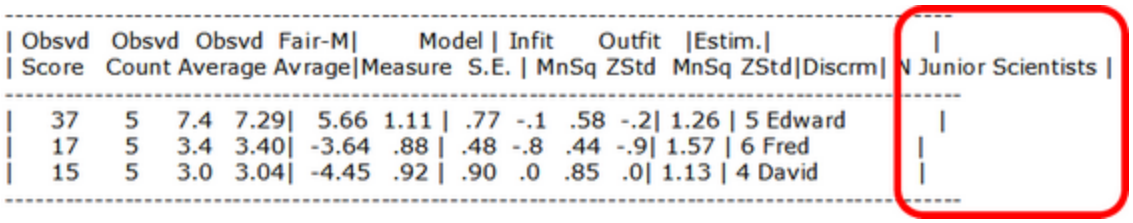
```
Subset connection O.K.
Output to C:\Facets\examples\Essays.out.txt
```

## 5.9 Output Tables

Output Tables are written directly into the report output file, which is displayed using NotePad or WordPad. Scroll down this report file. Examine it. Annotate it. Copy and paste it into your own report. If Tables do not display, check Changing the Text Editor or you may have specified Tables=No.




If your Output Tables display "ragged right" like this:



| Obsvd Score | Obsvd Count | Obsvd Average | Fair-M Avrage | Model Measure | Model S.E. | Infit MnSq | Infit ZStd | Outfit MnSq | Outfit ZStd | Estim. Discrm | N | Junior Scientists |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 37 | 5 | 7.4 | 7.29 | 5.66 | 1.11 | .77 | -.1 | .58 | -.2 | 1.26 | 5 | Edward |
| 17 | 5 | 3.4 | 3.40 | -3.64 | .88 | .48 | -.8 | .44 | -.9 | 1.57 | 6 | Fred |
| 15 | 5 | 3.0 | 3.04 | -4.45 | .92 | .90 | .0 | .85 | .0 | 1.13 | 4 | David |

Then change to a fixed-spaced font using NotePad's "Format" menu. Courier New displays like this:

| Obsvd Score | Obsvd Count | Obsvd Average | Fair-M Avrage | Model Measure | Model S.E. | Infit MnSq | Infit ZStd | Outfit MnSq | Outfit ZStd | Estim. Discrm | N | Junior Scientists |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 37 | 5 | 7.4 | 7.29 | 5.66 | 1.11 | .77 | -.1 | .58 | -.2 | 1.26 | 5 | Edward |
| 17 | 5 | 3.4 | 3.40 | -3.64 | .88 | .48 | -.8 | .44 | -.9 | 1.57 | 6 | Fred |
| 15 | 5 | 3.0 | 3.04 | -4.45 | .92 | .90 | .0 | .85 | .0 | 1.13 | 4 | David |

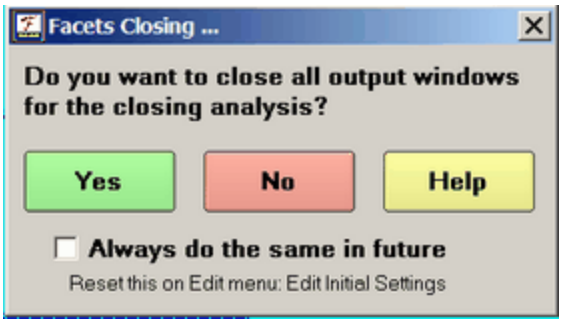
This can also be displayed as a webpage using a line-drawing font, see Ascii=

Descriptions of the output are linked from Output Table Index.

## 5.10 Stopping Facets

**Facets** exits immediately when:
You click on the top right ☒ box.
You select "Exit" on Files drop-down menu.

If there are open output windows when Facets exits, it asks:




**Yes**: the output windows are closed.
**No**: the output windows are left open.
**Help**: displays this page
**Always do the same in future:** make this the default action. It can be changed in Edit Initial Settings.

**Facets** terminates processing immediately when
1) errors are found in the specification file
2) the data do not match the specifications
3) you press **Ctrl+C** to force immediate termination.
Only partial output will be produced, but this may be helpful in diagnosing problems in the analysis.

**Facets** terminates its estimation phase at the end of the current iteration when
4) convergence is reached in estimation
5) the maximum number of iterations have been performed
6) you press **Ctrl+F** to stop the iteration process or press the ?? button.
Full results will be produced. Their value depends on how finely converged the estimates are. Check that the final set of values in the iteration table are all small. You may alter the convergence= criteria in order to obtain a faster, rougher, convergence or a more accurate, slower, convergence to the iterative estimation procedure. See also "My analysis does not converge."

## 5.11 Launching at the DOS prompt

Facets is a Windows-native program, so it will not run in "DOS Mode". But it can be started from the DOS prompt in an MS-DOS window under Windows. Facets can be included in batch files with Batch=Yes.

**To Run Facets:**
Either enter:
**C:> FACETS**
The program will ask you for the name of your specification file with
Specification file name:
then type in the name of the specification file you want, e.g.,
KCT.txt
the program will ask you for the name of your analysis output file with
Output file name:

then press enter or type the name of the output file you want, e.g.,
**KCTOUT.txt**
The analysis will commence and run to its conclusion.

Or enter:

**C:> FACETS specfile**
where specfile is the name of your specification file, e.g., KCT.txt
the program will ask you for the name of your analysis output file with
Output file name:
then type the name of the output file you want, e.g.,
**KCTOUT.txt**
The analysis will commence and run to its conclusion.

Or enter:

**C:> FACETS specfile outputfile**
where specfile is the name of your specification file, e.g., KCT.txt
and outputfile is the name of your analysis output file, e.g., KCTOUT.txt
The analysis will commence and run to its conclusion.

Or, for advanced users, enter:

**C:> FACETS specfile outputfile specification=value specification=value ..**
where
specfile is the name of your specification file,
outputfile (optional) is the name of the file to hold your results,
specification=value are a list of specifications like the specification file.

The specifications entered here take priority over those in the specification file. All specifications are allowed here except Labels=, Model= and Rating (or partial credit) scale=.
Data= must be followed by a file name.

For entries on the DOS command line, blanks are used as separators. Blanks are not allowed within specifications. This is correct:
C:> FACETS KCT.txt ITER=3 TITLE=SHORT-TEST
But, TITLE=SHORT TEST is incorrect due to the blank between SHORT and TEST.

**Example1:**

When doing multiple runs from different data sets with similar format, it can be convenient to set up one standard specification file. In this example, STANDARD.CTL is the name of my Facets specification file. I want to analyze 4 similarly formatted data sets, DATA1, DATA2, DATA3, DATA4. The write their results into 4 output files, DATA1OUT.txt, DATA2OUT.txt, DATA3OUT.txt, DATA4OUT.txt:

```
C:\FACETS> FACETS STANDARD.CTL DATA1OUT.txt DATA=DATA1 TITLE=DATA1
C:\FACETS> FACETS STANDARD.CTL DATA2OUT.txt DATA=DATA2 TITLE=DATA2
C:\FACETS> FACETS STANDARD.CTL DATA3OUT.txt DATA=DATA3 TITLE=DATA3
C:\FACETS> FACETS STANDARD.CTL DATA4OUT.txt DATA=DATA4 TITLE=DATA4
```

**Example 2: Batch-mode processing**

Use Notepad, or such like, to create a ".bat", file, e.g., myrun.bat
This is a text file (with line breaks) containing instructions, such as:

```
START /w FACETS STANDARD.CTL DATA1OUT.txt BATCH=YES DATA=DATA1 TITLE=DATA1
START /w FACETS STANDARD.CTL DATA2OUT.txt BATCH=YES DATA=DATA2 TITLE=DATA2
START /w FACETS STANDARD.CTL DATA3OUT.txt BATCH=YES DATA=DATA3 TITLE=DATA3
START /w FACETS STANDARD.CTL DATA4OUT.txt BATCH=YES DATA=DATA4 TITLE=DATA4
```

Note: **START /w** launches FACETS sequentially, not all at the same time. (If your version of Windows does not understand START /w, then remove it from each line.)

**Batch=Yes** is required to avoid Facets asking for user intervention.

Then launch this ".bat" file at the DOS prompt:

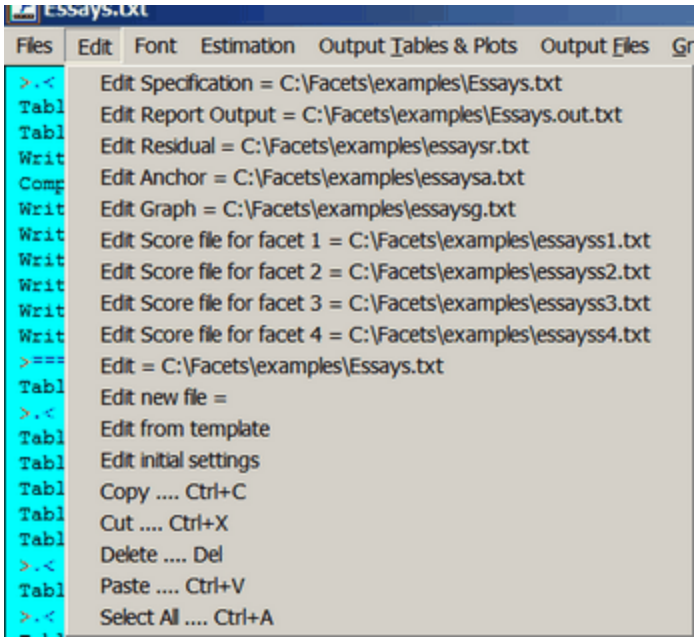
C:\FACETS> myrun (Enter)

Facets will run 4 times, in background (or as much in background as Windows permits.)

# 6 Menus

## 6.1 Edit pull-down menu

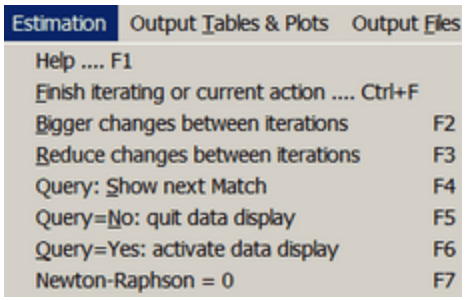
The Edit menu allows you to read and edit all input and output files, as well as previous specification files.

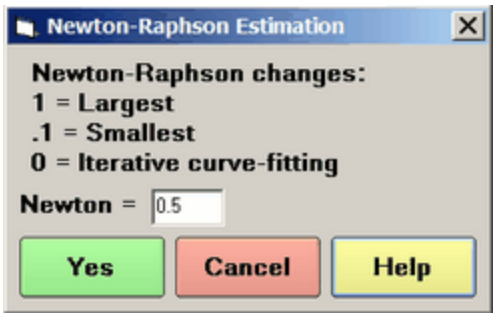



**Edit = filename** launches your text editor to edit the specified file. Editable files are added to this list as they are created.
**Edit new file =** launches your text editor with a blank file
**Edit from template** launches your text editor to edit the generic specification template file provided with Facets
**Edit initial settings** allows you to change your text editor and other initial settings
*Screen editing:*
**Copy from screen** (Ctrl+C keys) copies from the output screen
**Paste to screen** (Ctrl+V keys) writes to the output screen
and other short-cut keys

## 6.2 Estimation pull-down menu

The "Estimation" menu is intended to be accessed during the estimation phase.




| | |
|---|---|
| Help .... F1 | Displays this Help page: F1 is the short-cut key |
| Finish iterating of current action .... Ctrl+F | Completes estimation at the end of the current iteration. If estimation does not seem to be converging, or is converged enough for your purposes. This can also be used to finish some other long processes. |
| Bigger changes between iterations F2 | Increases the size of the changes during estimation. If the maximum residual does not decrease noticeably per iteration, Use the pull-down Estimation menu, and select "Bigger" changes, or press the F2 key. |
| Reduce changes between iterations F2 | Decreases the size of the changes during estimation. If the maximum residual oscillates between large positive and negative values, Use the pull-down Estimation menu, and select "Reduce" changes, or press the F3 key. |
| Query: Show next match F4 | When Query=Yes, this jumps to the next observation to be matched to a Models= specification. |
| Query=No: quit data display F6 | This suppresses reporting of data matching |
| Query=Yes: activate data display F6 | This shows diagnostic information about the matching of data to models. It enables you to check how your data are being matched to your Models= specifications. |
| Newton-Raphson = (value) | The current setting of Newton=. Newton = 0 for iterative curve fitting.  |

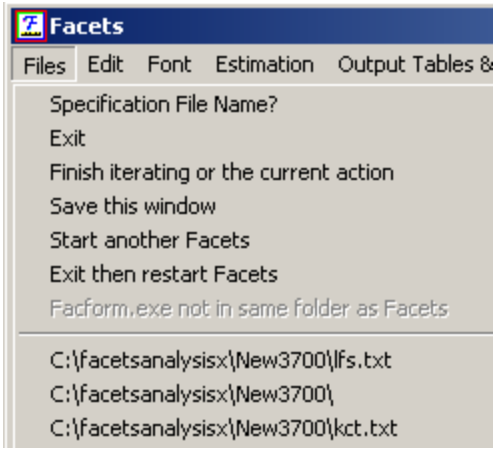


If the "max.=" residual values are large in the Iteration report, but changing only slowly between iterations (as in the example below), click on "**Bigger**" F2. If they are large, and jumping about in value, click on "**Reduce**" F3.

```
Iteration 10  Largest Residuals  |-------Score---------|   logit Measure
Facet     Name          No./Cat   Raw  Expected Residual  Value    Change
>=========================================================================<
.....
Iteration   10  JMLE                        max.=   1045.0     max.=   .018
.....
Iteration   11  JMLE                        max.=   1043.0     max.=   .016
```

## 6.3 Files pull-down menu

After regular reporting has completed, you are encouraged to explore your output using the pull-down menus. We will display them from left to right. Click on "Files" ...




**Specification File Name?** (also Ctrl+O)- click on this to start an analysis, or click on the specification file name in the list.

**Exit** stops Facets immediately, same as clicking on **x** in top-right corner of window.

**Finish iteration** (also Ctrl+F) forces the estimation procedure to halt at the end of the current iteration and move into the fit calculation and reporting phases.

**Save this window progress report (Ctrl+S)** writes the screen output to a text file.

**Start another Facets** launches another copy of Facets, leaving the currently running analysis active.

**Facform** launches the Facform data-formatter. Facform helps construct specification and data files, but these can also be constructed directly by hand, or by means of SAS, EXCEL etc.

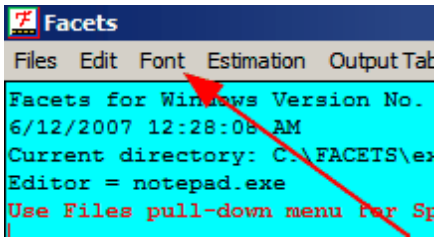
**Restart: facets (filename)** launches another Facets analysis with the stated specification file

**Exit, then Restart: facets (filename)** stops the current analysis and launches another Facets analysis with the stated specification file.

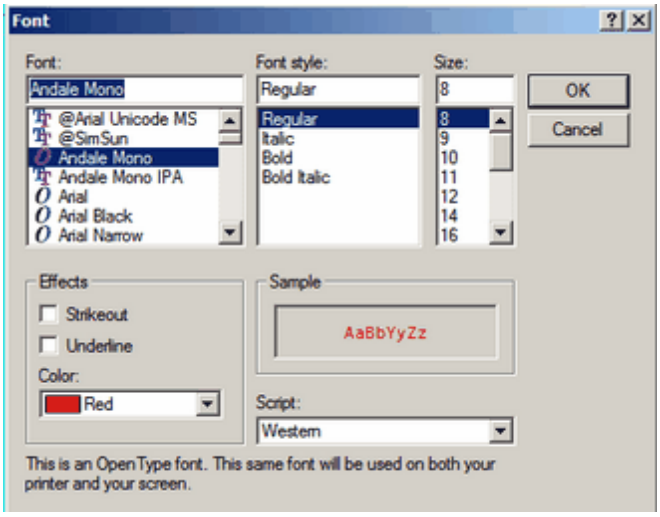
**C:\facetsanalysisx\New3700\lfs.txt** - Names of specification files are added to the bottom of the menu. Click on these when you want to start an analysis quickly.

**C:\facetsanalysisx\New3700\ -** - Names of folders are added to the bottom of the menu. Click on these when you want to locate a file quickly.

## 6.4 Font pull-down menu

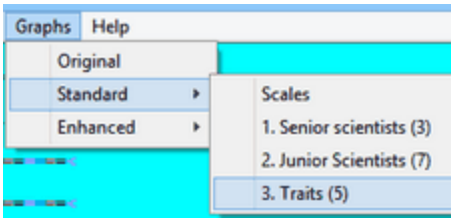



The font menu simplifies changing the font for some or all of the Facets Window display. Highlight the text you want to change, or press Ctrl+A to select all the text in the window. The text displays best with a fixed-space font like Courier New or Andale Mono.

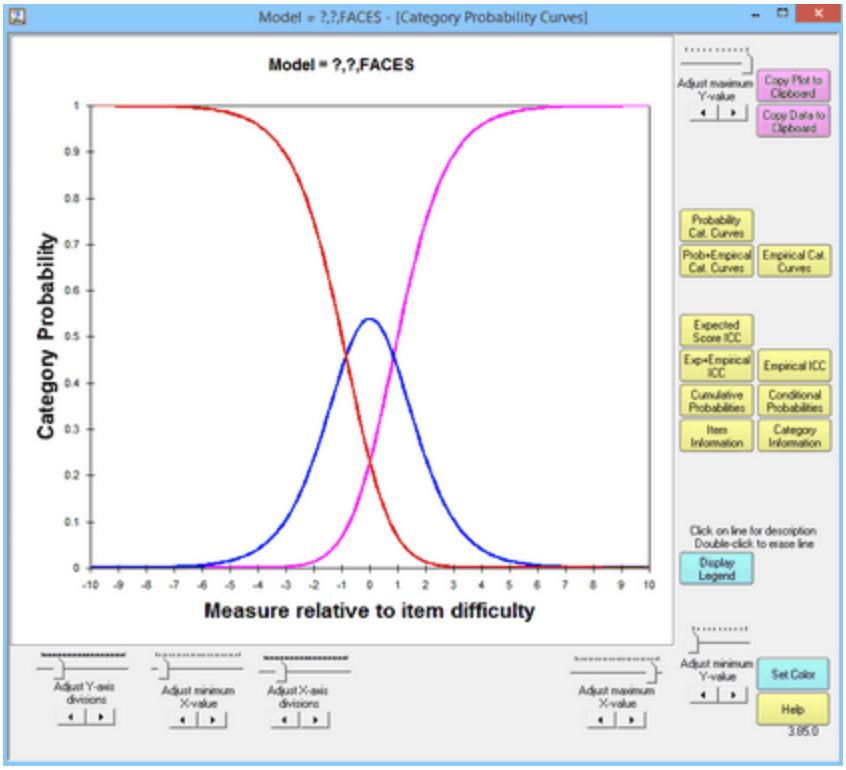



## 6.5 Graphs pull-down menu

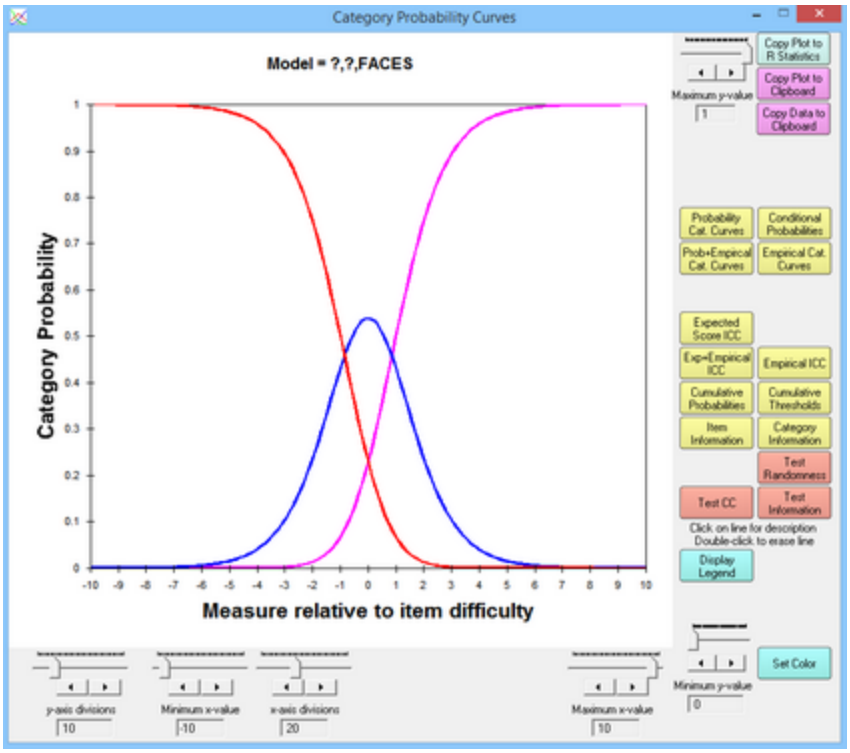
Click on "Graphs" on the Menu bar:



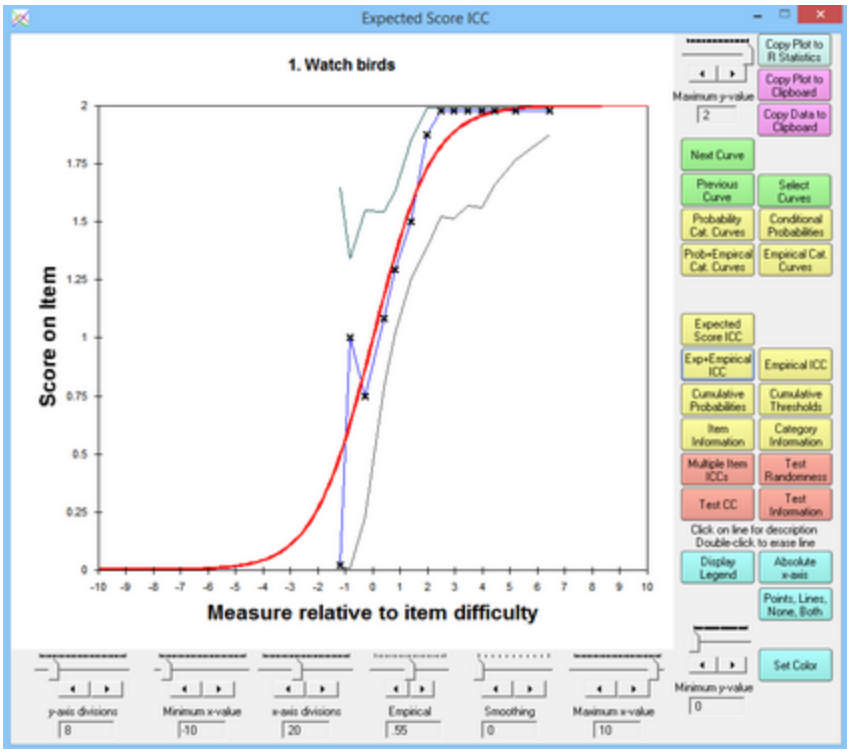

**1. Original. This is the Graph window implemented in *Facets* since 2002.**
Click on *Original*, then the Graph window will display graphs looking like:



**2. Standard. This uses the 32-bit (VB6) Winsteps graphing module.**
Click on *Standard,* and a sub-menu appears::
On sub-menu: *Scales* - this displays a Graph window similar to the *Original* window but with more functions:

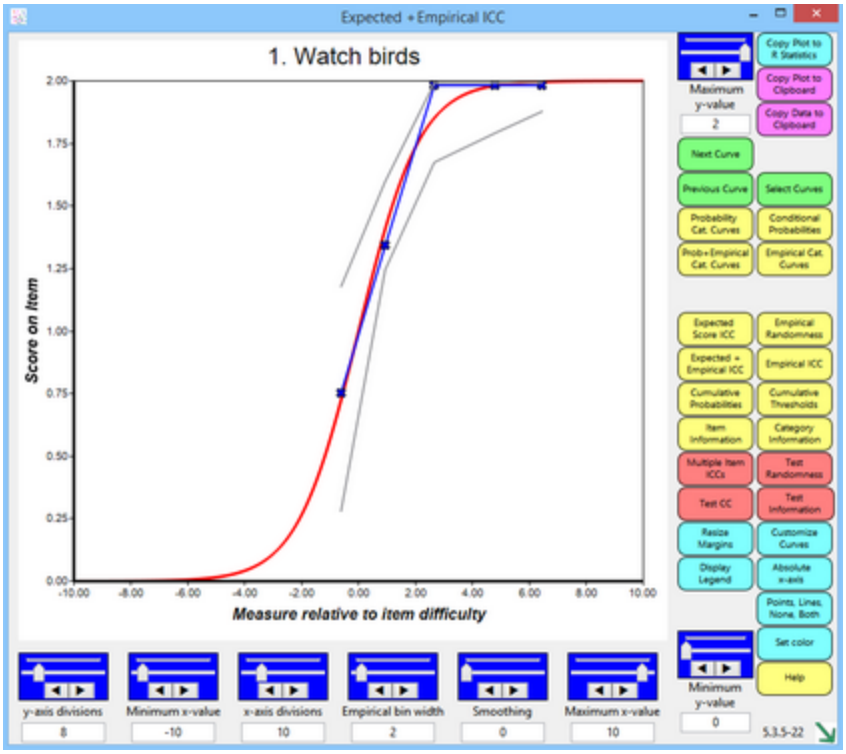



In analysis of Guilford.txt:
On sub-menu: *1. Child (75)* - Facet number 1. Facet name: Child. Count of active elements in the facet: 75
On sub-menu: *2. Activity (25)* - Facet number 2. Facet name: Activity. Count of active elements in the facet: 25
Click on your choice to display the model and empirical curves for each element of the facet.
Only elements with all observations on the same rating scale can be shown.



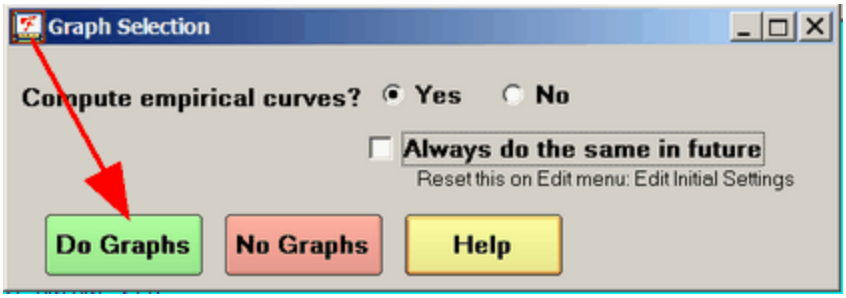

---

**3. Enhanced. This uses the 64-bit (Xojo) Winsteps graphing module.**
Click on *Enhanced,* and a sub-menu appears.This displays a Graph window similar to the *Standard* Graph window but with more functions:



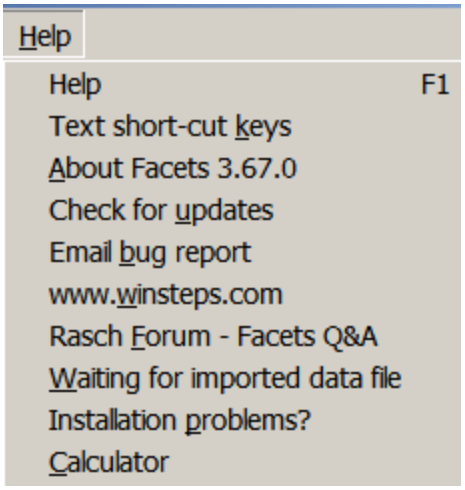

---

You may be asked if you want to compute the empirical curves, or only display the model curves. Check "Always do the same in future", and click "Yes" unless you have very large datasets. Click on "No" if you don't need to see the empirical curves. This can be reset in Edit Initial Settings.

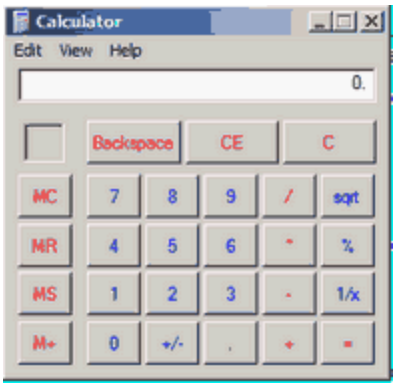



## 6.6 Help pull-down menu

The "Help" menu provides assistance.



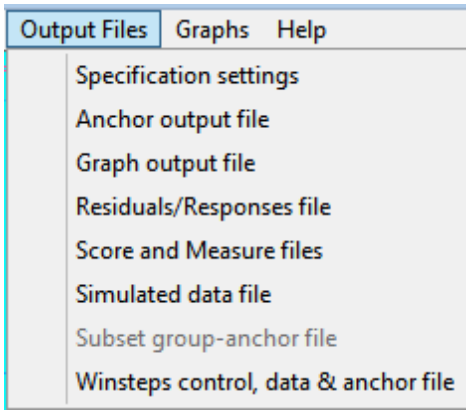

| | |
|---|---|
| **Help F1** | accesses this Help file from any Facets screen |
| **Text short-cut keys** | is the list of short-cut keys |
| **About Facets 3.87.0** | shows the Facets version number, which is also on the first line of the Iteration screen. And copyright details. |
| **Check for updates** | links to the Facets updates webpage of the Winsteps website: www.winsteps.com/facgood.htm |
| **Email bug report** | generates an email for you to report bugs or other suggestions to Winsteps.com |
| **www.winsteps.com** | links to the Facets webpage of the Winsteps website: www.winsteps.com/facets.htm |
| **Rasch Forum - Facets Q&A** | ask your colleagues for advice, and give them assistance, on our online Rasch Forum |
| **Waiting for imported data file** | suggestions for speeding up data file input |
| **Installation problems?** | links to the problem-solving webpage of the Winsteps website: www.winsteps.com/problems.htm |

| | |
|---|---|
| **Calculator** | displays the handy Windows calculator:  |



## 6.7 Output Files pull-down menu

The "Output Files" menu allows you to output the measures after the analysis is completed. These files are in addition to those generated in the main run. They are put into separate files or displayed on your screen.




Specification settings file: list of the settings of all specifications
Anchor output file: the original specification file, but with the measure estimates included as anchor values for all elements.
Graph output file: the coordinates for plotting probability curves
Residuals/Responses output file: the response-level data and statistics
Score output file: element scores and measures
Simulated data file: response-level data simulated to match the empirical data
Subset group-anchor file: This assists with resolving ambiguities in the measurement frame of reference due to subset disconnection. It is only active when this type of problem exists, e.g., with example file subsets.txt
Winsteps control & data file: output the Facets data in a format compatible with Winsteps®

These output files include those produced automatically from the main analysis, but with more options, e.g., SPSS format output.

## 6.8 Output Tables and Plots pull-down menu

The "Output Tables and Plots" menu allows you to report the measures after the analysis is completed. These Tables are in addition to those reported in the main run. They are put into separate files or displayed on your screen. Some Tables can also be displayed as Excel plots.

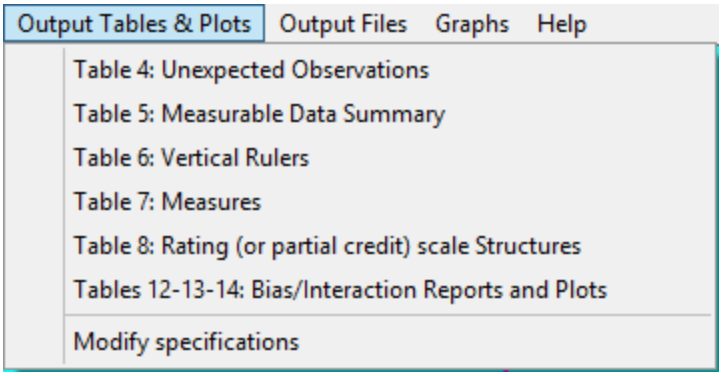



Table 4: Unexpected observations
Table 4 shows the most unexpected responses

Table 5: Measurable Data Summary
Table 5 reports summary statistics about the data for the analysis

Table 6: Vertical Rulers
Table 6.0 is the All facet summary - "rulers"
Table 6.0.0 is the Disjoint element listing
Table 6.2 Graphical description of facet statistics

Table 7: Measures
Table 7 is the facet measurement report
Table 7 reliability and chi-squared statistics
Table 7 agreement statistics

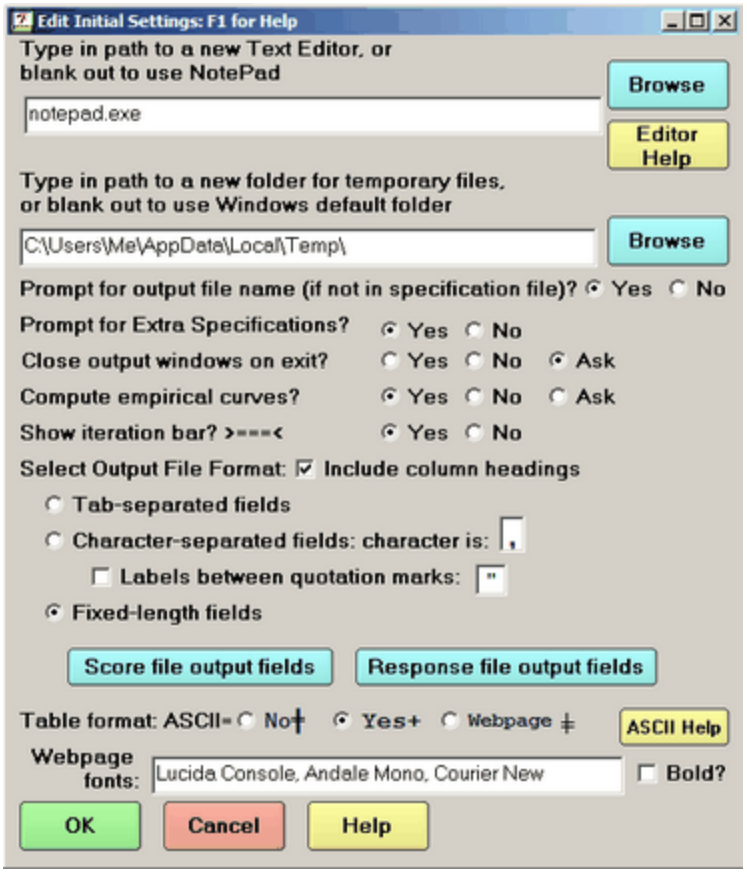
Table 8: Rating (or partial credit) scale Structures
Table 8.1 Dichotomy, binomial trial and Poisson statistics
Table 8.1 Rating scale (or partial credit) statistics
Table 8 Scale structure bar-chart
Table 8 Scale structure probability curves

Table 12-13-14: Bias/Interaction Report
Table 9 is the iteration report for bias estimation
Table 10 lists responses, still unexpected after allowing for bias
Table 11 shows how many responses were used to calculate bias estimates
Table 12 is the bias/interaction summary report
Table 13 is the bias/interaction report
and Excel DIF/bias plot (only from Output Tables pull-down menu)
Table 14 is the pairwise bias/interaction report

Modify Specifications allows modification of control specifications which affect general program output, but without requiring re-estimation.

## 6.9 Edit Initial Settings

From the Edit menu, Edit Initial Settings allows Facets standard functions to be pr-set or altered




**To change the text editor,** type in the path to your preferred editor (or "browse" to it) and click the OK box (or blank out the file name to use Notepad). Test the new editor by clicking on "Edit new file =" in the Edit pull-down menu. Useful replacements for Notepad include NotePad++ freeware: notepad-plus-plus.org

**The folder for temporary files** is usually the one assigned by Windows, but you may change this to the folder of your choice, for instance, on another disk drive.

**Prompt for output file name (if not in specification file)?** If no output file is specified with Output=, then a standard output file name is constructed by Facets. To have this automatically accepted, click on the "No" option.

**Prompt for Extra Specifications?** If Extra Specifications are new to be entered when Facets run, then specify No to remove unneeded user screen.

**Close output windows on exit?** If secondary windows are open when an analysis is ended, you are prompted as to whether the windows are to be closed. Click on the options to always close the windows (Yes) or leave them open (No) or be asked every time (Ask).

**Compute Empirical Curves?** When the Graphs menu is accessed, and there is a large amount of processing needed to produce the empirical response curves, then you are asked if you want this processing. There is no need to do it if you only want to look at the model curves.

**Show iteration bar?** >====< Usually "Yes", but there are some Unicode versions of Windows in which displaying the iteration bars noticeably slows down Facets.

**Select Output File Format:** this sets the choices for the graph, score, and residual output files: column headings, quotation marks, field format.

**Tab-separated fields:** fields in the output files are separated (delimited) by tab characters

**Character-separated fields:** fields in the output files are separated (delimited) by the specified characters
**Labels between quotation marks:** "Henry" or Henry (required by some data-input software).
**Fixed-length fields:** fields occupy fixed columns in the output files (easiest to read by eye).

**Score file output fields:** identifies which fields to include in the score file output.
**Response file output fields:** identifies which fields to include in the residual/response file output.

Table format: sets the default value for ASCII=. This formats the Output Tables for best display in ...

| | |
|---|---|
| **ASCII=Yes** | Windows, using a fixed-space font such as Courier New. |
| **ASCII=No** | MS-DOS or using a PC-CHAR font, such as Letter Gothic Legal. |
| **ASCII=Webpage** | HTML using a fixed-space font such as Courier New. |
| **Webpage fonts** | specifies the fonts to use for webpage display<br>**Bold? I**s the webpage font to be in bold face? |

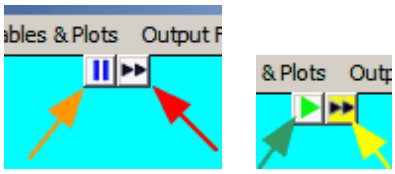
**OK:** actions the changes and saves as default values
**Cancel:** exits from this dialog box without changing the settings
**Help:** displays this page

---

**To change the "Current Directory" or "Working Folder"** when Facets launches, change the "Start in:" folder in the *Facets* shortcut on your Windows Desktop or Start menu:
Right-click on the shortcut
Type the folder name into "Start in:"
Click OK

## 6.10 Quick-action buttons

The Facets window has two useful push-buttons:




Click on the left-hand blue button || to pause or suspend the operation of Facets. This can be useful when you want to inspect the iteration screen or Facets is slowing down another application. Click on the green > button to resume Facets operation.

Click on the right-hand button » to finish iterating at the end of the current iteration. It is the same pressing as the Ctrl+F keys. The button turns yellow. It will return to white when the current processing phase completes.

## 6.11 Short-cut keys

| **Shortcut keys for the Facets operations window:** | |
|---|---|
| Ctrl+O | Specification file name? |
| Ctrl+F | Finish iterating |
| Ctrl+S | Save Facets window |
| Alt+Esc, Alt+Tab | Change the active application. |

| | |
|---|---|
| **Model checking:** | |
| F1 | Help topics |
| F2 | Bigger changes between iterations |
| F3 | Smaller changes between iterations |
| F4 | Query - show next match |
| F5 | Query=No, no more data display |
| F6 | Query=Yes, activate data display |
| | |
| **Copy, Delete, Cut:** | |
| Insert | Overwrite |
| Ctrl+A, Ctrl+Clear, Ctrl+Number Pad 5 | Select all |
| Ctrl+C, Ctrl+Insert | Copy |
| Ctrl+V, Shift+Insert | Paste |
| Ctrl+X | Cut |
| Ctrl+Z | Undo |
| Ctrl+Y | Redo |
| Backspace, F16 | Delete previous character. |
| Ctrl+Backspace | Delete previous word. |
| Ctrl+Delete | Delete the next word or selected characters. |
| Shift+Delete | Cut the selected characters. |
| Esc | Stop drag-drop. While doing a drag-drop of text. |
| | |
| **Move:** | |
| Ctrl+Left Arrow | Move cursor one word to the left. |
| Ctrl+Right Arrow | Move cursor one word to the right. |
| Ctrl+Up Arrow | Move to the line above. |
| Ctrl+Down | Arrow Move to the line below. |
| Ctrl+Home | Move to the beginning of the document. |
| Ctrl+End | Move to the end of the document. |

| | |
|---|---|
| Ctrl+Page Up | Move one page up. |
| Ctrl+Page Down | Move one page down. |
| | |
| **Text Alteration:** | |
| Ctrl+E | Center alignment |
| Ctrl+J | Justify alignment |
| Ctrl+R | Right alignment |
| Ctrl+L | Left alignment |
| Ctrl+1 | Line spacing = 1 line. |
| Ctrl+2 | Line spacing = 2 lines. |
| Ctrl+5 | Line spacing = 1.5 lines. |
| Ctrl+Shift+A | Set all caps. |
| Ctrl+Shift+L | Fiddle bullet style. |
| Ctrl+Shift+Right Arrow | Increase font size. |
| Ctrl+Shift+Left Arrow | Decrease font size. |

# 7 Examples of Specifications and Data

## 7.1 3mile.txt: 1 facet fixed effects analysis: Stress at Three Mile Island

After the nuclear accident at Three Mile Island in Spring 1979, mothers in the area were surveyed four times to determine their levels of stress. 115 mothers living within 5 miles of the plant were assigned to group LT5, another 152 mothers living between 5 and 10 miles away were assigned to group GT5. Data are from Fienberg SE, Bromet EJ, Follmann D, Lambert D, May SM 1985. Longitudinal analysis of categorical epidemiological data. Environmental Health Perspectives, 63, 241-248.

For this analysis, each mother at each time point is regarded as a random example of the fixed stress effect of one of the two groups. The analysis estimates the mean stress level of each group of mothers at each time point by averaging their stress levels as recorded in their ratings.
A better analytical technique is in the Arithmetic Test example.

Facets specifications and data (in file 3mile.txt):

```
Title = Three Mile Island - Mother's Stress (Fienberg et al.)
Facets = 1  ; only facet is group at time point
Noncenter=1  ; ratings on elements define their own origin
Convergence = .001,.0001 ; very precise convergence criteria
Vertical = 1A,1N,1* ; display facet 1 by name, number and distribution
Yardstick = 0,10 ; vertical columns as needed, lines 10 per logit.
Model = ?,Stress ; scale of stress for mothers
Rating (or partial credit) scale = Stress,R3 ; Stress has three levels
1=Low
2=Medium
```

```
3=High
*
Labels =
1=Group/time points
1=LT5 Winter 1979 ;Find out the effect for each group/time point
2=LT5 Spring 1980
3=LT5 Fall 1981
4=LT5 Fall 1982
5=GT5 Winter 1979
6=GT5 Spring 1980
7=GT5 Fall 1981
8=GT5 Fall 1982
*
data =
R2,1-4,1,1,1,1 ; 2 LT5 Mothers always low stress: R is the replication character: R2 indicates
         datum replicated twice
 |
1-4,2,1,1,2 ; 1 LT5 mother
R1,1-4,2,1,2,1 ; R1 is not needed here - shown for illustration
 |
R7,5-8,3,3,3,3 ; R7   7 GT5 Mothers always high stress
```

## 7.2 Baseball.txt: 1 facet paired comparison (BTL): League Baseball

Each baseball team in the American league in 1948 played every other team about 22 times. This is modeled with the Bradley-Terry-Luce (BTL) model for paired comparisons:
Team A strength (in logits) - Team B strength (in logits) = log (Probability Team A wins / Probability Team B wins)

For each pair of teams, the number of the victories of the first team against the second team is modeled as a count of successes on Bernoulli trials. There is only one facet to be measured: the teams. The probability of one team's success is obtained from the difference between its ability and the ability of the team it played.
This is modeled by specifying that the outcome is the result of the first team's measure, indicated by "?", from which is subtracted the second team's measure, indicated by "-?". Both teams are coded as elements in the same facet, 1.
To balance the data, enter each paired comparison twice: A vs. B, and B vs. A.
Adjust for this by giving the Models= a weighting of 0.5.
Teams did not all meet the same number of times. To allow for this, an extra facet, 2, is introduced: the number of games teams played against each other. This extra facet acts as a model selector, permitting pairs of teams to be matched with model statements specifying the correct number of Bernoulli trials. The elements of this extra facet serve no measurement purpose, so their measures are anchored at 0 (the umean= value), preventing them from affecting the teams' measures.

Facets specifications and data (in file Baseball.txt):

```
Title = American League Baseball 1948 (Fred Mosteller)
facets = 3 ; three facets in data: game count and two teams
arrange = m,N ; arrange tables by measure-descending, element number-ascending
entered in data = 2,1,1 ; number of encounters from facet 2, then two teams from facet 1
positive = 1 ; more wins - higher measure
non-centered = 0 ; the only active facet, teams, is centered
unexpected = 2 ; report ratings if standardized residual >=|2|
models = ; each model is weighted by 0.5 because each datapoint is entered twice.
;#,?, X,R23,1.0          ; do this analysis first: to check the raw scores
22,?,-?,B22,0.5 ; first team opposes second team 22 times
21,?,-?,B21,0.5 ; Washington-Chicago played 21 times
23,?,-?,B23,0.5 ; Cleveland-Boston played 23 times
*
labels =
1,Teams ; name of first facet
1=Cleveland ; names of elements, cities
 |
```

```
8=Chicago
*
2,Games played,A ; dummy facet - all elements anchored at 0
21,21 games,0 ; used for model selection only - anchored at 0
22,22 games,0
23,23 games,0
*
data =
23,1,2,12 ; Cleveland beat Boston 12 out of 23 games
|  ; all the other 26 paired comparisons
21,7,8,12 ; Washington beat Chicago 12 out of 21.
 ; same data again with teams reversed
23,2,1,11 ; Boston beat Cleveland 11 out of 23 games
|  ; all the other 26 paired comparisons
21,8,7,9 ; Chicago beat Washington 9 out of 21. Last line of data, and end of file
```

Table 7.1.2 Teams Measurement Report (arranged by N).

| Total Score | Total Count | Obsvd Average | Fair-M Avrage | Model Measure | Model S.E. | Infit MnSq | Infit ZStd | Outfit MnSq | Outfit ZStd | Estim. Discrm | Correlation PtMea | Correlation PtExp | N Teams |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 77.5 | 7 | 11.1 | 13.65 | .47 | .17 | .46 | -1.1 | .44 | -1.1 | 1.55 | .90 | .07 | 1 Cleveland |
| 77.5 | 7 | 11.1 | 13.52 | .45 | .17 | .50 | -.9 | .51 | -.9 | 1.47 | .88 | .07 | 2 Boston |

The reported standard error (.17 in this example) is relative to the latent variable. The standard error of the measure of each team for comparison with other teams is S.E./√2. So that the S.E. of the measure difference between Cleveland and Boston is $\sqrt{((.17/\sqrt{2})^2+(.17/\sqrt{2})^2)} = .17$ in this case.


Bradley, R.A. and Terry, M.E. (1952). Rank analysis of incomplete block designs, I. the method of paired comparisons. Biometrika, 39, 324–345.
Luce, R.D. (1959). Individual Choice Behaviours: A Theoretical Analysis. New York: J. Wiley.
Mosteller, F., 1951, Remarks on the method of paired comparisons: III, Psychometrika, 16/2, 207-218, published this dataset.


## 7.3 Kct.txt: 2 facet dichotomy: The Knox Cube Test

The Knox Cube Test measures short term memory (BTD p.28). There are two facets, Children and Tapping Items. Observations are scored as dichotomies, 1 for correct and 0 for incorrect.

**Facets** specifications and data (in file **Kct.txt**):

```
Title = Knox Cube Test (Best Test Design p.31) ; the report heading line
Facets = 2 ; two facets: children and items
Positive = 1 ; for facet 1, children, higher score = higher measure
Noncenter = 1 ; only facet 1, children, does not have mean measure set to zero
Pt-biserial = Yes ; report the point-biserial correlation
Vertical = 1*,1A,2N,2A ; show children by distribution and gender (sex), taps by number and
         name
Yard = 112,4 ; Vertical rulers 112 columns wide, with 4 lines per logit
Model = ?,?,D ; elements of the two facets interact to produce dichotomous responses
 ; log(Pni1/Pni0) = Bn - Di
 ; Bn = ability of child n, Di = Difficulty of item i,
 ; Pni1 = probability that child n on item i is scored 1.
Labels =
1,Children ; Children are facet 1
1-17 = Boy,,1 ; Pretend boys, in group 1, are numbered 1 through 17.
18-35 = Girl,,2 ; Pretend girls, in group 2, are numbered 18 through 35.
* ; end of child labels for facet 1
2,Tapping items ; Items are facet 2
1 = 1-4 ; Items labelled by the order in which the four blocks are tapped
2 = 2-3
3 = 1-2-4
4 = 1-3-4
```

```
5 = 2-1-4
6 = 3-4-1
7 = 1-4-3-2
8 = 1-4-2-3
9 = 1-3-2-4
10= 2-4-3-1
11= 1-3-1-2-4
12= 1-3-2-4-3
13= 1-4-3-2-4
14= 1-4-2-3-4-1
15= 1-3-2-4-1-3
16= 1-4-2-3-1-4
17= 1-4-3-1-2-4
18= 4-1-3-4-2-1-4
* ; end of item labels

Data = ; no data file name, so data follows immediately in this file
1 ,1   ,1 ; child 1 on item 1 scored 1  (blanks are ignored)
1 ,2-18,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0 ; child 1 on item 2 scored 1, on item 3 scored 1, etc
         to item 18
 | ; 594 more observations, 18 per data line
35,1-18,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0 ; child 35 scored 1 on items 1-3, 0 on items 4-18
```

Analysis of this data is initiated by:

C:> Facets Kct.txt

Computation of linear measures from these data requires many iterations (passes through the data) until convergence is reached. Convergence occurs when each difference between the observed marginal (total) score and the corresponding expected score for each child, and also for each item, based on the current estimates of the measures, is negligibly small. Progress towards convergence is shown on screen.

During each pass through the data, a line is drawn across the screen:

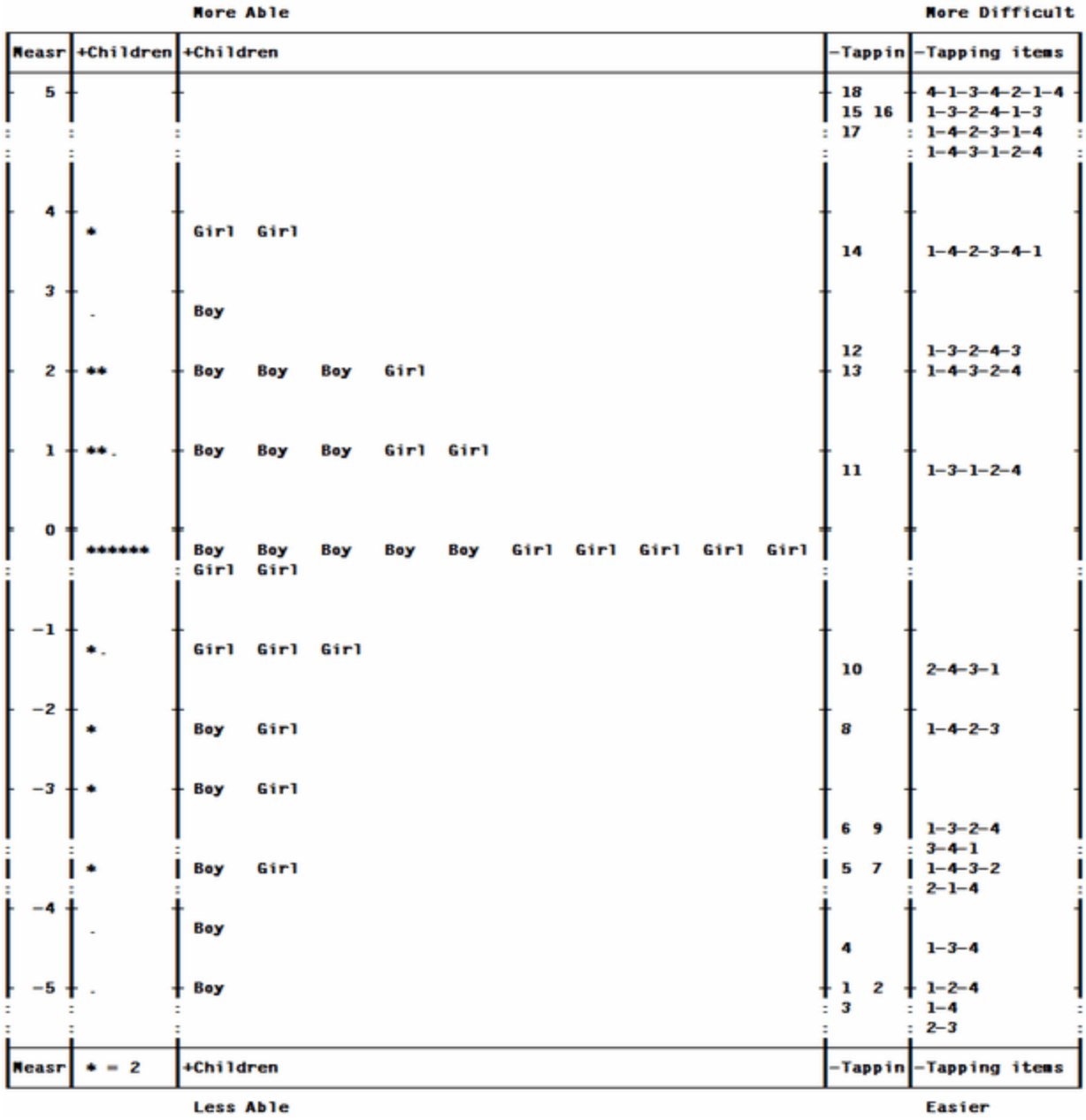
>==========================================<

This confirms that the program is running, and provides guidance as to how long an analysis is likely to take.

The output is written into file: KCT.OUT.txt. A useful table to examine initially to check for the success of the analysis is Table 6. Are the items positioned according to their difficulty in a way you would expect? Is the distribution of the children reasonable?

Table 6.0 All Facet Vertical "Rulers".

Vertical = (1*,1A,2N,2A,S) Yardstick (columns lines low high extreme)= 112,4,-5,5,End




## 7.4 Sportcas.txt: 2 facet partial credit/rank order: Sportscasting

In 1988, the abilities of seven baseball announcers were ranked on 6 items of sportscasting performance (Polskin, H., 1988, The best sportscasters in baseball. TV Guide, 30 July). These are discussed in MFRM Chapter 12 (but analyzed by a different method). A satisfactory method of analyzing rank ordered data is to treat each ordering as a test item, and the numerical rank placement as partial credit grades on that item.

Facets specifications and data (in file Sportcas.txt):

```
Title = Rank orderings of Sportscasters (TV Guide July 30 1988)
facets = 2 ; two facets: sportscasters and ranking items
arrange = 1m,N ; arrange sportscasters by measure-descending, and all facets by element number
positive = 0 ; no facet is positive - lower number=higher rank
```

```
noncenter = 2 ; center sportscasters
unexpected = 2 ; report rankings, when standardized residual >=|2|
model = ?,#,R7 ; sportscasters, and rankings each with its own scale ("partial credit")
labels =
1,Sportscasters ; name of first facet
1=Vin Scully ; names of sportscasters
2=Bob Costas
3=Al Michaels
4=Skip Caray
5=Harry Caray
6=Ralph Kiner
7=Steve Zabriskie
*
2,Items of Performance ; name of second facet
1=Calling the game ; names of items on which sportscasters were ranked
2=Working with analyst ; Note: all items will have the same score and calibration
3=Broadcasting ability ; if data are complete, i.e., every sportscaster ranked on every item
4=Quality of anecdotes
5=Knowledge of baseball
6=Enthusiasm level
*
data = ;  each column = item, has 7 rankings
1,1-6,1,3,1,1,1,4 ; first datum is 1,1,1  = Vin Scully - Calling - Rank 1
2,1-6,2,1,3,2,3,3
3,1-6,3,2,2,3,4,2
4,1-6,4,4,4,4,6,6
5,1-6,5,7,5,6,5,1
6,1-6,7,6,7,5,2,7
7,1-6,6,5,6,7,7,5
```

or

```
dvalues = 2, 1-6 ; put 1-6 in data facet 2 of all data records
1, 1,3,1,1,1,4 ; first datum is 1,1,1  = Vin Scully - Calling - Rank 1
 |
7, 6,5,6,7,7,5
```

The equivalent analysis using *Winsteps* is:

```
Title="Sportscasting"
NI=6
Item1=1
name1 = 8
codes=1234567
uscale=-1      ; reversed to give highest measure to lowest score
&end
Calling the game
Working with analyst
Broadcasting ability
Quality of anecdotes
Knowledge of baseball
Enthusiasm level
END NAMES
131114 Vin Scully
213233 Bob Costas
322342 Al Michaels
444466 Skip Caray
575651 Harry Caray
767527 Ralph Kiner
656775 Steve Zabriskie
```

## 7.5 Lfs.txt: 2 facet rating scale: Liking for Science

The Liking for Science Questionnaire measures children's attitudes to science-related activities (RSA p.18). It is an attitude survey with Likert scale ratings of 0=Dislike, 1=Don't know, 2=Like.

Facets specifications and data (in file Lfs.txt):

```
Title = Liking For Science (Rating Scale Analysis p.18) ; the report heading line
Arrange = F,m ; tables output in Fit-ascending (F) and Measure-descending (m) order
Facets = 2 ; two facets: children and science items
Pt-biserial = y ;compute point-biserial
Positive = 1 ; children are measured with higher score=greater measure
Noncenter = 1 ; mean child measure floats, but mean item measure (facet 2) is set to zero
Usort = U ; sort residuals by unexpectedness
Vertical = 2A,1* ; Vertical rulers for facet 2 by name, then facet 1 by distribution
Yardstick = 0,5 ; Vertical rulers: 0 = horizontally, as many columns per facet as required,
 ; 5 = vertically, 5 lines per logit
Model = ?,?,faces ; children interact with items to produce ratings on scale "faces"
 ; log(Pnik/Pnik-1) = Bn - Di - Fk
 ; Bn = ability of child n, Di = Difficulty of item i, Fk = Challenge of step k,
 ; Pnik = probability that child n on item i is scored k.

Rating (or partial credit) scale = faces,R2 ; definition of "faces" is "R2", a rating scale
       with categories 0 thru 2
0 = dislike ; category 0 of "faces" is labelled "dislike"
1 = don't know ; category 1 is "don't know"
2 = like ; category 2 is "like"
* ; end of scale definition for "faces"

Labels =
2,Activity ; facet 2 specified first, for convenience
1 = Watch birds ; first item
2 = Read about animals
| ; the other 21 items go here
24= Flowers live on ?
25= Talk about plants
* ; end of items
1,Child
1-75 = ; 75 children for whom there is no more information
* ; end of child labels
data =
1, 1-25, 1,2,1,1,1,0,2,0,1,2,2,2,2,0,2,1,1,2,2,0,2,1,0,2,0 ; person 1 rated item 1 with 1, but
       item 2 with 2 etc.
  |
75, 1-25, 1,2,0,0,1,1,0,1,1,2,1,2,2,2,2,1,0,2,2,0,2,0,0,1,0
```

Or:

```
Dvalues = 2, 1-25 ; place 1-25 in data facet 2 position in every line
1, 1,2,1,1,1,0,2,0,1,2,2,2,2,0,2,1,1,2,2,0,2,1,0,2,0 ; person 1 rated item 1 with 1, but item 2
       with 2 etc.
  |
75, 1,2,0,0,1,1,0,1,1,2,1,2,2,2,2,1,0,2,2,0,2,0,0,1,0
```

## 7.6 Kcta.txt: 3 facet dichotomy with item bias: The Knox Cube Test

In a standard two-facet Knox Cube analysis, respondent gender (sex) is indicated, but item bias (DIF, differential item functioning) is not computed. Here we add a DIF detection by coding a dummy "gender" Facet.

The Knox Cube Test measures short term memory (BTD p.28). There are two facets, Children and Tapping Items. Observations are scored as dichotomies, 1 for correct and 0 for incorrect.

**Facets** specifications and data (in file **Kcta.txt**):

```
Title = Knox Cube Test (Best Test Design p.31) ; the report heading line
Facets = 3 ; three facets: children and items and gender
Positive = 1 ; for facet 1, children, higher score = higher measure
Noncenter = 1 ; only facet 1, children, does not have mean measure set to zero
Pt-biserial = Yes ; report the point-biserial correlation
Vertical = 1*,1A,2N,2A ; show children by distribution and sex, taps by number and name
Yard = 112,4 ; Vertical rulers 112 columns wide, with 4 lines per logit
Model = ?,?B,?B,D ; look for bias/interaction between 2nd and 3rd Facets
Label
1,Children ; Children are facet 1
1-17 = Boy,,1 ; Pretend boys, in group 1, are numbered 1 through 17.
18-35 = Girl,,2 ; Pretend girls, in group 2, are numbered 18 through 35.
* ; end of child labels for facet 1
2,Tapping items ; Items are facet 2
1 = 1-4 ; Items labelled by the order in which the four blocks are tapped
2 = 2-3
3 = 1-2-4
4 = 1-3-4
5 = 2-1-4
6 = 3-4-1
7 = 1-4-3-2
8 = 1-4-2-3
9 = 1-3-2-4
10= 2-4-3-1
11= 1-3-1-2-4
12= 1-3-2-4-3
13= 1-4-3-2-4
14= 1-4-2-3-4-1
15= 1-3-2-4-1-3
16= 1-4-2-3-1-4
17= 1-4-3-1-2-4
18= 4-1-3-4-2-1-4
*
3,Gender,A ; Dummy gender facet is anchored
1=boys,0 ; anchor at 0 so do not affect analysis
2=girls,0 ; anchor at 0 so do not affect analysis
* ; end of item labels

Data = ; no data file name, so data follows immediately in this file
1,1,1, 1 ; child 1 on item 1 is a boy scored 1  (blanks are ignored)
1,2-18,1, 1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0 ; child 1 on item 2 is a boy scored 1, on item 3
       scored 1, etc to item 18
 | ; 594 more observations, 18 per data line
35,1-18,2, 1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0 ; child 35, a girl, scored 1 on items 1-3, 0 on
       items 4-18
```

Here is the DIF report:

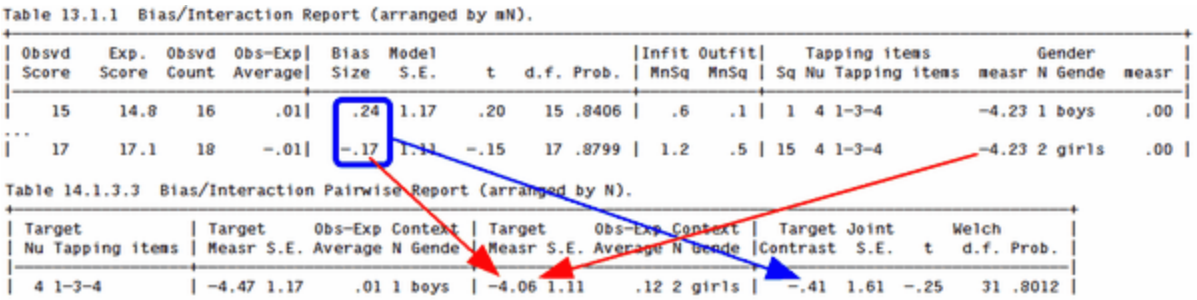


Table 13.1.1 Bias/Interaction Report (arranged by mN).

| Obsvd Score | Exp. Score | Obsvd Count | Obs-Exp Average | Bias Size | Model S.E. | t | d.f. | Prob. | Infit MnSq | Outfit MnSq | Tapping items Sq | Nu | Tapping items | Gender measr | N | Gende | measr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | 14.8 | 16 | .01 | .24 | 1.17 | .20 | 15 | .8406 | .6 | .1 | 1 | 4 | 1-3-4 | -4.23 | 1 | boys | .00 |
| ... | | | | | | | | | | | | | | | | | |
| 17 | 17.1 | 18 | -.01 | -.17 | 1.11 | -.15 | 17 | .8799 | 1.2 | .5 | 15 | 4 | 1-3-4 | -4.23 | 2 | girls | .00 |

Table 14.1.3.3 Bias/Interaction Pairwise Report (arranged by N).

| Target Nu | Tapping items | Target Measr | S.E. | Obs-Exp Average | N | Gende | Target Measr | S.E. | Obs-Exp Context Average | N | Gende | Target Contrast | Joint S.E. | t | Welch d.f. | Prob. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 1-3-4 | -4.47 | 1.17 | .01 | 1 | boys | -4.06 | 1.11 | .12 | 2 | girls | -.41 | 1.61 | -.25 | 31 | .8012 |

The reported DIF is against the overall item difficulty. The boys vs. girls DIF size on item 4, tapping pattern 1-3-4, is:
DIF size = Target Contrast Target measure for boys - Target measure for girls = -.4.47 - -4.08 = -.41.

We can see by comparing the observed and expected scores, that the boys scored 15 when 14.8 was expected, and the girls 17 when 17.1 was expected. So the item was .41 logits easier for the boys than for the girls. The t-test for this DIF size is approximately DIF size / joint standard error: t = -.41 / sqrt(SE(boy)² + SE(girl)²) = -.41/ sqrt(1.17²+1.11²) = .43/1.61 = -.25 with (boys count - 1 + girls count - 1) = 32 degrees of freedom. P=.80 for a 2-sided t-test. So this is not statistically significant. The hypothesis of "No DIF" is not rejected.

## 7.7 Woodcut.txt: 3 facet Poisson count: Woodcutting

Four teams of woodcutters were timed cutting three types of log, with and without bark. One way of modeling these data is as Poisson counts of tenths of seconds. The better teams take less time.

Facets specifications and data (in file Woodcut.txt):

```
title = Woodcutting Experiment (David Wallace)
facets = 3 ; trees, bark, teams
arrange = N,3f ; "N": output tables of measures arranging all facets by element number
 ; "3f": also arrange facet 3, teams, by fit
positive = 1,2 ; higher time, harder to cut. Lower time, better team - facet 3
noncenter=3 ; non-center the teams
vertical=1A,2A,3N ; elements reported on vertical rulers by alphabetical name, where known
model=?B,?B,?,Chops ; time to cut through log in 10ths seconds. Bias interactions between trees
       and bark.
*
rating scale=Chops,P ; model as Poisson counts with estimated step difficulty
*
Labels =
1,Tree
1=Spruce ; 3 types of log
2=Pine
3=Larch
*
2,Bark
1,Without bark ; each log, with and without bark
2,With bark
*
3,Wood cutting team
1-4= ; names of 4 teams unknown
*
Data =
1,1,1,64 ; the spruce without bark was cut by team 1 in 6.4 seconds
| ; 22 more data points
3,2,4,61 ; the larch with bark was cut by team 4 in 6.1 seconds
```

Poisson scales imply an infinite number of categories, 0,1,2,3,..., but with a known relationship. Facets reports the degree to which the data meets this categorization:

Table 8.1 Category Statistics.

| DATA Category Score | Counts Used | % | Cum. % | QUALITY CONTROL Avge Meas | Exp. Meas | OUTFIT MnSq | Obsd-Expd Diagnostic Residual |
|---|---|---|---|---|---|---|---|
| 40 | 1 | 4% | 4% | 2.44 | .41 | .5 | .8 |
| 43 | 1 | 4% | 8% | 2.53 | .48 | .0 | .8 |
| 46 | 1 | 4% | 13% | 2.59 | .53 | .1 | .8 |
| 60 | 1 | 4% | 17% | 2.72 | .81 | .3 | .7 |
| 61 | 3 | 13% | 29% | 2.76 | .28 | 5.1 | 2.7 |
| 62 | 1 | 4% | 33% | 2.92 | .85 | 6.7 | .7 |
| 64 | 2 | 8% | 42% | 2.89+ | .45 | 9.9 | 1.7 |
| 68 | 1 | 4% | 46% | 2.83+ | .95 | .0 | .7 |
| 73 | 1 | 4% | 50% | 3.04 | .97 | 9.9 | .7 |
| 74 | 1 | 4% | 54% | 2.92+ | .97 | .2 | .7 |
| 75 | 1 | 4% | 58% | 2.87+ | .97 | .6 | .7 |
| 86 | 1 | 4% | 63% | 2.95 | .89 | .9 | .7 |
| 88 | 1 | 4% | 67% | 3.06 | .86 | 1.1 | .7 |
| 100 | 1 | 4% | 71% | 3.00+ | .71 | 6.2 | .8 |
| 102 | 1 | 4% | 75% | 3.14 | .68 | .8 | .8 |
| 109 | 1 | 4% | 79% | 3.06+ | .58 | 8.3 | .8 |
| 113 | 1 | 4% | 83% | 3.13 | .52 | 2.7 | .8 |
| 120 | 1 | 4% | 88% | 3.32 | .42 | 9.9 | .9 |
| 127 | 1 | 4% | 92% | 3.15+ | .33 | 9.9 | .9 |
| 131 | 1 | 4% | 96% | 3.23 | .28 | 4.2 | .9 |
| 134 | 1 | 4% | 100% | 3.23 | .25 | 8.4 | .9 |

Poisson discrimination: .67 S.E. .00

## 7.8 Guilford.txt: 3 facet rating scale: Creativity (with Excel input data file)

In a study commissioned by the US Navy (J.P. Guilford, 1954, Psychometric Methods, 2nd Ed., New York: McGraw-Hill, p.282, and MFRM Chapter 10), three senior scientists rated seven junior scientists on five traits of creativity. Ratings ranged from 1 to 9, with 9 meaning most creative.

Facets specifications and data (in file Guilford.txt):

```
Title = Ratings of Scientists (Psychometric Methods p.282 Guilford 1954)
; measurement model
Facets = 3  ; three facets: judges, examinees, items
Non-centered = 1 ; examinees and items are centered on 0, judges are allowed to float
Positive = 2  ; the examinees have greater creativity with greater score
Model = ?B,?B,?,Creativity ; judges, examinees and items produce ratings on "Creativity".
 ; A bias/interaction analysis, ?B,?B,?, will report interactions between facets 1,
 ; senior scientists (judges,) and 2, junior scientists (examinees).
 ; log(Pnijk/Pnijk-1) = Bn - Di - Cj - Fk
 ; Bn = ability n, Di = difficulty i, Cj = Severity j, Fk = Challenge k,
 ; Pnijk = probability that child n on item i is rated by judge j with score of k.
Rating (or partial credit) scale = Creativity,R9 ;Creativity is a rating scale with possible
       categories 0 to 9
1 = lowest ; name of lowest observed category
5 = middle ; no need to list unnamed categories
9 = highest ; name of highest observed category
*
; for Table 4: unexpected observations
Unexpected = 2 ; report ratings if standardized residual >=|2|
Usort = (1,2,3),(3,1,2),(Z,3) ; sort and report unexpected ratings several ways (1,2,3) is
       Senior, Junior, Trait

; for Table 6: map of elements
Vertical = 2N,3A,2*,1L,1A ;define rulers to display and position facet elements

; for Table 7: table of measures
Arrange = m,2N,0f ; arrange tables by measure-descending for all facets,
   ; 2N = element number-ascending for facet 2,
   ; and 0f = t-statistic-descending for facet 0 (bias interactions)
Inter-rater = 1 ; facet 1 is the rater facet

; for Table 13: bias/interaction report
Zscore = 1,2  ;report biases greater in size than 1 logit or with z>2

; Output files
Score file = GUILFORD.SC ; score files .SC.1.txt, SC.2.txt and .SC.3.txt produced
```

```
; facet and element identification
Labels =
1, Senior scientists ; name of first (judge) facet
1 = Avogadro ; names of elements within senior scientist facet
2 = Brahe
3 = Cavendish
*
2,Junior Scientists ; name of second (examinee) facet
7=George  ; element order does not matter
1=Anne
2=Betty
3=Chris
4=David
5=Edward
6=Fred
*
3,Traits ; name of third (item) facet
1=Attack
2=Basis
3=Clarity
4=Daring
5=Enthusiasm
*

Data=
1,1,1-5,5,5,3,5,3
1,2,1-5,9,7,5,8,5
1,3,1-5,3,3,3,7,1
1,4,1-5,7,3,1,3,3
1,5,1-5,9,7,7,8,5
1,6,1-5,3,5,3,5,1
1,7,1-5,7,7,5,5,5
2,1,1-5,6,5,4,6,3
2,2,1-5,8,7,5,7,2
2,3,1-5,4,5,3,6,6
2,4,1-5,5,6,4,5,5
2,5,1-5,2,4,3,2,3
2,6,1-5,4,4,6,4,2
2,7,1-5,3,3,5,5,4
3,1,1-5,5,5,5,7,3
3,2,1-5,7,7,5,7,5
3,3,1-5,3,5,5,5,5
3,4,1-5,5,3,3,3,1
3,5,1-5,9,7,7,7,7
3,6,1-5,3,3,3,5,3
3,7,1-5,7,7,7,5,7 ; last line of data, and end of file
```

or:

```
Dvalues = 3, 1-5 ; place 1-5 in data facet 3 location
Data=
1,1,5,5,3,5,3
 |
3,7,7,7,7,5,7
```

or:

```
Data = Creativity.txt ; Text data file
or:
Data = Creativity.xls ; Excel data file
or:
Data = Creativity.sav ; SPSS data file
```

## 7.9 Subsets.txt: 3 facet Subset reconnection

Here is an example, using Guilford's data, of data that lack sufficient connection to yield unambiguous measure estimates. It reports subsets. In this example, pairs of judges rate subsets of examinees, but without crossing (overlap). Did Betty get the highest score because she was the most able or because her judges were more lenient? The data can't tell us. There is more about this in the discussion of subset connectedness.

Facets specifications and data (in file Subsets.txt):

```
; subsets.txt
Title = Ratings of Scientists (edited to illustrate ambiguity in measurement)
Facets = 3           ; three facets: judges, examinees, items
Umean = 50, 10             ; user-scaling = 50 +logit*10
Positive = 1,2,3     ; the examinees have greater creativity with greater score
Non-centered = 1     ; examinees and items are centered on 0, judges are allowed to float
pt-biserial = measure    ; point-measure correlation
omitunobserved = No      ; so we anchored elements which have no ratings
Arrange=N                ; Table 7 in element-number order
Model = ?,?,?,R9     ; judges, examinees and items produce ratings on "Creativity".
Labels=                  ; to name the components
1,Judges, A          ; name of first facet
1=Avogadro (1)           ; name of element within facet (subset)
2=Brahe (1)          ; element numbers must be in Labels=, or they are treated as missing
3=Cavendish (2)
4=Davey (2)
5=Anchored judge, 60 ; has no ratings: anchored at 60 units (umean=50)
10=Lone judge (3)        ; rates only lone examinee
*
2,Examinees
1=Anne (1)
2=Betty (1)
3=Chris (1)
4=David (2)
5=Edward (2)
6=Fred (2)
7=George (2)
10=Lone examinee (3)     ; rated only by lone judge
*
3,Items
1=Attack (1+2)           ; Attack is in subsets 2 and 3
2=Basis (1+2)
3=Clarity (1+2)
4=Daring (1+2)
5=Enthusiasm (1+2)
10=Lone item (3)
*

Data=
; Subset 1
1,1, 1-5, 5,5,3,5,3  ; typical paired-rater design
1,2, 1-5, 9,7,5,8,5  ; raters 1 and 2 rate examinees 1, 2, 3
1,3, 1-5, 3,3,3,7,1      ; everyone is rated on items 1,2,3,4,5
2,1, 1-5, 6,5,4,6,3
2,2, 1-5, 8,7,5,7,2
2,3, 1-5, 4,5,3,6,6

; Subset 2
3,4, 1-5, 5,3,3,3,1  ; raters 3 and 4 rate examinees 4, 5, 6, 7
3,5, 1-5, 9,7,7,7,7      ; everyone is rated on items 1,2,3,4,5
3,6, 1-5, 3,3,3,5,3
3,7, 1-5, 7,7,7,5,7
4,4, 1-5, 5,6,4,5,5
```

```
4,5, 1-5, 2,4,3,2,3
4,6, 1-5, 4,4,6,4,2
4,7, 1-5, 3,3,5,5,4

; Subset 3
10,10,10, 5              ; lone rater rates lone examinee on lone item
```

In the Iteration Report to the screen:

```
Consolidating 2 subsets..
 2 subsets remain
Warning (6)! There may be 2 disjoint subsets
```

In the Output file:

```
Table 6.0.0  Disjoint Subset Element Listing.

Subset number: 1
Facet: 1. Judges  2 Elements: 1 2
Facet: 2. Examinees  3 Elements: 1-3

Subset number: 2
Facet: 1. Judges  2 Elements: 3 4
Facet: 2. Examinees  4 Elements: 4-7
```

Table 7.1.3 Judges Measurement Report (arranged by N).

| Total Score | Total Count | Obsvd Average | Fair-M Average | Measure | Model S.E. | Infit MnSq | Infit ZStd | Outfit MnSq | Outfit ZStd | Estim. Discrm | Correlation PtMea | PtExp | N | Judges | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 72 | 15 | 4.8 | 4.62 | 49.48 | 1.70 | .68 | -.9 | .73 | -.7 | 1.57 | .90 | .09 | 1 | Avogadro | in subset: 1 |
| 77 | 15 | 5.1 | 4.99 | 50.89 | 1.69 | .72 | -.8 | .76 | -.6 | .76 | .60 | .09 | 2 | Brahe | in subset: 1 |
| 102 | 20 | 5.1 | 5.17 | 51.58 | 1.44 | 1.02 | .1 | 1.03 | .1 | 1.56 | .73 | .08 | 3 | Cavendish | in subset: 2 |
| 79 | 20 | 4.0 | 3.94 | 46.63 | 1.54 | 1.29 | .9 | 1.33 | 1.0 | .30 | -.23 | .07 | 4 | Davey | in subset: 2 |

**Investigating the ambiguity:**

Click on the Output Files Menu pull-down menu
Click on Winsteps control & data file
This enables you to construct data files showing the pattern of responses of one facet against another. Subsetting will show as distinctive patterns.

Here is what it looks like with Examinees as rows, and Judges as columns:

```
1 Anne     ; item-column labels
2 Betty
3 Chris
4 David
5 Edward
6 Fred
7 George
END LABELS
 5 9 3 . . . . 1 Avogadro    ; row data + label
 6 8 4 . . . . 2 Brahe
 . . . 5 9 3 7 3 Cavendish
 . . . 5 2 4 3 4 Davey
```

**Resolving the ambiguity:**

Click on the Output Files Menu pull-down menu.
Click on Subset group-anchor file.
This produces:

```
To resolve subset problems, copy-and-paste after Labels=
Non-center= must reference a facet that is not anchored or group-anchored.

Group anchor this facet:
```

```
1,Judges, G ; group-anchoring at Umean = 50
1,Avogadro,50, 1
2,Brahe,50, 1
3,Cavendish,50, 2
4,Davey,50, 2
*

And/or group anchor this facet:

2,Examinees, G ; group-anchoring at Umean = 50
1,Anne,50, 1
2,Betty,50, 1
3,Chris,50, 1
4,David,50, 2
5,Edward,50, 2
6,Fred,50, 2
7,George,50, 2
*
```

**Action:**
To establish an unambiguous measurement structure,
(a) We can assert that the rater pairs have the same mean severity (i.e., are randomly equivalent) using group-anchoring:

```
Non-center=2   ; facet 2 is allowed to float, because facet 1 will be group-anchored.
1,Judges, G ; group-anchor
1=Avogadro,50, 1  ; rater 1 is in group 1 with a mean value of 0.
2=Brahe,50, 1
3=Cavendish,50, 2  ; rater 3 is in group 2 with a mean value of 0.
4=Davey,50, 2
*
```

or (b) We can assert that the examinee samples have the same mean ability (i.e., are randomly equivalent):

```
Non-center=1   ; facet 1 is allowed to float, because facet 2 will be group-anchored.
2,Examinees, G ; group-anchor
1=Anne,50, 1   ; examinee 1 is in group 1 with a mean value of 0
2=Betty,50, 1
3=Chris,50, 1
4=David,50, 2   ; examinee 2 is in group 1 with a mean value of 0
5=Edward,50, 2
6=Fred,50, 2
7=George,50, 2
*
```

## 7.10 Measure.txt: 4 facet Arithmetic Test: Measuring, Anchoring and Describing

Many measurement problems are best tackled with two Facets runs:

1. Measurement of student ability, judge severity, item difficulty, etc.
The actual participant elements that interacted to provide the observations are measured, and then the elements anchored at those measures. This provides a stable frame of reference for further analyses.

2. Descriptive summaries of various effects identified by demographics, judge training, sub-test content, etc.
Participating elements are replaced by their hypothesized components, and further analyses are performed, with care to provide connected, i.e, unambiguous, measurement conditions.

Using R. Mislevy's data set (extracted from the Armed Services Vocational Aptitude Battery - ASVAB), it is hypothesized that 776 students (black females, black males, white females, and white males) each have an arithmetic ability (a fixed effect). The analyst wants to decompose these abilities into gender (sex) and race (ethnicity) effects. First the item difficulties are calibrated and the 776 examinees measured:

Facets specifications and data (in file Measure.txt):

```
title = Arithmetic Competency - R. Mislevy
anchorfile = measanc.txt ; an anchored file is written out
facets = 4 ; demographics included in the facets
pt-biserial = y ; point-biserial as a rough fit statistics
vertical = 1N,2A,3A,4* ; for communication
yard = 0, 4 ;
models =
?, , ,?,D ; use this model for measuring items and students
; ?,?,?, ,D ; use this model for demographic summaries - commented out here
*
positive = 2,3,4 ; all facets except item difficulties are abilities
noncenter = 2,  4 ; noncenter students and one demographic
labels =
1,Arithmetic
1-4 ; 4 arithmetic items
*
2,Race
1=Black
2=White
*
3,Sex
1=Female
2=Male
*
4,Students
1-776 ; no more information about the students
*
Data =
1-4,2,2,1,0,0,0,0  ; on the 4 items, white male student 1, failed
 |
1-4,1,1,776,1,1,1,1 ; on the 4 items, black female student 776, succeeded
```

or

```
Dvalues = 1, 1-4  ; place 1-4 in first data facet location of all data records
Data=
 2,2,1,0,0,0,0 ; on the 4 items, white male student 1, failed
 |
 1,1,776,1,1,1,1 ; on the 4 items, black female student 776, succeeded
```

Then the descriptive phase is performed to estimate the demographic effects. This is set up by editing the anchor file produced in the measurement phase, and changing the model statement:

Facets specifications and data (in file Measanc.txt, also in file Meas2anc.txt):

```
title = Arithmetic Competency - R. Mislevy
; anchorfile = measanc.txt
facets = 4
pt-biserial = y
vertical = 1N,2A,3A,4*
yard = 0, 4
positive =   2,3,4 ; facets 1, 2, 3 this time
noncenter=   2,4 ; facet , Race, 2 floats
Models=
;?,,,?,RS1,1, (D) ; comment out the measurement model
 ?,?,?,,RS1,1, (D) ; use the summarizing model
*
Rating (or partial credit) scale=RS1,D ; use the measurement run rating scale
 0=,0,A, ; Rasch-Andrich thresholds in the description run
*
Labels=
```

```
1,Arithmetic,A ; these are anchored at their Rasch-Andrich thresholds
1,1,-.6079245
2,2,-.1628681
3,3,.4518689
4,4,.9676156
*
2,Race,A ; the A is inoperative, because there are no logit values
1,Black,
2,White,
*
3,Sex,A
1,Female,
2,Male,
*
4,Students,A ; this facet is ignored this run
1,1,-2.537946 ; these anchored measures are ignored this run
 |
1-4,1,1,776,1,1,1,1
```

## 7.11 Dives.txt: 4 facet rating scale with missing data: Diving

At the 1988 Illinois Boys Diving Competition, 12 divers performed 3 dives which were rated by 7 judges. The divers chose which types of dive to perform, so each diver did not perform most types of dive. Not all ratings were recorded. Thus, most of the complete data is "missing"! The original ratings were in the range 0 to 10 with half point increments, so all ratings have been doubled to make them integers. Previous analyses have been performed so that good starting logit values are known. One diver did the wrong dive by mistake - this is deliberately treated as missing data in this analysis. Include clearly aberrant data, e.g., lucky guesses, if your purpose is to describe the past, e.g., for awarding a prize. Omit such data if your purpose is to predict the future, e.g., to discover examinee's ability to perform surgery or to obtain item difficulty calibrations for an item bank.

Facets specifications and data (in file Dives.txt):

```
Title = 1988 Illinois Boys Diving Competition (Anne Wendt)
Facets = 4 ; four facets are diver, dives, round, judges
Inter-rater = 4 ; facet 4 is the rater facet
Noncenter = 1 ; diver measures float
Positive = 1 ; only for divers does greater score mean greater measure
Models = ; these models will be scanned in order until a match is found for each datum:
 7,8,?,?,M ; make diver 7 on dive 8 missing data - He did the wrong dive!
 ?,?,X,?,DoublePoints ; Round excluded for this analysis by X
*
Rating scale = DoublePoints,R20,Keep ; Keep unobserved intermediate categories
 0  = 0.0
 10 = 5.0 ; 5 on the original scoring
 20 = 10.0
*
Labels=
1,Diver
;  diver          previous scores included in name for reference
 1=Marty Turek    292.85 425.65        , 2.08 ; logit starting value of 2.08 after second ","
|   ; 10 other divers
12=Bryan Hanania   251.15              ,-1.19
*
2,Dives
 ; Each dives official weighted difficulty, e.g., 1.4, used as dive "name" after "="
1=1.4, -0.60 ;starting logit values after ","
| ; 6 other dives
8=2.6,  0.99
9=2.4? ; unclear what dive this was
*
3,Round
1-3 ; there were 3 rounds
```

```
*
4,Judges
1-7 ; there were 7 judges
*
Data =
1,8,1,1,14 ; Diver 1, Marty, made a dive 8, a "2.6" dive, in round 1 rated by judge 1 with
       7*2=14
1,8,1,2-4,11,12,12 ; judges 2, 3, 4 rated 11, 12, 12
1,1,2,1-7,16,13,13,16,14,13,14
2,8,2,1-7,12,12,10,11,10,11,13
3,1,1,1-4,10,11,10,11
3,6,2,1-7,11,10,09,08,09,09,11
| ; 131 other ratings
12,2,3,7,11 ; Bryan made a "1.7" dive in round 3 rated by judge 7 with 5.5*2=11

or
dvalues = 4, 1-7 ; facet 4 is the 7 judges
Data = Dives.xls ; Excel data file format
```

## 7.12 Divesb.txt: 4 facet Interactions and Dummy Facets: Diving + Creativity

Dummy facets are facets intended only for investigating interactions, not for measuring main effects. All the elements of a dummy facet are anchored at 0.

Example 1: At the 1988 Illinois Boys Diving Competition, 12 divers performed 3 dives which were rated by 7 judges. The divers chose which types of dive to perform, so each diver did not perform most types of dive. Not all ratings were recorded. Thus, most of the data is "missing"! The original ratings were in the range 0 to 10 with half point increments, so all ratings have been doubled to make them integers. Previous analyses have been performed so that good starting logit values are known. One diver did the wrong dive by mistake - this is deliberately treated as missing data in this analysis.
Include clearly aberrant data, e.g., lucky guesses, if your purpose is to describe the past, e.g., for awarding a prize. Omit such data if your purpose is to predict the future, e.g., to discover examinee's ability to perform surgery or to obtain item difficulty calibrations for an item bank.

**We are investigating if there is any judge-by-round (= rater by time-point) interaction.**

Facets specifications and data (in file Divesb.txt)

```
title = 1988 Illinois Boys Diving Competition (Anne Wendt)
facets = 4 ; four facets are diver, dives, round, judges
Inter-rater = 4 ; facet 4 is the rater facet
noncenter = 1 ; diver measures float
positive = 1 ; only for divers does greater score mean greater measure
models = ; these models will be scanned in order until a match is found for each datum:
 7,8,?,?,M ; make diver 7 on dive 8 missing - He did the wrong dive!
 ?,?,?B,?B,R20 ; "round"is a dummy facet for interaction with judges
  ; but round is checked for validity
*
labels=
1,Diver
;  diver          previous scores included in name for reference
 1=Marty Turek    292.85 425.65        , 2.08 ; logit starting value of 2.08 after second ","
|  ; 10 other divers
12=Bryan Hanania   251.15             ,-1.19
*
2,Dives
 ; Each dives official weighted difficulty, e.g., 1.4, used as dive "name" after "="
1=1.4
| ; 6 other dives
8=2.6
```

```
9=2.4? ; unclear what dive this was
*
3,Round,A ; anchored at zero means dummy facet, does not affect measurement
1-3=,0 ; there were 3 rounds
*
4,Judges
1-7 ; there were 7 judges
*
data =
1,8,1,1,14 ; Diver 1, Marty, made a dive 8, a "2.6" dive, in round 1 rated by judge 1 with
       7*2=14
| ; 136 other ratings
12,2,3,7,11 ; Bryan made a "1.7" dive in round 3 rated by judge 7 with 5.5*2=11
```

Largest judge by round interaction is:

1988 Illinois Boys Diving Competition (Anne Wendt)
Table 14.2.2.3 Bias/Interaction Pairwise Report

| Target N J | Target Measr | S.E. | Obs-Exp Average | Context N | Round | Target Measr | S.E. | Obs-Exp Average | Context N | Round | Target Joint Contrast | S.E. | t | d.f. | Prob. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 1 | -.57 | .24 | .50 | 2 | round 2 | .08 | .24 | -1.46 | 3 | round 3 | -.65 | .34 | -1.93 | 21 | .0678 |

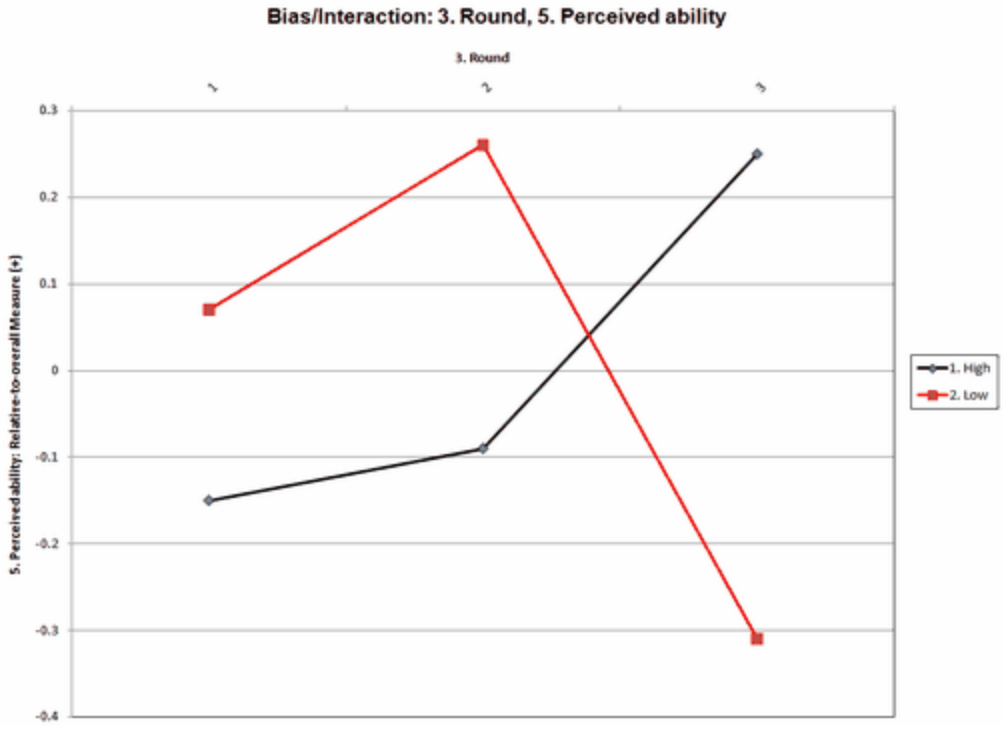
Judge 1 was relatively lenient in Round 2 (+0.50 score points), but severe in Round 3 (-1.46 score points).

Example 2: In the Dives data, we suspect that there is an interaction between perceived diver ability (High and Low) and dive round. This is example dataset: Divesint.txt

```
Title = 1988 Illinois Boys Diving Competition (Anne Wendt)
Facets = 4 ; four facets are diver, dives, round, judges
Noncenter = 1 ; diver measures float
Positive = 1 ; only for divers does greater score mean greater measure
Models = ; these models will be scanned in order until a match is found for each datum:
 7,8,?,?,?,M ; make diver 7 on dive 8 missing - He did the wrong dive!
; interaction between facet 3 (round) and facet 5 (perceived ability)
 ?,?,?B,?,?B,DoublePoints ; Round excluded for this analysis by X
*
Rating scale = DoublePoints,R20,Keep ; Keep unobserved intermediate categories
 0  = 0.0
 10 = 5.0 ; 5 on the original scoring
 20 = 10.0
*
labels=
1,Diver
;  diver          previous scores
1, H: Marty Turek     292.85  ; previous scores
2, H: Tom Wright      279.95
3, L: Mike Gotkowski  249.9
4, L: Matt Paulson    244.55
5, L: Scott Ternovits 252.8
6, L: Ross Moyer      243.4
7, H: Curt Billings   266.25
8, H: Steve Hutchings 267.15
9, L: Larry Kirk      258.35
10, H: Kurt Becker     284.4
11, L: Lance Kleffman  259.6
12, L: Bryan Hanania   251.15
*
2,Dives
1-9
*
3,Round, D ; a dummy facet
1-3
*
```

```
4,Judges
1-7
*
5, Perceived ability, D ; a dummy facet
1, High previous score 260 up
2, Low previous score 259 down
*
Dvalues =
4, 1-7 ; the observations for the 7 judges are in Excel
5, 1, 1, 1 ; element label for facet 5 is first character of label in facet 1
*
Data = Dives.xls
; has this format:
  ; diver    dive     round    judge 1   judge 2   judge 3   judge 4   judge 5   judge 6   judge 7
     1        8        1        14        11        12        12        .         .         .
     1        1        2        16        13        13        16        14        13        14
     2        8        2        12        12        10        11        10        11        13
```

The interactions indicate a jump in ability in round 3 for those with perceived high ability. Do they handle the pressure of the final round better? Are the judges more lenient with them in the third round? Or ... This could be the start of an exciting investigation into competitive behavior of great interest to diving coaches.




Example 3: We want to investigate whether there is an interaction between judges and score-bands for the Creativity (Guilford) data. A dummy facet "score-band" is introduced into the data. It has 3 elements: 1, 4, 7, corresponding to the observed rating ranges 1-3, 4-6, 7-9.

Facets specifications and data (in file Guilfordsb.txt):

```
; Guilfordsb.txt
Facets = 4 ; four facets judges, examinees, items, score-band
Inter-rater = 1 ; facet 1 is the rater facet
Model = ?B,?,?,?B, R9 ; interaction between judges and score-band

Labels=                ;to name the components
```

```
1,Senior scientists    ;name of first facet raters
1=Avogadro             ;names of elements within facet
2=Brahe                ;these must be named, or they are treated as missing
3=Cavendish
*
2,Junior Scientists
2=Betty
5=Edward
7=George
1=Anne
3=Chris
4=David
6=Fred
*
3,Traits
1=Attack
2=Basis
3=Clarity
4=Daring
5=Enthusiasm
*
4, Score-band, A   ; anchored
1, Low 1-3, 0      ; 1+2+3 ; anchored at 0 to avoid distorting the measures
4, Medium 4-6, 0   ; 4+5+6
7, High 7-9, 0     ; 7+8+9
*

; rater, examinee, item, score band, response
Data=
1,1,1,4,5
1,1,2,4,5
....
```

## 7.13 Kctinter.txt: 4 facet item-type interactions: Knox Cube Test

The Knox Cube Test is examined for bias between gender (sex) and item type

**Facets** specifications and data (in file **Kctinter.txt**):

```
; kctinter.txt
TITLE='Knox Cube Test with interactions'
Facets = 4   ; four facets: children, items, child gender, item length
Positive = 1   ; for facet 1, children, higher score = higher measure
Noncenter = 1   ; only facet 1, children, does not have mean measure set to zero
Model =
?,?,? ,? ,D   ; elements of the four facets interact to produce dichotomous responses
?,?,?B,?B,D   ; interactions between facets 3 and 4, gender and item length
*
Labels =
1, Children   ; Children are facet 1
1-17   = 1 Boy,,1 ; Boys, in group 1, are numbered 1 through 17. Gender indicated by first
       character in label
18-35 = 2 Girl,,2 ; Girls, in group 2, are numbered 18 through 35.
*    ; end of child labels for facet 1
2,Tapping items ; Items are facet 2
1   = 2 1-4   ; Items labelled by the order in which the four blocks are tapped
2   = 2 2-3   ; first character in label is the number of taps in the tapping pattern
3   = 3 1-2-4
4   = 3 1-3-4
5   = 3 2-1-4
6   = 3 3-4-1
7   = 4 1-4-3-2
8   = 4 1-4-2-3
```

```
9  = 4 1-3-2-4
10 = 4 2-4-3-1
11 = 5 1-3-1-2-4
12 = 5 1-3-2-4-3
13 = 5 1-4-3-2-4
14 = 6 1-4-2-3-4-1
15 = 6 1-3-2-4-1-3
16 = 6 1-4-2-3-1-4
17 = 6 1-4-3-1-2-4
18 = 7-1-3-4-2-1-4
*
3 = Gender, A  ; Dummy facet for child gender, only for interactions
1 = Boy, 0  ; elements anchored at 0 so do not enter into the measure estimation
2 = Girl, 0
*
4 = Number of Taps, A ; Dummy facet for number of taps in an item tapping pattern
2 = 2 taps, 0
3 = 3 taps, 0
4 = 4 taps, 0
5 = 5 taps, 0
6 = 6 taps, 0
7 = 7 taps, 0
*
Dvalues = ; The child gender and number of taps are in the labels, so can be omitted from the
       data
3, 1, 1, 1 ; Child gender element for Facet 3 is indicated in the label of Facet 1, column 1
       length 1
4, 2, 1, 1 ; Item length element for Facet 4 is indicated in the label of Facet 2, column 1
       length 1
*

; in the data:
; facet 1 element number
; facet 2 element number (a range)
; facet 3 not in data, obtained from Dvalues=
; facet 4 not in data, obtained from Dvalues=
; observations (as many as match range)

Data =  ; no data file name, so data follows immediately in this file
1 ,1   ,1 ; child 1 on item 1 scored 1: Facet elements for facets 3 and 4 provided in Dvalues=
1 ,2-18,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0
2 ,1-18,1,1,1,1,1,1,0,0,1,1,1,0,0,1,0,0,0,0
3 ,1-18,1,1,1,1,1,1,1,1,1,1,0,1,0,0,0,0,0,0,0
4 ,1-18,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0
5 ,1-18,1,1,1,1,1,1,1,1,1,1,1,0,0,1,1,0,0,0,0
6 ,1-18,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0
7 ,1-18,1,1,1,1,1,0,0,1,0,1,0,0,0,0,0,0,0,0,0
8 ,1-18,1,1,1,1,1,1,0,1,0,1,1,0,0,0,0,0,0,0,0
9 ,1-18,1,1,1,1,1,1,1,1,1,1,1,1,0,0,1,0,0,0,0,0
10,1-18,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0
11,1-18,1,1,1,1,1,1,1,1,1,1,1,1,0,0,1,0,0,0,0,0
12,1-18,1,1,1,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0
13,1-18,1,1,1,1,1,1,1,1,1,1,0,1,0,0,0,0,0,0,0,0
14,1-18,1,1,1,1,1,1,1,1,1,1,1,0,1,0,1,0,0,0,0
15,1-18,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0
16,1-18,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
17,1-18,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0
18,1-18,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0
19,1-18,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0
20,1-18,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0
21,1-18,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0
22,1-18,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0
23,1-18,1,1,1,1,1,0,0,1,0,0,1,0,0,0,0,0,0,0,0
24,1-18,1,1,1,1,1,1,0,0,1,1,1,1,1,0,0,0,0,0,0
```

```
25,1-18,1,1,1,0,1,1,1,1,1,0,0,0,0,0,0,0,0,0
26,1-18,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0
27,1-18,1,1,1,1,1,1,1,1,1,1,1,1,0,1,0,0,1,1,0
28,1-18,1,1,1,1,1,0,1,1,1,1,1,0,0,0,0,0,0,0,0
29,1-18,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0
30,1-18,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,1,0,0,0
31,1-18,1,1,1,1,1,1,1,1,1,1,1,1,0,1,0,0,0,0,0
32,1-18,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0
33,1-18,1,1,1,1,1,1,1,1,1,1,1,0,0,1,0,0,0,0,0
34,1-18,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0
35,1-18,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0
```

This produces interaction Tables enabling us to investigate the relative performance of the genders on different tapping lengths.

```
'Knox Cube Test with interactions'  06-16-2008 18:35:33
Table 14.2.3.3  Bias/Interaction Pairwise Report (arranged by N).

Bias/Interaction: 3. Gender, 4. Number of Taps
```

| Target N Number | Target Measr | S.E. | Obs-Exp Average | Context N | Gende | Target Measr | S.E. | Obs-Exp Average | Context N | Gende | Target Joint Contrast | S.E. | t | d.f. | Prob. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 3 taps | -.38 | .63 | .02 | 1 | Boy | .30 | .53 | .11 | 2 | Girl | -.68 | .82 | -.82 | 100 | .4130 |
| 4 4 taps | .37 | .42 | -.03 | 1 | Boy | -.33 | .41 | .14 | 2 | Girl | .70 | .59 | 1.19 | 134 | .2370 |
| 5 5 taps | -.07 | .40 | .01 | 1 | Boy | .10 | .43 | -.01 | 2 | Girl | -.17 | .59 | -.29 | 100 | .7708 |
| 6 6 taps | -.26 | .63 | .01 | 1 | Boy | .29 | .68 | .01 | 2 | Girl | -.55 | .93 | -.59 | 134 | .5571 |

## 7.14 Pair.txt: 4 facet judge pairs: Language test

In this Foreign Language reading test, examinees read two text passages. The first passage (chosen at random from many available passages) is heard by a judge. The examinee than moves to an intermediate waiting room. Then the examinee is heard by a different judge on a different, randomly selected, passage. Examinees are assigned to the next available judge at each stage. Rating is on a holistic 1-5 category scale.

Facets specifications and data (in file Pair.txt):

```
Title = Paired-judge Reading Test  ; the report heading line
Facets = 4   ; examinees, texts, raters, time-point
Converge = .5, .1  ; converged when biggest score residual is .5 and biggest change is .1
        logits
Positive = 1, 3  ; examinees are able, raters are lenient
Noncenter = 1   ; examinee measures relative to texts, raters, etc.
Umean = 50, 5   ; user-friendly logit-rescaled measures: logit = 50 + 5 * logit
Arrange = F,m   ; tables output in Fit-ascending (F) and Measure-descending (m) order
Usort = U   ; sort residuals by unexpectedness
Vertical = 1A, 3A, 2A ; Display examinees, raters, texts by name
Model =
?,?,?,?,reading  ; basic model: raters share the 5-point rating scale
?,?B,?,?B,reading  ; text x time-point interactions: does practice help?
*
Rating scale = reading,R9,K ; 5 category scale, K=keep intermediate unobserved categories
1 = incoherent  ; sounding out the words, or worse
5 = fluent   ; as a native speaker
*
Labels=
1, Examinees
12, Stan
14, Amanda
23, Victor
30, Leslie
41, Don
52, Chin
```

```
62, Anne
73, David
82, Jacinta
90, Kallen
99, Ethne
*
2, Passages
1, Buchan
2, Sapper
3, Fleming
4, Doyle
5, Montagu
6, Blyton
*
3, Judges
1, Glass
2, Draper
3, Karnes
4, Borich
5, Chase
6, Gredler
*
4, Order, A  ; these are anchored at umean=50 because for interactions only
1, First, 50
2, Second, 50
*
data=
12,1,1,1, 2 ; Stan reads Buchan to Glass first, and is rated 2
12,3,5,2, 3 ; Stan reads Fleming to Chase second, and is rated 3
14,6,4,1, 1
14,4,6,2, 2
23,1,4,1, 1
23,3,2,2, 3
30,1,5,1, 1
30,5,4,2, 2
41,3,5,1, 4
41,6,4,2, 3
52,4,6,1, 1
52,5,2,2, 3
62,4,3,1, 1
62,5,2,2, 3
73,4,3,1, 4
73,6,2,2, 3
82,1,6,1, 2
82,5,1,2, 5
90,3,1,1, 2
90,2,4,2, 4
99,6,3,1, 3
99,2,1,2, 5
```

In Table 7, the "Fair Average" reports a standard rating for a standard context:

Table 7.1.2 Examinees Measurement Report (arranged by mN).

| Total Score | Total Count | Obsvd Average | Fair-M Avrage | Measure | Model S.E. | Infit MnSq | Infit ZStd | Outfit MnSq | Outfit ZStd | Estim. Discrm | Displ. | Correlation PtMea | Correlation PtExp | Nu | Examinees |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 2 | 2.0 | 2.74 | 45.66 | 7.49 | .06 | -.8 | .08 | -.4 | 1.64 | .17 | .99 | .42 | 23 | Victor |
| 3 | 2 | 1.5 | 2.59 | 43.73 | 6.26 | .20 | -.9 | .23 | -.7 | 1.95 | -.14 | 1.09 | .11 | 14 | Amanda |
| 6 | 2 | 3.0 | 2.14 | 38.57 | 5.46 | .12 | -1.3 | .12 | -1.3 | 2.01 | -.71 | 1.00 | .29 | 90 | Kallen |

Here, Kallen has a higher observed average score of 3.0, but lower "Fair Average" score of 2.14 because he read out easier passages to more lenient judges.

## 7.15 Essays.txt: 4 facet rating scale with bias analysis: Essays

A carefully conducted grading of English essays written for the Advanced Placement Program was conducted by H.I. Braun (Understanding score reliability: Experiments in calibrating essay readers. *Journal of Educational Statistics, Spring 1988, 13/1,* 1-18). The four facets are examinee, essay topic, reader, and grading session. In this analysis, bias interactions between essay topic and reader are reported. Also interactions are reported for reader by grading session, which is entered as a "dummy" facet, with all elements anchored at 0.

Facets specifications and data (in file Essays.txt):

```
; this is file essays.txt
title = AP English Essays (College Board/ETS)
convergence = 0.1 ; size of largest remaining marginal score residual at convergence
unexpected = 3.0 ; size of smallest standardized residual to report
arrange = M ; arrange output tables in Measure ascending order
facets = 4 ; there are 4 facets in this analysis
noncenter = 1 ; examinee facet floats
positive = 1 ; for examinees, greater score = greater measure
Inter-rater = 3 ; facet 3 is the rater facet
usort = 2,3,1 ; sort residuals by 2=Essay, 3=Reader, 1=Examinee
Model=
?,?B,?B,?,R9 ; observations are ratings in range 1-9.
  ; look for interaction/bias between reader and essay type
?,?,?B,?B,R9 ; look for rater x grading session interaction
*
Labels=
1,examinee
 1-32 ; 32 otherwise anonymous examinees
*
2,Essay
 1,A ; 3 essays
 2,B
 3,C
*
3,Reader
 1-12 ; 12 otherwise anonymous readers
*
4,Session,A ; this is a dummy facet, used only for investigating interactions
 11,day 1 time 1 ,0 ; 8 sessions - all anchored at 0
 12,day 1 time 2 ,0
 21,day 2 time 1 ,0
 22,day 2 time 2 ,0
 31,day 3 time 1 ,0
 32,day 3 time 2 ,0
 41,day 4 time 1 ,0
 42,day 4 time 2 ,0
*
data =
05,1,1,11,4 ; first rating: examinee 5, essay 1, reader 1, session 11, rating of 4
 | ; more data
09,1,1,11,3 ; last rating
```

## 7.16 Olympics.txt: 4 facet rating scale with unobserved categories: ice-skating

The Pairs ice-skating competition at the 2002 Winter Olympics in Salt Lake City generated considerable controversy. Allegations were made that the French judge was influenced by outsiders. Ultimately Gold Medals were awarded to both a Russian and a Canadian pair.

Here are the data set up as a 4-facet analysis in Olympics.txt. The original ratings are numbers like 5.4. These have been multiplied by 10 to produce the integer ratings that Facets expects. Not all intermediate categories in the

functional range of 2.9 to 5.9 were observed, but the unobserved intermediate category 3.0 is maintained in the rating scale.

Use the Output Tables menu to look for interactions between judges and skater pairs.

For a two-facet Rasch analysis of these data, investigating rater behavior, see www.winsteps.com/winman/example15.htm

Facets specifications and data (in file Olympics.txt):

```
Title  = "Pairs Skating: Winter Olympics, SLC 2002"
Facets = 4 ; Facets are Judges, Skating Pairs, Programs, Skills
Convergence= 12.0, .001 ; convergence when largest residual is less than 12.0
                        ; and biggest logit change is less than .001
Positive = 1, 2, 3, 4 ; all Facets positive:
Models =
?,?,?,?, ISU ; keep unobserved intermediate categories with scores up to 6.0
*
Rating scale = ISU, R60, Keep ; 'Keep" maintains unobserved intermediate categories in the
         rating scale
59 = 5.9  ; all observed ratings multiplied by 10
*

Labels=
1, Judge ; leniency
1= Rus ;Mrs. Marina SANAIA : RUSSIA
....
*
2, Pair ; ability
 1= BS-Rus ;  1 BEREZHNAYA Elena / SIKHARULIDZE Anton : RUS
 2= SP-Can ;  5 SALE Jamie / PELLETIER David : CAN
....
*
3, Program ; easiness
1 = Short
2 = Free
*
4, Skill ; easiness
1 = Technical Merit
2 = Artistic Impression
*
Dvalues=
1, 1-9  ; Facet 1 is the 9 judges
*
Data =
 1 1 1 58 58 57 58 58 58 58 58 57
 1 1 2 58 58 58 58 59 58 58 58 58
....
```

## 7.17 Essayday.txt: 5 facet rating scale with bias analysis: Essays+Days

Using the Essays dataset, judge-day interactions are now to be examined. Since the day number is already coded in Facet 4, this can be implemented with Dvalues= without reformatting the data.

Facets specifications and data (in file Essayday.txt):

```
; this is file essayday.txt
title = AP English Essays (College Board/ETS)
facets = 5 ; there are 5 facets in this analysis
Model=
```

```
?,?B,?,?,?B, R9 ; look for interaction/bias between reader and day
*
Dvalues=
5, 4, 5, 1 ; day number: facet 5 of each observation from label for element in facet 4
           ; starting in column 5 of the label with a length of one.
*
noncenter = 1 ; examinee facet floats
positive = 1 ; for examinees, greater score = greater measure
inter-rater = 3 ; facet 3 is the rater facet
usort = 2,3,1 ; sort residuals by 2=Essay, 3=Reader, 1=Examinee
convergence = 0.1 ; size of largest remaining marginal score residual at convergence
unexpected = 3.0 ; size of smallest standardized residual to report
arrange = M ; arrange output tables in Measure ascending order

Labels=
1,examinee
 1-32 ; 32 otherwise anonymous examinees
*
2,Essay
 1,A ; 3 essays
 2,B
 3,C
*
3,Reader
 1-12 ; 12 otherwise anonymous readers
*
4,Session,A ; this is a dummy facet
 11,day 1 time 1 ,0 ; 8 sessions - all anchored at 0
 12,day 1 time 2 ,0
 21,day 2 time 1 ,0
 22,day 2 time 2 ,0
 31,day 3 time 1 ,0
 32,day 3 time 2 ,0
 41,day 4 time 1 ,0
 42,day 4 time 2 ,0
*
5, Day, A ; this is a dummy facet, used only for investigating interactions
1, Day one, 0 ; Dvalues= obtains these from Facet 4.
2, Day two, 0
3, Day three, 0
4, Day four, 0
*

data =
05,1,1,11,4 ; first rating: examinee 5, essay 1, reader 1, session 11, rating of 4
 | ; more data
09,1,1,11,3 ; last rating
```

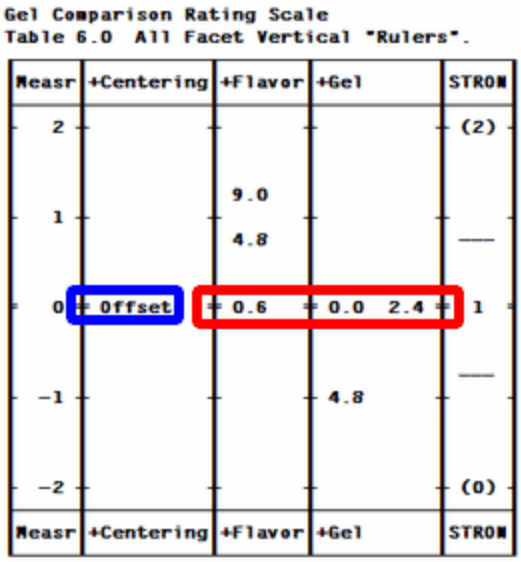
## 7.18 Gel.txt: 7 facet Paired comparison with ties: Flavor Strength of Gels

In order to determine the effect of varying gel and flavor concentration on the subjective evaluation of flavor strength, assessors were asked to state which sample of each pair had the greater flavor strength. A decision of "no difference" was allowed. The data is from A. Springall (Applied Statistics, 22/1, p. 66). This analysis models the comparison as a three category rating scale: 2 = prefer first sample, 1 = tie, 0 = prefer second sample. Enter each datapoint twice in your data: A vs. B and B vs. A. Model weighting is 0.5 to adjust for this.

Note: If a rater facet is used to identify the judge, then anchor the rater elements at 0. Their fit statistics will characterize their behavior relative to the consensus.

Facets specifications and data (in file Gel.txt):

```
title = Gel Comparison Rating Scale ; report heading
facets = 7 ;7 facets in data set up for several analyses
positive = 1,2,3,4 ; measure everything in the same direction
non-center = 1 ; allow for skew in data due to arbitrary selection of "first" and "second"
         samples
entry = 1,4,4,2,3,2,3 ; several pairs in the comparisons
models =

;X,?, X,?,?, X, X,Stronger,1.0 ; to check the raw scores

 ?, ,  ,?,?,-?,-?,Stronger,0.5 ; for the components
;?,?,-?,X,X, X, X,Stronger,0.5 ; for the treatments - commented out this time
*
Rating (or partial credit) scale = Stronger,R2
2 = Prefer first  ; the possible outcomes
1 = Tie
0 = Prefer second
*
Labels =
1,Centering ; adjusts for lack of symmetry in choice of origin
1=Offset ; since Facets 2 & 3 are anchored so that there is
; a clear meaning to the 0 of the scale,
; this element, named "Offset" is specified for every
; observation and allowed to float.
*
2=Flavor,A ; anchoring for the local origin
1=0.6,0 ; choose origin for convenience in comprehension
2=4.8
3=9.0
*
3=Gel,A ; anchoring for local origin
1=0.0,0 ; choose origin for convenience in comprehension
2=2.4
3=4.8
*
4=Treatment ; ignored for main analysis
1= 0.6 0.0 ; name reflects flavor-gel combination
2= 4.8 0.0
 |
9= 9.0 4.8
*
data = ;885 comparisons in total (entered twice, weighted 0.5)
; count, offset, 1st treatment, 2nd treatment, 1st flavor, 1st gel, 2nd flavor, 2nd gel,
         preference
R2,1 ,2,1, 2,1, 1,1, 0 ; 1st treatment 4.8/0.0 compared with 2nd treatment 0.6/0.0, 2nd
         preferred
 |
R8,1 ,9,8, 3,3, 2,3, 2 ; last recorded comparison
; data entered again in reverse sequence
R2,1 ,1,2, 1,1, 2,1, 2 ; 1st treatment 0.6/0.0 compared with 2nd treatment 4.9/0.0, 1st
         preferred
 |
R8,1 ,8,9, 2,3, 3,3, 0 ; last recorded comparison - reversed
```




Anchoring flavor strength of "0.6" at 0 logits, and gel strength "0.0" at 0 logits, gives the measurement system a substantive zero, similar to zero on the Celsius temperature measurement system. Since both flavor and gel facets are anchored, we introduce another facet, "Centering" and its element, "Offset", which compensates for the misalignment between flavor and gel.

## 7.19 Chinese.txt: example of UTF-8 multibyte character codes

*Facets* support Chinese and other character sets (alphabets) in UTF-8 format. Here is example file: chinese.txt

```
; Chinese.txt
title = "Facets Chinese 1!会"
ascii=Y
output=chinese-facets.out.txt
facets=2
models=?B,?B,中英
rating scale=中英,R4
1=时通过
2=投递
3=3
*
utf8sub = "$"
labels=
1, 中英文均可
1, 及时通过
2, 邮箱投
3, 及时
*
2, a中英
1, 及时通过
2, 邮箱投递
3, 123456
*
data=
1,1-2, 3, 4
2,1-2, 1, 2
1,1-2, 3, 3
2,1-2, 4, 2
3,1-2, 3, 4
3,1-2, 3, 3
1-3,3, 1, 2, 1
```

This file was created with Notepad++ and UTF8-encoding.

When this specification and data file is analyzed, we see in the Facets analysis window:

```
....
 Facets = 2
 Labels =
  1,$$$$$ ; (elements = 3)
  2,a$$ ; (elements = 3)
 Opening work files
....
```

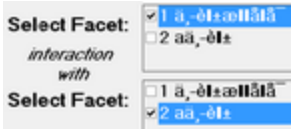
Multibyte characters are not displayed, but are replaced with a substitute ASCII character: $ which can be changed with the specification UTF8sub=

In the Facets output Tables, we see:

```
Facets Chinese 1!会 8/14/2019 2:01:20 PM
Table 4.1 Unexpected Responses (15 residuals sorted by u).

+-------------------------------------------------------------+
| Cat  Score  Exp.  Resd StRes| N 中英文均可     N a中英        |
|-----------------------------+-------------------------------|
|  2     2     1.1    .9  2.2 | 2 邮箱投         3 123456       |
|  4     4     2.4   1.6  1.9 | 2 邮箱投         1 及时通过     |
|  1     1     2.4  -1.4 -1.6 | 2 邮箱投         1 及时通过     |
|  4     4     3.2    .8  1.1 | 1 及时通过       2 邮箱投递      |
```

The vertical bars on the right-hand side do not align exactly. This is because the Chinese characters do not conform with the monospaced font.

In some output and menus, the multibyte characters appear as ANSI characters:




# 8 Building Specification and Data files

## 8.1 Facets requires specifications and data

Facets requires specifications and data. The data are qualitative observations. These are numbered in ascending sequence along the latent variable. For instance:

| **Qualitatively-ordered observations** must be exclusive, extensive, exhaustive | **Numerical (scored, coded) observations:** higher number = more of the latent variable must be 0 or positive integers |
|---|---|
| Yes-No | 1-0 |
| Correct-Incorrect | 1-0 |
| None-Some-All | 0-1-2 |
| Likert scale: Strongly Disagree-Disagree-Neutral-Agree-Strongly Agree | 1-2-3-4-5 |
| Essay rating | 1-2-3-4-5-6-7-8-9 |

| | |
|---|---|
| Attitude survey | 1-2-3-4-5-6-7-8-9-10 |
| Visual-analog scale or Percent | 0-1-2-3-4-….-96-97-98-99-100 |

The specifications define how these data are modeled to be the outcome of a linear combination of elements. Each element belongs to a facet. All facets and elements are numbered. For instance:

| Facet (number) | Elements (numbers) |
|---|---|
| Examinees (1) | Albert (1), Beryl (2), Claudia (3), ... |
| Raters (2) | Prof. Cilic (1), Dr. Nadal (2), Ing. Laver (3), ... |
| Items (3) | Item 1 (1), Item 2 (2), Item 3 (3), ... |
| Time-points (4) | Admission (1), Discharge (2) |

The measurement model becomes:

*Examinee ability + Rater leniency - Item difficulty - Time-point challenge → Observation*

a) **Specifications** are instructions on how to analyze the data. They are in the Facets specification file. They look like:
*Specification = value*
Most specifications have standard default values, suitable for most analyses.

Example 1: an excerpt from a list of specifications:

```
Title = Paired-judge Reading Test  ; the report heading line
Facets = 4   ; facets are 1=examinees, 2=raters, 3=items, 4=time-point
Positive = 1, 2  ; examinees are able, raters are lenient
Noncenter = 1   ; examinee measures float relative to raters, items, time-points
```

b) The **data** are the qualitative observations. Their location is specified by Data=. These can be in the same file as the specifications or in separate files. Each data point, response, observation or rating is analyzed as the interaction of one (or more) elements from each of the specified facets.

Example 2: an excerpt of Facets data:
Examinee 23 is rated by rater 4 on item 6 at time-point 2 with a rating of 7

```
Data=
23, 4, 6, 2, 7
```

Example 3: Here is a typical data file, after data collection:
123 A 25467 ; Examinee 123 is rated by judge A on items 1-5 with ratings 2, 5, 4, 6, 7
452 C 34245 ; Examinee 452, judge C, ratings 3, 4, 2, 4, 5
312 B 34452 ; Examinee 312, judge B, ratings 3, 4, 4, 5, 2
123 B 24443 ; Examinee 123, judge B, ratings 2, 4, 4, 4, 3

The Facets specification and data file:

| Specification | Explanation |
|---|---|
| `Title = Sample Analysis` | The title of the analysis appears on every output report. |
| `Facets = 3` | There are 3 facets: examinees, judges, items. |
| `Model = ?,?,?,R` | The facets measurement model is: any examinee (?,) can combine with any rater (?,) and with any item (?,) to produce a rating (R) |

| | |
|---|---|
| Noncenter = 1 | Center the elements all the facets except facet 1. This established the zero-point (local origin) of the measurement frame-of-reference. |
| Positive = 1, 2 | Facets with "higher score = higher measure" (ability, leniency, easiness) are positive.<br>Facets with "higher score = lower measure" (disability, severity, difficulty) are negative. |
| Umean = 0, 1 | The user-defined mean for the measures is zero. The user-defined scaling is 1. The reported measures will be in logits. |
| Labels = | The list of labels (identifiers) follows: |
| 1,Examinees<br>123 =<br>452 =<br>312 =<br>* | Facet 1 is the Examinees.<br>123 is the first examinee element.<br>453 is the second examinee<br>312 is another examinee<br>* is the end of Facet 1 |
| 2,Judges<br>1 = A<br>2 = B<br>3 = C<br>* | Facets 2 is the Judges<br>A is the label for the first judge, with element number 1. |
| 3,Items<br>1-5<br>* | Facet 3 is the Items.<br>The item element numbers are 1, 2, 3, 4, 5 |
| Data= | The data follows, or the data can be in a separate file: Data = filename.txt |
| 123, 1, 1-5, 2, 5, 4, 6, 7 | (comma-separated) Examinee 123 is rated by judge 1 (A) on items 1-5 with ratings 2, 5, 4, 6, 7 |
| 452 3 1-5 3 4 2 4 5 | (blank separated) Examinee 452, judge 3 (C), items 1-5, ratings 3, 4, 2, 4, 5 |
| 312→2→1-5→3→4→4→5→2 | (→tab separated) Examinee 312, judge2 (B), items 1-5, ratings 3, 4, 4, 5, 2 |
| 123, B, 1-5, 2, 4, 4, 4, 3 | Element labels are also allowed in the data. They will be automatically replaced by their element numbers |
| ... more data | |

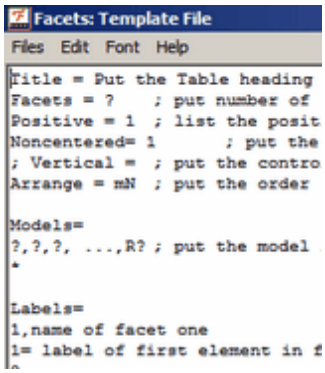
**Look for a problem like yours in our examples. Then copy-and-edit it!**

You can also use the template, Excel, SAS, SPSS, STATA and Facform.

*End of Introductory Overview: Now start on your own! Run one of the Example files: how about the Diving Competition?*

## 8.2 Template for specifications and data

*The Template File:* a skeleton specification and data file is provided to guide you in the construction of your own specification and data file.
This can be accessed from the Edit menu, "Template".
Edit it to suit your own needs. Here is Template.txt:




```
Title = Put the Table heading title here
Facets = ?  ; put number of facets here
Positive = 1 ; list the positively oriented facets here
Noncentered= 1 ; put the one (usually) floating facet here
; Vertical = ; put the control for the "rulers" in Table 6 here, if not a standard layout is
shown.
Arrange = mN ; put the order for the measure Table 7 here

Models=
?,?,?, ...,R? ; put the model statement for your Facets here.
*

Labels=
1,name of facet one
1= label of first element in facet one
2 =

*
2, name of facet two
1= label of first element in facet two
*

data=
; enter in format:
; element number for facet one, element number for facet two, , , , observation
```

## 8.3 Excel data file

Excel datasets, with their data correctly formatted, can be specified directly to Facets.

Data= Exceldata.xls

Dvalues= is useful for constant values in the data

**Excel-format data files.**

Excel worksheets, with their data correctly formatted, can be specified directly to Facets.

```
Data = Excel filename.xls
Data = Creativity.xls
```

or, if the data worksheet is not the first worksheet, usually Sheet,

```
Data = Excel filename.xls(sheet number)
Data = Creativity.xls(2)
```

Dvalues= is useful for constant values in the data

Facets analyzes integer data in the range 0-255. Excel is a good means of reformatting your data, if necessary.

(a) Set up the Excel spreadsheet as follows:

Imagine a three-facet situation: judges, examinees and items.

In Row 1, put ;judge - the semi-colon is so that Facets will treat this as a comment. Then examinees and items

If you see:

```
First active data line is
judge examinee items ratings
             Processed as judge, examinee, items, ratings, ,,
>Check (2)? Invalid datum location R0,0,0,0, at or near line 1. Datum "" is too big or not a
positive integer, treated as missing.
```

Then change "judge" to ";judge" to make the heading row a comment for Facets.

In column A, put the judge number

In column B, put the examinee number

Either:

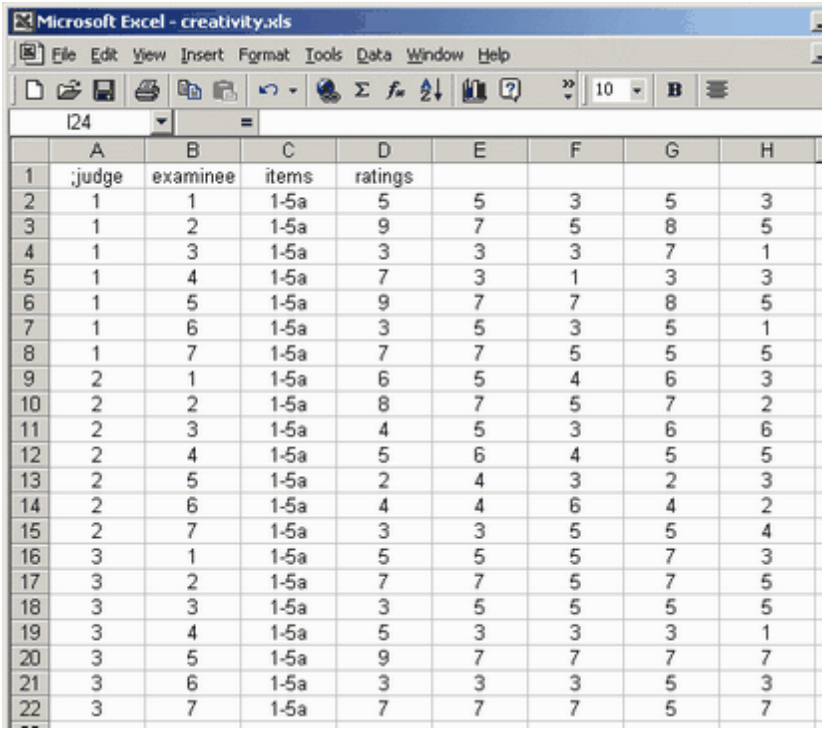
In column C, put the item number range as a character constant, e.g., 1-5
to enter this as text, type: '1-5 , i.e., apostrophe 1-5
If this doesn't work,
type 1-5a the "a" will be ignored by Facets
or
type 1_5 an underline can be used instead of a minus sign.

In columns D-H put the judges' ratings of the examinee on items 1,2,3,4,5
Entries beyond column H will be ignored.

Or:

Alternatively, put the judge number in columns A, the examinee number in column B, include in the specifications:
Dvalues = 3, 1-5 ; place 1-5 in data facet 3
and in columns C-G put the judges' ratings of the examinee on items 1,2,3,4,5

Here is what this looks like for the Creativity data. It is in a file called Creativity.xls.




| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | ;judge | examinee | items | ratings | | | | |
| 2 | 1 | 1 | 1-5a | 5 | 5 | 3 | 5 | 3 |
| 3 | 1 | 2 | 1-5a | 9 | 7 | 5 | 8 | 5 |
| 4 | 1 | 3 | 1-5a | 3 | 3 | 3 | 7 | 1 |
| 5 | 1 | 4 | 1-5a | 7 | 3 | 1 | 3 | 3 |
| 6 | 1 | 5 | 1-5a | 9 | 7 | 7 | 8 | 5 |
| 7 | 1 | 6 | 1-5a | 3 | 5 | 3 | 5 | 1 |
| 8 | 1 | 7 | 1-5a | 7 | 7 | 5 | 5 | 5 |
| 9 | 2 | 1 | 1-5a | 6 | 5 | 4 | 6 | 3 |
| 10 | 2 | 2 | 1-5a | 8 | 7 | 5 | 7 | 2 |
| 11 | 2 | 3 | 1-5a | 4 | 5 | 3 | 6 | 6 |
| 12 | 2 | 4 | 1-5a | 5 | 6 | 4 | 5 | 5 |
| 13 | 2 | 5 | 1-5a | 2 | 4 | 3 | 2 | 3 |
| 14 | 2 | 6 | 1-5a | 4 | 4 | 6 | 4 | 2 |
| 15 | 2 | 7 | 1-5a | 3 | 3 | 5 | 5 | 4 |
| 16 | 3 | 1 | 1-5a | 5 | 5 | 5 | 7 | 3 |
| 17 | 3 | 2 | 1-5a | 7 | 7 | 5 | 7 | 5 |
| 18 | 3 | 3 | 1-5a | 3 | 5 | 5 | 5 | 5 |
| 19 | 3 | 4 | 1-5a | 5 | 3 | 3 | 3 | 1 |
| 20 | 3 | 5 | 1-5a | 9 | 7 | 7 | 7 | 7 |
| 21 | 3 | 6 | 1-5a | 3 | 3 | 3 | 5 | 3 |
| 22 | 3 | 7 | 1-5a | 7 | 7 | 7 | 5 | 7 |

Columns to the right of the data are usually ignored by Facets.

(b) Specify this Excel worksheet directly with the Data=

```
Data = Creativity.xls
```

or, **particularly if Windows reports "Low memory" or "Out of memory",**

Use "save as" CSV (Comma-delimited) (*.csv) to obtain a file in the Data= format, e.g, creativity.csv

```
;examinee,;judge,items,ratings,,,,
1,1,1-5,5,5,3,5,3
2,1,1-5,9,7,5,8,5
3,1,1-5,3,3,3,7,1
4,1,1-5,7,3,1,3,3
5,1,1-5,9,7,7,8,5
6,1,1-5,3,5,3,5,1
7,1,1-5,7,7,5,5,5
1,2,1-5,6,5,4,6,3
2,2,1-5,8,7,5,7,2
3,2,1-5,4,5,3,6,6
4,2,1-5,5,6,4,5,5
5,2,1-5,2,4,3,2,3
6,2,1-5,4,4,6,4,2
7,2,1-5,3,3,5,5,4
1,3,1-5,5,5,5,7,3
2,3,1-5,7,7,5,7,5
3,3,1-5,3,5,5,5,5
4,3,1-5,5,3,3,3,1
5,3,1-5,9,7,7,7,7
6,3,1-5,3,3,3,5,3
7,3,1-5,7,7,7,5,7
```

c) For the specifications, set up an outline skeleton of the specification file:

```
Title="Creativity using Excel"
Facets=3
Positive=1,2,3
Noncenter=1
Model=?,?,?,R9
Labels=Build  ; this will tell Facets to build a list of elements
1,Judges
*
2,Examinees
*
3,Items
1=Attack
2=Basis
3=Clarity
4=Daring
5=Enthusiasm
*
Data=creativity.xls
```

d) Save your specification file as Cspec.txt and analyze it with Facets.

The output reports:

```
Table 2. Data Summary Report.
....
```

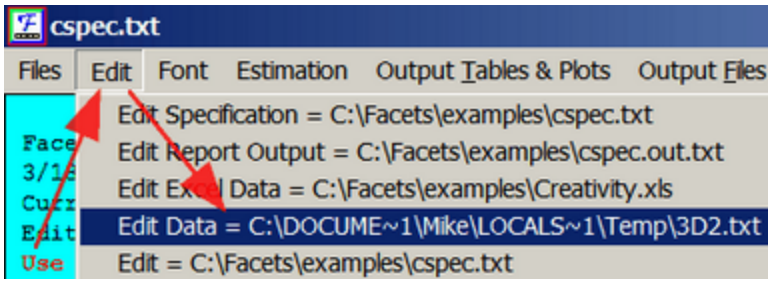
```
This list is generated by Labels=Build. Please copy and paste into your specification file,
       where needed
Labels=
 1, Judges,    ; facet 1
 1 =
 2 =
 3 =
 *
 2, Examinees,    ; facet 2
 1 =
 2 =
 3 =
 4 =
 5 =
 6 =
 7 =
 *
```

e) Open Cspec.txt and paste the extra element numbers into the Labels=list

```
Labels=
 1, Judges,    ; facet 1
 1 =
 2 =
 3 =
 *
 2, Examinees,    ; facet 2
 1 =
 2 =
 3 =
 4 =
 5 =
 6 =
 7 =
 *
3,Items
1=Attack
2=Basis
3=Clarity
4=Daring
5=Enthusiasm
*
```

f) Save your revised Cspec.txt specification file, and analyze it with Facets.
The analysis should report correctly.

g) If the data appear to be incorrect, verify that Facets has imported your Excel datasheet correctly, look at the temporary "Data =" file in the Edit pull-down menu.



## 8.4 R data file

R datasets, with their data correctly formatted, can be specified directly to Facets.

Data= Rdataset.rda
Dvalues= is useful for constant values in the data

Here is an example of data in an R data file.

Please edit Guilford.txt so that the following specifications are not commented ";"

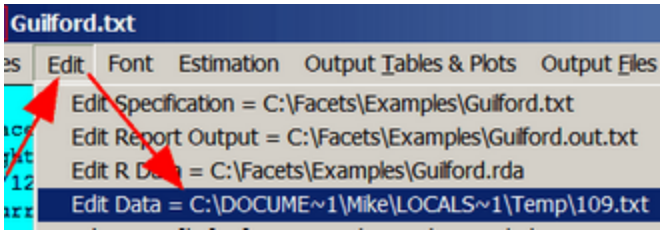
```
dvalues = 3, 1-5
data = Guilford.rda ; R file omitting 1-5 for the 3rd facet
```

Comment ";" all the other data= specifications.

Launch Facets and analyze *Guilford.txt*. You should see:

```
Table 2. Data Summary Report
Assigning models to "Guilford.rda"
Importing datafile= C:\Facets\Examples\Guilford.rda
Waiting for imported datafile ...
Continuing ...
First active data line is: 1,1,5,5,3,5,3
             Processed as: 1, 1, 1-5, 5,5,3,5,3
>============================================================================<
Total lines in data file = 22
Total data lines = 22
Responses matched to model: ?B,?B,?,CREATIVITY,1 = 105
```

If the data appear to be incorrect, verify that Facets has imported your R Statistics data file correctly, look at the temporary "Data =" file in the Edit pull-down menu.




We see a text-format version of the R file. The values are comma-separated.

```
;"Judge","Examinee","Item.1","Item.2","Item3","Item.4","Item.5"
1,1,5,5,3,5,3
1,2,9,7,5,8,5
1,3,3,3,3,7,1
1,4,7,3,1,3,3
1,5,9,7,7,8,5
1,6,3,5,3,5,1
1,7,7,7,5,5,5
2,1,6,5,4,6,3
2,2,8,7,5,7,2
2,3,4,5,3,6,6
2,4,5,6,4,5,5
2,5,2,4,3,2,3
2,6,4,4,6,4,2
2,7,3,3,5,5,4
3,1,5,5,5,7,3
3,2,7,7,5,7,5
```

```
3,3,3,5,5,5,5
3,4,5,3,3,3,1
3,5,9,7,7,7,7
3,6,3,3,3,5,3
3,7,7,7,7,5,7
```

## 8.5 SAS data file

SAS datasets, with their data correctly formatted, can be specified directly to Facets.

Data= SASdata.sas7bdat
Dvalues= is useful for constant values in the data

Facets analyzes integer data in the range 0-255.

Follow the format of SASdata.sas7bdat:
This data set has seven facets: Id, Female, Race, SES, Schtyp, Prog, Items
The five Items are: Read, Write, Math, Science, Socst
There are 11 SAS variables: id, female, race, ses, schtyp, prog, read, write, math, science, socst
The data are dichotomous: are on a rating scale 0-100

Here is SASspecs.txt, the Facets specification file which builds the list of element labels from SASdata.sas7bdat:

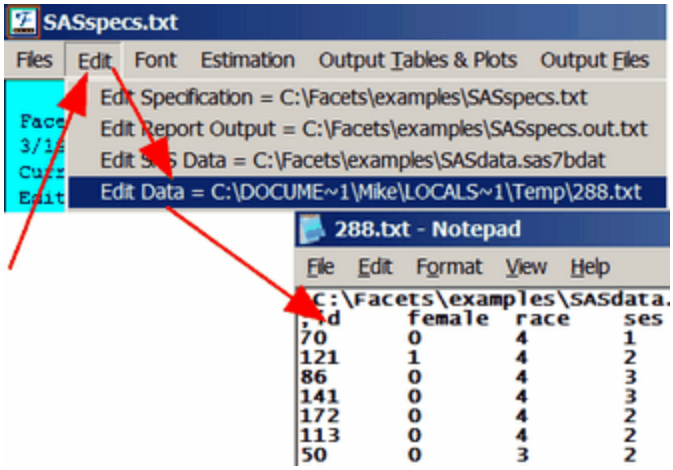
```
Title = "Facets analysis with SAS file: Build list of element labels"
Facets = 7  ; Id, Female, Race, SES, Schtyp, Prog, Items
Models = ?, ?, ?, ?, ?, ?, ?, R100
Positive = 1, 2, 3, 4, 5, 6, 7 ; all positive
Noncentered = 5 ; we want Schtyp to float
Umean = 100, 50 ; the user mean = 100, user scale = 50 units per logit
Yardstick = 0, 5 ; 5 rows per user unit in Table 6: rulers.
Labels=Build
1, Id, D ; a dummy facet for this analysis
*
2, Female
*
3, Race
*
4, SES
*
5, Schtype
*
6, Prog
*
7, Item
1 = Read
2 = Write
3 = Math
4 = Science
5 = Socst
*
Dvalues=*
7, 1-5  ; the data reference for facet 5 is always items 1 to 5
*
Data = SASdata.sas7bdat ; Example dataset from www.ats.ucla.edu
```

After a Facets analysis, the output file SASspecs.out.txt contains:

```
Table 2. Data Summary Report.
.....
This list is generated by Labels=Build. ....
Labels=
 1, Id, D ; facet 1
```

```
 1 =
 2 =
.....
 199 =
 200 =
 *
 2, Female,    ; facet 2
 0 =   ; this is ignored. It is Keepasnull=
 1 =
 *
 3, Race,    ; facet 3
 1 =
 2 =
 3 =
 4 =
 *
 4, SES,    ; facet 4
 1 =
 2 =
 3 =
 *
 5, Schtype,    ; facet 5
 1 =
 2 =
 *
 6, Prog,    ; facet 6
 1 =
 2 =
 3 =
 *
```

If this fails to produce the correct list of labels, check that Facets has input the SAS file correctly:
Click on: "Edit" menu, "Edit Data=".
The contents of the SAS file are displayed in tab-separated format:




Copy-and-paste the Labels= into SASspecs.txt and save it with a new name: SASgood.txt
We also want to change the Keepasnull= value to an unused element number, 999.

```
Title = "Facets analysis with SAS file: Build list of element labels"
Facets = 7   ; Id, Female, Race, SES, Schtyp, Prog, Items
Models = ?, ?, ?, ?, ?, ?, ?, R100
Positive = 1, 2, 3, 4, 5, 6, 7 ; all positive
Noncentered = 5 ; we want Schtyp to float
Umean = 100, 50 ; the user mean = 100, user scale = 50 units per logit
Yardstick = 0, 5 ; 5 rows per user unit in Table 6: rulers.
```

```
Keepasnull = 999 ; we want element 0 to be an active element
Labels=
 1, Id, D ; facet 1
 1 =
 2 =
.....
 199 =
 200 =
 *
 2, Female,    ; facet 2
 0 =
 1 =
 *
 3, Race,    ; facet 3
 1 =
 2 =
 3 =
 4 =
 *
 4, SES,    ; facet 4
 1 =
 2 =
 3 =
 *
 5, Schtype,    ; facet 5
 1 =
 2 =
 *
 6, Prog,    ; facet 6
 1 =
 2 =
 3 =
 *
Dvalues=*
7, 1-5  ; the data reference for facet 5 is always items 1 to 5
*
Data = SASdata.sas7bdat ; Example dataset from www.ats.ucla.edu/stat/
```

Analysis of SASgood.txt is successful:

```
Table 6.0  All Facet Vertical "Rulers".

Vertical = (1*,2A,3A,4A,5A,6A,7A,S) Yardstick (columns lines low high extreme)=
0,5,97,103,End
+-------------------------------------------------------------------------------+
|Measr|+Id         |+Female|+Race|+SES|+Schtype|+Prog|+Item             |Scale|
|-----+-------------+-------+-----+----+--------+-----+------------------+-----|
| 103 +             +       +     +    +        +     +                  +(76) |
|     |             |       |     |    |        | 2   |                  | --- |
|     |             |       |     |    |        |     |                  |     |
|     |             |       |     |    |        |     |                  |  57 |
|     |             |       |     |    |        |     |                  | --- |
| 102 +             +       + 4   +    +        +     +                  +     |
|     |             |       |     |    |        |     |                  |  56 |
|     |             |       | 2   |    |        |     |                  |     |
|     |             |       |     | 3  |        |     |                  | --- |
|     |             |       |     |    |        |     |                  |  55 |
| 101 +             +       +     +    +        +     +                  +     |
|     |             |       |     |    |        |     |                  | --- |
|     |             |       |     |    |        |     |                  |     |
|     |             |       |     |    |        |     |                  |  54 |
|     |             | 1     |     |    |        |     | Math     Write   | --- |
| 100 + ********** +       +     + 2  +        + 1   + Socst            +     |
```

```
|     |            | 0     |     |     |        |     | Read            |  53 |
|     |            |       |     |     |        |     | Science         | --- |
|     |            |       |     |     |        |     |                 |     |
|     |            |       |     |     |        |     |                 |  52 |
|  99 +            +       +     +     +        +     +                 +     |
|     |            |       |     |     |        |     |                 | --- |
|     |            |       |     | 1   |        |     |                 |  51 |
|     |            |       | 3   |     |        |     |                 |     |
|     |            |       | 1   |     |        |     |                 | --- |
|  98 +            +       +     +     +        +     +                 +  50 |
|     |            |       |     |     |        |     |                 |     |
|     |            |       |     |     | 1      |     |                 | --- |
|     |            |       |     |     |        |     |                 |  49 |
|     |            |       |     |     |        | 3   |                 |     |
|  97 +            +       +     +     + 2      +     +                 +(26) |
|-----+------------+-------+-----+-----+--------+-----+-----------------+-----|
|Measr| * = 20     |+Female|+Race|+SES|+Schtype|+Prog|+Item            |Scale|
+-----------------------------------------------------------------------------+
```

## 8.6 SPSS data file

SPSS datasets, with their data correctly formatted, can be specified directly to Facets.

Data= SPSSdata.sav

Dvalues= is useful for constant values in the data

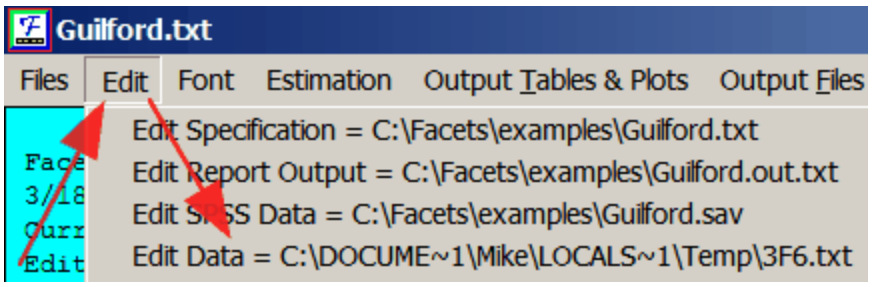
Here is an example of data in an SPSS file.

Please edit Guilford.txt so that the following specifications are not commented ";"
dvalues = 3, 1-5
data = Guilford.sav ; SPSS file omitting 1-5 for the 3rd facet
Comment ";" all the other data= specifications.

Launch Facets and analyze Guilford.txt. You should see

```
Table 2. Data Summary Report
Assigning models to "Guilford.sav"
Importing datafile= C:\Facets\examples\Guilford.sav
Waiting for imported datafile ...
Continuing ...
First active data line is: 1 1 5 5 3 5 3
              Processed as: 1, 1, 1-5, 5,5,3,5,3
>=============================================================================<
Total lines in data file = 24
Total data lines = 24
Responses matched to model ?B,?B,?,CREATIVITY,1 = 105
```

If the data appear to be incorrect, verify that Facets has imported your SPSS data file correctly, look at the temporary "Data =" file in the Edit pull-down menu.




We see a text-format version of the SPSS file. The values are tab-separated:

```
; c:\facets\examples\guilford.sav guilford Ratings of Scientists (Psychometric Methods p.282
        Guilford 1954)
; Cases: 21 Variables: 7 Date: 5/15/2008 1:17:43 PM
;Judge Examinee Attack Basis Clarity Daring Enthusiasm
;Judge number Examinee Number Item 1. Attack Item 2. Basis Item 3. Clarity Item 4. Daring Item
        5. Enthusiasm
1 1 5 5 3 5 3
1 2 9 7 5 8 5
1 3 3 3 3 7 1
1 4 7 3 1 3 3
1 5 9 7 7 8 5
1 6 3 5 3 5 1
1 7 7 7 5 5 5
2 1 6 5 4 6 3
2 2 8 7 5 7 2
2 3 4 5 3 6 6
2 4 5 6 4 5 5
2 5 2 4 3 2 3
2 6 4 4 6 4 2
2 7 3 3 5 5 4
3 1 5 5 5 7 3
3 2 7 7 5 7 5
3 3 3 5 5 5 5
3 4 5 3 3 3 1
3 5 9 7 7 7 7
3 6 3 3 3 5 3
3 7 7 7 7 5 7
```

## 8.7 STATA data file (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 Facets, first in STATA,**
**. saveold filename13, version(13)**
Then Data=filename13.dta can be processed by Facets.

STATA datasets, with their data correctly formatted, can be specified directly to Facets.
Data = STATAdata.dta
Dvalues= is useful for constant values in the data

Facets analyzes integer data in the range 0-255.

Follow the format of STATAdata.dta:
This data set has three facets: States, Years, Items
The two Items are "pop", and "area".
There are four STATA variables: state, year, pop, area.
The data are dichotomous: 0, 1

Here is STATAspecs.txt, the Facets specification file which builds the list of element labels from STATAdata.dta:

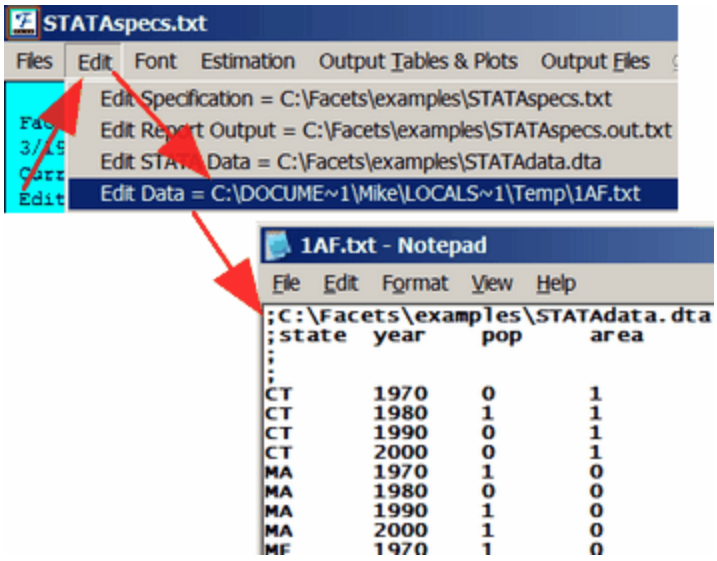
```
Title = "Facets analysis with Stata file: Build list of element labels"
Facets = 3  ; State, Year, Item
Models = ?, ?, ?, D
Positive = 1
Noncentered = 1
Labels=Build
1, State
*
2, Year
*
3, Item
```

```
1, Pop
2, Area
*
Dvalues=*
3, 1-2  ; the data reference for facet 3 is always items 1 and 2
*
Data = STATAdata.dta ; Example dataset from www.stata-press.com
```

After a Facets analysis, the output file STATAspecs.out.txt contains:

```
Table 2. Data Summary Report.
.....
This list is generated by Labels=Build. ....
Labels=
 1, State,    ; facet 1
 1 = CT
 2 = MA
 3 = ME
 4 = NH
 5 = RI
 6 = VT
 *
 2, Year,    ; facet 2
 1970 =
 1980 =
 1990 =
 2000 =
 *
```

If this fails to produce the correct list of labels, check that Facets has input the STATA file correctly:
Click on: "Edit" menu, "Edit Data =".
The contents of the STATA file are displayed in tab-separated format:




Copy-and-paste the Labels= into STATAspecs.txt and save it with a new name: STATAgood.txt

```
Title = "Facets analysis with Stata file: element labels added"
Facets = 3  ; State, Year, Item
Models = ?, ?, ?, D
Positive = 1
Noncentered = 1
Labels=
 1, State,    ; facet 1
```

```
 1 = CT
 2 = MA
 3 = ME
 4 = NH
 5 = RI
 6 = VT
 *
 2, Year,    ; facet 2
 1970 =
 1980 =
 1990 =
 2000 =
 *
3, Item
1, Pop
2, Area
*
Dvalues=*
3, 1-2  ; the data reference for facet 3 is always items 1 and 2
*
Data = STATAdata.dta ; Example dataset from www.stata-press.com
```

Analysis of STATAgood.txt is successful:

```
"Facets analysis with Stata file: element labels added"
Table 6.0  All Facet Vertical "Rulers".

Vertical = (1A,2A,3A,S) Yardstick (columns lines low high extreme)= 0,4,-2,2,End
+---------------------------------------+
|Measr|+State  |-Year       |-Item       |
|-----+--------+------------+------------|
|   2 + NH     +            +            |
|     |        |            |            |
|     |        |            |            |
|     | VT     |            |            |
|   1 +        +            +            |
|     |        |            |            |
|     | CT     | 2000       |            |
|     |        |            |            |
*   0 *        * 1980       * Area  Pop  *
|     |        | 1970  1990 |            |
|     | MA  ME |            |            |
|     |        |            |            |
|  -1 +        +            +            |
|     |        |            |            |
|     |        |            |            |
|     |        |            |            |
|  -2 + RI     +            +            |
|-----+--------+------------+------------|
|Measr|+State  |-Year       |-Item       |
+---------------------------------------+
```

## 8.8 Text data file

Text (.txt) datasets, with their data correctly formatted, can be specified directly to Facets.

Data= Textdata.txt

Dvalues= is useful for constant values in the data

For more information, please see Data=

## 8.9 Facform data formatter

The Facform program is provided to help you with awkward file conversions. It can be downloaded from www.winsteps.com/facets.htm or it may be already installed in the Facets folder: Facform.exe with user-manual Facform.pdf.

*Example: A three-facet flat data file*
Here is the data for an assessment of patient functioning. The performances of ten patients on a specific task were videotaped and rated by a number of raters on 70 items of performance. These data are in file "PATDAT.txt":

```
001BMT67MWP28                          5555555555555  555  354545555555551          555 55 555555
       35554
001DCS67MWP28                          55555 555555 3 555  355445555555553        555 55 555 55
         55554
|
(and so on down to)
|
010MMO                           555  55555  55555  5 555  555 555 5555  55555 5  5555 5  55555
         555 5
010PBG                          45554 555554 555554 5 5554 554 555 55555 5555545   555 54 555554
         554 5
^  ^                            ^
1  4                            31  Columns
```

The keyword list (in file PATIENTkey.txt) converts the flat file into a Facets file:

```
; File: PATkey.txt
; This converts a flat data file into a Facets file (Comments start with ;)
$Input = PATDAT.txt  ; the flat file
$Spoutput=PATCON.txt ; the file for Facets specifications
$Output = PATFAC.txt ; the file of Facets-formatted data
; three facets
$Facets=3
; labels for the facets
$Flabel=1,clients
$Flabel=2,raters
$Flabel=3,tasks
; client id number
$Label=1,$S1E3    client number starts in column 1, ends in column 3
; rater label (non-numeric)
$Label=2,,$S4E6    rater "name" starts in column 4, ends in column 6
; count of task number
task=0      ; a user-defined variable to keep track of task number
; location in data record
pointer=30 ;    a user-defined variable to keep track of location in record
; 70 tasks
$Do=70  ; perform the  $Do-$Again loop 70 times
task=task+1 ; increment our "task" number
pointer=pointer+1 ; increment location in record
; current task number
$Label=3,task ; we want this to be the task element number
; current rating
$Rating=$S(pointer)W1 ; rating at location "pointer" has a width of 1 character
$Again
```

After running the following at the DOS prompt:
C:>FACFORM PATIENTkey.txt<

The Facets-format output file is PATCON.txt

```
; FACFORM Version No. 1.21
```

```
; from Keyword file: PATkey.txt
Facets = 3
Data file = PATFAC.txt
Labels =
1,clients
1=
2=
|
(and so on down to)
|
9=
10=
*
2,raters
1=BMT ; Facform assigns element numbers to these "names"
2=DCS
|
(and so on down to)
|
14=REP
15=RLP
*
3,tasks
1=
2=
|
(and so on down to)
|
69=
70=
*
```

The resulting Facets specification data file is PATFAC.txt:
Facform has combined observations in sequence by element number into single data records

```
1,1,8-
       70,5,5,5,5,5,5,5,5,5,5,5,5,5,,,5,5,5,,,3,5,4,5,4,5,5,5,5,5,5,5,5,5,1,,,,,,,,,,5,5,5,,5,5
       ,,5,5,5,5,5,5,,3,5,5,5,4
|
(and so on down to)
|
10,13,1-
       70,4,5,5,5,4,,5,5,5,5,5,4,,5,5,5,5,5,4,,5,,5,5,5,4,,5,5,4,,5,5,5,,5,5,5,5,5,,5,5,5,5,5,4
       ,5,,,,5,5,5,,5,4,,5,5,5,5,5,4,,5,5,4,,5
```

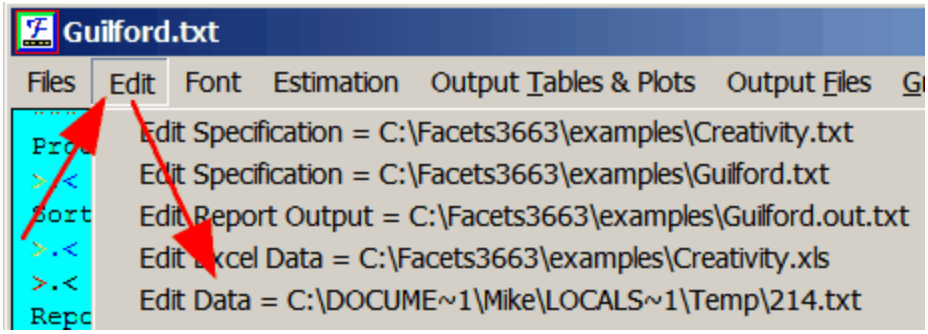
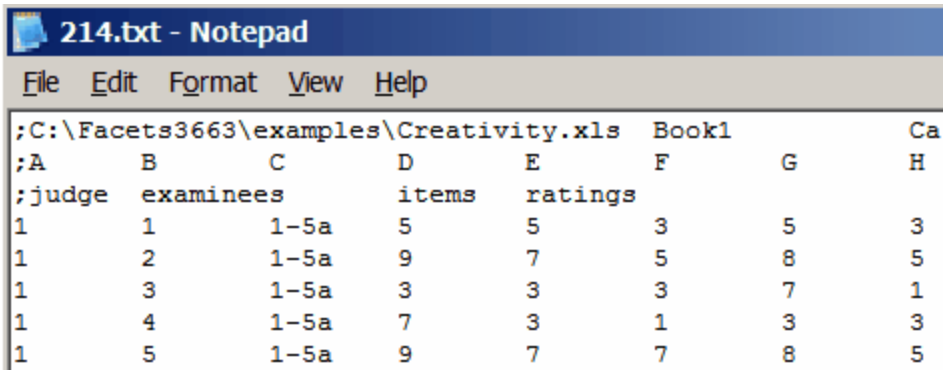
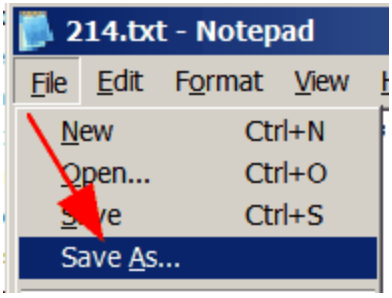
## 8.10 Waiting for imported datafile

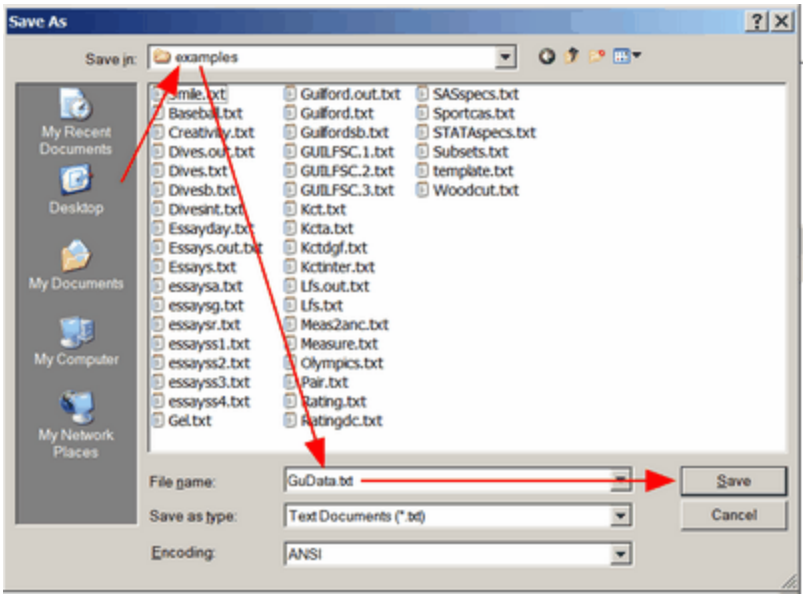
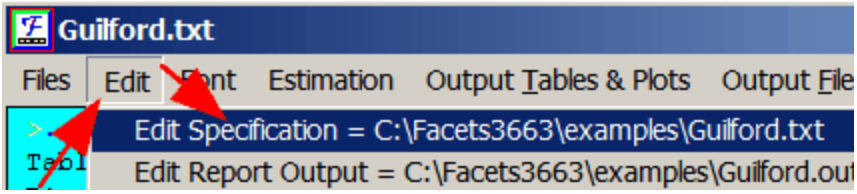
Facets can import data in many different file formats, but some of these require that the file-format's originating software be launched to decode and transmit the data to Facets. This can be a slow process. So we see messages such as:
*Waiting for imported datafile ...*

This wait can be avoided on subsequent analyses of the same dataset if the data is saved in Facets format, and then accessed directly.

Example: Guilford.txt is has a data file in Excel format.

| Guilford.txt | `; 3. Data from an Excel spreadsheet`<br>`Data= Creativity.xls  ; Excel file with 1-5` |
|---|---|

| | |
|---|---|
| with data in Creativity.xls | |
| Analyze Guilford.txt | Table 2. Data Summary Report<br>Assigning models to "Creativity.xls"<br>Importing datafile= C:\Facets3663\examples\Creativity.xls<br>Waiting for imported datafile ... (see Help to avoid this wait)<br>Continuing ...<br><br>**Notice the "Waiting" message ...** |
| Edit menu<br><br>Edit Data = |  |
| The contents of the temporary file are the values imported from Creativity.xls |  |
| "Save as" |  |

| | |
|---|---|
| GuData.txt in the same folder as Guilford.txt |  |
| Facets<br>Edit menu<br><br>Edit specification file |  |
| Change the name of the Data= file<br><br>then Save it | 4=Daring<br>5=Enthusiasm<br>*<br><br>Data = Gudata.txt ; the saved .xls contents |
| Next analysis of Guilford.txt | Table 2. Data Summary Report<br>Assigning models to "Gudata.txt"<br>First active data line is: 1 1<br><br>**Very fast processing with saved data in Gudata.txt** |

# 9 Specifications *

## 9.1 %include file (included specification and data files)

Additional Facets specifications can be placed in separate text files, and then included with your specification file. This is done with the specification:
% include filename
which can be anywhere in the Facets specification and data files, and can also be nested.

Example:
Standard format: Knox Cube Test

Using %include files:

```
Title = Knox Cube Test (Best Test Design p.31) ; the report heading line
Facets = 2 ; two facets: children and items
Positive = 1 ; for facet 1, children, higher score = higher measure
Noncenter = 1 ; only facet 1, children, does not have mean measure set to zero
```

```
Pt-biserial = Yes ; report the point-biserial correlation
Vertical = 1*,1A,2N,2A ; show children by distribution and gender (sex), taps by number and
        name
Yard = 112,4 ; Vertical rulers 112 columns wide, with 4 lines per logit
Model = ?,?,D ; elements of the two facets interact to produce dichotomous responses
 ; log(Pni1/Pni0) = Bn - Di
 ; Bn = ability of child n, Di = Difficulty of item i,
 ; Pni1 = probability that child n on item i is scored 1.

Labels =
%labels.txt
* ; end of item labels

Data = ; no data file name, so data follows immediately in this file
1 ,1   ,1 ; child 1 on item 1 scored 1  (blanks are ignored)
1 ,2-18,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0 ; child 1 on item 2 scored 1, on item 3 scored 1, etc
        to item 18
 | ; 594 more observations, 18 per data line
35,1-18,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0 ; child 35 scored 1 on items 1-3, 0 on items 4-18
```

and in file: labels.txt

```
1,Children ; Children are facet 1
1-17 = Boy,,1 ; Pretend boys, in group 1, are numbered 1 through 17.
18-35 = Girl,,2 ; Pretend girls, in group 2, are numbered 18 through 35.
* ; end of child labels for facet 1
2,Tapping items ; Items are facet 2
1 = 1-4 ; Items labelled by the order in which the four blocks are tapped
2 = 2-3
3 = 1-2-4
4 = 1-3-4
5 = 2-1-4
6 = 3-4-1
7 = 1-4-3-2
8 = 1-4-2-3
9 = 1-3-2-4
10= 2-4-3-1
11= 1-3-1-2-4
12= 1-3-2-4-3
13= 1-4-3-2-4
14= 1-4-2-3-4-1
15= 1-3-2-4-1-3
16= 1-4-2-3-1-4
17= 1-4-3-1-2-4
18= 4-1-3-4-2-1-4
```

## 9.2 Anchor output file =

When Anchorfile= is specified, Facets writes out a new specification file containing the input specifications for this analysis. This file omits any specifications entered as extra specifications, but includes the final estimated measures for all elements and scales. This file also has all estimated measures marked as ",A" for "anchored".

These anchor values can be used as starting values for slightly different later runs. Edit the ",A"'s out of the first line of each facet in the Labels= specifications and out of the category lines of Rating (or partial credit) scale=.

This can also be specified from the Output Files menu, Anchor Output file.

Example: Using an edited version of Guilford.txt.
Anchorfile=Guilford.anc

```
Title = Ratings of Scientists (Psychometric Methods p.282 Guilford 1954)
Score file = GUILFSC ; score files GUILFSC.1.txt, SC.2.txt and SC.3.txt produced
```

```
Facets = 3    ; three facets: judges, examinees, items
Inter-rater = 1      ; facet 1 is the rater facet
Arrange = m,2N,0f    ; arrange tables by measure-descending for all facets,
Positive = 2  ; the examinees have greater creativity with greater score
Non-centered = 1     ; examinees and items are centered on 0, judges are allowed to float
Unexpected = 2       ; report ratings if standardized residual >=|2|
Usort = (1,2,3),(3,1,2),(Z,3)      ; sort and report unexpected ratings several ways (1,2,3) is
Senior, Junior, Trait
Vertical = 2N,3A,2*,1L,1A   ;define rulers to display and position facet elements
Zscore = 1,2  ;report biases greater in size than 1 logit or with z>2
Pt-biserial = measure ; point-measure correlation
Models =
?B,?B,?,RS1,1 ; CREATIVITY
*
Rating (or partial credit) scale = RS1,R9,G,O ; CREATIVITY
```
; Facets has renamed the rating scale in order to avoid ambiguity. Please edit the rating-scale name in the Anchorfile if you wish.
```
 1=lowest,0,A  ; this is a place-holder for the bottom category
 2=,-.6441868,A
 3=,-2.317694,A
 4=,.8300989,A
 5=middle,-1.477257,A
 6=,1.710222,A
 7=,-1.001601,A
 8=,2.358206,A
 9=highest,.5422113,A
 ; Rasch-Andrich Thresholds =  0, -.6441868 ,-2.317694 ,.8300989, -1.477257 ,1.710222 ,-
1.001601 ,2.358206 .5422113
*
Labels =
1,Senior scientists,A
 1=Avogadro,0.038439
 2=Brahe,.2356183
 3=Cavendish,-0.091576
*
2,Junior Scientists,A
 1=Anne,-0.068648
 2=Betty,.637455
 3=Chris,-.2451435
 4=David,-.4621402
 5=Edward,.4228902
 6=Fred,-.5607015
 7=George,.2762882
*
3,Traits,A
 1=Attack,-.2684814
 2=Basis,-.1409857
 3=Clarity,.2020537
 4=Daring,-.2898233
 5=Enthusiasm,.4972367
*
Data=
1,1,1-5,5,5,3,5,3
....
```

*Example 1:* Two time points and a rating scale:

(1) Obtain the common structure for the rating scale:
Construct the entire data set. Put in a dummy "time-point" facet with the two elements anchored at 0.
If the same person appears twice, then give them two related id-numbers, so that they are easy to pair up later.
Analyze the dataset. Write out an Anchorfile=

(2) Time-point 1 is the "gold standard":
For the time-1 data, edit the Model= statements for element 1 of the time-point facet.
Copy-and-paste into the specification file the parts you want to anchor from the (1) anchor file.
Analyze the time-1 data and output another anchorfile.

(3) Time-point 2 is measured in the Time-point 1 "frame of reference":
For the time-2 data, edit the Model= statements for element 2 of the time-point facet.
Copy-and-paste into the specification file the parts you want to anchor from the (2) anchor file.
Analyze the time-2 data.

Example 2: I want generate a report for each criterion entered in the analysis. I am interested in getting fair averages for the different criteria used by the raters.
1. Perform the complete analysis.
2. Output an Anchorfile=
3. In the Anchorfile=, keep everything anchored but comment out all except one criterion.
4. Analyze the Anchorfile=.
Now all the reporting will be for only that one criterion.
5. Return to 3. for the next criterion.

Example 3: Fair Average score for Item-by-Candidate combinations.
i. Do a standard analysis:
   Facets = 3 ; Candidates, Raters, Items
   Non-centered = 1 ; candidates - we need this
   Model = ?, ?, ?, R
   Anchorfile = anc.txt ; we need this
ii. In anc.txt,
   Edit the model specification:
   Model = ?, X, ?, .... ; inactivates the raters
   Remove "Anchorfile="
   Residual file = residual.xls ; Excel file - we need this
iii. Analyze anc.txt
iv. Open Residual.xls ; in Excel
   Delete all columns except: Candidate, Item, Expected score
   Sort the file: Candidate major, Item minor.
   All the expected scores for each candidate-item combination should be the same
   Used the Excel "Advanced filter" to remove duplicate lines.
v. The "expected score" is the "Fair Average score for Item-by-Candidate"

*Example 2:* I have 10 raters and one of them is stringent. I want to see the stringent rater's adjusted ratings and score.

Facets automatically adjusts for rater stringency/leniency. If you want to see what the stringent rater would have done, then here is an approach:

1. Analyze your data with Facets in the usual way.
Output the Anchorfile= from the Output Files menu

2. Edit the Anchorfile=.
a. Delete all the raters except the stringent one.
b. Change the anchor value for the stringent rater to 0.0

3. Analyze the edited Anchorfile.
Output the Residualfile= to Excel using the Output Files menu

4. In the Excel file, the "Expected" response values are the adjusted ratings.

## 9.3 Arrangement by Number/Alpha/Measure/Fit/PBS (Asc./Desc.) = N

Arrange= controls the ordering of the elements in the tables of measures: Table 7 and Table 14..

| Arrange= Table arrangements. Specifies ordering of elements within tables | | |
|---|---|---|
| **Sort order** | | |
| **Ascending** | **Descending** | **Arrangements for facets 1, 2, 3, 4, ...** |
| **A** | **a** | Element labels (alphabetically) |
| **M** | **m** | Element Measure |
| **F** | **f** | Fit - more extreme of INFIT and OUTFIT (also Z, T) |
| **N** | **n** | Element numbers |
| **P** | **p** | Point-biserial correlation |
| | | |
| | | Arrangements for bias/interaction: use facet 0 |
| **M** | **m** | Bias Measure |
| **T** | **t** | t-statistic of bias measure (also F, Z) |
| **N** | **n** | Bias term sequence numbers:<br>use the Output Tables menu, Tables 12-13-14, to sort elements by facet.<br>Top facet is minor, bottom facet is major for sort order. |

These letters can be combined to produce many rearranged copies of each report, e.g.,

Arrange = Ma,F

means produce two copies of each report:

for the first copy,
Ma specifies: measure descending (major), alphabetic ascending (minor).
for the second copy
F specifies: fit ascending (major) for the main reports.
t-statistic ascending (major) for the Bias reports.

The rearrangements can be specified to apply only to a particular facet by prefacing that facet's number to its rearrangement instructions. The Bias tables are indicated as facet 0.

Example: "Arrange = 1m,2m,2P,3A,0M,N" means

1m report the elements of facet 1, sorted by measure, descending
2m report the elements of facet 2, sorted by measure, descending
2P report the elements of facet 2, sorted by point-biserial, ascending
3A report the elements of facet 3, sorted alphabetically, ascending
0M report the bias terms, sorted by bias size (measure), ascending
N report all facets, sorted by entry number order, ascending

Parentheses () are optional: Arrange = (1f, 2m, F)

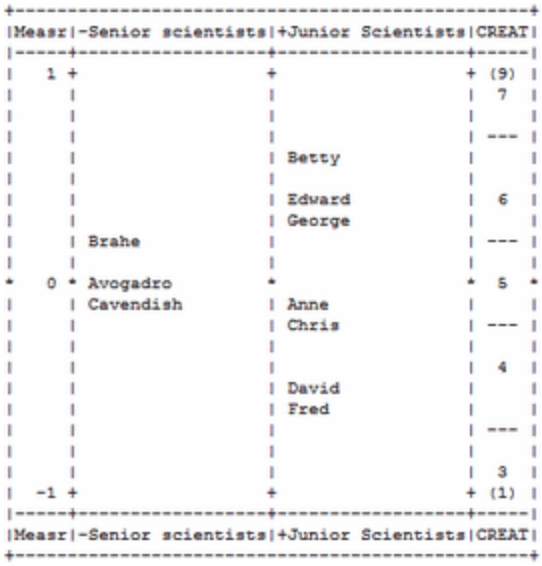
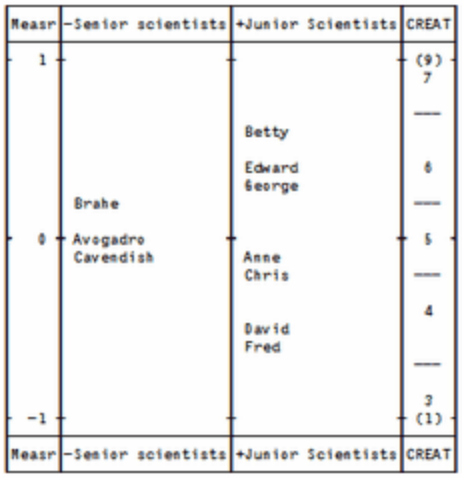
*Arrange*= by itself suppresses Tables 7 and 14.

For more elaborate sorting and column select, the easiest approach is to copy-and-paste the Measure or Bias/Interaction Table into an Excel worksheet, then use Excel's "Data", "Text to Columns" to put the Facets columns into Excel columns. You can then sort the columns and delete unneeded ones.

## 9.4 ASCII output table display format = Yes

ASCII= controls the font and format in which output tables are displayed. ASCII= can be set in Edit Initial Settings, the Extra Specifications box, your Specification file, and in Modify Output Specifications

The Facets output Tables are produced in a format compatible with most widely-used fixed-space (monospace) font `Courier New`. But this does not produce pretty table displays. The displays can be changed:

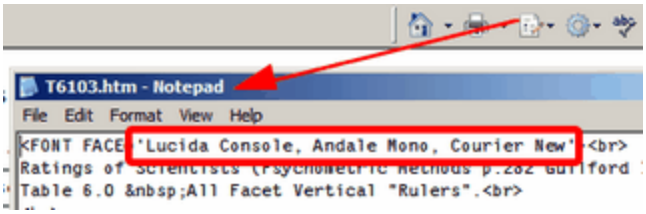
<table>
<tr>
<td>ASCII = Yes<br>Display using standard ASCII character set. Can be copied-and-pasted from NotePad to Word. Format Font in Word.<br><br>Microsoft Word compatible fonts: Letter Gothic line, Courier New, Andale Mono, Lucida Console, Consolas, SImPL, SimSUN, Bitstream Vera Sans Mono, Crystal, Onuava.<br><br>To change font in NotePad, click on "Format", "Font" and the "Font:" name.<br><br>MS Word, Options, Advanced:<br>Uncheck "Balance SBCS characters and DBCS characters"<br><br>If you lack a font, please Google the Internet to find and download it.</td>
<td> 
</td>
</tr>
<tr>
<td>ASCII = No<br>Display using PC-CHAR (MS-DOS) character set. Can be copied-and-pasted from NotePad to Word. Format Font in Word.<br><br>Microsoft Word compatible fonts: Letter Gothic Line. MS Line Draw<br><br>To change font in NotePad, click on "Format", "Font" and the "Font:" name.<br><br>To make the font change permanent in NotePad:<br>Windows "Start", "Run", "Regedit",<br>"HKEY_CURRENT_USER\Software\Microsoft\Notepad"<br>Double-click on "lfFaceName"<br>"Letter Gothic Line"<br>OK<br>Close registry x</td>
<td> 
</td>
</tr>
</table>

**ASCII = Webpage**
Display using Unicode (HTML) character set as a webpage. Can be copied-and-pasted to Word.

Microsoft Word compatible fonts: Lucida Console, Andale Mono, Courier New
To change the default webpage font, see Edit Initial Settings

| Measr | -Senior scientists | +Junior Scientists | CREAT |
|---|---|---|---|
| 1 | | | (9) 7 |
| | | Betty | |
| | | Edward George | 6 |
| | Brahe | | |
| 0 | Avogadro Cavendish | Anne Chris | 5 |
| | | | 4 |
| | | David Fred | |
| -1 | | | 3 (1) |
| Measr | -Senior scientists | +Junior Scientists | CREAT |

To **change the font** for the current webpage in Internet Explorer:
Click on Page icon (Alt+P).
View Source.
Enter the desired font after "FACE=", e.g., face='Courier New'
File menu: Save.
Internet Explorer: Click on Refresh icon (F5).

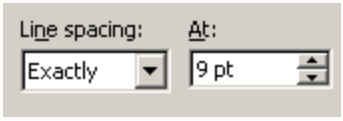


This must be a fixed-space font.

**Unwanted gaps in lines** in Microsoft Word: "Format paragraph", "Line spacing", "Exactly", and then set the line spacing one point less than the point-size of the text.

In my plot, the point-size is 10pt, so the "Exactly" line space is 9pt.

**Was:**



**At: one point less than your font size**

**Now:**

## 9.5 Barcharts output = Yes

Barcharts of the distribution of measures and fit statistics are produced for each facet as part of the output in Table 6.

| | |
|---|---|
| Barcharts = No | Omit Table 6.1 bar-charts |
| Barcharts = Yes | Display most Table 6.1 bar-charts: logit, infit mean-square, outfit mean-square, infit standardized, outfit standardized |
| Barcharts = All | Display all Table 6.1 bar-charts: standard error, observation count, observation raw score, average observation, fair-score average, point-biserial |

---

```
Table 6.1  Senior scientists Facet Summary.
Logit:
                                    1     1      1
 +-------------------------------Q----S--+-M-----S----Q--------------------------+
```

```
-1                                        0                                        1

S.E.:
           1
           9931
 +--------QSMQ-------------------------------------------------------------------+
 0                                                                                1
```

---

## 9.6 Batch file processing = Yes

To avoid Facets requesting user intervention when it is run in batch mode from the DOS prompt or otherwise, include Batch=Yes on the command line in your.bat or .cmd file.

---

1. Construct a Facets specification and data file that works correctly when launched in the usual way. Let's say this is "c:\myfolder\myspec.txt".

2. Let say that facets.exe is in "c:\facets\facets.exe"

3. Construct a text file which contains:
START /WAIT "c:\facets\facets.exe" BATCH=NO "c:\myfolder\myspec.txt" "c:\myfolder\myoutput.txt"

4. Save the text file and rename it "c:\myfolder\mybatch.bat"

5. Double-click on "c:\myfolder\mybatch.bat"

6. Facets will run in batch mode and you will see its operation on your screen. The output is in c:\myfolder\myoutput.txt

7. Exit from Facets.

8. Edit "c:\myfolder\mybatch.bat" to be
START /WAIT "c:\facets\facets.exe" BATCH=YES "c:\myfolder\myspec.txt" "c:\myfolder\myoutput.txt"

9. Double-click on "c:\myfolder\mybatch.bat"

10. Facets will run in hidden batch mode. Its output is in c:\myfolder\myoutput.txt

---

The general format is: Facets BATCH=YES specification-file output-file extra-specifications

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

Here is a typical .CMD file. It is Facetsxp.cmd in the Examples folder. Double-click on it to launch it.

```
echo  This is the version for Windows-NT, 2000
echo  This is a batch file to run Facets in batch mode
echo  Edit the next lines and add more.
echo  Format of lines is:
echo  START /WAIT ..\Facets BATCH=YES Control-file Output-file Extra=specifications
START /WAIT ..\Facets BATCH=YES dives.txt dives.out.txt
START /WAIT ..\Facets BATCH=YES lfs.txt lfs.out.txt
```

```
START /WAIT ..\Facets BATCH=YES kct.txt kct.out.txt
```

---

Example 1: Simulate and analyze 1000 Facets data sets from lfs.txt

Create text file "Simul.bat" in the Examples folder, copy this text into it, save and double-click on Simul.bat

```
Rem Simulate and analyze 1000 data sets
rem perform analysis of the original data and write out the anchor values
START /WAIT c:\facets\ BATCH=YES lfs.txt specfile.out.txt anchorfile=lfsanc.txt
SET /A COUNT=1
:LOOP
echo Loop number %COUNT%
rem do this 1000 times
IF %COUNT% == 1001 GOTO END
rem generate simulate data file from lfsanc.txt to avoid re-estimation each time
START /WAIT c:\facets\Facets BATCH=YES lfsanc.txt specfile.out.txt simul=s%COUNT%.txt
rem analyze original file replacing data in lfs.txt with the simulated data
START /WAIT c:\facets\ BATCH=YES lfs.txt s%COUNT%.out.txt data=s%COUNT%.txt scorefile=sc%
       COUNT%.txt
rem accumulate scorefiles by facet
type sc%COUNT%1.txt >> cand.results.txt
type sc%COUNT%2.txt >> rater.results.txt
type sc%COUNT%3.txt >> paper.results.txt
type sc%COUNT%4.txt >> task.results.txt
SET /A COUNT=COUNT+1
GOTO LOOP
:END
PAUSE
```

Example 2: Simulate and analyze multiple Facets data sets from lfs.txt using %include files:

At the top of the lfs.txt file, add this line:

```
%include.txt
```

Create text file "Simul.bat" in the Examples folder, copy this text into it, save and double-click on Simul.bat

```
Rem Simulate and analyze 1000 data sets
rem perform analysis of the original data and write out the anchor values
START /WAIT c:\facets\ BATCH=YES lfs.txt specfile.out.txt anchorfile=lfsanc.txt
SET /A COUNT=1
:LOOP
echo Loop number %COUNT%
rem do this 1000 times
IF %COUNT% == 1001 GOTO END
rem generate simulate data file from lfsanc.txt to avoid re-estimation each time
START /WAIT c:\facets\Facets BATCH=YES lfsanc.txt specfile.out.txt simul=s%COUNT%.txt
rem analyze original file replacing data in lfs.txt with the simulated data

echo "s%COUNT%.out.txt data=s%COUNT%.txt scorefile=sc%COUNT%.txt" >include.txt
START /WAIT c:\facets\ BATCH=YES lfs.txt

rem accumulate scorefiles by facet
type sc%COUNT%1.txt >> cand.results.txt
type sc%COUNT%2.txt >> rater.results.txt
type sc%COUNT%3.txt >> paper.results.txt
type sc%COUNT%4.txt >> task.results.txt
SET /A COUNT=COUNT+1
GOTO LOOP
:END
PAUSE
```

## 9.7 Beep sound = No

Facets usually operates silently. Facets will beep when processing is completed with
Beep = Yes

## 9.8 Bias (direction: difficulty/ability/omit) = Ability

The size of the estimated bias in Table 12 and Table 13 is reported in logits(or Umean= units), and its significance is reported as a t-statistic. The direction, i.e., sign, of the bias depends on how it is to be interpreted. In Table 14 and on the Excel plots, the bias direction accords with the reported facet direction, Positive=.

In the specific observations being analyzed for local bias:
Bias Direction = Plus or + or Positive or Ability or Leniency or Easiness (higher score = higher measure)
Bias Direction = Minus or - or Negative or Difficulty or Severity (higher score = lower measure)

Observed score > Expected score
Person is more able here than overall. Element is less difficult here than overall

Observed score < Expected score
Person is less able here than overall. Element is more difficult here than overall

Mathematically:
local measure = global measure + bias measure
local measure = global measure - bias measure

To suppress bias analysis: Bias direction = Omit

Example: Table 13:
Bias = Plus

```
-----------------------------------
|Obsvd Exp.  Obsvd Obs-Exp| Bias
|Score Score Count Average| Measure
-----------------------------------
| 9791 10506.0 6771  -.11 | -.10
```

<= observed-expected is negative (lower person performance than expected), so person ability bias is negative.

Bias = Minus

```
-----------------------------------
|Obsvd Exp.  Obsvd Obs-Exp| Bias
|Score Score Count Average| Measure
-----------------------------------
| 9791 10506.0 6771  -.11 |  .10
```

<= observed-expected is negative (higher difficulty than expected), so item difficulty bias is positive.

## 9.9 Box show around Tables = Yes

This removes the line-drawing character from all the Output Tables so that Tables are no long in tabular boxes. This is useful for importing the Tables into other software, spreadsheets or documents.

*Example:*
With Boxshow=Yes, the default:

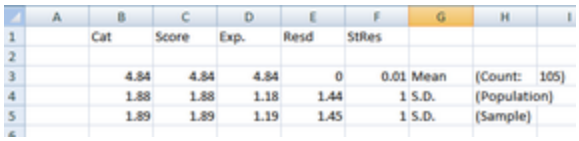
```
+-------------------------------------------------+
| Cat  Score  Exp.  Resd StRes|                   |
|-----------------------------+-------------------|
| 4.84  4.84  4.84   .00  .01 | Mean (Count: 105) |
| 1.88  1.88  1.18  1.44 1.00 | S.D. (Population) |
| 1.89  1.89  1.19  1.45 1.00 | S.D. (Sample)     |
```

```
+-------------------------------------------------+
```

With Boxshow=No, removing the boxes:

```
  Cat   Score   Exp.   Resd StRes

  4.84   4.84   4.84    .00   .01   Mean (Count: 105)
  1.88   1.88   1.18   1.44 1.00   S.D. (Population)
  1.89   1.89   1.19   1.45 1.00   S.D. (Sample)
```

This makes it easier to import a Table into Excel or other software. For Excel,
1) Copy-and-paste the Table into a blank worksheet
2) Excel "Data", "Text to Columns", "Delimited", "Space"
3) Here it is ....



| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | | Cat | Score | Exp. | Resd | StRes | | | |
| 2 | | | | | | | | | |
| 3 | | 4.84 | 4.84 | 4.84 | 0 | 0.01 | Mean | (Count: | 105) |
| 4 | | 1.88 | 1.88 | 1.18 | 1.44 | 1 | S.D. | (Population) | |
| 5 | | 1.89 | 1.89 | 1.19 | 1.45 | 1 | S.D. | (Sample) | |

## 9.10 Center facet =

**One facet must be non-centered in most analyses.** Center= specifies which facets are to be centered, so that the mean of the element measures is zero (or Umean=). Other facets are non-centered, unless they are anchored or group-anchored. If both Center= and Noncenter= and Center= are specified, only the first encountered in the specification file is actioned. When any element within a facet is anchored, or group-anchored, then the origin of the scale for that facet is set by the anchoring. Both Center= and Noncenter= are ignored for that facet.

a) Center= facet number, facet number, ...
a mean element measure of zero (or Umean=) for element measures with non-extreme scores
Example: Center= 2, 3, 4

b) Center= facet number Z, facet number Z, ...
a mean element measure of zero (or Umean=) for all element measures, including extreme scores
Example: Center= 2Z, 3Z, 4Z

c) Center= facet number, facet number Z, ...
a combination of (a) and (b)
Example: Center= 2, 3Z, 4

Extreme measures (corresponding to minimum and maximum possible scores) are included in the Means in Table 7, so that these may not be 0 (or Umean=) for centered facets unless: Center= facet number Z

In most analyses, if all facets are centered or anchored, then the estimates are over-constrained and displacements are reported. If more than one facet is *non-centered,* the frame of reference is not sufficiently constrained and the element measures are ambiguous (non-identified). The usual convention is to center the frame of reference on the agents of measurement: items, tasks, judges. This establishes the origin. Then the objects of measurement, persons, are positioned with respect to this origin. Technically, centering is an identifiability constraint, so that the estimates are uniquely defined.

Example 1: facets 2 (items) and 3 (raters) are to be centered; facet 1 is non-centered.

- Center= 2, 3 ; non-extreme element measures are centered in facets 2 and 3.
- Center= 2Z, 3Z ; all element measures are centered in facets 2 and 3
- Center= 2, 3Z ; non-extreme element measures are centered for facet 2; all element measures are centered for facet 3

Example 2: No facets are to be centered. This is an unusual situation such as when there is only one active facet so that the item difficulties are defined relative to an implicit origin as in the "Three Mile Island" example.

Center=0

## 9.11 Convergence (criteria to end iteration) = .5, .01, 0, 0

This sets the convergence criteria for the iterative joint JMLE (unconditional UCON) maximum likelihood estimation procedure (also the *Facets* implementation of PMLE). Both criteria (of size and change) must be satisfied for iteration to cease automatically. Select "Finish Iterating" from Files pull-down menu to override automatic operation. You may be able to speed up convergence by using the "Estimation" pull-down menu, and requesting Bigger changes.

The four criteria are:

| | |
|---|---|
| element: maximum residual<br>default: 0.5 | the maximum size of the marginal score-point residual (i.e., difference between observed and expected "total" raw score after omission of extreme scores) for any element. The standard convergence value is 0.5 score points, half the smallest observable difference between raw scores. |
| element: maximum logit change<br>default: 0.01 | the maximum size of the largest logit change in any estimated measure for an element during the previous iteration (regardless of Umean=) . The standard convergence value is .01 logits, the smallest useful or printable difference. |
| rating-scale category: maximum residual<br>default: 0 (ignored: element maximum residual used instead) | the maximum size of the largest marginal score point residual (i.e., difference between observed and expected "total" raw score) for any category. |
| Andrich threshold: maximum logit change<br>default: 0 (ignored: element: maximum logit change used instead) | the maximum size of the largest logit change in any estimated measure for a Rasch-Andrich threshold (step calibration) during this iteration. |
| If a criterion value is not specified, then its value is not changed.<br>If a criterion value is set to 0, then that criterion is ignored. | |

Example 1: In some situations, a pass-fail or other "high stakes" decision may hinge upon a difference of hundredths of a logit between a person measure and a criterion measure. For the final, decisive analysis, set the convergence criteria very tightly, e.g.,

Iterations=0 ; unlimited number of iterations

Convergence=.01, .0001 ; exaggerated accuracy: .01 score points and .0001 logits

*Be prepared to let your computer run a long time!*

Example 2: You want convergence to occur when no marginal score point residual (e.g., difference between any element's observed and expected raw score) is greater than 1.0 score points, and the default logit change is left at its default value:

Convergence = 1.0 ; 1.0 score points and .01 logits (the default). Category defaults are unchanged at 0 and so are ignored.

Example 3: You want to apply the convergence criterion applied in Facets 3.38.

Convergence = 0.5, .01, 0.5, 0 ; category and item residuals apply

Example 4: You want to match the convergence criterion applied in Facets 3.22, the last DOS version.

Convergence = .01, .001 ; run apparently tighter convergence criteria

Example 5: Convergence = 0.1, 0.001, 0.5, 0.01

Element measure: we want John's (and all elements) expected score to be within 0.1 of his observed score, and his measure not to change by more than 0.001 logits each iteration.
Category: we want category 1 (and 2, 3, 4) to have an expected frequency count within 0.5 of its observed frequency count, and the Rasch-Andrich threshold between categories 1 and 2 (also 2 and 3, 3 and 4) not to change by more than 0.01 logits each iteration.

## 9.12 CSV output format = " "

Numerical values written to the graph, score, and residual output files, files can be in fixed column format (the standard format) or separated by commas or tabs (variable width spaces used to align columns). The standard format can be changed with Edit Initial Settings.

| | |
|---|---|
| CSV=No<br>CSV=Fixed<br>CSV= | fixed column formatting |
| CSV=Tabs | values separated by Tabs |
| CSV= Commas<br>CSV= Yes | values separated by commas |
| CSV= Semicolons | values separated by semicolons |

Example:Provide coordinates so I can plot beautiful probability curves using my graphics package:
Graphfile=plot.dat; the name of an output file with Tab-separated values.
CSV=Tabs

## 9.13 Data for analysis =

This tells Facets where to find the data. Observations must be ordered integers in the range 0-254 for analysis. Facets can recode other data into integers using *Rating Scale=*

Either in the same file as the specifications:
Data =
followed by the data in the Facets data-format. This must be the last control specification, e.g.,
Data=
2,23,6,4 ;(The data)
0,13,9,1
3,4-6,8,1,5,4

Or in a separate data file or data files:
Data = diskfile.txt
where diskfile.txt is the name of the file containing the responses in the Facets data-format.

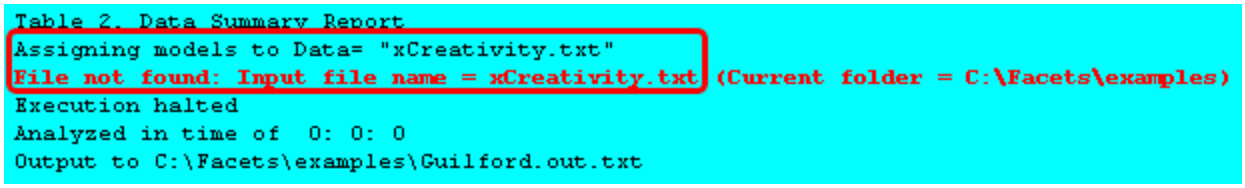
This can be extended to several data files with "+" signs:
Data = diskfile1.txt + diskfile2.txt + diskfile3.txt

or use %include file
Data =
%diskfile1.txt
%diskfile2.txt
%diskfile3.txt

The format of the diskfile.txt is determined by the filename suffix:

| Suffix | Data= Format |
|---|---|
| .txt | text file (MS-DOS or Windows) |
| .xls .xlsx | Excel workbook: first or only worksheet |
| .rda .rdata | R data file |
| .sdata | SAS data file |
| .sav | SPSS data file |
| .dta | STATA 13 and earlier data file |
| (other) | text file (MS-DOS or Windows) |

If the data file is not found, a file dialog box is displayed:




Please Edit Specification file, and Exit then restart the analysis.

---

*Facets data format: Single observations*

| Facet 1 | Facet 2 | Facet ... | Observation |
|---|---|---|---|
| Element number<br>or Element label<br>or Dvalues= | Element number<br>or Element label<br>or Dvalues= | ... | number<br>or value to recode with Rating-scale= |

*Facets data format: Multiple observations*

| Facet 1 | Facet 2 | Facet ... | Observation |
|---|---|---|---|
| Element number<br>or Element label<br>or Dvalues= | Element number range<br>(one facet, any facet) | ... | numbers to match the element number range<br>or values to recode with Rating-scale= |

Every line of data must have the same number of facet references (element numbers or element labels), usually given by Facets=. These facet references always appear in the same order. The element numbers are given in order by facet number, as specified after Labels=, unless overridden by Entered=. The facets references are then followed by one or more observed responses.

Example 1. Here is the flat file of observations:

```
123 665 25467
452 339 34245
312 034 34452
^   ^   ^
```

```
Examinee
    Judge
        Items 1 through 5
```
where format of flat file record is:
columns 1-3, "123" in record 1, are examinee identifiers
columns 5-7, "665" in record 1, are judge identifiers
columns 9-13, "25467" in record 1, are the ratings on items 1 to 5.

Convert this to Facets format. Insert commas and "1-5" added:
Labels=
1, Examinee
123
452
312
*
2, Judge
665
339
034
*
3, Item
1-5
*
Dvalues = 3, 1-5 ; put 1-5 in third data facet
; Extra blanks in the data file are ignored, also comments starting ";"
Data=
```
123 665 2 5 4 6 7 ; blanks are separators
452 339 3 4 2 4 5 ; "1-5" added from Dvalues=
312 034 "3" 4 4 5 2 ; values can be in quotes
```

Example 2: Data references recorded in facet number order with element labels
Labels=
2, Person ; Facet number 2 in the data
23=Fred
*
1, Judge ; The facet number, "1", not the order of definition, second, is what matters.
2=Hugo
*
3, Item
4=Sincerity
*
Data=
Hugo,Fred,4,1 ; Judge 2 (Hugo) rated Person 23 (Fred) on Item 4 (Sincerity) with a 1.

Example 3: Person 13 is rated on Item 4 by Judge 6 with a "3". This is data record is
Data=
13,4,6,3

Example 4: The data are letter grades, which are converted to numeric ratings by a Rating (or partial credit) scale= specification. Student 13's grade by Teacher 4 on Essay 1 is "B+"
Rating (or partial credit) scale=Grades,R9
9,Best,,A
7,B-Plus,,B+
*
Data=
13,4,1,B+

Example 5: **Data references not in facet order.** Entry= supplies both the order of the facet references and their identification in the data records:

Facets=4 ; 4 facets defined
Entry = 3,2,1 ; only 3 in data records, in different order
Model = ?,?,?,D ; first ? matches first reference, i.e., facet 3, etc.
Data =
4,23,2,1 ; Judge 2 (Hugo) rated Person 23 (Fred) on Item 4 (Sincerity) with a 1.

Example 6: **Element number 0** of each facet is used to indicate that no element of that facet participated, e.g., Judge number 0 is used when no judge intermediated. Keepasnull= enables you to use 0 as an active element number. Thus, if a facet does not take part in an observation, element number 0 is recorded in the position in the data line for that facet, e.g.,

2,0,27,3 ; item 27 (facet 3) did not require a judge from facet 2 to rate it.

Example 7: Some data do not require references to some facets. Person 13 (facet 2) scored a 1 on item 9 (facet 3), which was not judged (facet 1). The judge element number is entered as 0 for that observation in the data:

0,13,9,1 ; element number in facet 1 is 0 (not participating)

Example 8: **Multiple observations can be entered on the same data line**, provided that the elements are in ascending or descending sequence by element number in one facet, and all other facets are the same. The extreme element numbers of the sequence are given, e.g., items 1 through 10 are indicated by 1-10, Judge 2 (facet 1) gave person 23 (facet 2) on item 6 (facet 3) a response of 4, then

Judge 2 (facet 1) gave person 24 (facet 2) on item 6 (facet 3) a response of 5.
This can be recorded as
2,23,6,4
2,24,6,5
or, this can be recorded in one line as:
2,23-24,6,4,5 or 2,23_24,6,4,5
where the data points are entered to match the elements in the sequence.
The data points are also separated by commas ",".

Example 9: Person 3 (facet 1) took a 10 item test (items 1 through 10 of facet 2). He succeeded, scored 1, on items 1 to 5, and failed (scored 0) on items 6 to 10.

3,1-10,1,1,1,1,1,0,0,0,0,0 ; items ascending order
or
3,10-1,0,0,0,0,0,1,1,1,1,1 ; items descending order
or
3,1-4,0,0,0,0 ; some items ascending order
3,5-10,0,1,1,1,1,1 ; more items ascending order

Example 10: **Missing observations:** Person 13 is rated by Judge 8 on item 2 with a "3", item 3 with a "4", and item 6 with a "2". This is entered in the data file as

13,2-3,8,3,4 ; two observations in sequence
13,6,8,2 ; item not in sequence
or
13,2-6,8,3,4,,,2 ; items 4 and 5 entered as missing

Example 11: **Dvalues= simplifies data formatting.** The data are in a fixed format:

Person Judge Responses to items 1 to 6
13 4 134243
5 17 243223
These can be converted to Facets format by a word processor (Facets ignores blanks):
13 , 4, 1-6, 1,3,4,2,4,3
5 , 17, 1-6, 2,4,3,2,2,3

or
Dvalues = 3, 1-6 ; data facet 3 is 1-6
Data=

```
13  4 1 3 4 2 4 3 ; blanks are separators
 5 17 2 4 3 2 2 3 ; data facet 3 is 1-6
```

Example 12: **Replication (weighting) of a data point** can be specified by R (or another replication character) and the number of replications, for instance:

R3,2,23,6,4 means that the value of 4 was observed in this context 3 times.

Fractional replication permits flexible observation-weighting, see also Fractional models.

R3.5,2,23,6,4 means that the value of 4 was observed in this context 3.5 times.

Example 13: Organizing the ratings. There are three facets: examinees, tasks, raters. Each examinee performs two tasks that are rated by two raters:

Simplest:
examinee, task 1, rater A, rating
examinee, task 1, rater B, rating
examinee, task 2, rater A, rating
examinee, task 2, rater B, rating

or if the task element numbers are adjacent, then:
examinee, task 1-2, rater A, rating for task 1, rating for task 2
examinee, task 1-2, rater B, rating for task 1, rating for task 2

or if the rater element numbers are adjacent:
examinee, task A, rater 1-2, rating for rater 1, rating for rater 2
examinee, task B, rater 1-2, rating for rater 1, rating for rater 2

but NOT:
examinee, task 1-2, rater 1-2, rating, rating, rating, rating

## 9.14 Delements element type in data =

Delements= specifies the element identifier in the data file for the Labels= facet.

| **Format 1: Delements = type** | |
|---|---|
| Delements = N | all element identifiers in the data are element numbers in Labels=. Element identifiers like "1-4" are treated as ranges of elements: 1,2,3,4 |
| Delements = L | all element identifiers in the data are element labels in Labels=. Element identifiers like "1-4" are treated as the labels of single elements: "1-4". Element numbers without labels in Labels= are matched on the numerical text of the numbers. Null elements can be matched on the Null element number or its element label. |
| Delements = NL<br>*Warning! This can cause unintended matches. It is safer to use Format 2.* | all element identifiers in the data can be both element numbers and element labels in Labels=. Element numbers are matched first. If the element identifier in the data does not match any element number for the specified facet, the identifier is matched with the |

| | |
|---|---|
| | element labels. Element identifiers like "1-4" are treated as ranges of elements: 1,2,3,4 |
| Delements = LN<br>*Warning! This can cause unintended matches. It is safer to use Format 2.* | all element identifiers in the data can be both element numbers and element labels in Labels=. Element labels are matched first. If the element identifier in the data does not match any element label for the specified facet, the identifier is matched with the element numbers. Element identifiers like "1-4" are treated as the labels of single elements: "1-4". If the element label "1-4" is not found, then the element identifier 1-4 are treated as ranges of elements: 1,2,3,4 |

| **Format 2: Delements = Facet number + type, Facet number + type, ...**<br>Example: Delements = 1N, 2L, 3NL,4LN,0L | |
|---|---|
| 1N | facet 1 element identifiers in the data are element numbers in Labels= |
| 2L | facet 2 element identifiers in the data are element labels in Labels= |
| 3NL | facet 3 element identifiers in the data can be both element numbers and element labels in Labels=. Element numbers are matched first. If the element identifier does not match any element number for the specified facet, the identifier is matched with the element labels. |
| 4LN | facet 4 element identifiers in the data can be both element numbers and element labels in Labels=. Element labels are matched first. If the element identifier in the data does not match any element label for the specified facet, the identifier is matched with the element numbers |
| 0L | When a facet in the data is skipped with Entered=1,0,2, the element identifiers in the data for facet 0 may be numbers, labels or both. 0L specifies that they are labels so that element identifiers like "1-4" are to be treated as single elements, not ranges of elements. |

| **Format 3: Delements = Facet number + type + label-matching location**<br>Example: Delements = 3LS2W3 | |
|---|---|
| 1NS2W3, (substring is ignored) | facet 1 element identifiers in the data are element numbers in Labels= |
| 2LS2W3 or 2NLS2W3 or 2LNS2W3 | facet 2 element identifiers in the data are in the element labels in Labels=. They are S2W3, starting in column 2 of the element label with a width of 3 columns. |

| **Delements = L or N or NL or LN - Matching element identifiers in the data with Labels= element labels** |
|---|

| Delements= | When matching an element identifier in the data with element numbers or labels in Labels=, the priority is: | | | |
|---|---|---|---|---|
| N or NL<br>or LN (if no element label match) | The identifier in the data is matched with the element number in the facet list | Facet 1: "7" in the data matches element number "7" so the element number for analysis is 7 | Labels=<br>1, Grade level<br>**7**= Grade seven<br>*<br>Data=<br>**7,** x21, boy, run, M, 1 | |
| 2LS3E5 or 2LS3W3<br>or<br>2LNS3E5 or 2LNS3W3<br>or<br>2NLS3E5 or 2NLS3W3 (if no element number match) | The identifier is an exact match to the segment of an element label specified with S..W.. or S..E.. | Facet 2: "x21" in the data matches "x21" in columns 3,4,5, so the element number is 8 | Labels=<br>2, cohort<br>8= my**x21**pq<br>*<br>Data=<br>7, **x21**, boy, run, M, 1 | |
| L or LN<br>or NL (if no element number match) | The identifier is an exact match to the part of an element label before : (a colon) | Facet 3: "boy" in the data matches "boy:". so the element number is 2 | Labels=<br>3, gender facet<br>2 = **boy**: male student<br>*<br>Data=<br>7, x21, **boy**, run, M, 1 | |
| L or LN<br>or NL (if no element number match) | The identifier is an exact match with an element label without a : (a colon) | Facet 4: "run" in the data matches "run:". so the element number is 4 | Labels=<br>4, activity facet<br>4 = **run**<br>*<br>Data=<br>7, x21, boy, **run**, M, 1 | |
| L or LN<br>or NL (if no element number match) | The identifier is a leading-character match with an element label that has no : (a colon) | Facet 5: "M" in the data matches the leading M in "Marymount", so the element number is 8 | Labels=<br>5, school facet<br>8= **Marymount**<br>*<br>Data=<br>7, x21, boy, run, **M**, 1 | |

```
Facets = 5

Labels=
1, Grade level
7= Grade seven
*
2, cohort
8= myx21pq
*
3, gender facet
2 = boy: male student
*
4, activity facet
4 = run
*
5, school facet
8= Marymount
*
Data=
7, x21, boy, run, M, 1
```

*Example 1:* in the data, the facet 1 element identifiers are element labels, for facet 2, element numbers, and for facet 3 are element labels. The observations are 0-1 dichotomies:

Facets = 3
Delements = 1L, 2N, 3L
Labels =
1, Gender
1= Female
2= Male
*
2, Zip code

60601 = Central Chicago
60637 = Hype Park Chicago
...
*
3, Income
1 = Low
2 = Medium
3 = High
*
Data =
F 60637 2 0  ; Female, 60637 Zip code, Medium income, No
M 60601 3 1  ; Male, 60601 Zip code, High income, Yes

*Example 2:* chess tournament: paired comparisons with element labels. Facet 1 element identifiers are labels. Facet 2 element identifiers are numbers.

Facets = 3 ; player, player, round
Entered in data = 1,1, 2  ; the players are facet 1, round is facet 2
Models = ?,-?, ?, D
Delements = 1L, 2N  ; element identifiers for facet 1 are element labels, for facet 2 are element numbers
Labels=
1, Players
1= John
2= Mary
3= ...
*
2 = Round
1= Monday morning
2= Monday afternoon
3= Monday evening
.....
*
Data =
John Mary 1 0 ; John played Mary on Monday morning. John lost.

## 9.15 Dvalues for data facets =

If a facet entry in the data file has the **same value** for all data lines, then it can be omitted from the data, and replaced by a Dvalue= statement, e.g., a set of items administered to all persons.

If a facet entry in the data file has its **value in the label of another element**, then it can be omitted from the data, and replaced by a Dvalue= statement, e.g., a demographic indicator in the person element label.

Facets are always their locations in the data, before Entered=
Data=
facet 1, facet 2, facet 3, facet 4, ..... , observation

| **Dvalues =** element details for element not specified in data file | | |
|---|---|---|
| **Element value** | **Examples** | **Meaning in the data** |
| element number in facet position in data | Dvalues = 3, 1 | data facet 3 element is always element 1 |

| | | |
|---|---|---|
| element number in facet - element number in facet | Dvalues = 3, 1-4 | facet 3 elements are always elements 1-4 |
| $Line | Dvalues = 3, $Line | facet 3 element is the line number in this data file |
| $Cumulative | Dvalues = 3, $Cumulative | facet 3 element is the line number in all the data files combined |
| reference facet number, $Group | Dvalues = 3, 1, $Group | facet 3 element is the group-number of the data facet 1 element |
| reference facet number, start column, width in columns | Dvalues = 3, 1, 2, 3 | data facet 3 element is in the element label of data facet 1 element, starting in column 2 with a width of 3 columns |

**A. Elements numbers the same for every data line.** This occurs when, for instance, every examinee is rated on items 1-5, or when every patient is observed on items 1-13.

The format for one Dvalue= is:
**Dvalue = facet entry number in data, constant value**

The format for several Dvalues= is:
Dvalue=
facet entry number in data, constant value
facet entry number in data, constant value
....
*

Example 1: In the Guilford.txt example, the standard layout is:
Facets = 3
Data =
1, 1, 1-5, 5,5,3,5,3
1, 2, 1-5, 9,7,5,8,5
....
3, 7, 1-5, 7,7,7,5,7

This is the same as:
Facets = 3
Dvalue = 3, 1-5
Data =
1, 1, 5,5,3,5,3
1, 2, 9,7,5,8,5
....
3, 7, 7,7,7,5,7

**B. Elements numbers based on line number.** This occurs when, for instance, the examinee is the line number in the data

The format for one Dvalue= is:
**Dvalue = facet entry number in data, program variable**

The format for several Dvalues= is:
Dvalue=
facet entry number in data, program variable

facet entry number in data, program variables
....
*

Program variables are:
$Line = Data-line number in current data file (excluding comment lines and blank lines)
$Cumulative = Data-line number in current data file (including previous data files, if any)

Example: for the Knox Cube Test:
dvalues=
1, $Line  ; facet 1: person elements are same as data line number
2, 1-18   ; facet 2: every person responds to 18 items
*
Data =
1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0 ; child 1, items 1-18
1,1,1,1,1,1,0,0,1,1,1,0,0,1,0,0,0,0
1,1,1,1,1,1,1,1,1,0,1,0,0,0,0,0,0,0
1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0

**C. An element identifier for one facet** (the target) **is in the element label of another facet** (the reference) for the same observation. If the segment in the reference element label is a number, that becomes the element number for the target facet. If non-numeric, then the segment is matched to the start of the elements labels given for the target facets, up to :. If ":" is omitted, then a match is made to the left-end of the target element label.
The format is:
**Dvalue= facet entry number for target facet, reference facet location in data, start column in the label, length of field in the label**
The format for several Dvalues= is:
Dvalue=
facet entry number for target facet, reference facet location in data, start column in the label, length of field in the label
facet entry number for target facet, reference facet location in data, start column in the label, length of field in the label
.....
*

Example 1. The Knox Cube Test with interactions example data set contains replacements for facets 3 and 4:

```
Dvalues =  ; The child gender (sex) and number of taps are in the labels,
           ; so can be omitted from the data
3, 1, 1, 1 ; Child gender element for Facet 3 is indicated in
           ; the label of Facet 1, column 1 length 1
4, 2, 1, 1 ; Item length element for Facet 4 is indicated in
           ;the label of Facet 2, column 1 length 1
*
```

Example 2. Ethnicity is indicated by a numeric value in column 1 of the student label. Gender is in column 2.

```
Facets=4  ; persons, item, ethnicity, gender
Models = ?,?B,?,?B,D ; we want to look for item-gender DIF
Labels=
1, Student
234 = 61 Jose  ; Jose is a Hispanic (6 in column 1), male (column 2 of label=1)
235 = 32 Mary  ; Mary is a Caucasian (3 in column 1), female (column 2 of label =2)
.....
*
2, Items
1-4   ; 4 dichotomous items, scored 0-1.
*
```

```
3, Ethnicity, A  ; elements anchored at 0 to avoid distorting student measures
3, Caucasian, 0
6, Hispanic, 0
*
4, Gender, A  ; elements anchored at 0 to avoid distorting student measures
1, Male, 0
2, Female, 0
*
Dvalues=
2, 1-4   ; all students administered items 1-4
3, 1, 1, 1  ; element number for facet 3 (ethnicity) is in element label for facet 1, column 1 with a length of 1
4, 1, 2, 1  ; element number for facet 4 (gender) is in element label for facet 1, column 2 with a length of 1
*
Data=
234, 0,1,1,0  ; facet 1 element 234 then 4 observations. would be 234, 1-4, 6, 1, 0,1,1,0 without Dvalues =
235, 1,1,0,0  ; facet 1 element 235 then 4 observations. would be 235, 1-4, 3, 2, 1,1,0,0 without Dvalues =
```

Example 3. Gender is indicated by a non-numeric value in column 2 of the student label

```
Facets=3 ; persons, item, DIF indicator
Models = ?,?B,?B ; we want to look for gender-related item DIF
Labels=
1, Student
234 = 6M Jose  ; Jose is a male (column 2 of label=1)
235 = 3F Mary  ; Mary is a female (column 2 of label =2)
.....
*
2, Items
1-4   ; 4 dichotomous items, scored 0-1.
*
3, Gender, A  ; elements anchored at 0 to avoid distorting student measures; only for DIF analysis
1, M:Male, 0  ; M: - colon, so M is used to match the code in the student label
2, Female, 0  ; no :, so match is made to left-end of label for length of dvalues= field (width 1 in this example).
*
Dvalues=
2, 1-4   ; all students administered items 1-4
3, 1, 2, 1  ; element number for facet 3 is in column 2 of element label for facet 1.
*
Data=
234, 0,1,1,0  ; facet 1 element 234 with 4 observations. would be 234, 1-4, 1, 0,1,1,0 without Dvalues =
235, 1,1,0,0  ; facet 1 element 235 with 4 observations. would be 235, 1-4, 2, 1,1,0,0 without Dvalues =
```

Example 4: To test that Dvalues= is functioning correctly:

```
title= "dvalue test"
facets = 2
residualfile=dvalueresiduals.txt
models=?,?,R
labels=
1, Facet 1
1=Fred
2=Mary
*
2, Facet 2
1=M:supervisor
2=F:janitor
*
dvalues=
2,1,1,1 ; facet 2 is the first character of the element for facet 1
*
```

```
data=
1-2, 0,1
1-2, 1,0

; this analyzes as (see residualfile=)
; 1 1 2 2 Obs
; 1 Fred 2 F:janitor 0
; 2 Mary 1 M:supervisor 1
; 1 Fred 2 F:janitor 1
; 2 Mary 1 M:supervisor 0
```

Example 5: Demonstrates element numbers in element labels:

```
facets = 3
model=?,?,?,R
labels=
1, Facet 1
1-10
*
2, Facet 2 ; this is not in the data line
1 = fruit ; these element numbers are in the element labels of Facet 3
2 = vegetable
*
3, Facet 3 ; this is in the data line
1,1 apple  ; the element label contains the element number for Facet 2
2,1 banana
3,1 orange
4,2 carrot
5,2 potato
6,2 cabbage
7,2 pea
*
dvalue=*
2, 3,1,1  ; element numbers for Facet 2 are in the element labels of Facet 3
*
data=
1,1-7, 1,0,1,0,1,0,1  ; Facet 1, Facet 3, 7 responses. Facet 2 specified by Facet 3
2,1-7, 0,1,0,1,0,1,0
```

**D. An element identifier for one facet** (the target) **is in the group number of an element of another facet.**

The format is:

**Dvalue= facet entry number for target facet, reference facet location in data, $G**

The format for several Dvalues= is:

Dvalue=
facet entry number for target facet, reference facet location in data, $G
facet entry number for target facet, reference facet location in data, $G
.....
*

Example 6: The element number of facet 3 is the group number of the element in facet 1:

```
Facets=3 ; person, item, ethnicity
Models=?,?,?,R
Labels=
1, Person
1 = George, , 1  ; in ethnic group 1
2 = Maria, , 2 ; in ethnic group 2
....
*
2, Item
1 = Arithmetic
2 = Social Studies
....
```

```
*
3, Ethnic group
1 = African-American
2 = Asian
.....
*
Dvalues = 3, 1, $Group  ; element number for facet 3 is group number in facet 1
Data =
1, 1, 4  ; person 1 (George) is rated on item 2 (Arithmetic) with a rating of 4.
         ; the element number for facet 3 is the group number for facet 1 (1, African-American).
```

Example 7: with an Excel spreadsheet
Facets = 4 ; facet 4 is the item facets
Dvalues= 4, 1-20 ; there are 20 item elements. Then 20 ratings
Data = exceldata.xls

Exceldata.xls:

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | ; Persons | Tasks | Raters | Ratings | | |
| 2 | 1 | 1 | 5 | 1 | 1 | 2 |
| 3 | 2 | 2 | 5 | 1 | 1 | 2 |

---

Example 7. **Dvalues= applies to the data lines before Entered= is applied.**

Facets = 4
Entered = 1,2,2,3
Models = ?, ?,-?,?,D
Dvalues =
4, 1-4  ; this is facet 4 in the data, and will match the elements of facet 3 in Labels= after Entered=
*

## 9.16 Elements in Labels=

## 3. Element Identification

| 3A. Element number = 1 to 2147483646, 0, see Keepasnull= | 3B. Element label, | 3C. Measure value, | 3D. Group number, 1 to 32767 | 3E. Weight, | 3F. Target element number |
|---|---|---|---|---|---|

**3A. Element numbers**
Assign each element within each facet its own integer element number in the range 1 to 2147483646 for each facet. Numbers may be skipped over or omitted, and listed in any order. If the same element number is listed twice in Labels= then the details of the element are combined.. The maximum number of elements across all facets is 100,000,000.
Thus,
each judge must be assigned a unique element number in the range 1 to 2147483646.
each person must be assigned a unique element number in the range 1 to 2147483646.
each item must be assigned a unique element number in the range 1 to 2147483646.
The ranges of judge, person, and item numbers may overlap, so that judge 1 may rate person 1 on item 1.

Your elements may already labeled with numbers that are too big, such as social security numbers, or with non-numeric names, such as Mary. These elements must be re-identified with unique element numbers in the range 1 - 2147483646. Facform can assist with this. It can be downloaded from www.winsteps.com/facets.htm.

Element 0 is the null element, indicating that the facet does not apply to the observation. But 0 can be used as a regular element number if Keepasnull= has been used to set a different element number as the null element.

If a desired element number is too big (such as a Social Security number), then use it as an element label with a shorter element number. The element identifier in the data can be the element label: Delements= NL

Observations in Data= referencing elements not included in the Labels= list are omitted from Facets analysis. You can deliberately omit an element (and its observations) from an analysis by typing a ";" before its element number, turning that specification line into a comment.

**3B. Element labels**
Each element number can have an element label (including the Null= element). They are shown in Table 7 The element labels can be used as element identifiers in the data (with Delements=) or as references to the elements in other facets (with Dvalues=).

**How to construct a list containing all facets and elements:**
a) Start with Labels=

b) List for yourself all the facets, in whatever order you like:
e.g., raters, persons, items

c) For each facet, the facet number and its identifying label are entered first.
Labels=
1, Judges ; the rater facet

d) For that facet, the element numbers and their labels are entered, in any order.
Labels=
1, Judges ; the rater facet
10 = Rehnquist
1 = John Jay
123456789 = Jose ; this is Jose

If the element numbers are in a range, which share features, then they can be summarized:
Labels=
1, Judges ; the rater facet
1-100 = Males
101-200 = Females
Specifying long ranges containing many unused number will slow down Facets, e.g,,
1 - 5000000 = Males ; when there are only 10 males
Use the Excel "unique records" feature to list the numbers used

If the labels are to be used for reference by Dvalues=, then the code letter can be indicated by ":".
Dvalues=
3, 1, 2, 1 ; element number for facet 3 in the data is in the for element label for facet 1 at column 2 with a length of 1.
*
Labels=
1, Persons
1, 3M Fred ; M is used by Dvalues=
2, 2F Mary ; F is used by Dvalues=
3, 2M Jose ; M is used by Dvalues=
...

*
; either
3, Gender ; Sex
1, F: Girls ; F prior to : is matched to F of Mary
2, M: Boys ; M prior to : is matched to M of Fred and Jose
*
; or
3, Gender
1, Female ; F is matched to F of Mary
2, Male  ; M is matched to M of Fred and Jose
*

e) An "*" marks the end of each facet's elements.

Labels=
1, Judges ; the rater facet
10 = Rehnquist
 1 = John Jay
*

f) For each further facet, repeat c) through e).

Labels=
1, Judges ; the rater facet
10 = Rehnquist
 1 = John Jay
*
2, Persons

Example 1:

```
Labels= ; specifying that facet/element details follow:

1,Judges ; This labels the first facet
1=John Jay ; Judge no. 1
2=Roy Beam ; Judge no. 2
; Judge 3 to be omitted
4=Wapner ; Judge 4
 | ; the other judges go here
15=Scalia ; Judge no. 15
* ; an asterisk terminates the elements for this facet

2,Persons ; facet 2
1=Ben ; first person
2=7021596 ; person 2 as a telephone number
 |
23=Zabado ; the last person
*

3,Items ; facet 3
1= 2+2 ; the first item addition
2= Tennis ; second item
3= Attitude ; third item
4 ; the fourth item has no label
* ; end of labels for last facet also ends Labels=
```

When there is no description, Facets uses the facet or element number instead, e.g.,

```
Labels=
1, Rater
4  ; this means 4=4
*
```

g) Big element identification numbers are no problem.

1) There must be no more that 4,000,000 element numbers.

2) element numbers must be in the range: 0 - 2,147,438,646
If your element identifiers are outside that range, please use those element identifiers as element labels, not numbers

3) elements can be identified by element numbers or element labels in the data file.

4) if you don't put the element identifiers in Labels=, then Facets will construct a list of elements in its Output Table 2. You can paste this list into Labels=

Example 2: student IDs are numbers like 9999999999999999999 and you want to use them in the data file

Step 1) Construct the data file with student IDs as element identifiers for Facet 1

Step 2) In your Facets specification file:
```
Delements= 1L  ; element identifiers in the data file are element labels
Labels=
1=Students
; no element identifiers here for the Students
*
2=Items
1=What is the capital of Germany?
......
```

Step 3) Analyze your data file with Facets. Output Table 2 will contain a list of element numbers and labels:
List of elements not specified in Labels=. Please copy and paste into your specification file, where needed
```
Labels=Nobuild ; to suppress this list
 1, Students,  ; facet 1
 1 = 210100002
 2 = 210100008
 3 = 210100013
 4 = 210100025
 5 = 210100026
....
```

Step 4) Copy and paste these element numbers and labels into your Facets specification file:
```
Delements= 1L  ; element identifiers in the data file are element labels
Labels=
1=Students
1 = 210100002
2 = 210100008
3 = 210100013
4 = 210100025
5 = 210100026
....
*
2=Items
1=What is the capital of Germany?
......
```

Step 5) Do your Facets analysis

**3C. Establish pre-set measure values for starting values or anchor values.**

Ignore this for initial analyses.
When logit anchor values or starting values are to be provided, a comma followed by a third value is appended after each element label. These values have one of three meanings:

**1) Starting values** (the standard when values are provided)
this is useful for speeding up analyses when you have a good idea of what some of the measures will be. These values will not be changed during the PROX phase of the estimation. Whenever the element label is followed by a logit value, this is used as the starting estimate for that element in the analysis.

```
Labels=
1,Persons ; no code after facet label
23=Joe,2.3 ; means Estimation starts with Joe at 2.3 logit
```

**2) Anchor values: A and D**
Use this when you know what you want some of the measures to be from another analysis or from an item bank. Elements representing demographic facets such as gender can be anchored at 0 logits. This excludes them from estimation, but includes them for fit statistics and bias analysis.
When the facet label is followed by ",**A**", then the measure of each element in that facet is anchored (fixed) at the value following its label, whenever such a value is provided.
When the facet label is followed by ",**D**", then the measure of each element in that facet is anchored (fixed) at the Umean= value. The anchor value following its label (if any) is ignored.

```
Labels=
2=Persons,A ; anchoring wanted
1=Ben ; no value - anchoring does not apply
23=Joe,2.3 ; Joe anchored at 2.3 for the entire analysis
```

When Umean= is used, then the anchor values must align with it:

```
Umean=50,10 ; User mean is 50, user scaling is 10 per logit.
2=Persons,A ; anchoring wanted
1=Ben ; no value - anchoring does not apply
23=Joe, 73 ; Joe anchored at 50 + 10*2.3 for the entire analysis
```

This feature is useful for dummy facets removing classification elements from the measurement. These may be used to partition fit, detect bias or select rating scale (or partial credit) models.

```
Labels=
2=Classifier,A ; anchoring wanted
1=Type A,0 ; anchor at 0, so doesn't affect measurement
2=Type B,0 ; anchor at 0, so doesn't affect measurement
```

or, and easier,

```
Labels=
2=Classifier,D ; dummy-facet anchoring wanted
1=Type A ; no anchor value required
2=Type B ; no anchor value required
```

When Umean= is used, use the Umean= value for these "dummy" elements

```
Umean=50,10 ; User mean is 50, user scaling is 10 per logit.
Labels=
2=Classifier,D ; dummy-facet anchoring wanted
1=Type A ; no anchor value required
2=Type B ; no anchor value required
```

**3) Group-anchoring: G (all elements, including extreme elements) and X (exclude extreme elements)**
Use this to equate by groups of elements, e.g., to maintain the same average severity of a group judges from one test analysis to the next, or to anchor the average difficulty calibration of a set of 6th grade math items. An easy way to specify a particular group mean is to give all group elements that same value.

When the facet label is followed by ",G" or ",X", groups of elements are anchored so that their mean is fixed, though they float individually relative to that mean. "G" includes all measurable elements, extreme and not extreme. "X" excludes measures corresponding to extreme scores from the group-anchor computation. This is how "G" behaved in earlier version of *Facets.*

*Example 3.1:* If several groups are to be defined, enter each element's group number as a fourth entry:

```
2,Raters, G  ; Raters are facet 2
1=Ben  ; no value, normal estimation, does not belong to a group
2=Mike, 2.7,1 ; Mike contributes 2.7 logits to group 1
3=Mary, 3.3,2 ; Mary contributes 3.3 logits to group 2
5=Anne,-1.7,1 ; Anne contributes -1.7 logits to group 1
6=Joel, 0.5,2 ; Joe contributes 0.5 logits to group 2
7=Fred,  ,2 ; Fred is reported in group 2, but he does not participate in the group-anchoring.
 ; The double comma, ", ,", omits the anchor value.
8=Kent, 3.1 ; No group number, so 3.1 is a starting value
9=Irma,-2.8,0 ; Group 0 means "anchor at this value"
10=Jim, 1.3,0 ; Group 0 means "anchor at this value"
11=Abe, 1.5,3 ; Only one in a group-anchor, so treated as if anchored, "A", but reported as a group
*
```

Mike, Anne and Joe are in group 1, so their individual estimated measures will alter, but their mean measure will be fixed at (2.0 -1.0 + 0.5)/3 = 0.5 logits. Mary is the only one anchored in group 2, so her measure will be fixed at 3.0.

*Example 3.2:* An easy way to specify a particular group mean is to give all group elements that same value. In this example we want Group 1 to have average measures of 0 and Group 2 to have an average measure of 1.5

```
Labels=
2=Persons,G ; group-anchoring wanted
1=Ben,0,1 ; Ben is in group 1
2=Mike,0,1 ; Mike is in group 1
3=Mary,1.5,2 ; Mary is in group 2
5=Anne,0,1 ; Anne is in group 1
6=Joe,0,1 ; Joe is in group 1
7=Fred,1.5,2 ; Fred is in group 2
```

*Example 3.3:* Group 0 is for assigning standard anchor values. It acts as a group of 1 element for anchoring.

```
Labels=
2=Persons,G ; group-anchoring wanted
1=Ben,0,1 ; Ben is in group 1
2=Mike,0,1 ; Mike is in group 1
3=Mary,1,0 ; Mary is in group 0 - so is anchored at 1.0
4=John,2,0 ; John is in group 0 - so is anchored at 2.0
```

See Nested Designs for more examples.

**3D. Grouping**

Grouping of elements is controlled by a fourth parameter at the element level. This is used for reporting as well as group-anchoring. Thus, to report people classified into two groups, males and females:

```
Labels=
2 = Person ; No anchoring specified, so values (if any) will be starting estimates. We could also specify
    ",A" (anchoring) or ",G" (group-anchoring).
1 = Ben, , 1 ; in group 1, note use of double commas, ",,"
2 = Mike, , 1 ; in group 1
3 = May, 3.4, 2 ; in group 2, with a starting logit value of 3.4
5 = Anne, , 2 ; in group 2
```

6 = Harold, ,1 ; in group 1

**Group number in element label:** Group numbers can be numeric (1,2,3,..) or can specify a number in the element label: $(starting position)W(width).

1 = 1 Ben, ,$1W1 ; the group number will be the first character of "1 Ben" = 1.

This can be conveniently applied to all the elements of a facet:

Labels=
2, Person
1, 123 Ben
2, 235 Mary
3, 123 George
....
1003, 347 Jose
1-1003, , , $1W3 ; the group number for all Persons is in the first 3 columns of the person label.

Table 7 contains:

1) all elements of each facet with summary statistics for the facet.
2) each group of elements within each facet with summary statistics for the group.

Grouping simplifies analysis of subsets of elements in a facet. Boys and girls can be separately listed. The mean measures of the two groups can be directly compared. Grouping of test items by strand, and arranging the elements by measure, aids in verifying the construct validity of each strand. It also makes the strand information more available for use in curriculum and assessment decision-making.

**3E. Element weighting**: default 1.0
Observations specified for this element are weighted with this element weight multiplied by model weights and observation weights. See Weighting the data. If you want to report an element, but not allow it to influence other estimates, then give it a very small weight.

**3F. Target element number**: default none.
Two or more elements with different element numbers can be reassigned to the same element number. The observations with the element number of this specification are reassigned to the element with the Target element number. Duplicate element numbers can be combined. For instance:

Labels=
1, Tasks
1 = Task A, , , , 4 ; observations with element 1 are reassigned to element 4 for analysis.
.....
4 = Task AB ; observations for element 1 are included with observations for with element 4

## 9.17 Entry order of facets in data = 1,2,3,4,5,..

This specification identifies to which facet each element number in a data line belongs. If you want persons as facet 1, items as facet 2, and judges as facet 3, but the data has been entered with facets in the order of persons (facet 1), judges (facet 3), items (facet 2), then

Facets = 3
Labels =
1 = Person
|
*
2 = Item
|
*
3 = Judge
|

*
Entry = 1,3,2
Data =
4,6,5,2 ; Person 4 was rated by judge 6 on item 5 with a "2"

If there are more facets in the data than you wish to include in this analysis, extra facets may be ignored by specifying them as 0.
e.g., if whatever is in the second data facet is to be ignored,
Facets = 3 ; number of facets in the data
Entry = 2,0,1 ; only look at facets 1,2 in Labels=

Then for all other purposes, there are only two facets in each data line, so that
Model = ?,?,D ; so only two facets in the model statement: facet 2, facet 1, Dichotomous

A facet may be entered twice for paired comparison analysis.
For example, when a judge (facet 1) compares the performance of person A (facet 2) with that of person B (also facet 2), then
Facets = 3
Entry = 1,2,2
Model = ?,?,-?,B1,0.5 ; "-?" reverses the sign of the second person measure
; enter each data point twice: A vs. B, and then B vs. A. so weighting is 0.5

## 9.18 Facets in data = 2

Place Facets= before any other specifications referencing facet numbers. This specifies how many different facets interact to make the set of observations.

Facets in data = label facets, data facets
Label facets = highest numbered facet in Labels=
Data facets = number of facets in the data file (including facets implied by Dvalues=) - not required if same as Label facets or inferred from Entry-order=.

Example 1: For judges, persons, items:
Facets=3

In complex data sets, every facet may not be involved in every observation. For instance, some observations may involve persons with items and tasks, while other observations involve persons with judges and items. Since there are four facets altogether (persons, items, tasks, judges), specify four facets, even though no single observation involves all four:
Facets=4

A rating scale (or partial credit) category or response is not a facet. Data are the result of interactions between the facets. Thus "Mary responds to item 7 with a `3'" indicates two facets (persons and items), not three (persons, items and responses).

Example 2: Two facets in Labels=, but three in the data: paired comparisons.
Facets = 2, 3 ; the "3" is optional because it can be inferred from Entry=
Entry = 1,2,2 ; 3 facets positions in the data, with element numbers from Labels= facets, 1, 2 and 2.
Model = ?,?,-?,D ; "-?" reverses the sign of the second person measure

## 9.19 Fair averages based on = Mean

Fair average scores are reported in the output for each element. These are the scores that correspond to the logit measures as though each element of that facet encountered elements of similar difficulty in the other facets. Fair averages are intended for communicating the measures as adjusted ratings. This is useful when the audience have a strong conceptualization of the rating scale, but little interest in, or understanding of, the measurement system.

Fair average = Mean

This provides a norm-referenced average the measures for all elements (except this element) are set to the average values of the elements in their facets. It uses mean measure of the elements of each facet (except the current element) as the reference for computation. This is the default option. It is shown as Fair(M) in Table 7.

Fair average = Zero

This provides a criterion-referenced average the measures for all elements (except this element) are set to zero (logits or on user-scaling). It uses the origin of the measurement scale for each facet (except the current element) as the reference for computation. This was the default option in early version of Facets. It is shown as Fair(Z) in Table 7.

For the non-centered facet (typically persons), these two fair averages are usually the same. For a centered facet (e.g., items or raters) they are different. So for your non-centered rater facet, do you want the "fair-average" for a rater to be the rating given by this rater to a person with an "average" measure, or to a person with a "zero" measure? You may need to try both to identify which is actually what you want to report.

Look at your non-centered facet. Do you want the fair averages for all elements to be determined by a person at the Umean= value (Fair=zero) or a person at the person-sample mean (Fair=mean). If you are describing performances on this test then (fair=mean).

Example 1: An examination board wishes to use criterion-referenced fair scores for rater comparisons, because a "zero" logit person is at the pass-fail point. If students are the non-centered facet, then the fair scores for the students should be the same for fair=mean and fair=zero. For the raters, items, etc., fair=zero would be more student-sample-independent.

Fair score = Zero

Example 2: An examination board wishes to use ratings based on an average task rated by an average rater:

Fair score = Mean

Example 3: I wanted to use a fair average (with Fair=zero) of 2
as a cutscore. No person has exactly this. How can I find the person measure?

One approach:

1. Analyze your data and output an Anchorfile=
2. look for person measures with Fair Average near 2
3. In the Anchorfile=, change the anchored person measures so they cover the range discovered in 2. No need to change the data.
4. Analyze the modified anchor file and see which person measure has a Fair Average near enough to 2.0
5. redo 2, 3, 4 if needed.

Another approach:

1. Analyze your data and output the Scorefile= for the persons to Excel
2. Sort on the Fair Average column
3. Delete all values far from a Fair Average of 2.0
4. Plot Measureagainst Fair Average
5. Tell Excel to draw the trend line and display the equation
6. Put value of 1.33 into the equation.

## 9.20 Glabel group labels =

This specification enables element groups to be labeled. This appear in Table 7.

Labels=

```
4,Session,A      ; this is a dummy facet, used only for investigating interactions
 11,day 1 time 1 ,0, 1    ; group-day 1
 12,day 1 time 2 ,0, 1    ; group-day 1
 21,day 2 time 1 ,0, 2    ; group-day 2
 22,day 2 time 2 ,0, 2    ; group-day 2
 31,day 3 time 1 ,0, 3    ; group-day 3
 32,day 3 time 2 ,0, 3    ; group-day 3
 41,day 4 time 1 ,0, 4    ; group-day 4
 42,day 4 time 2 ,0, 4    ; group-day 4
*

Glabel =
; facet, group, label
4,1, group-day 1
4,2, group-day 2
4,3, group-day 3
4,4, group-day 4
*
```

Table 7.4.1.1 [group-day 1]  Session Measurement Report  (arranged by MN).

| Total Score | Total Count | Obsvd Average | Fair-M Avrage | Group | Nu | Session | |
|---|---|---|---|---|---|---|---|
| Group 1 [group-day 1] | | | | | | | |
| 684 | 144 | 4.8 | 4.73A | | 1 | 11 day 1 time 1 | |
| 674 | 144 | 4.7 | 4.73A | | 1 | 12 day 1 time 2 | |
| 679.0 | 144.0 | 4.7 | 4.73 | | 1 | Mean (Count: 2) | [group-day 1] |
| 5.0 | .0 | .0 | .00 | | 1 | S.D. (Population) | [group-day 1] |
| 7.1 | .0 | .0 | .00 | | 1 | S.D. (Sample) | [group-day 1] |

Example: high-low split of persons:
1. from Facets, produce the output table of the person measures in measure order
2. do a rectangular-copy (alt+mouse) using Microsoft Word or Notepad++ creating this list:
Person element number = ,,,, 1  ; for all the persons in the high group
....
Person element number = , , , 2  ; for all the persons in the low group
.....

3. append this list to the person element labels under Labels=. Keep the existing person element labels. *Facets* will combine the group numbers with the other person details.

4. Add group labels:
```
Glabel =
; facet, group, label
1,1, high   ; person facet is facet 1
1,2, low
*
```

Applying this to kct.txt:

Table 7.1.4  Children Measurement Report  (arranged by N).

```
|  Total    Total    Obsvd  Fair(M)|   +     Model | Infit      Outfit    |Estim.| Corr. |        |
                    |
|  Score    Count  Average Average|Measure  S.E. | MnSq ZStd  MnSq ZStd |Discrm| PtBis | Group |
Nu Children         |
|-------------------------------+--------------+----------------------+------+-------+-------
+---------------------|
|      9.7     18.0      .54     .46 |    -.38  1.04 | 1.00  -.2   .70  1.0 |      |   .82 |       |
Mean (Count: 35)     |
|      2.4      .0      .13     .31 |    2.24   .18 |  .96  1.2  1.33  1.1 |      |   .13 |       |
S.D. (Population)    |
|      2.4      .0      .13     .32 |    2.27   .18 |  .97  1.2  1.35  1.2 |      |   .13 |       |
S.D. (Sample)        |
|-------------------------------+--------------+----------------------+------+-------+-------
+---------------------|
|     11.0     18.0      .61     .63 |     .82  1.06 | 1.07  -.1   .78   .5 |      |   .84 |     1 |
Mean (Count: 24)     |
|      1.3      .0      .07     .21 |    1.30   .07 | 1.12  1.3  1.56   .9 |      |   .12 |     1 |
S.D. (Population)    |
|      1.3      .0      .07     .22 |    1.33   .07 | 1.14  1.3  1.59   .9 |      |   .12 |     1 |
S.D. (Sample)        |
|-------------------------------+--------------+----------------------+------+-------+-------
+---------------------|
|      7.0     18.0      .39     .09 |   -2.99   .98 |  .85  -.3   .49  1.9 |      |   .75 |     2 |
Mean (Count: 11)     |
|      1.8      .0      .10     .08 |    1.52   .29 |  .30   .7   .29  1.2 |      |   .12 |     2 |
S.D. (Population)    |
|      1.9      .0      .11     .08 |    1.59   .30 |  .31   .8   .30  1.2 |      |   .13 |     2 |
S.D. (Sample)        |
-------------------------------------------------------------------------------------------------
----------------------
```

## 9.21 Graph plotting file name = " "

**A better plot may result from copying Facets graph coordinates, see "Copy Plot Data to Clipboard".**

Provides the values plotted in Table 8 for the probability curves and expected score ogive in fixed column format. If comma-separated values are preferred, specify "Graphfile=filename" and CSV=Y. For Tab-separated values, specify "Graphfile=filename" and CSV=Tab

filenames ending .xls and .xlsx are written as Excel workbooks
filenames ending .sav are written as SPSS save files
filenames ending .rda and .rdata are written as R Statistics data files
filenames ending with any other suffix or no suffix are written as DOS text files.

Heading lines= controls display of heading lines. QM= quotation marks, controls whether labels are within quotation marks.

From the Output Tables menu, this Graph Output file controls the production of the Graph output file

| Field | Columns | Contents |
|---|---|---|
| 1 | 1-10 | Scale number(matches Sub-Table number for Table 8) |
| 2 | 11-20 | Measure (X-Axis) |
| 3 | 21-30 | Expected Score |
| 4 | 31-40 | Category value corresponding to expected score |
| 5 | 41-50 | Probability for bottom category |

| | | |
|---|---|---|
| 6 | 51-60 | Probability for second category |
| 7 ... | 61-70, ... | Probabilities for higher categories |

```
Scale    Measure  Expected    ExpCat    Prob:0    Prob:1    Prob:2
    1      -3.00       .10         0       .90       .10       .00
    1      -2.91       .11         0       .89       .11       .00
    1      -2.82       .12         0       .88       .11       .00
```

Example 1: Provide coordinates so I can plot beautiful probability curves using my graphics package: Graphfile=plot.dat,CSV; the name of an output file with comma-separated values, produces, for a 3 category rating scale (or partial credit):

```
1,-2.00,.23,0,.79,.20,.01
1,-1.92,.24,0,.77,.21,.02
1,-1.85,.26,0,.76,.22,.02
```

Example 2: You want a computed raw rating (score) on each criterion for each essay. The computed raw rating matches the observed raw rating (score). The observed rating is awarded by Rater X. The computed rating is awarded by Rater Z, who has the mean rater leniency, which is usually zero.

1. Output the Facets Graph file.
2. Compute the difference between the measure for Rater X and the measure for Rater Z.
3. Find the Measure for the observed rating (in the Expected Score column) in the Graph file
4. Adjust this measure by the measure difference from 2.
5. The computed rating is the Expected Score at the adjusted Measure.

If the rater is lenient, then the computed rating will be lower than the observed rating (except for the bottom category).
If the rater is severe, then the computed rating will be higher than the observed rating (except for the top category).

## 9.22 Hardware = 80

Specifying a value with Hardware= sets the range of the exponents calculated by Facets. Reduce this value in steps of 10 to overcome numeric overflow problems.

Example: For an early model IBM-PC with low precision computation:
Hardware = 40

## 9.23 Heading lines = Yes

This controls the inclusion of heading lines in the graph, score, and residual output files. This setting can be changed with Edit Initial Settings.

| | |
|---|---|
| **Heading lines = Yes** | include heading lines |
| **Heading lines = No** | omit heading lines |

Example: you will read the residual file directly into your statistics program
Residual file = datafile.txt
Heading lines = no

## 9.24 Inter-rater Agreement Coefficients = 0

Specify the rater facet number to report agreement among raters.

Inter-rater = 0
Rater agreement statistics are not reported.

Inter-rater = rater facet number
This is the facet number of the rater facet, e.g., Inter-rater = 2 ; Facet number 2 is the rater facet. See Agreement Statistics for more details. Matching of ratings is done using all other facets.

Inter-rater = rater facet number, agreement facet number, agreement facet number, .....
Inter-rater agreement statistics are matched only on facets listed as agreement facets.

Example: There are 4 facets. The rater facet is 3. Only Facet 1, students, and Facet 2, tasks, are to be used for inter-rater agreement. Facet 4, rating-session, is ignored for inter-rater agreement.

```
Facets = 4 ; 1=students,2= tasks, 3=raters, 4=rating-session
Inter-rater = 3, 1, 2 ; rater facet is 3, agreements for facets 1 and 2.
Data=
11,72,63,3, 7 ; rater 63 rates student 11 on task 72. Session 3 is ingored.
matches
11,72,34,2, 7 ; rater 34 rates student 11 on task 72. Session 7 is ignored.
```

## 9.25 Iterations (stop after) = 0 (unlimited)

This specifies the maximum number of iterations to be performed, regardless of the other convergence criteria. A value of 0 permits an unlimited number of iterations. See also "My analysis does not converge."

Iteration can always be made to terminate by pressing the Ctrl+F keys simultaneously. See Short-cut keys.

To set the maximum iterations to 20, type:

```
Iterations=20
```

To allow the iterative process to continue until convergence is reached on the other criteria, type:

```
Iterations=0
```

## 9.26 Juxtapose (column headings after every this many lines) = 0

This is useful for setting output page length. This is the number of rows in reports before column headings are repeated. Headings always repeat just before table summary lines, when there are more than 15 element lines to report.

Juxtapose = 0 means do not repeat the headings.
Juxtapose = 30 means repeat the heading every 30 rows of output.

## 9.27 Labels of facets and elements = (required)

This specification names each facet and follows it by a list of the element numbers and labels (names) to be included in the analysis.

| **Labels =** (Nobuild) | | | |
|---|---|---|---|
| Facet number , | Facet label or name, | Anchoring, (A,D,G,X) + No group reporting (N) | Estimation priority |

| Element number = | Element label or name, | Measure, starting value, anchor value or group-anchor value | Group number, for group-reporting or group-anchoring | Element weight, defaults to 1.0 | Target element number for merged elements: observations in the data with the Element number are assigned to the Target element number. Example: 1= , , , , 2 so that 1's become 2's in the data for this facet |
|---|---|---|---|---|---|
| * | | | | | |

**1. Labels = (or if Labels= is omitted)**

*Facets* adds to Output Table 2 a list of the elements in the data that are not specified in Labels=.
These can be copy-and-pasted into your specification file.
Specify **Labels=Nobuild** to omit the list of elements not specified in Labels= from Table 2.

## 2. Facet Identification

**2A. Facet numbers**

Choose the order you prefer to think about your facets, and number them in that order. The number you assign to each facet will determine the order in which the facets are reported.

If your facets are persons, judges and items, you may choose:
Facet 1: judges
Facet 2: persons
Facet 3: items

**2B. Facet Anchoring**

Choose the order you prefer to think about your facets, and number them in that order. The number you assign to each facet will determine the order in which the facets are reported.

If your facets are persons, judges and items, you may choose:
Facet 1: judges
Facet 2: persons
Facet 3: items

| Facet Anchor/Group code | Element Specification | Description |
|---|---|---|
| (None) | Element number = (label), (value), **Group number** | Group number is used for reporting groups of elements in Table 7 |
| A | Element number = (label), **Value**, (group number) | Element measure is anchored at Value. Groups reported if there is a group number. |
| D | Element number = (label), (value), (group number) | Element measure is anchored at **Umean=** value. Groups reported if there is a group number. |
| G | Element number = (label), **Value**, **Group number** | Each element measure is measured independently but the sum of the measures |

| | | |
|---|---|---|
| | | of the group of elements are constrained to equal the sum of their Values. |
| X | Element number = (label), **Value**, **Group number** | The same as G, but elements with extreme scores (zero, perfect) are excluded. |
| ...N | Element number = (label), (value), (group number) | This suppresses group reporting for a facet in Table 7. Groups are ignored for measure reporting.<br>Labels=<br>2=Persons,GN ; group-anchoring wanted, but Table 7 does not report measures for each group separately<br>1=Ben,0,1 ; Ben is in group 1<br>2=Mike,0,2 ; Mike is in group 2<br>..... |

| **Valid Numerical Ranges** | | |
|---|---|---|
| Facets | 1 - 255 | Any number in the range is valid. They do not need to start at 0 or 1, nor do they need to be in sequence. |
| Elements | Element numbers are in the range 1 - 2,147,438,648<br><br>Element 0 is usually the null element number, meaning "this facet does not apply to this observation." Element 0 can be used as an element number, provided Null element= is used to change the null element number from 0 to another number. | |
| Group numbers | 0-255<br>0 means "no group" | |
| Observations | 0-255 (can be rescored with Rating scale=) | |

**2C. Estimation priority -** Ignore this for most analyses.
This allows the measures for the elements of some facets to be estimated before the elements of other facets. For instance, facets modeling interaction terms can be estimated after their main effects have been estimated.
Priority 0: do not estimate the measures. This is equivalent to anchoring (A or D).
Priority 1: estimate the measures (the default)
Priority 2: estimate the measures after the Priority 1 measures have been estimated.
Priority 3: estimate after Priority 1 and 2.
Priority 4 and higher: estimate after higher priority measures have been estimated.

## 3. Element Identification

See Elements.

# 9.28 Left-hand placement of row labels = No

Element labels, category labels and other row identifiers are placed at the right-hand end of each output line in Tables 7,8 and the Bias Tables. These may be placed at the left-hand end:
Left = yes

Left=No

| Total Score | Total Count | Obsvd Average | Fair(M) Average | Agree. Exp % | N | Senior scientists |
|---|---|---|---|---|---|---|
| 35 | 35 | 1.00 | 1.01 | 1.4 | 1 | Avogadro |
| 159 | 35 | 4.54 | 4.48 | 12.0 | 2 | Brahe |
| 181 | 35 | 5.17 | 5.15 | 11.5 | 3 | Cavendish |

Left=Yes

| N | Senior scientists | Total Score | Total Count | Obsvd Average | Fair(M) Average | Agree. Exp % |
|---|---|---|---|---|---|---|
| 1 | Avogadro | 35 | 35 | 1.00 | 1.01 | 1.4 |
| 2 | Brahe | 159 | 35 | 4.54 | 4.48 | 12.0 |
| 3 | Cavendish | 181 | 35 | 5.17 | 5.15 | 11.5 |

## 9.29 Missing data codes =

Codes that indicate "missing" whenever they appear in the data may be specified with Missing=. They will then be ignored whenever they appear. "Missing" refers to missing observations. It means "observation unknown or observation not collected, so ignore it".

There cannot be "missing element numbers". If the element number is unknown (for instance, you don't know which item was being rated), then the observation cannot be included in the analysis - so omit it from the data or comment it out with ";". If the element number does not apply (for instance, an item that does not require a rater), then the element number is null, usually 0, the Keep as null= value.

Facets treats as "missing data", any data value (or space or omitted) outside the specified numerical range for a valid observation.

Example: Models=?,?,?,R2
Valid data for R2 are 0,1,2 so anything else is treated as missing data.

The first time Facets encounters a value outside the valid range (other than blank or omitted), it issues a warning message, just in case you have a data entry error, but proceeds with the analysis:
Check (2)? Invalid datum location: "___" at or near line: ___. Datum "_" is too big or non-numeric, treated as missing.

I recommend being explicit about missing data. Use the same code every time, e.g., ".".
If you do this, then you can specify:
Missing = .
so that Facets will not issue a warning message for ".".

So in your file you might have:
Facets=4
Models=?,?,?,?,R6
1,2,3,1-4, 3, 2, ., 4 ; where the response to item 3 is missing

Example 1: Missing observations have been entered in the data file with code "9". These are to be ignored and bypassed.

```
Missing = 9
Facets = 2
Data=
22, 36, 1 ; this observation of "1" is analyzed
```

23, 37, 9 ; this observation of "9" is ignored as missing data

Example 2: Codes "^", "." and "#" are all to be ignored as observation values.
Missing = ^, ., #
Facets = 3
Data=
22, 0, 17, 1 ; this observation of "1" is analyzed. Element "0" means that facet 2 does not apply to this observation.
23, 3, 13, # ; this observation of "#" is ignored as missing data

Example 3: To make specific observations into missing data use the missing data model.
Models=
23, 37, M ; all data points with elements 23 and 37 are treated as missing

## 9.30 Model to be used in the analysis = ?,?,D,1

This specifies how the facets interact to produce the data. Its form parallels that of the data. One indicator, such as "?", is specified for each facet in the data, followed by another indicator, such as "D", for the measurement model which is specified to produce the data. Additional models can be listed after the Model= statement and followed by an "*". Model weighting can be specified after the model-type indicator.

How Models= functions with data: Matching data with measurement models
and Model statement examples

---

The process is:

*Suppose 6 experts rate 19 items for quality of manufacture on a rating scale from 1 to 4:*

**A. Decide how many facets you have.**
There are 2 facets: experts and items
Let's call experts facet 1 and items facet 2.

Then, in your Facets specification file, you will have
Facets=2

**B. Identify the individual elements in the two facets**
Then, in your Facets specification file, you will have
Labels=
1, experts
1, first expert
.....
6, sixth expert
*
2, items
1, first item
2, second item
....
19, nineteenth item
*

**C. Decide how they interact**
Any expert - this is indicated by "?" can interact with any item indicated by "?"

**D. Decide on the response structure.**
It is a rating scale, indicated by "R", with the highest category "4"

So in your Facets specification file you will have
Models = ?,?,R4 ; and element in facet 1 (expert) can interact with any element in facet 2 (item) to produce a rating on a scale whose highest category is 4.

**E. The data will look like:**
element from facet 1, element from facet 2, rating
3, 18, 2 ; expert 3 gave item 18 a rating of 2.

---

A wide variety of models can be constructed to enable measures to be estimated from many types of qualitative data. Facets are specified in the same order in the Model= specification as they are in the data lines. Each model definition includes one entry for each facet specified in the Labels= specification, unless overridden by an Entry= specification. Zero terms, "0", in the Entry= specification are bypassed and not referenced in Model= specifications.

Each model specification includes

a) control characters, such as "?", except for "0" facets in an Entry= specification.

b) a code specifying the type of scale (dichotomy, rating scale, partial credit, etc.), or giving the name of a scale explicitly defined by a Rating (or partial credit) scale= specification.

c) Optionally, a weight to be assigned to data matching this model. The standard value is 1.
Weights are always arbitrary, based on other information and value judgements external to the data. Use weights only when non-measurement considerations have a specific, justifiable priority, e.g., when a 100 item MCQ test and one essay graded on a 5 point scale are to be given equal weight in the final, combined measure.

d) Optionally, as a final parameter following the weight, a scale description,
Model=?,?,R,,Farley stress scale ; ",," indicates the standard weight of 1 applies

---

Control characters can be difficult to understand at first. On first reading, skip down to the "Examples" to get the feel of what this is all about.

Facets are positioned in the same order in the Model= specification as they are positioned in the data lines. Each model definition includes one entry for each facet in a data line. Zero terms, "0", in the Entry= specification are bypassed.

| **Model= Facet control characters** | *Meaning for this facet* |
|---|---|
| **? or ?? or $** | Model can match a datum with any element number in this facet, e.g., any examinee. |
| **# or ##** | Model can match a datum with any element number in this facet, e.g., any item. Also, each element of this facet matched to this model has its own rating scale, i.e., "#" specifies a "partial credit" model. Usually only one facet has # in a PCM model specification. |
| **blank** | Ignore this facet when matching the Model= statement to a datum, but verify that the element number for this facet for a datum that matches this model statement is listed after Labels=. Typed as ",,". |

| | |
|---|---|
| **X** | Ignore this facet when matching data to this Model= statement. Do not check the element number in this facet for validity when a match occurs. |
| **0** or Keep zero= value | Model can only match a datum in which this facet does not participate, i.e., when element number 0 is used for this facet in the datum reference. |
| **-** e.g., -? or -# | Reverses the orientation of the measure of the element of facet, when combined with other facet control characters. "-?" means can match any element, but with the element's measure reversed in direction and sign. When using -?, it is recommended that the data be entered twice, once with each facet as -?, and the models= be weighted 0.5. |
| **an element number**, e.g., 23 | Model can only match a datum with exactly this element number in this facet. **Element labels** are not allowed. |
| **number-number,** e.g., 23-36 | Model can match a datum with any element number from the specified range in this facet. |
| **number#** 23# or 23-36# | Model can match a datum with a matching element number from the specified number, but each element number is associated with a unique "partial credit" scale. |
| **@ or @number or @number-number** | This facet is used for reference by Dvalues= or for model selection. It is ignored (not used) for measurement. |
| **B** e.g., Model= ?B,?,?B,R | Generate Bias interaction estimates for combinations of the elements of each facet marked by "B". At least two "B" terms are needed. The "B" is appended to one of the other facet control characters, e.g., "?B". The bias interactions are coded in one or more model statements, but act as though they are coded in all model statements. Model statements with different combinations of facets marked by "B" each produce separate sets of bias estimates for all the data.<br>For example: for 2-way interactions between 4 facets (also from Output Tables menu):<br>Models=<br>?,?,?,?,R4<br>?B,?B,?,?,R4 ; interaction between facets 1 and 2<br>?B,?,?B,?,R4 ; interaction between facets 1 and 3<br>?B,?,?,?B,R4 ; interaction between facets 1 and 4<br>?,?B,?B,?,R4 ; interaction between facets 2 and 3<br>?,?B,?,?B,R4 ; interaction between facets 2 and 4<br>?,?,?B,?B,R4 ; interaction between facets 3 and 4<br>; 3-way interaction, only from Specification file<br>?B,?,?B,?B,R4 ; interaction between facets 1, 3 and 4<br>* |

More than one model can be specified. See: Matching data with models.

| **Model= and Rating Scale= Scale codes** | *Meaning for this model* |
|---|---|

| | |
|---|---|
| **D** | Dichotomous data. Only 0 and 1 are valid. |
| **Dn** | Dichotomize the counts. Data values 0 through n-1 are treated as 0. Data values n and above are treated as 1. E.g., "D5" recodes "0" to "4" as 0 and "5" and above as 1. |
| **R** | The rating scale (or partial credit) categories are in the range 0 to 9. The actual valid category numbers are those found in the data. RK to maintain unobserved intermediate categories in the category ordering. |
| **Rn** | The rating scale (or partial credit) categories are in the range 0 to "n". Data values about "n" are missing data. The actual valid category numbers are those found in the data. If 20 is the largest category number used, then specify "R20". |
| **RnK** | Suffix "K" (Keep) maintains unobserved intermediate categories in the category ordering. If K is omitted, the categories are renumbered consecutively to remove the unobserved intermediate numbers |
| **M** | Treat all observations matching this model as Missing data, i.e, a way to ignore particular data, effectively deleting these data. |
| **Bn** | Binomial (Bernoulli) trials, e.g., "B3" means 3 trials. In the Model= statement, put the number of trials. In the Data= statement, put the number of successes. Use Rating Scale= for anchored discrimination. |
| **B1** | 1 binomial trial, which is the same as a dichotomy, "D". |
| **B100** | Useful for ratings expressed as percentages %. Use Rating Scale= for anchored discrimination. |
| **P** | Poisson counts, with theoretical range of 0 through infinity, though observed counts must be in the range 0 to 255. Use Rating Scale= for anchored discrimination. |
| **the name of a user-defined scale** | A name such as "Opinion". This name must match a name given in a Rating (or partial credit) scale= specification. |

Example: A test in which every item has a different rating scale (partial credit) with a different numbers of categories.
The highest numerical category of any item "6".
Model = ?, ?, #, R6 ; this allows items with categories such as 0,1,2 and also 1,2,3,4,5,6

There are more examples at Model statement examples.

---

**Data weighting:** This specifies the weight to be assigned to each datum in estimating measures, fit statistics and bias analyses. This is entered in the Model= specification after the scale code, e.g., **Model=?,12,D,2** specifies a weight of 2 for responses to item 12.

| **Model= Weighting control** | Meaning for this model |
|---|---|
| **1 (the standard)** | Give the datum the standard weight of 1 in estimating measures, fit statistics and bias sizes. |
| **n** | Give the datum a weight of "n", e.g., 2.5, in estimating measures and fit statistics. This gives it greater influence than data with lesser weights. |
| **0** | Give the datum zero weight, i.e., treat the datum as missing (but report, if possible, in the residual file.) |

Adjust the weights as a set so that the standard errors reported for the persons by weighted and unweighted analyses are about the same. This prevents the weighting misleading you about test reliability, etc.

**Data replication:** Data point replication or weighting can be done by prefixing R (or another replication character) + weight before the data point after Data=, e.g.,
R12.5 , 1 , 2, 3 means weight 12.5 times, the observation of "3" for element 1 of facet 1 and element 2 of facet 2.

Multiple identical sets of observations can be replicated on the same line, by preceding the data for one observation by R and the number of replications, e.g, 20 replications are indicated by R20.

Example 1: Survey data has been summarized by response rating. 237 people responded in category 3 on item 27.
Data=
R237,27,3

---

Example 1: The basic Rasch model for dichotomous interactions between objects and agents is specified by:
Model=?,?,D

"?,?,D" specifies that any element of the first facet (the first "?") can interact with any element of the second facet (the second "?") to produce a dichotomous observation (the "D"). Record a dichotomous observation as a "1" for success/right/more, or a "0" for failure/wrong/less. This implements the basic Rasch dichotomous model:
log(Pni1/Pni0) = Bn - Di
where
Pni1 is the probability of person n getting item i right
Pni0 is the probability of person n getting item i wrong
Bn is the ability of person n
Di is the difficulty of item i.

Example 2: The Andrich rating scale model for judges, persons and items is specified by:
Model=?,?,?,R
"?,?,?,R" states that any judge,"?", can rate any person, "?", on any item, "?", using a common rating scale, "R".
This implements an Andrich rating scale model:
log(Pnijk/Pnijk-1) = Bn - Di - Cj - Fk
where
Pnijk is the probability that person n is awarded, on item i by judge j, a rating of k
Pnijk-1 is the probability that person n is awarded, on item i by judge j, a rating of k-1
Bn is the ability of person n
Di is the difficulty of item i
Cj is the severity of judge j
Fk is the Rasch-Andrich threshold (step calibration) of step k of the rating scale. This is the location on the latent variable (relative to the item difficulty) where categories k and k-1 are equally probable.

Example 3: More than one model can be specified. See: Matching data with models. A multiple-model analysis of only items 4 and 5. Item 4 is a true/false dichotomous item, but item 5 is a Likert rating scale (or partial credit) item. The examinees are facet 1, and the items are facet 2:
Model=?,4,D
?,5,R
*

or, all models may be specified on lines following Model=,
Model=
?,4,D

?,5,R
*

Example 4: I have a 2 rating scale instrument of 32 items. The first 19 items are on one 6 category rating scale and the remaining 13 items are on a different 6 category rating scale. There are 4 facets, and the items are the 4th facet.

Facet = 4
Models =
?,?,?,1-19,R6 ; items 1-19 are on a rating scale with highest category numbered "6"
?,?,?,20-32,R10K ; items 20-32 are on another rating scale with highest category numbered "10". Unobserved intermediate categories are to be maintained in the ordering "K".
*

Example 5: If one item is to have more weight than another, e.g., a correct answer on item 31 is worth 2 points.

Models=
?, ?, 31, D, 2 ; weight 2
?, ?, ? , D, 1 ; default weighting of 1
*

Example 6: Some responses are to be treated as missing data

Facets=3
Models=
2,1,20,M ; this is the "missing data" model
2,1,24,M
2,2,20,M
?,?,?,R
*

Example 7: Two different items to the same rating scale:

Facets=3
Models =
1,?,?,MyScale ; item 1 uses MyScale
4,?,?,MyScale ; item 4 uses MyScale
?,?,?,D ; everything else is a dichotomy
*
Rating Scale = MyScale,R, G ; G means all General, so items 1 and 4 share the same scale
or
Rating Scale = MyScale,R, S ; S means all Specific, so items 1 and 4 have different versions of MyScale

Example 8: More examples of model statements

Example 9: Weighting:
Two Cases: A and B. Four aspects: Taste, Touch, Sound, Sight.
Case A Taste weight twice as important as the rest.
Case B Sound weight twice as important as the rest.

Labels =
1, Examinees
1-1000
*
2, Case
1=A
2=B
*
3, Aspect
1=Taste

2=Touch
3=Sound
4=Sight
*
Models=
?, 1, 1, MyScale, 2 ; Case A Taste weighted 2
?, 2, 3, MyScale, 2 ; Case B Sound weighted 2
?, ?, ?, MyScale, 1 ; everything else weighted 1
*
Rating scale = MyScale, R9, General ; this rating scale is the same for all models

If you want to keep the "reliabilities" and standard errors meaningful then adjust the weights:

Original total weights = 2 cases x 4 aspects = 8
New total weights = 2 + 2 + 6 = 10
Weight adjustment to maintain total weight is 8/10.

So adjusted weighting is:
Models=
?, 1, 1, MyScale, 1.6 ; Case A Taste
?, 2, 3, MyScale, 1.6 ; Case B Sound
?, ?, ?, MyScale, 0.8 ; everything else
*

## 9.31 Negatively-oriented facet =

This command specifies which facets are negatively oriented. The standard in *Facets* is for only the first facet to be measured positively. If both Positive= and Negative= are specified, only the first one encountered is actioned. See Positive= for more details.

Example 1: Negative = 2, 3
Persons (facet 1) respond to items (facet 2) which are scored by raters using rating scales (facet 3). Persons are measured positively, i.e., a high ability measure means the raw score was high, a low measure means the raw score was low. On the other hand, items are measured negatively so the higher the Rasch measure, the greater difficulty of the item; so are raters, such that the higher the Rasch measure, the greater severity of the rater. This is a convention in educational measurement.

Example 2: Negative = 0
No facets are negatively-oriented. All facets are oriented positively. For instance, Patients (facet 1) are rated on tasks (facet 2) by raters (facet 3). All facets are measured positively, so that higher Rasch measures mean that patients scored higher, that tasks were easier, and that raters were more lenient. This is the convention in health care measurement.

## 9.32 Newton-Raphson step size = 0

Facets uses Joint Maximum Likelihood Estimation (JMLE) or a *Facets* implementation of PMLE to estimate the Rasch measures from ordinal data. This requires an iterative process. Initial estimates are imputed for all the element measures. The expected observations are computed based on these estimates and totaled for each element. Then for each element, the observed and expected total scores are compared, and the element measure re-estimated to a value intended to make the expected total score the same as the observed total score. This process is repeated until the differences between the observed and the expected total scores are too small to matter. This is called convergence.

Initial estimates are obtained using the PROX (normal approximation) algorithm.

More exact estimates are obtained using iterative curve fitting (when Newton=0) or the Newton-Raphson method. Newton= can be set using the Estimation menu.

| | |
|---|---|
| Newton=0 | specifies iterative curve fitting. The expected scores follow logistic ogives. The improved measure estimates are the locations on the logistic ogives predicted to produce the observed total scores. |
| Newton=0.1<br>Newton=0.5<br>Newton=1<br>Newton= ... | specifies Newton-Raphson method. The expected scores follow local curves defined by their first and second derivatives. When Newton=1, the improved measure estimates are the locations predicted to produce the observed scores. When Newton=0.5, the improved estimates are halfway between the current estimates at the Newton=1 estimates, and similarly for other values of Newton=. |

## 9.33 Noncenter facet = 1

**One facet must be non-centered in most analyses**, otherwise the estimates are over-constrained.

This specifies which facets are not to be centered. Other facets are centered, unless they are anchored. Using Noncenter=, centered facets are constrained to have a mean element measure of zero for *non-extreme* element measures. Center= can constrain facets to have a mean element measure of zero for *all* element measures.

In most analyses, if no facet is non-centered then the estimation is over-constrained and displacements are reported. If more than one facet is non-centered, the estimation is not sufficiently constrained and the estimates are ambiguous (non-identifiable). The usual convention is to center the frame of reference on the agents of measurement: items, tasks, judges. This establishes the origin. Then the objects of measurement, persons, are positioned with respect to this origin.

Extreme measures (corresponding to minimum and maximum possible scores) are included in the Means in Table 7, so that these may not be 0 for centered facets unless Center=(facet number)Z

If both Noncenter= and Center= are specified, only the first encountered is actioned.

When any element within a facet is anchored, or group-anchored, then the origin of the scale for that facet is set by the anchoring. Both Center= and Noncenter= are ignored.

Example 1: Persons are facet 2 and are not to be centered, then
Noncenter=2

Example 2: facet 1 contains the boys, and facet 3 the girls:
Noncenter=1,3

Example 3: All facets are anchored or centered or self-defined. Self-definition occurs when there is only one active facet so that the item difficulties are defined relative to an implicit origin, as in the "Three Mile" example.
Noncenter= or Noncenter=0

## 9.34 Null element = 0

| *Specification* | *For all facets ....* |
|---|---|
| Keep as Null= 0<br>or<br>Null element= 0 | Element number 0 is entered as the element number in a data record when a facet does not participate in generating an observation. |

| | |
|---|---|
| Keep as Null= positive integer, e.g., 999<br>or<br>Null element= positive integer, e.g., 999 | Element number (positive integer) is entered as the element number in a data record when a facet does not participate in generating an observation. Element 0 can be used as an active element. |
| Keep as Null= -1<br>or<br>Null element= -1 | All facets participate in every observation. Element 0 can be used as an active element. |

Example 1: Element number 0 means "Short", and "1" means "Tall". "9" means "this facet is not applicable".

```
Null = 9
Labels=
4, Height
0 = short
1 = tall
*
Data=
23,9,14,0,2 ; facet 2 does not participate, facet 4 has element number 0
23,7,14,1,4 ; all facets active
23,9,14,9,3 ; facets 2 and 4 not participating
```

Example 2: Null elements when element labels are identifiers in the data.

```
Facets =3 ; students, items, raters
Delements = N, 3L  ; element identifiers are numbers except for facet 3 when the identifiers are the rater names
Null = 0  ; the "does not apply" element is element number 0
Labels =
1, Students
100-150   ; 50 students, numbered 100 to 150
*
2, Items
1-10 ; these are MCQ items
21-25 ; these are essay-writing items
*
3, Raters
0 = None  ; no rater = the Null element
1 = George
2 = Martha
3 = Betty
*
Data =
52, 3, None, 1 ; student 52 succeeded on MCQ item 3 which has no rater
67, 24, Martha, 3 ; student 67 on essay item 24 was rated by Martha with a 3.
.....
```

## 9.35 Omit unobserved elements = Yes

| | |
|---|---|
| | Elements with observations are always reported. |
| Omit unobserved elements = Yes<br>Show unobserved elements = No | Do not report elements that have no observations in the current dataset |
| Omit unobserved elements = No<br>Show unobserved elements = Yes | Table 7 reports anchored elements that have no observations in the current analysis |

| | |
|---|---|
| | Scorefile= : reports all elements specified in Labels= whether they have observations or not. |

The setting of "Omit unobserved elements =" can be changed with Edit Initial Settings and in the Table 7 and Scorefile= dialogs.

Example 1: There are unobserved anchored raters, because the whole rater bank is automatically included in every analysis. We don't want to see them in Table 7
Omit unobserved elements = Yes ; don't display the unobserved anchored elements

Example 2: we want to see which elements have not yet been rated in Scorefile=
Omit unobserved elements = No ; display all the unobserved elements

## 9.36 Output results/anchors/residuals =

The output file name can be entered at the prompt when Facets starts, or be entered as part of the specification file.

The name of the file that will contain the main results of the analysis is the first entry. This is the name of file into which output is written, e.g.,
Output = results.txt

An output file name given on the DOS command line supersedes the results file name given here, e.g, C:>FACETS SPECS.txt OUTPUT.txt, causes output to go to OUTPUT.txt rather than RESULTSOUT.txt. If two files names are given on the DOS line, they are output and anchor files, three names are the output, anchor and residual files.

The anchor output file name for a file of anchoring values is the second entry (or specified with Anchorfile=. Whenever a file name is given, then Facets writes out a new specification file containing the input specifications for this analysis. This file omits any specifications entered on the DOS command line, but includes the final estimated measures for all elements and scales. This file also has all estimated measures marked as ",A" for "anchored".

These anchor values can be used as starting values for slightly different later runs. Edit the ",A"'s out of the first line of each facet in the Labels= specifications and out of the category lines of Rating scale=.

The residual output file name for a file of observations, their expected values and their residuals is the third entry, or specified with Residualfile=. Whenever a file name is given, the standardized residuals for all the data points will be written to it.

Example 1: A file of anchor values is to be produced to be used in an equating study. This file is to be named "Anchor.txt":

```
Output = Results.txt,Anchor.txt
 or
Outputfile = Results.txt
Anchorfile = Anchor.txt
 or
C:>Facets SPECS.txt Results.txt Anchor.txt
```

Example 2: A file of residual values is needed for fit analysis, but no anchored file is wanted.

```
Output = Results.txt, ,Residual.txt ; Note: ", ," means no anchor output file
 or
Outputfile = Results.txt
Residualfile = Residual.txt
```

## 9.37 Positively-oriented facet = 1

This command specifies which facets are positively oriented. The standard in *Facets* is for only the first facet to be measured positively, so consult the table below to understand how to interpret the Rasch measures which are output by *Facets.* The facet orientations specified in your command file are indicated in the output on the all-facet ruler (Table 6.0) with a + or - sign at the top of the ruler. *(This section incorporates text from William Bonk.)*

If both Positive= and Negative= are specified, only the first one encountered is actioned.

Example 1: Positive=1 (Only facet 1 is positively-oriented. *Facets* uses this as standard)
Persons (facet 1) respond to items (facet 2) which are scored by raters using rating scales (facet 3). Persons are measured positively, i.e., a high ability measure means the raw score was high, a low measure means the raw score was low. On the other hand, items are measured negatively so the higher the Rasch measure, the greater difficulty of the item; so are raters, such that the higher the Rasch measure, the greater severity of the rater. This is a convention in educational measurement.

| Facet | Interpretation |
|---|---|
| Examinee (positive) | Higher Rasch measures mean greater ability |
| Test items (negative) | Higher Rasch measures mean items are harder |
| Survey items (negative) | Higher Rasch measures mean items are more difficult to endorse |
| Raters (negative) | For examinees: Higher Rasch measures mean raters are more severe<br>For standard-setting: Higher Rasch measures mean raters are less demanding |
| Prompt (negative) | Higher Rasch measures mean prompts are harder (tend to yield lower raw scores) |

Example 2: Positive=1,2,3 (All facets are positively-oriented)
Another method is to orient all facets positively. For instance, Patients (facet 1) are rated on tasks (facet 2) by raters (facet 3). All facets are measured positively, so that higher Rasch measures mean that patients scored higher, that tasks were easier, and that raters were more lenient. This is the convention in health care measurement.
Positive=1,2,3

| Facet | Interpretation |
|---|---|
| Examinee (positive) | Higher Rasch measures mean greater ability |
| Test items (positive) | Higher Rasch measures mean items are easier |
| Survey items (positive) | Higher Rasch measures mean items are easier to endorse |
| Raters (positive) | For examinees: Higher Rasch measures mean raters are more lenient<br>For standard-setting: Higher Rasch measures mean raters are more demanding |
| Prompt (positive) | Higher Rasch measures mean prompts are easier (tend to yield higher raw scores) |

Positive= specifies which facets are to be measured positively, i.e., higher score means higher logit value. The other facets will be negatively oriented. Positively-oriented facets are indicated by "+" in Table 6.0. Positive-orientation and Negative-orientation" are complex. We must answer:

For a particular facet,

**Do increasing raw scores indicate more or less of what we are looking for?**
counting right answers = more capability
counting mistakes = less capability
longer times in performing a task = less capability
higher Likert rating = more agreement or less agreement (depending how the categories are numbered).
higher p-value = more easiness for items
higher ratings = more leniency for raters
etc.

For a particular facet,
**Do we want higher Rasch measures to indicate higher or lower levels of what we are looking for:**
more ability as in educational testing
more disability as in some medical applications
more leniency for raters
more severity for raters
more difficulty (to perform, to answer correctly, to agree with, etc.) for items
more easiness (to perform, to answer correctly, to agree with, etc.) for items
or
etc.

Then, for a particular facet,
**if higher raw scores align with what we want higher Rasch measures to indicate, then the facet must be coded as "Positive=",**
**but if lower raw scores align with what we want higher Rasch measures to indicate, then the facet must be omitted from "Positive=".**

| | **Orientation of facet** | |
|---|---|---|
| **Meaning of raw scores** | **Positive** | **Negative** |
| Examinees: higher raw score indicate higher capability | Higher Rasch measures mean greater ability | Higher Rasch measures mean greater disability *(not recommended because tends to be misunderstood)* |
| Items: higher p-values or average ratings indicate greater easiness | Higher Rasch measures mean greater item easiness | Higher Rasch measures mean greater item difficulty |
| Raters: higher average ratings indicate greater leniency for examinees or more demanding for standard setting | Higher Rasch measures mean greater rater leniency or more demanding for standard setting | Higher Rasch measures mean greater rater severity or less demanding for standard setting |
| Prompt: higher average raw scores indicate greater easiness | Higher Rasch measures mean greater prompt easiness | Higher Rasch measures mean greater prompt difficulty |

Example 3: The usual convention in educational testing is that only those facets corresponding to the objects of measurement, typically persons, are measured positively (higher score means higher ability measure), but the agents of measurement, items, tasks and judges are measured negatively (lower score means higher difficulty measure or higher severity measure). For instance, Persons (facet 1) take items (facet 2) with some items rated by raters (facet 3). Persons are to be measured positively, i.e., higher score corresponds to higher ability measure. Items are to be measured negatively (lower score corresponds to higher difficulty measure) and raters are to be measured negatively (lower score corresponds to higher severity).

Positive=1

Example 4: In Example 3, the raters (facet 3) are to be reported by leniency (higher score means higher leniency measure), not by severity (lower score means higher severity measure).
Positive=1,3

Example 4: The usual convention in healthcare is that all facets corresponding are measured positively: higher score means higher ability measure, higher easiness measure or higher leniency measure. For instance, Patients (facet 1) are rated on items (facet 2) by raters (facet 3). All facets are to be measured positively, i.e., higher score corresponds to higher measure.
Positive=1,2,3

Example 5: Both facet 2, boys, and facet 4, girls, are to be reported as positively oriented (higher scores = higher measures). All other facets negatively (higher scores = lower measures):
Positive=2,4

Example 6: No facets are to be positively oriented. For all facets, higher scores mean less of what we are looking for.
Positive=

## 9.38 Pt-biserial correlation = Measure

This reports the correlation between the raw-score or measure for each element.

| | |
|---|---|
| Pt-Biserial = Yes or P or B or Omit | Point-biserial correlation of observation for the current element with its average observation omitting the observation. |
| Pt-Biserial = Include or All | Point-biserial correlation of observation for the current element with its average observation including the observation. |
| Pt-Biserial = Measure | Point-measure correlation of observation for the current element with the measure sum for the observation, and also outputs the expected value of the correlation |
| Pt-Biserial = No or blank | No correlation is reported, but the point-measure correlation is computed for Scorefile= |

The point-biserial correlation is a many-facet version of the Pearson point-biserial correlation, $r_{pbis}$. It takes an extra iteration to calculate, but is useful on new data to check that all element scores work in the same direction along the variable. Negative point-biserial correlations usually signify miskeyed or miscoded data, or negatively worded items. More of a variable (however defined) is always intended to correspond to a higher score.

Let's start with a 2-facet (rectangle, Winsteps) situation. The point-biserial for an item is the correlation of the person scores with their responses to the item. The point-biserial for a person is the correlation of the item scores (p-values) with the responses by the person to those items.

Then to 3 facets (persons, items, raters), we have to extrapolate from the 2-facet procedure.
The point biserial for element x in facet 1 is the correlation of the total score for each element in facet 2 with the response by element x to that element in facet 2, combined with the correlation of the total score for each element in facet 3 with the response by element x to that element in facet. Notice that the element's responses are scanned twice. For the point-measure correlation, they are scanned once: the element's response correlated with sum of the measures for the response.

When a Model= statement specifies measure reversal with "-", i.e., Model=?,-?,R4, then the category value is reversed for the reversed facet, by subtracting the observation from the specified maximum. Thus a value of "3" is treated as "3" for the first facet, "?", but as "4-3"="1" for the second facet, "-?", when computing the point-biserial.

Pt-biserial = Yes: For three facets, i,j,k, the formula for this product-moment correlation coefficient is

```
Ai = (Ti - Wijk*Xijk) / (Ci - Wijk)
```

where Ti is the weighted total score for element i = sum (Wijk*Xijk). Ci is the weighted count of observations for element i = Sum(Wijk). Ai is the average observation for element i omitting element Wijk*Xijk. Ai is correlated with Wijk*Xijk

Wijk = 1 unless a weight is specified in Models=, Labels= or Data=. For instance,

```
Facets = 3
Data=
R5.24 , 1, 2, 3,   7
```

then

```
X123=7 ; the observation is 7
W123=5.24 ; its weight is 5.24
```

Pt-biserial = Include: For three facets, i,j,k, the formula for this product-moment correlation coefficient is
Ai = Ti / Ci
Then the point-biserial correlation for element k is:
PBSk = Correlation ( {Ai, Aj}, Xijk ) for i = 1,Ni and j = 1,Nj

Since the point-biserial is poorly defined for missing data, rating scales (or partial credit items) and multiple facets, please regard this correlation as an indication, not definitive.

See also: Linacre, J.M. (2003). Computing the “Single Rater-Rest of Raters” (SR/ROR) Correlations. Appendix A in C. Myford & E. Wolfe: Detecting and Measuring Rater Effects . *Journal of Applied Measurement, 4,* 421-2.

*Example:* A complete 3-facet dataset. We want the point-biserial and point-measure correlations for element j1.

Data:

| | j1 | | | j2 | | | j3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | i1 | i2 | i3 | i1 | i2 | i3 | i1 | i2 | i3 |
| p1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| p2 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| p3 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 |
| p4 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |

Element totals and counts:

| Element | Ti Total | Ci Count | Measures (+ve) |
|---|---|---|---|
| p1 | 8 | 9 | 3.61 |
| p2 | 4 | 9 | -0.42 |
| p3 | 6 | 9 | 1.43 |
| p4 | 3 | 9 | -1.39 |
| i1 | 11 | 12 | 2.78 |
| i2 | 7 | 12 | -0.09 |
| i3 | 3 | 12 | -2.69 |
| j1 | 7 | 12 | -0.02 |
| j2 | 6 | 12 | -0.77 |

| | | | |
|---|---|---|---|
| j3 | 8 | 12 | 0.79 |

Computation of Ai and the point-biserial correlation for element j1:

| For j1 | | Observation | Total | Count | Pt-biserial= Yes, Exclude | Pt-biserial= All, Include |
|---|---|---|---|---|---|---|
| for persons | | Xnij | Ti | Ci | Ai | Ai |
| p1 | i1 | 1 | 8 | 9 | 0.88 | 0.89 |
| p2 | i1 | 1 | 4 | 9 | 0.38 | 0.44 |
| p3 | i1 | 1 | 6 | 9 | 0.63 | 0.67 |
| p4 | i1 | 1 | 3 | 9 | 0.25 | 0.33 |
| p1 | i2 | 1 | 8 | 9 | 0.88 | 0.89 |
| p2 | i2 | 1 | 4 | 9 | 0.38 | 0.44 |
| p3 | i2 | 0 | 6 | 9 | 0.75 | 0.67 |
| p4 | i2 | 0 | 3 | 9 | 0.38 | 0.33 |
| p1 | i3 | 1 | 8 | 9 | 0.88 | 0.89 |
| p2 | i3 | 0 | 4 | 9 | 0.5 | 0.44 |
| p3 | i3 | 0 | 6 | 9 | 0.75 | 0.67 |
| p4 | i3 | 0 | 3 | 9 | 0.38 | 0.33 |
| for items | | | | | | |
| p1 | i1 | 1 | 11 | 12 | 0.91 | 0.92 |
| p2 | i1 | 1 | 11 | 12 | 0.91 | 0.92 |
| p3 | i1 | 1 | 11 | 12 | 0.91 | 0.92 |
| p4 | i1 | 1 | 11 | 12 | 0.91 | 0.92 |
| p1 | i2 | 1 | 7 | 12 | 0.55 | 0.58 |
| p2 | i2 | 1 | 7 | 12 | 0.55 | 0.58 |
| p3 | i2 | 0 | 7 | 12 | 0.64 | 0.58 |
| p4 | i2 | 0 | 7 | 12 | 0.64 | 0.58 |
| p1 | i3 | 1 | 3 | 12 | 0.18 | 0.25 |
| p2 | i3 | 0 | 3 | 12 | 0.27 | 0.25 |
| p3 | i3 | 0 | 3 | 12 | 0.27 | 0.25 |
| p4 | i3 | 0 | 3 | 12 | 0.27 | 0.25 |
| | | ^ | Facets Table 7: | | PtBis = Yes 0.34 | Ptbis=Inc 0.51 |

Computation of point-measure correlation for element j1:

| For j1 | | Observation Xnij | Sum of Measures | Expected Observation Enij | Model Variance of Xnij around Enij |
|---|---|---|---|---|---|
| p1 | i1 | 1 | 6.37 | 1 | 0 |
| p1 | i2 | 1 | 3.5 | 0.97 | 0.03 |
| p1 | i3 | 1 | 0.91 | 0.71 | 0.2 |
| p2 | i1 | 1 | 2.34 | 0.91 | 0.08 |
| p2 | i2 | 1 | -0.53 | 0.37 | 0.23 |
| p2 | i3 | 0 | -3.12 | 0.04 | 0.04 |
| p3 | i1 | 1 | 4.19 | 0.99 | 0.01 |
| p3 | i2 | 0 | 1.32 | 0.79 | 0.17 |
| p3 | i3 | 0 | -1.27 | 0.22 | 0.17 |
| p4 | i1 | 1 | 1.38 | 0.8 | 0.16 |
| p4 | i2 | 0 | -1.49 | 0.18 | 0.15 |
| p4 | i3 | 0 | -4.09 | 0.02 | 0.02 |
| PtBis=Measure: PtMea = 0.73 | | | | Variance of Enij = 0.14 | Average Model Variance = 0.11 |
| Correlation of Enij with Measures: | | | | | 0.94 |
| Attenuation of correlation due to error = Sqrt( Variance of Enij / (Variance of Enij + Average Model Variance) )= | | | | | 0.75 |
| Expected Point-measure Correlation = 0.94 * 0.75 = PtExp = | | | | | 0.71 |

## 9.39 QM quotation marks around labels =

In the graph, score, and residual output files, labels can be enclosed within quotation marks for tab or comma-separated files. This setting can be changed with Edit Initial Settings.

QM = Yes or QM = Double or QM = Quotes
Use double quotation marks: "label"

QM = Single or QM = Apostrophe
Use single quotation marks, apostrophes: `label`

QM = No or QM =
Don't use quotation marks

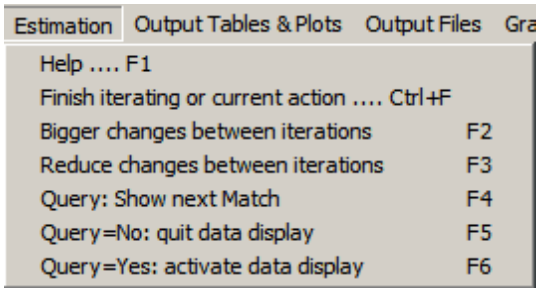
QM = character
Use this character as a quotation mark

Example 1: My database requires character values to be in quotes:
QM = Double
produces: "Watch bugs"

Example 2: My score output file is for display. No quotation marks please:
QM = No
produces: Watch bugs

## 9.40 Query to monitor matching of data to models = No

Specifying Query=Yes in your specification file or from the "Estimation" menu causes Facets to report the decisions it makes in the matching process for each datum.




When **Query=Yes** is in effect, the following is displayed on the screen:
Query=Y in effect! Data matching displayed:

Then, for each data point a message is displayed,
press **Ctrl+Z** to continue.
press **Ctrl+Q** to cancel Query=Y operation.

| Query=Y: Message | Meaning |
|---|---|
| datum: 1,1,0 | The precise Facets-format data element to be matched to a model |
| blank | The entire data line is blank, or commented out with ; |
| replications less than 1 | The datum looks like R0,1,1,0 with a replication of 0. So the datum is not used. |
| datum matched model: _ | Matching was successful. In this case, the datum matched the first model specified. |
| match with "?" failed for element _ in facet _ of model _ | Matching to this model failed. |
| no matching model - datum ignored | Datum matched no model |
| unspecified element: _____ in facet: _ | Datum contained an element not specified in Labels=, and so datum is bypassed. |
| no datum value - datum ignored | The datum value was blank, and so this datum is bypassed, e.g, ,, in a string of data elements. |
| model is Missing | Datum treated as missing, and so ignored. |
| data recoded to ___ | Datum has been given a new value |
| negative datum | negative values are treated as missing |
| non-numeric datum | non-numeric values are treated as missing or invalid |

## 9.41 Rating scale (or Response model) =

The Rating (or partial credit) scale= statement provides a simple way to provide further information about the scoring model beyond that in the Model= specification. You can name each category of a scale, provide Rasch-Andrich threshold-values (step calibrations) for anchoring or starting values, and recode observations.

| **Components of Rating (or partial credit) scale=** | |
|---|---|
| Format: | *Rating scale = user name, structure, scope, numeration* |
| *user name of scale or response model* | any set of alphanumeric characters, e.g., "Likert". To be used, it must match exactly a user name specified in a Model= statement. |
| *structure* | D, R, B, P = any scale code in table below, except a user name |
| *scope* | **S = Specific** (or # in a Model= specification) means that each occurrence of this scale name in a different Model= specification refers to a separate copy of the scale, with its own Rasch-Andrich thresholds (step calibrations), though each has the same number of categories, category names, etc.<br>**G = General** means that every reference to this scale in any Model= specification refers to the same, single manifestation of the scale. |
| *numeration* | O = **Ordinal** means that the category labels are arranged cardinally, representing ascending, but adjacent, qualitative levels of performance regardless of their values.<br>K =**Keep** means that the category labels are cardinal numbers, such that all intermediate numbers represent levels of performance, regardless as to whether they are observed in any particular data set. |

| **Model= and Rating Scale= Scale codes** | *Meaning for this model* |
|---|---|
| **D** | Dichotomous data. Only 0 and 1 are valid. |
| **Dn** | Dichotomize the counts. Data values 0 through n-1 are treated as 0. Data values n and above are treated as 1. E.g., "D5" recodes "0" to "4" as 0 and "5" and above as 1. |
| **R** | The rating scale (or partial credit) categories are in the range 0 to 9. The actual valid category numbers are those found in the data. RK to maintain unobserved intermediate categories in the category ordering. |
| **Rn** | The rating scale (or partial credit) categories are in the range 0 to "n". Data values about "n" are missing data. The actual valid category numbers are those found in the data. If 20 is the largest category number used, then specify "R20". |
| **RnK** | Suffix "K" (Keep) maintains unobserved intermediate categories in the category ordering. If K is omitted, the categories are renumbered consecutively to remove the unobserved intermediate numbers |
| **M** | Treat all observations matching this model as Missing data, i.e, a way to ignore particular data, effectively deleting these data. |
| **Bn** | Binomial (Bernoulli) trials, e.g., "B3" means 3 trials. In the Model= statement, put the number of trials. In the Data= statement, put the number of successes. Use Rating Scale= for anchored discrimination. |

| | |
|---|---|
| **B1** | 1 binomial trial, which is the same as a dichotomy, "D". |
| **B100** | Useful for ratings expressed as percentages %. Use Rating Scale= for anchored discrimination. |
| **P** | Poisson counts, with theoretical range of 0 through infinity, though observed counts must be in the range 0 to 255. Use Rating Scale= for anchored discrimination. |
| **the name of a user-defined scale** | A name such as "Opinion". This name must match a name given in a Rating (or partial credit) scale= specification. |

| **Components of category description lines** | |
|---|---|
| Format: | Rating scale = myscale, R5<br>*category number, category name, measure value, anchor flag, recoded values, reordered values*<br>*category number, category name, measure value, anchor flag, recoded values, reordered values*<br>.....<br>* |
| category number | quantitative count of ordered qualitative steps, e.g., 2.<br>-1 is treated as missing data and is used when recoding data. |
| category name | label for category, e.g., "agree" |
| measure value | These provide starting or pre-set fixed values.<br>For rating scales (or partial credits items): Rasch-Andrich threshold-values (step calibrations).<br>For binomial trials and Poisson counts: scale discrimination, but only when entered for category 0, and with value greater than 0. |
| anchor flag | For rating scales (or partial credit items), ",A" means Anchor this category at its pre-set Rasch-Andrich threshold (step calibration) value. If omitted, or any other letter, the logit value is only a starting value. Anchoring a category with a pre-set Rasch-Andrich threshold forces it to remain in the estimation even when there are no matching responses. Anchor ",A" the lowest category with "0" to force it to remain in the estimation. For binomial trials and Poisson counts: ",A" entered for category 0 means anchor (fix) the scale discrimination at the assigned value. |
| recoded values | Data values to be recoded, separated by "+" signs (optional). Numeric ranges to be recoded are indicated by "-". *Examples:*<br>5+6+Bad recodes "5", "6" or "Bad" in the data file to the category number.<br>"5-8" recodes "5", "6", "7" or "8" to the category number.<br>1+5-8 recodes "1", "5", "6", "7", "8" to the category number. |

*Example 0.* Facets data must be integers, but mine is decimal.

1. multiply all the data by 2: 3.5, 4, 4.5, 5 -> 7,8,9, 10
Use Rating Scale= to recode the data this way

Rating (or partial credit) scale=MyScale,R10 ;
7=3.5,,,3.5
8=4,,,4
9=4.5,,,4.5

```
10=5,,,5
*

2. weight the Models by 0.5
Models = ?,?,?, MyScale, 0.5
```

*Example 1:* Anchor a rating scale (or partial credit) at pre-set Rasch-Andrich thresholds (step calibrations.)

```
Model=?,?,faces,1 ; the "Liking for Science" faces
*
Rating (or partial credit) scale=faces,R3
1=dislike,0,A ; always anchor bottom category at "0"
2=don't know,-0.85,A ; anchor first step at -0.85 Rasch-Andrich threshold
3=like,0.85,A ; anchor second step at +0.85 Rasch-Andrich threshold
* ; as usual, Rasch-Andrich thresholds sum to zero.
```

*Example 2:* Center a rating scale (or partial credit) at the point where categories 3 and 4 are equally probable. Note: usually a scale is centered where the first and last categories are equally probable. More detailed scale rating scale anchoring example.

```
Model=?,?,friendliness,1 ; the scale
*
Rating (or partial credit) scale=friendliness,R4
1=obnoxious
2=irksome
3=passable
4=friendly,0,A ; Forces categories 3 and 4 to be equally probable at a relative logit of 0.
```

*Example 3:* Define a Likert scale of "quality" for persons and items, with item 1 specified to have its own Rasch-Andrich thresholds (scale calibrations). Recoding is required.

```
Model=
?,1,quality,1 ; a scale named "quality" for item 1
?,?,quality,1 ; a scale named "quality" for all other items
*

Rating (or partial credit) scale=quality,R3,Specific ; the scale is called "quality"
0=dreadful
1=bad
2=moderate
3=good,,,5+6+Good ; "5","6","Good" recoded to 3.
 ; ",,," means logit value and anchor status omitted
-1=unwanted,,,4  ; "4" was used for "no opinion", recoded to -1 so ignored
* ; "0","1","2","3" in the data are not recoded, so retain their values.
```

*Example 4:* Define a Likert scale of "intensity" for items 1 to 5, and "frequency" for items 6 to 10. The "frequency" items are each to have their own scale structure.

```
Model=
?,1-5,intensity ; "intensity" scale for items 1-5
?,6-10#,frequency ; "frequency" scale for items 6-10 with "partial credit" format
*

Rating (or partial credit) scale=intensity,R4 ; the scale is called "intensity"
1=none
2=slightly
```

3=generally
4=completely
*
Rating (or partial credit) scale=frequency,R4 ; the scale is called "frequency"
1=never
2=sometimes
3=often
4=always
*

The components of the Rating (or partial credit) scale= specification:
Rating (or partial credit) scale=quality,R3,Specific ; the scale is called "quality"

"quality" (or any other name you choose)
is the name of your scale. It must match the scale named in a Model= statement.

*R3* an Andrich rating scale (or partial credit) with valid categories in the range 0 through 3.

*Specific* each model statement referencing quality generates a scale with the same structure and category names, but different Rasch-Andrich thresholds (step calibrations).

*Example 5:* Items 1 and 2 are rated on the same scale with the Rasch-Andrich thresholds. Items 3 and 4 are rated on scales with the same categories, but different Rasch-Andrich thresholds:

```
Model=
?,1,Samescale
?,2,Samescale
?,3,Namesonly
?,4,Namesonly
*
Rating (or partial credit) scale=Samescale,R5,General
; only one set of Rasch-Andrich threshold is estimated for all model statements
; category 0 is not used ; this is a potentially 6 category (0-5) rating scale (or partial credit)
1,Deficient
2,Satisfactory
3,Good
4,Excellent
5,Prize winning
*
Rating (or partial credit) scale=Namesonly,R3,Specific
; one set of Rasch-Andrich thresholds per model statement
0=Strongly disagree ; this is a 4 category (0-3) rating scale (or partial credit)
1=Disagree
2=Agree
3=Strongly Agree
*
```

*Example 6:* Scale "flavor" has been analyzed, and we use the earlier values as starting values.

Rating (or partial credit) scale=Flavor,R
0=weak ; bottom categories always have 0.
1=medium,-3 ; the Rasch-Andrich threshold from 0 to 1 is -3 logits
2=strong,3 ; the step value from 1 to 2 is 3 logits
* ; The sum of the anchor Rasch-Andrich thresholds is the conventional zero.

*Example 7:* Collapsing a four category scale (0-3) into three categories (0-2):

Rating (or partial credit) scale=Accuracy,R2
0=wrong ; no recoding. "0" remains "0"

```
1=partial,,,2 ; "2" in data recoded to "1" for analysis.
 ; "1" in data remains "1" for analysis, ",,," means no pre-set logit value and no anchoring.
2=correct,,,3 ; "3" in data recoded to "2" for analysis.
 ; "2" in data already made "1" for analysis.
*
data=
1,2,0 ; 0 remains category 0
4,3,1 ; 1 remains category 1
5,4,2 ; 2 recoded to category 1
6,23,3 ; 3 recoded to category 2
13,7,4 ; since 4 is not recoded and is too big for R2, Facets terminates with the message:
    Data is: 13,7,4
    Error 26 in line 53: Invalid datum value: non-numeric or too big for model
    Execution halted
```

*Example 8:* Recoding non-numeric values.

Categories do not have to be valid numbers, but must match the data file exactly, so that, for a data file which contains "R" for right answers, and "W" or "X" for wrong answers, and "M" for missing:

```
Rating (or partial credit) scale=Keyed,D ; a dichotomous scale called "Keyed"
0=wrong,,,W+X ; both "W" and "X" recoded to "0", "+" is a separator
1=right,,,R ; "R" recoded to "1"
-1=missing,,,M ; "M" recoded to "-1" - ignored as missing data
*
data=
1,2,R ; R recoded 1
2,3,W ; W recoded 0
15,23,X ; X recoded 0
7,104,M ; M recoded to -1, treated as missing data
```

*Example 9:* Maintaining the rating scale (or partial credit) structure with unobserved intermediate categories. Unobserved intermediate categories can be kept in the analysis.

```
Model=?,?,Multilevel
Rating (or partial credit) scale=Multilevel,R2,G,K ; means that 0, 1, 2 are valid
    ; if  0 and 2 are observed, 1 is forced to exist.
    Dichotomies can be forced to 3 categories, to match 3 level partial credit items, by scoring the
    dichotomies 0=wrong, 2=right, and modeling them R2,G,K.
```

*Example 10:* An observation is recorded as "percents". These are to be modelled with the same discrimination as in a previous analysis, 0.72.

```
Model=?,?,Percent
Rating (or partial credit) scale=Percent,B100,G ; model % at 0-100 binomial trials
0=0,0.72,A ; Anchor the scale discrimination at 0.72
```

*Example 11:* Forcing structure (step) anchoring with dichotomous items. Dichotomous items have only one step, so usually the Rasch-Andrich threshold is at zero logits relative to the item difficulty. To force a different value:

```
Facets = 2
Model = ?,?,MyDichotomy
Rating scale = MyDichotomy, R2
0 = 0, 0, A ; anchor bottom category at 0 - this is merely a place-holder
1 = 1, 2, A ; anchor the second category at 2 logits
```

2 = 2 ; this forces Facets to run a rating scale model, but it drops from the analysis because the data are 0, 1.
*
If the items are centered, this will move all person abilities by 2 logits. If the persons are centered, the item difficulties move by 2 logits.

*Example 12:* The item-person alignment is to be set at 80% success on dichotomous items, instead of the standard 50% success.
Model = ?,?,?, Dichotomous
Rating scale = Dichotomous, R1 ; define this as a rating scale with categories 0,1 rather than a standard dichotomy (D)
0 = 0, 0, A ; Place-keeper for bottom category
1 = 1, -1.39, A ; Anchor Rasch - Andrich threshold for 0-1 threshold at -1.39 logits
*

```
Table 6 Standard 50% Offset - kct.txt
----------------------------
|Measr|+Children|-Tapping i|
----------------------------
+   1 + **.     +          +
|     |         | 11       |
|     |         |          |
|     |         |          |
*   0 *         *          *
|     | ******  |          |  ---
|     |         |          |   |
|     |         |          |   80% probability of succes
+  -1 +         +          +   |
|     | *.      |          |   |
|     |         | 10       |  ---
|     |         |          |
+  -2 +         +          +
----------------------------
|Measr| * = 2   |-Tapping i|
----------------------------
```

Table 6 80% offset -1.39 logits

```
----------------------------
|Measr|+Children|-Tapping i|
----------------------------
+   1 +         +          +
|     |         | 11       |
|     | **      |          |
|     |         |          |
*   0 *         *          *
|     |         |          |
|     | **.     |          |
|     |         |          |
+  -1 +         +          +
|     |         |          |
|     | ******  | 10       |  <- Item with 80% probability of success targeted
|     |         |          |
+  -2 +         +          +
----------------------------
|Measr| * = 2   |-Tapping i|
----------------------------
```

*Example 13.* Data has the range 0-1000, but Facets only accepts 0-254. Convert the data with the Rating Scale= specification:
models = ?,?,...,spscale
rating scale=spscale,R250,Keep ; keep unobserved intermediate categories in the rating scale structure

```
0,0-1,,,0+1 "0-1" is the category label.
1,2-5,,,2+3+4+5 ; this can be constructed in Excel, and then pasted into the Facets specifications
2,6-9,,,6+7+8+9
3,10-13,,,10+11+12+13
....
248,990-993,,,990+991+992+993
249,994-997,,,994+995+996+997
250,998-1000,,,998+999+1000
*
```

*Example 14:* The rating-scale anchor values are the relative log-odds of adjacent categories. For instance, if the category frequencies are

0 20
1 10
2 20

and all other measures (person abilities, item difficulties, etc.) are 0. Then Facets would show:

```
Rating (or partial credit) scale=RS1,R2,G,O
0=,0,A,,          ; this "0" is a conventional value to indicate that this is the bottom of the rating scale
1=,0.69,A,,       ; this is log(frequency(0)/frequency(1)) = loge(20/10)
2=,-.0.69,A,,     ; this is log(frequency(1)/frequency(2)) = loge(10/20)
```

In rating scale applications, we may want to impose the constraint that the log-odds values increase, if so, we we will only accept rating scale which conceptually have category structure similar to:

0 10
1 20
2 10

This would produce:

```
Rating (or partial credit) scale=RS1,R2,G,O
0=,0,A,,          ; this "0" is a conventional value to indicate that this is the bottom of the rating scale
1=,-0.69,A,,      ; this is log(frequency(0)/frequency(1)) = loge(10/20)
2=,.0.69,A,,      ; this is log(frequency(1)/frequency(2)) = loge(20/10)
```

*Example 15:* The test has a 7-category rating scale, but some items are oriented forward and others are reversed, inverted, negative:

```
Facets =3  ; Facet 3 are the items
Models =
?, ?, 1, Forward
?, ?, 2, Reversed
?, ?, 3-8, Forward
?, ?, 9-12, Reversed
*

Rating scale= Forward, R7, General
1 =
2 =
3 =
4 =
5 =
6 =
7 =
*

Rating scale= Reversed, R7, General
```

```
1 = , , , 7
2 = , , , 6
3 = , , , 5
4 = , , , 4
5 = , , , 3
6 = , , , 2
7 = , , , 1
*
```

*Example 16:* The rating scales do not have the same number of rating categories: fluency 40-100, accuracy 0-70, coherence 0-30, etc.

Let's assume the items are facet 3,

1. Every category number between the lowest and the highest is observable

```
Models=
?, ?, #, R100K   ; K means "keep unobserved intermediate categories"
*
```

2. Every category number between the lowest and the highest category is not observable, but every observable category has been observed in this dataset

```
Models=
?, ?, #, R100   ; unobserved categories will be collapsed out of the rating scales
*
```

3. Only some categories are observable, but not all of those have been observed

```
Models=
?, ?, 1, Fluency
?, ?, 2, Accuracy
?, ?, 3, Coherence
*

Rating scale = Fluency, R100, Keep
0 = 0
1 = 10, , , 10  ; rescore 10 as "1"
2 = 20, , , 20
....
10 = 100, , , 100
*

Rating scale = Accuracy, R70, Keep
0 = 0
1 = 10, , , 10  ; rescore 10 as "1"
2 = 20, , , 20
....
7 = 70, , , 70
*

Rating scale = Coherence, R30, Keep
0 = 0
1 = 5, , , , 5
2 = 10, , , 10  ; rescore 10 as "1"
3 = 15, , , 15
```

4 = 20, , , 20
5 = 25, , , 25
6 = 30, , , 30
*

And there are more possibilities ...

## 9.42 Replication character = R

Replication of observations in the data can be specified with R (or the specified character).

Replication character = R or one of A-Z # $ % &
Upper and lower case characters are the same. Other characters may be available. If you need to use one, please try it. More than one character is also allowed: Replication character = R#
There is no character between this and the number of replications (integer R6 or decimal R3.5).

Example 1: player 1 won 3 times against player 2, and lost 4 times:
Replication character = R ; this is the default value
R3 1 2 1 ; R3 "replicate this observation 3 times"
R4 1 2 0

Example 2: Roger won 3 times against Alfred, and lost 4 times:
Facets = 2 ; 2 facets in the data
Entered = 1, 1 ; both facets in the data are facet 1 in the labels
Models = ?,-?,D ; any element in facet 1 - any facet in element 1 -> dichotomous observation
Labels=
1, players
1 = Roger
2 = Alfred
...
*
Delements = LN ; element identifiers in the data are labels
Replication = # ; the replication character is # and we use R for the first letter of some element labels in the data.
Data=
#3 Roger Alfred 1
#4 Roger Alfred 0

Example 3: For one element, the data looks like 2.7, 5.6, 8.9, ranging from 0 to 10. but other items are integers.
In Facets, every advancing score must mean one higher level of the latent variable. In your data, 0,1,2,3,4,5,6,7,8,9,10 are 11 levels of performance. What does 2.7 mean? From the viewpoint of Facets it means:
an observation of 2 weighted 0.3 (= 0.6) + an observation of 3 weighted 0.7 (= 2.1 so that 0.6+2.1=2.7)
so, if the original Facets observation was:
facet1, facet2, facet3, 2.7
then the Facets data is:
Data=
R0.3, facet1, facet2, facet3, 2
R0.7, facet1, facet2, facet3, 3

## 9.43 Residuals / Responses Output File = ""

If a residual/response output filename is specified by Residual file=filename, a file of responses and residuals from the main analysis is produced. This file is designed for input into other programs with one line per measurable observation. It can be used for calculating other fit statistics and producing specialized diagnostic reports.

Heading lines= control the output of the heading line. CSV= allows tab-delimited and other formats. QM quotation marks, controls whether labels are within quotation marks.

filenames ending .xls and .xlsx are written as Excel workbooks
filenames ending .sav are written as SPSS save files
filenames ending .rda and .rdata are written as R Statistics data files
filenames ending with any other suffix or no suffix are written as DOS text files.

This file can be produced from the Output Files menu by clicking on Residual/Response Output file. This has additional options.

Here is an example of the format with 4 decimal places in the "Select fields" dialog box. The precise format depends on the number of facets in your data:

```
Obs       Stp       Exp       Res       Var     StRes        Wt     LProb   Measure     Displ    Status     MPCat        El
      E2        M1        M2 Children  Tapping_it
  1        1      .97       .03       .02       .16      1.00      -.03      3.66      1.03         2         1         1
         1    -2.98     -6.64 Boy       1-4
  1        1      .97       .03       .02       .16      1.00      -.03      3.66      1.03         2         1         1
         2    -2.98     -6.64 Boy       2-3
```

The columns are:

| Fixed field columns | Abbreviation | Description |
|---|---|---|
| 1-10 | Obs | response as observed in the data file |
| 11-20 | Stp | observed response as renumbered into a count of ordered steps |
| 21-30 | Exp | expected score for this response (decimal places set in selection dialog box) |
| 31-40 | Res | score residual: (observed Stp - expected Exp) |
| 41-50 | Var | model variance of observed score around the expected score for this response, the statistical information in this response |
| 51-60 | StRes | standardized residual: residual / sqrt (variance) |
| 61-70 | Wt | weighting (model weight * observation weight * item weight) |
| 71-80 | LProb | natural logarithm of the probability of the observation |
| 81-90 | Meas | sum of the measures of the elements producing the observation. User-scaled: Meas = sum(element measures - umean) + umean |
| 91-100 | Disp | displacement = measure residual = (score residual / variance)*(user-scaling). The measure of element 1 according to this observation is "element measure" for element + "displacement" * (orientation of facet 1). This |

<table>
<tr><td></td><td></td><td colspan="2">is limited to the range -10 to +10 logits.</td></tr>
<tr><td rowspan="9">101-110</td><td rowspan="9">Status</td><td>Status Code</td><td>Meaning</td></tr>
<tr><td>-6 (not used for estimation)</td><td>Response in two multiple-observation ranges, such as 1-4, 2-6,...</td></tr>
<tr><td>-5 (not used)</td><td>Responses after end-of-file.</td></tr>
<tr><td>-4 (not used)</td><td>Responses only in extreme scores.</td></tr>
<tr><td>-3 (not used)</td><td>Responses with invalid elements. Elements for these observations are not defined. See Table 2.</td></tr>
<tr><td>-2 (not used)</td><td>Responses in two extreme scores</td></tr>
<tr><td>-1 (not used)</td><td>Responses invalid after recounting<br>A dichotomy or rating scale has less than two categories, so it cannot be estimated.</td></tr>
<tr><td>1 (used for estimation)</td><td>Responses used for estimation</td></tr>
<tr><td>2 (used)</td><td>Responses in one extreme score</td></tr>
<tr><td>111-120</td><td>MPCat</td><td colspan="2">most probable category to be observed. If two categories are equally probable, then the higher category is shown here</td></tr>
<tr><td>121-130</td><td>E(facet number)</td><td colspan="2">element number for facet 1 or null element, usually 0</td></tr>
<tr><td>|</td><td></td><td colspan="2"></td></tr>
<tr><td>|</td><td>M(facet number)</td><td colspan="2">element measure for facet 1 from Table 7 (user-scaled)</td></tr>
<tr><td>|</td><td></td><td colspan="2"></td></tr>
<tr><td>|</td><td>(facet label)</td><td colspan="2">element label for facet 1</td></tr>
<tr><td>|</td><td></td><td colspan="2"></td></tr>
</table>

For "Category implies Measure" (C->M) and "Measure implies Category" (M->C) statistics, for each observation in the Facets Residualfile=,
"expected score for this response" - round this to the nearest category number = expected average category
if "expected average category" = "observed response as renumbered into a count of ordered steps" 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

Example: The "Obs" (observed) is the original data. The "Stp" (step) is the ordinal version of the original data. This version is used for analysis, and is the version on which the "Exp" (expected) and the "Res" (residual) are based.

This version may be the same as the original data, or the original data may be transformed either due to explicit instructions by the analyst, or by default operation of Facets.

For instance, suppose that the original data are observations of these three values: 10, 20 and 30. Then, by default, Facets will analyze these observations as the "steps": 10, 11, 12. If the original data are intended to be 10,11,12,13,14,....,28,29,30. Then please specify this is in your Models= statement:
Models=
?,?,..., R30K  ; where "K" means "Keep" the original numeration.

*Example 1:* I need the the S.D. of the observed ratings for each rater.
Facets does not output this statistic. So, output the Residualfile= to Excel. Sort by Rater number. Then compute the S.D. for each rater separately or use the Excel SUBTOTAL function:
1. On the Data tab, in the Outline group, click **Subtotal**. The Subtotal dialog box is displayed.
2. In the At each change in box, click the nested subtotal column. ...
3. In the Use function box, click the summary function that you want to use to calculate the subtotals. ...
4. Clear the Replace current subtotals check box.
5. Click OK

*Example2 :* Where I can find the adjusted ratings after incorporating the differences in the leniency/severity measures on the raw ratings of the corresponding raters?

1. Estimate your data using *Facets.*
Output an Anchorfile= with data.
2. In the anchored file, change the anchored rater severities to 0 logits. Keep everything anchored.
3. Analyze the anchored file with *Facets.* Everything should be anchored, and there should be displacements.
4. Output the Residual File. The "expected score for this response" is the adjusted rating.

Akaike Informaton Criterion: **AIC** = 2k - 2ln(L)
where k = number of free parameters, and ln(L) = log-likelihood of the data.
ln(L) = the sum of the Lprob field in the Residual file. From the Facets "Output Files" menu, output the Residual file to Excel, then sum the LProb column.
The number of free parameters is difficult to estimate. You could use the total number of elements as an approximation.

Bayesian Information Criterion: **BIC** = ln(n)k - 2ln(L)
where n is the number of observations = rows in the residual file

## 9.44 Reversed scoring = Yes

For models with reversed facets, such as Models= ?,-?,R3

| In Specification and Data file: | Meaning |
|---|---|
| ; Symmetric scoring<br>Reversed=Yes<br>Models= ?,-?,R2<br>Data =<br>1,2,2 ; win for 1 against 2<br>3,4,1 ; draw for 3 against 4 | ; Symmetric scoring<br>Reversed scoring in effect<br>Model has a "-" facet<br>In the data file: scoring is 2 win, 1 draw, 0 loss<br>element 1 is scored 2, element 2 is scored 2-2 = 0<br>element 3 is scored 1, element 4 is scored 2-1 = 1 |
| ; Asymmetric scoring<br>Reversed=No<br>Models= ?,-?,R3<br>Data =<br>1,2,3 ; win for 1 against 2 in soccer | ; Asymmetric scoring<br>Reversed scoring not active<br>Model has a "-" facet<br>In the data file: scoring is 3 win, 1 draw, 0 loss<br>element 1 is scored 3, element 2 is not scored |

| | |
|---|---|
| 2,1,0 ; loss for 2 against 1<br>3,4,1 ; draw for 3 against 4 in soccer<br>4,3,0 ; draw for 4 against 3 | element 2 is scored 0, element 1 is not scored<br>element 3 is scored 1, element 4 is not scored<br>element 4 is scored 1, element 3 is not scored |

## 9.45 Score/Measure Output Files = ""

Score and measure output files simplify the transfer of measures and calibrations to other computer software. you may specify a file name that will be used to hold the numeric results presented in Facets measure tables. This is called interactively from the Output Files menu as Score and measure file output.

| Score file name | What happens |
|---|---|
| Scorefile= filename.xls<br>Scorefile= filename.xlsx<br>Excel file | For Excel-formatted output files, the scores and measures for the element of each facet are written to a separate workbook. |
| Scorefile= filename.sav<br>SPSS file | For SPSS-formatted output files, the scores and measures for the elements of all facets are written to one SPSS .sav file, which includes facet number as a variable. |
| Scorefile= filename.rda<br>Scorefile= filename.rdata<br>R Statistics file | For R Statistics-formatted output files, the scores and measures for the element of each facet are written to a separate R data file. |
| Scorefile= filename with any other suffix or no suffix<br>CSV=<br>CSV=Tab<br>CSV=CSV<br>DOS Text file | The measures for each facet are written to a separate file, whose name is that specified by Scorefile=, with the facet number appended. E.g., if Scorefile=Results, then the measures for facet 1 are written to Results.1, for facet 2 to Results.2 etc.<br>The file format is specified by CSV |

All fields, except the Label, are reported with two decimal places, and leading blanks. To select fields, use the Select Fields screen.

Heading lines are included or omitted, depending on Heading lines=.
QM quotation marks, controls whether labels are withing quotation marks.

Example: Write out score files for each facet in fixed field format:
Score file= kct

This puts the measures for facet 1 in file "kct.1.txt", for facet 2 in "kct.2.txt", etc. The results depend on the score file field selection, but can be:

```
                Child
1

                                                                                                                                                F-      F-
T.Sco T.Cou Obs.A FairM Measu       Infit       Outfi       Outfi       PtMeE Discr Displ       Statu Group 1       Child
                          S.E.  Infit       PtMea

                                                                                                                                                Child
7     18    0.29  0.05  -0.96 0.83  0.61  -1.19 0.29  1.99  0.8   0.75  1.71  -0.03 -1    -1    1     Boy   1
10    18    0.5   0.43  1.74  1.11  4.1   2.47  6.15  1.94  0.65  0.87  -0.96 0     -1    -1    2     Boy   1     Child
```

The fixed field format is:

| Field | Columns | Notes |
|---|---|---|
| 0. Element number | 241-250 | Element number is the first field when "Element number after statistics" is *not checked* in the Score File Field Selection dialog box. Otherwise it is before the Element label. |
| 1. Raw score | 1-10 | T.Score includes extreme scores, when Totalscore=Yes<br>Score excludes extreme scores, when Totalscore=No |
| 2. Response count | 11-20 | T.Count includes extreme scores, when Totalscore=Yes<br>Count excludes extreme scores, when Totalscore=No<br>Elements with no observations are omitted unless Omit-unobserved=No |
| 3. Observed average score | 21-30 | (Raw Score)/(Raw Count) |
| 4. Fair average score | 31-40 | FairMAvge: Fair=Mean computes from the mean of the other facets<br>FairZAvge: Fair=Zero computes from the local zero of the other facets |
| 5. Measure | 41-50 | Decimal places set with Umean=x,x,x |
| 6. Standard error | 51-60 | Decimal places set with Umean=x,x,x |
| 7. Infit mean-square | 61-70 | |
| 8. Infit standardized | 71-80 | |
| 9. Outfit mean-square | 81-90 | |
| 10. Outfit standardized | 91-100 | |
| 11. Point-measure correlation | 101-110 | or Point-biserial correlation. See Pt-biserial= |
| 12. Point-measure expectation | 111-120 | .00 if not computed. |
| 13. Discrimination | 121-130 | |
| 14. Displacement | 131-140 | Decimal places set with Umean=x,x,x. This is limited to the range -10 to +10 logits. |
| 15. Status code | 141-150 | See **Status Code** list below |
| 16. Group number | 151-160 | |
| 17. Weight | 161-170 | |

| | | |
|---|---|---|
| 18. Sign | 171-180 | |
| 19. Infit Chi-squared | 181-190 | = infit information-weighted mean-square * d.f. |
| 20. Infit Chi-squared d.f. | 191-200 | |
| 21. Infit Chi-squared 2-sided probability | 201-210 | 2-sided probability of the chi-squared value. <.05 is significant underfit or overfit. |
| 22. Outfit Chi-squared | 211-220 | = outfit conventional mean-square * d.f. |
| 23. Outfit Chi-squared d.f. | 221-230 | |
| 24. Outfit Chi-squared 2-sided probability | 231-240 | 2-sided probability of the chi-squared value. <.05 is significant underfit or overfit. |
| 25. Element number | 241-250 | Element number is here when "Element number after statistics" is checked in the Score File Field Selection dialog box. Otherwise it is the first field. |
| 26. Element label | 251- ? | field length depends on the length of the element labels |
| 27. Facet number | 10 columns | field positions depends on the length of the element labels |
| 28. Facet label | ? | field length depends on the length of the facet label |

| Status code | Meaning for element measure |
|---|---|
| 1 | Anchored measure |
| 0 | No data available to estimate measure |
| -1 | Measure estimated from data |
| -2 | Measure estimated for extreme low score: XTREME= |
| -3 | Measure estimated for extreme high score: XTREME= |
| -4 | Unmeasurable, because all relevant data are generated by more than one extreme element |
| -5 | Only one active data point for this element |

Example: Compute the Spearman Reliability and Separation Index for score file.

1. Compute the sample S.D. of the Measures, then square the S.D. = Measure Variance (MV)
2. Square all the Standard errors and average the squares = Error Variance (EV)
3. Spearman (Kuder-Richardson, Cronbach) Reliability = (MV-EV) / MV
4. Separation Index = square-root ( (MV-EV) / EV )

## 9.46 SE (standard error is Model/Real) = Model

The "Model" standard error is the precision when of the Rasch measures when the data fit the model. The "Real" (enlarged) standard error increases the Model standard error to accord with the misfit in the data.

The Model SE is the smallest possible standard error (highest possible precision), and implies that the data accord with the measurement model, i.e, that misfit is random. If you hypothesize that noise in the data is systematic and not random (the worst case), and so wish to see standard errors (precisions) enlarged to encompass this systematic lack of fit discovered in the current analysis, then report the "Real" SE:

SE = Real ; "Real" can be abbreviated to "R", or specify SE=Enlarged

SE = Real is the more conservative approach. It is recommended for exploratory analysis.

Real enlarged S.E. = Model S.E. * sqrt(Max(INFIT MnSq, 1))

When SE = Real, the reported Reliabilities in Table 7 are calculated using the Real standard errors. The "Real" reliabilities are lower than the default "Model" reliabilities,

The setting of SE= can be changed for *post-hoc* reporting in Modify Specifications.

## 9.47 Show unobserved elements = No

| | |
|---|---|
| | Elements with observations are always reported. |
| Omit unobserved elements = Yes<br>Show unobserved elements = No | Do not report elements that have no observations in the current dataset |
| Omit unobserved elements = No<br>Show unobserved elements = Yes | Table 7 reports anchored elements that have no observations in the current analysis<br><br>Scorefile= : reports all elements specified in Labels= whether they have observations or not. |

The setting of "Omit unobserved elements =" can be changed with Edit Initial Settings and in the Table 7 and Scorefile= dialogs.

Example 1: There are unobserved anchored raters, because the whole rater bank is automatically included in every analysis. We don't want to see them in Table 7

Omit unobserved elements = Yes ; don't display the unobserved anchored elements

Example 2: we want to see which elements have not yet been rated in Scorefile=

Omit unobserved elements = No ; display all the unobserved elements

## 9.48 Simulated data file = ""

A file of simulated data can be constructed from the measures estimated (or anchored) for the main analysis. It will have one simulated observation for each observation in the original data file. Each simulation is unique, so that multiple different simulations can be obtained with the Output Files menu clicking on Simulated Data file.

Simulated data file = filename

The simulated data can be analyzed using Data=filename in the original specification file (or enter at Extra specifications? prompt). Comment out any Dvalues= specifications in the original specification file.

The simulated data file has the basic Facets data format:

```
; Simulated data matching the empirical data structure
; Ratings of Scientists (edited to illustrate ambiguity in measurement)
```

```
; matching: C:\FACETS\examples\subsets.txt
1,2,1,7 ; 9   ; 1,2,1 are facet elements. 7 is simulated. 9 is the original data value.
1,2,2,7 ; 7
1,2,3,4 ; 5
1,2,4,9 ; 8
1,2,5,3 ; 5
```

*Example 1:* We need to compute the S.E. for every element including sampling error. This is "parametric bootstrapping". Also
Simulate and analyze 1000 Facets data sets from lfs.txt in one folder.
Save the following as x.bat in c:\Facets\examples, and then double-click on x.bat.
If you are using Minifac, then change \Facets to \Minifac -
The S.D. of the estimates for each element is its total S.E.

```
SET /A COUNT=1
:LOOP
echo Loop number %COUNT%
rem do this 1000 times
IF %COUNT% == 1001 GOTO END
rem generate simulate data file from lfs.txt
START /WAIT ..\Facets BATCH=YES lfs.txt specfile.out.txt simul=s%COUNT%.txt
rem analyze simulated data using in original specification file
START /WAIT ..\Facets BATCH=YES lfs.txt s%COUNT%.out.txt data=s%COUNT%.txt
SET /A COUNT=COUNT+1
GOTO LOOP
:END
PAUSE
```

*Example 2:* Simulate data corresponding to various types of rater behavior.
1. We conceptualize the rater effects we want to investigate, for instance "halo effect".
2. We formulate statistical models corresponding to each of the rater effects, for instance "halo effect" = all observations by a rater of a person are the same as the first observation.
3. We propose the parameter values which would correspond to each of those rater effects.
4. We use the statistical models of 2. and the parameter values of 3. to generate the data.

If the models in 2. are Rasch models that can be simulated by Facets, then we can use the parameter values in 3. as anchor values in Facets analyses. Then generate data in 4. using the "simulate data file" option in Facets. In order to make the Facets program run, we give it some data, but it does not matter what the data are, because Facets will use the anchor values, not values estimated from the data, to generate the new data.

If the models in 2. are not models that can be simulated by Facets, then we can formulate the data directly that match what we intend, for instance 3 3 3 3 3 3 could be one rater-person data string for "halo effect", and 4 4 4 4 4 4 could be another data string. Or we can use general-purpose simulation software, such as Simfit, simfit.usal.es/english/default.htm

*Example 3:* Discover the estimation bias in a set of *Facets* estimates.
Use the batch file in Example 1, to simulate data matching your dataset. Then compare the standard deviations of the estimates. Here are the numbers from an dataset with 3 facets and 5 simulations:

| Population S.D. | Facet 1 | Facet 2 | Facet3 |
|---|---|---|---|
| Original S.D. | 1.04 | 1.17 | 0.18 |
| Simulation 1 S.D. | 1.09 | 1.21 | 0.19 |
| Simulation 2 S.D. | 1.10 | 1.21 | 0.19 |
| Simulation 3 S.D. | 1.11 | 1.22 | 0.18 |

| | | | |
|---|---|---|---|
| Simulation 4 S.D. | 1.10 | 1.22 | 0.18 |
| Simulation 5 S.D. | 1.11 | 1.21 | 0.18 |
| Average of simulation S.D.s | 1.10 | 1.21 | 0.18 |
| Estimation bias = Average S.D./Original S.D. | 1.06 | 1.04 | 1.02 |

*Example 4:* The simulated data file is to have a different data pattern than the original data file, but be based on the same element measures.

1. Analyze your original data.
2. Output an anchor file (Anchorfile=) with no data but all elements anchored.
3. Construct a dummy data file of all the same response values, such as "1", with the data pattern you want. You can use Excel to do this.
4. Analyze the anchorfile as your specification file and the dummy data file as your data file.
5. Output the simulated data file. This will now have the data pattern that you want, and match the measures of the original dataset.

*Example 5:* Simulate a data file with more persons: One approach:

1. From your current Facets analysis, output an Anchorfile=
2. In the anchor file, add to the person facet in Labels=, the new person elements and their anchor values. These can be generated using Excel and your desired mean and S.D. of the additional person measures.
3. Add to the data new dummy observations, such as "1", with the new persons, the raters, the items. You can use Excel to do this
4. Analyze the anchorfile with Facets, using the new data file. Everything should be anchored.
5. Ignore the output of the analysis.
6. Output file. Simulate a new simulated data file.
7. Unanchor everything in the anchorfile. Save it as your new specification file.
8. Analyze the unanchored specification file with the simulated data file.

*Example 6:* Simulate a data file with more persons: Another approach:

You want to go from 25 candidates to 100 candidates.
Say your original candidate numbers are 1-25
Copy that dataset and change the candidate numbers to 101-125
Copy the dataset again and change the candidate numbers to 201-225
Copy the dataset again and change the candidate numbers to 301-325
Put all four datasets together.
Change your Facets Labels= from 1-25 to 1-325 (missing element numbers do not matter)
Analyze the combined dataset of 4x25 = 100 candidates, and simulate data from it.

*Example 7:* Estimate the person "test" reliability for an incomplete design:

1. Analyze the original data with *Facets*. A generic reliability will be reported for the persons, but without considering the specifics of the design
2. Simulate multiple datasets to match the original design - see Example 1. No anchoring
3. Analyze each of the dataset as though it is the original data.
4. Compute the mean and variance of the estimates for each element.
5. Better reliability estimate = (observed variance of the means - average of the element variances) / observed variance of the means

## 9.49 Subset detection = Yes

Subsets in the data indicate a lack of identifiability of the estimates. An infinity of different sets of estimates would produce the same fit to the Rasch model. Subsets = Yes requests Facets to report this problem of "connectedness".

Raw scores provide a Procrustean solution to the problem of connectedness: a rating of "1" implies the same level of performance everywhere, i.e, all judges are equally severe. Rasch says that the meaning of a "1" depends on its

context. This enables more meaning to be extracted from the data, but also requires more care of the analyst and test designer. In Facets, Procrustean solutions are still available through the use of anchoring.

| | |
|---|---|
| Subset detection = No or Bypass | do not detect subsets. Warning message: "Warning (2)! Subset checking bypassed" |
| Subset detection = List | detect subsets. Show list in Table 6. Subset numbers in Table 7. |
| Subset detection = Report | detect subsets. Show rulers and list in Table 6. Subset numbers in Table 7. |
| Subset detection = Yes (the default) | detect subsets, show details in Table 7, but do not show details in Table 6 unless checked in the Table 6 dialog box |

Facets attempts to discover if the data permit the construction of one unambiguous measurement system. Specify Subset detect=No to bypass detection. Use this to speed up later runs, once data connectivity has been verified.

Subsets are listed in Table 6.0.0 and identified with the elements in Table 7. If subsets exist in the data, use the Output Files menu, Subset group-anchor file option to obtain group-anchor values to resolve the subsetting.

*Example:* Analyzing Subsets.txt

Subsets = No or Subsets = Bypass: "Warning (2)! Subset checking bypassed"

Subsets = Yes

Table 6.0
There are 2 disconnected subsets identified in Table 7.

Table 7.1.1 Senior scientists Measurement Report
in subset: 2

Subsets = List

Table 6.0.0 Disjoint Subset Element Listing.
Subset number: 1
Facet: 1. Senior scientists 2 Elements: 1 2
Facet: 2. Junior Scientists 3 Elements: 1-3

Table 7.1.1 Senior scientists Measurement Report
in subset: 2

Subsets = Report

```
Table 6.0.1 Facet Vertical "Rulers".
+-------------------------------------------------------------------+
|Measr|+Senior scientists    |+Junior Scientists   |+Traits   |Scale|
|-----+----------------------+---------------------+----------+-----|
|  65 +                      + Betty  - subset 1   +          + (9) |
|  55 +                      + David - subset 2    + Attack   +  5  |
|  45 + Avogadro - subset 1  + Anne  - subset 1  2 + Clarity  +  4  |
|  35 + Davey - subset 2     +                     +          + (1) |
|-----+----------------------+---------------------+----------+-----|
|Measr|+Senior scientists    |+Junior Scientists   |+Traits   |Scale|
+-------------------------------------------------------------------+
```

Table 6.0.0 Disjoint Subset Element Listing.

Subset number: 1
Facet: 1. Senior scientists  2 Elements: 1 2
Facet: 2. Junior Scientists  3 Elements: 1-3

Table 7.1.1  Senior scientists Measurement Report
in subset: 2

## 9.50 T3onscreen show only one line on screen iteration report = Y

Table 3, the iteration report can be shown in two levels of detail on screen:

T3onscreen = Yes is the summary report

```
Table 3. Iteration Report
+-----------------------------------------------------------+
| Iteration      Max. Score Residual      Max. Logit Change |
|             Elements    %  Categories   Elements    Steps |
|-----------------------------------------------------------|
......
| JMLE   4      3.4105   2.3    15.8580      .1925   -.0060 |
>===========================================================<
```

T3onscreen = No is the detailed report

```
Iteration 4   Largest Residuals  |----------Score-----------|   logit Measure
Facet     Name          No./Cat   Raw  Expected Residual Resd%  Value    Change
>=========================================================================<
Scale:    ?,?,FACES           2    852    836.1     15.9   1.9    .703     .006
Scale:    ?,?,FACES           1    620    621.8     -1.8   -.3   -.703    -.006
Activity  Animal lives whe   11    119    115.6      3.4   2.3   -.813    -.039
Child     2                   2     50     49.6       .1    .3   5.501     .192
Iteration    4  JMLE                    max.=     3.4   2.3    max.=    .192
```

## 9.51 T4maximum number of residuals in Table 4 = 100

Table 4 can become long. T4maximum= limits the number of residuals to be reported. These residuals are the most conspicuous and must meet the requirements set by Unexpected=. They are reporting in the order specified by Usort=

T4maximum = (number) ; number of unexpected residuals to report.
T4maximum = 0 ; do not report Table 4

Example: limit Table 4 to report the 4 most alarming observations:
T4maximum = 4

```
Table 4.1 Unexpected Responses (4 residuals sorted by 2,3,1).
+-----------------------------------------------------------------+
|Cat   Step   Exp. Resd  StRes| Nu ex N E Nu Re Nu Session        |
|------------------------------+----------------------------------|
|  9     9     5.0   4.0  2.8 | 17 17 3 C 11 11 22 day 2 time 2 |
|  1     1     5.0  -4.0 -2.7 | 17 17 2 B  8 8  32 day 3 time 2 |
|  1     1     4.8  -3.8 -2.6 | 17 17 2 B  9 9  31 day 3 time 1 |
|  1     1     4.7  -3.7 -2.6 | 17 17 2 B  2 2  12 day 1 time 2 |
|------------------------------+----------------------------------|
|Cat   Step   Exp. Resd  StRes| Nu ex N E Nu Re Nu Session        |
+-----------------------------------------------------------------+
```

## 9.52 T7LGS Table 7 List-Group details-Group Summary

Table 7 reports details of the elements. T7LGS= controls what components of the Table are reported. Options are L = List of all elements, G = List element details by group, S = Subtotals of Group elements

T7LGS = LGS (the default) reports the elements, elements by group and group subtotals.

Example 1: for element list only in numerical order
Arrange = N
T7LGS = L

Example 2: for group summary subtotals only:
Arrange = N
T7LGS = S

## 9.53 T8NBC Table 8 Numbers-Barcharts-Curves = NBC

Table 8 reports details of the rating scale structures. T8NBC= controls what components of the Table are reported. Options are N = Numbers, B = Barcharts, C = Curves. Any combination is allowed.

T8NBC = NBC (the default) reports the numbers, the barcharts and curves.

T8NBC= N reports only the numbers. The curves can be displayed in prettier format from the Graphs menu.

## 9.54 Tables output from main analysis = Yes

The Output Tables are usually automatically written into the Report Output file.

| | |
|---|---|
| Tables = No | Output Tables above 5 are not automatically constructed and displayed after the main analysis. Only basic summary and operational Tables are displayed. Other output tables can be requested from the Output Tables menu. |
| Tables = Yes | All relevant output Tables are automatically displayed after the main analysis. |
| Tables = 011010... (up to 14 tables)<br>Not table 1, do table 2, 3, not table 4, do table 5, not table 6, .... | Counting by position in 01 string, 0 or absent = table not displayed, 1 = table displayed if available. |

*Example:* Only show basic summary Tables automatically.
Tables = No
Only Tables 1,2,3,5 are displayed.

## 9.55 Title for each results table =

This allows you to enter a piece of text to give a title to your report, e.g.,
Title = 8th Grade Essays

Specify "^" as the first character of the title when you want each table to start on a new page, e.g.,
Title = ^8th Grade Essays

## 9.56 Total scores reported = Yes

Elements with extreme scores (zero, perfect) do not provide information about the measures of other elements and so are dropped from the estimation. Estimates for extreme scores are done separately with Xtreme score=.

| | |
|---|---|
| Totalscore = No (default) | Report scores after omission of responses that are part of extreme scores |
| Totalscore = Yes | Report scores including all responses. |

Example: For the KCT data.
*Totalscore* = *No* omits observations in other extreme scores. It produces:

```
Table 7.1.1  Children Measurement Report  (arranged by mN).
-------------------------------------------------------------------------------------------------
       --
|  Obsvd  Obsvd  Obsvd  Fair-M|        Model | Infit      Outfit   |Estim.|
       |
|  Score  Count Average Avrage|Measure  S.E. | MnSq ZStd  MnSq ZStd|Discrm| Nu Children
       |
-------------------------------------------------------------------------------------------------
       --
|     11     14       .8    .98|   3.72   .94 | 1.95  1.6   .88  2.4| -.04 | 27 Girl
       |
```

*Totalscore* = *Yes* includes extreme observations in all scores. It produces:

```
Table 7.1.1  Children Measurement Report  (arranged by mN).
-------------------------------------------------------------------------------------------------
       --
|  Total  Total  Obsvd  Fair-M|        Model | Infit      Outfit   |Estim.|
       |
|  Score  Count Average Avrage|Measure  S.E. | MnSq ZStd  MnSq ZStd|Discrm| Nu Children
       |
-------------------------------------------------------------------------------------------------
       --
|     14     18       .8    .98|   3.72   .94 | 1.95  1.6   .88  2.4| -.04 | 27 Girl
       |
```

## 9.57 Udecimals (user-chosen decimal places) = 2

**Improving Communication:**
The standard unit of measurement is the logit (log-odds unit). These measurement units are inconvenient for reporting results to examinees, parents, administrators etc. They can be linearly rescaled to make the reported results more useful to the user. The standard origin (mean) is the mean of the item difficulties, the standard scaling is 1, i.e., 1 reported unit per logit. These values may be changed by

Umean = user origin value, user units per logit, decimal places to report
also, for convenience, these can be specified separately
Uscale = user units per logit
Udecimals = user decimal places

**Anchoring and Umean=**
It is always safest to include the Umean= (mean),(logit scaling) of the analysis that produced the anchor values. Umean=(umean of anchoring analysis) is needed so that Facets knows what is the baseline for Fair average=zero or if an unanchored facet is to be centered or Dummy anchoring (,D) is used for a facet.

Example 1: A useful transformation is to take the high and low measures on the vertical rulers in Table 6.0 or Table 7 and assign that range to 0-100. Then linearly transform all other measures.

Suppose the highest measure is +8.0 logits and the lowest measure is -7.0 logits on Table 6.0 of an analysis.
This gives a range of 15 logits, i.e., 100/15 = 6.67 points per logit.

For reporting purposes, transform all measures by Measure* 6.67 + 46.7 and report with 1 decimal place.
Umean = 7, 6.67, 1
A measure of +2.0 logits becomes 2+*6.67 + 46.7 = 60.0 user-scaled points
Standard errors become (Standard error logits) * 6.67.
A standard error of .03 logits becomes (.03)*6.67 = .2 user-scaled points

Example 2: Set the origin of the measurement scale at 50 units, with 10 units per logit, and report 0 decimal places. This usually gives a measure range somewhat like 0-100.
Umean = 50, 10, 0

Example 3: From an earlier run, the lowest person measure is -6.23 logits, and highest 7.45 logits. Rescale so that the person measure range is 0 - 100.
The new scale is 100 units. The old range was 6.23 + 7.45 = 13.68 logits, so scaling factor = 100 / 13.68 = 7.31 units per logit.
The new origin is offset 6.23 logits = 6.23*7.31 units = 45.54 units
Umean = 45.54, 7.31 ; the standard number of decimal places, 2, is reported.

Example 4: You wish to treat an element as a dummy. It is used for selection, fit analysis and bias detection only.
If Umean = 0 (the standard)
Labels=
4=dummy facet,A
1=dummy element,0 ; anchor this at the umean= value
....
*

If Umean = 50 (user setting)
Labels=
4=dummy facet,A
1=dummy element,50 ; anchor this at the umean= value
...
*

Example 5: -5 to +3 is to be reported as 0 to 25.
Usually the "previous Umean=" is zero.
The computation generally is:
Existing range of measures = -5 to +3 (or whatever is in your Facets report)
Desired range = 0 to 25
Then:
Scaling = (25 - 0) / (3 - -5) = 25/8 = 3.125
Umean= is given by equivalent values: 25 matches 3
New value = old value * scaling + Umean
25 = 3.125 * 3 + Umean
Umean = 15.625
So that the Facets specification is:
Umean = 15.625, 3.125
Checking this for the 0 point:
0 = -5*3.125 + 15.625

Example 6: You want a meaningful baseline for classrooms or Grade-levels across schools. Prof. Everett Smith suggests: determining the classroom means first. Then using the mean and SD of those means, determine UMEAN and USCALE so that the new metric would be 100/10 for the classroom means.

## 9.59 Unexpected (standardized residuals reported at least) = 3

This controls the size of the smallest standardized residual listed as unexpected in Table 4 of the output results file. When the data fit the model, about 5% of standardized residuals are outside ±2, and about 1% are outside ±3.

Residuals concentrated on particular elements show local misfit, which may not be widespread enough to affect the summary fit statistics. At most, T4max= residuals will be reported.

Example 1: To request all residuals with absolute value greater than or equal to 2.5 in Table 4, sorted by in descending order, type
Unexpected=2.5
Usort= u

Example 2: To output residuals for all elements in Table 4, type
Unexpected=0

*Caution:* Requesting all residuals can produce voluminous output! It may be more useful to produce a Residual file, see Residualfile=

## 9.60 Uscale (user rescaling of measures) = 1

**Improving Communication:**
The standard unit of measurement is the logit (log-odds unit). These measurement units are inconvenient for reporting results to examinees, parents, administrators etc. They can be linearly rescaled to make the reported results more useful to the user. The standard origin (mean) is the mean of the item difficulties, the standard scaling is 1, i.e., 1 reported unit per logit. These values may be changed by
Umean = user origin value, user units per logit, decimal places to report
also, for convenience, these can be specified separately
Uscale = user units per logit
Udecimals = user decimal places

**Anchoring and Umean=**
It is always safest to include the Umean= (mean),(logit scaling) of the analysis that produced the anchor values. Umean=(umean of anchoring analysis) is needed so that Facets knows what is the baseline for Fair average=zero or if an unanchored facet is to be centered or Dummy anchoring (,D) is used for a facet.

Example 1: A useful transformation is to take the high and low measures on the vertical rulers in Table 6.0 or Table 7 and assign that range to 0-100. Then linearly transform all other measures.
Suppose the highest measure is +8.0 logits and the lowest measure is -7.0 logits on Table 6.0 of an analysis.
This gives a range of 15 logits, i.e., 100/15 = 6.67 points per logit.
For reporting purposes, transform all measures by Measure* 6.67 + 46.7 and report with 1 decimal place.
Umean = 7, 6.67, 1
A measure of +2.0 logits becomes 2+*6.67 + 46.7 = 60.0 user-scaled points
Standard errors become (Standard error logits) * 6.67.
A standard error of .03 logits becomes (.03)*6.67 = .2 user-scaled points

Example 2: Set the origin of the measurement scale at 50 units, with 10 units per logit, and report 0 decimal places. This usually gives a measure range somewhat like 0-100.
Umean = 50, 10, 0

Example 3: From an earlier run, the lowest person measure is -6.23 logits, and highest 7.45 logits. Rescale so that the person measure range is 0 - 100.
The new scale is 100 units. The old range was 6.23 + 7.45 = 13.68 logits, so scaling factor = 100 / 13.68 = 7.31 units per logit.
The new origin is offset 6.23 logits = 6.23*7.31 units = 45.54 units
Umean = 45.54, 7.31 ; the standard number of decimal places, 2, is reported.

Example 4: You wish to treat an element as a dummy. It is used for selection, fit analysis and bias detection only.
If Umean = 0 (the standard)
Labels=
4=dummy facet,A

1=dummy element,0 ; anchor this at the umean= value
....
*

If Umean = 50 (user setting)
Labels=
4=dummy facet,A
1=dummy element,50 ; anchor this at the umean= value
...
*

Example 5: -5 to +3 is to be reported as 0 to 25.
Usually the "previous Umean=" is zero.
The computation generally is:
Existing range of measures = -5 to +3 (or whatever is in your Facets report)
Desired range = 0 to 25
Then:
Scaling = (25 - 0) / (3 - -5) = 25/8 = 3.125
Umean= is given by equivalent values: 25 matches 3
New value = old value * scaling + Umean
25 = 3.125 * 3 + Umean
Umean = 15.625
So that the Facets specification is:
Umean = 15.625, 3.125
Checking this for the 0 point:
0 = -5*3.125 + 15.625

Example 6: You want a meaningful baseline for classrooms or Grade-levels across schools. Prof. Everett Smith suggests: determining the classroom means first. Then using the mean and SD of those means, determine UMEAN and USCALE so that the new metric would be 100/10 for the classroom means.

## 9.61 Usort sort order of unexpected responses = u

Unexpected observations (with standardized residuals greater than Unexpected=) are listed in Table 4. Usort= controls the order in which the unexpected observations are listed.

| **Usort=** | | |
|---|---|---|
| facet number: 1,2,3,.. | Sort residuals by this facet location, ascending | |
| , | separator between facet numbers and lists | |
| ( ) | sort and list delimiters | |
| | | |
| **Sort order:** | | |
| **Ascending** | **Descending** | **Sorted by:** |
| C | c | Category or Data or Observation, i.e., the original scored response (also D, O) |
| X | x | Step Value or Xni, i.e., the recounted value used for estimation (also V) |

| | | |
|---|---|---|
| E | e | Expected value or Pni, i.e., the modeled value computed from the measures (also P) |
| R | r | Residual or Yni, i.e., the difference between the step value and the modeled value, Yni = Xni - Pni (also Y) |
| Z | z | Standardized residual or Zni, i.e, the statistical unexpectedness of the observation (also S, F, T) |
| U | u | Unexpectedness: Absolute value of standardized residual |

Example 1: In Table 4, it is often useful to sort the unexpected residuals by their facet location and size. Sort the residuals by Facet 1 major, Facet 2 minor and then Facet 3.

Usort=1,2,3

Multiple sorts and multiple listings of the residuals are obtained by placing each sort order in (). Produces two listings, one sorted by Facet 2 major, the other sorted by Facet 3.

Usort=(2,1,3),(3,2,1)

Example 2: Persons are facet 1, items are facet 2 and judges are facet 3. We want to investigate the judges one at a time, paying attention to the most unexpected residuals for each item in Table 4, then identify grossly unexpected person ratings:

Usort = (3,2,u),(u,1)

Unexpected= 3 ; only for very unexpected ratings

## 9.62 UTF8enc 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

## 9.63 UTF8sub substitute character = .

UTF-8 multibyte characters (such as Chinese) are not displayed in the Facets analysis window. Instead a subsitute character is displayed. This can be changed with UTF8sub=

## 9.64 Vertical rulers in Table 6 = 1A,2A,3A,4A,..

Vertical rulers present the measures in graphical form in Table 6. The vertical rulers usually present the facets in columns in numerical order from left to right. Each column lists that facet's element alphabetical labels (or numbers if no labels) vertically by measure. To the right of the facet columns are the rating and other scales, showing the expected scores at each measure.

When you wish to suppress this part of the output, enter

Vertical = No

The standard representation of the facets may be overridden as follows:

Vertical=4N,2A,2N,1*

The facet numbers are written in the order they are to appear. Here, the columns, from left, to right, will be Facet 4, then Facet 2, then Facet 2 again, then Facet 1. Facet 3 and any other facets will not appear.

Yardstick= controls some aspects of the formatting of Table 6.

The presentation of elements of facets is controlled by letters after the facet numbers:

| **Vertical=** | |
|---|---|
| N | show element numbers |
| 4N | show the element numbers of Facet 4 |
| * | show frequency distribution of the elements |
| C | show count of the elements in this position |
| A | (or any character other than N nor *, e.g., L) show element labels (alphabetically) |
| S | show the rating scale category numbers (default) |
| SL | show the rating scale category numbers and labels |
| #S | do not show scoring (rating scale or partial credit) values |

Example 1:

In analysis of the Knox Cube Test, facet 1 is the Children, facet 2 is the Items.

The vertical rulers are to show the distribution of the Children, then the item numbers and their alphabetical names:

Vertical=1*,2N,2A

```
-------------------------------------------------------------------------------
|Measr|+Childre|-Tapping it|-Tapping items                                    |
-------------------------------------------------------------------------------
|   5 +        + 18        + 4-1-3-4-2-1-4                                   +
|     |        | 15 16 17  | 1-3-2-4-1-3    1-4-2-3-1-4    1-4-3-1-2-4    |
|     |        |           |                                                  |
|   4 +        +           +                                                  +
|     | *      |           |                                                  |
|     |        | 14        | 1-4-2-3-4-1                                      |
-------------------------------------------------------------------------------
|Measr| * = 2  |-Tapping it|-Tapping items                                    |
-------------------------------------------------------------------------------
*=2 indicates number of elements per *
```

+Children: Children is a positive facet (Positive=), such that greater raw score means higher measure.
-Tapping: Tapping is a negative facet (Negative=), such that greater raw score means lower measure.

Example 2:
Disjoint subsets (detected by the connectivity algorithm) are always reported with Vertical=(N,#S), because the scale information is not relevant.

```
Table 6.0.1 Disjoint Subset 1 Vertical Summary.

Vertical = (1N,2N,3N,#S) Yardstick (columns, lines) = 0,4
-----------------------------------
|Measr|+supervisors|+nurses|+items|
-----------------------------------
+   2 +            +       +      +
|     |            |       |      |
|     |            |       |      |
|     |            |       |      |
```

```
+    1 +              + 18     +       +
|      |              |        |       |
|      |              |        |       |
|      |              |        |       |
*    0 *              *        * 1  2  *
|      |              |        |       |
|      | 18           |        |       |
|      |              |        |       |
+   -1 +              +        +       +
-------------------------------------
|Measr|+supervisors|+nurses|+items|
-------------------------------------
```

Example 3: Show the "Liking for Science" rating scale with labels:
Vertical = 1C, 2A, SL

```
Vertical = (1C,2A,SL) Yardstick (columns lines low high extreme)= 0,3,-4,5,End
-------------------------------------------------------------------------------
|Measr|+Child|-Activity                                                |FACES|
-------------------------------------------------------------------------------
+   5 +    1 +                                                        + (3) +
|     |    1 |                                                        | like |
|     |      |                                                        |      |
+   4 +    1 +                                                        +      +
|     |    1 |                                                        |      |
|     |    1 |                                                        |      |
+   3 +    1 +                                                        +      +
|     |    1 |                                                        |      |
|     |    4 | Find cans                                              |      |
+   2 +    5 + Watch rat                                              +      +
|     |    4 | Look in cracks      Watch bugs          Watch grass change |      |
|     |    7 |                                                        | ---  |
+   1 +    9 + Watch same animal                                      +      +
|     |    5 | Make map            Name weeds          Talk about plants |      |
|     |   14 | Encyclopedia        Pictures of plants  Read about plants |      |
*   0 *   11 * Animals eat what ?   Watch animals eat                 *  1   *
|     |    2 | Animal stories      Flowers live on ?   Watch birds       |don't|
|     |    2 | Read about animals                                     |      |
+  -1 +    2 + Animal lives where   Watch bird nest                    +      +
|     |    2 | Grow garden         Listen bird sing                   | ---  |
|     |    1 |                                                        |      |
+  -2 +      + Go museum                                              +      +
|     |      | Go zoo                                                 |      |
|     |      |                                                        |      |
+  -3 +      + Go picnic                                              +      +
|     |      |                                                        |      |
|     |      |                                                        | (0)  |
+  -4 +      +                                                       +disli+
-------------------------------------------------------------------------------
|Measr| Count|-Activity                                                |FACES|
-------------------------------------------------------------------------------
```

Example 4: Show all the element labels for the Guilford data:

```
Vertical = (1A,2A,3A,S) Yardstick (columns lines low high extreme)= 0,10,-1,1,End
+-------------------------------------------------------------------------+
|Measr|-Senior scientists|+Junior Scientists|-Traits                |CREAT|
|-----+------------------+------------------+-----------------------+-----|
|   1 +                  +                  +                       + (9) |
|     |                  |                  |                       |  7  |
|     |                  |                  |                       |     |
|     |                  |                  |                       | --- |
|     |                  | Betty            |                       |     |
|     |                  |                  | Enthusiasm            |     |
|     |                  | Edward           |                       |  6  |
|     |                  | George           |                       |     |
|     | Brahe            |                  | Clarity               | --- |
|     |                  |                  |                       |     |
*   0 * Avogadro         *                  *                       *  5  *
|     | Cavendish        | Anne             | Basis                 |     |
|     |                  | Chris            |                       | --- |
|     |                  |                  | Attack      Daring    |     |
|     |                  |                  |                       |  4  |
```

```
|      |                    | David             |                          |      |
|      |                    | Fred              |                          |      |
|      |                    |                   |                          | ---  |
|      |                    |                   |                          |      |
|      |                    |                   |                          |  3   |
|  -1  +                    +                   +                          + (1)  |
|------+--------------------+-------------------+--------------------------+------|
|Measr|-Senior scientists|+Junior Scientists|-Traits                      |CREAT|
+-----------------------------------------------------------------------------+
```

## 9.65 Whexact Wilson-Hilferty standardization = Yes

*Some versions of Facets are restricted to WHEXACT=NO.*

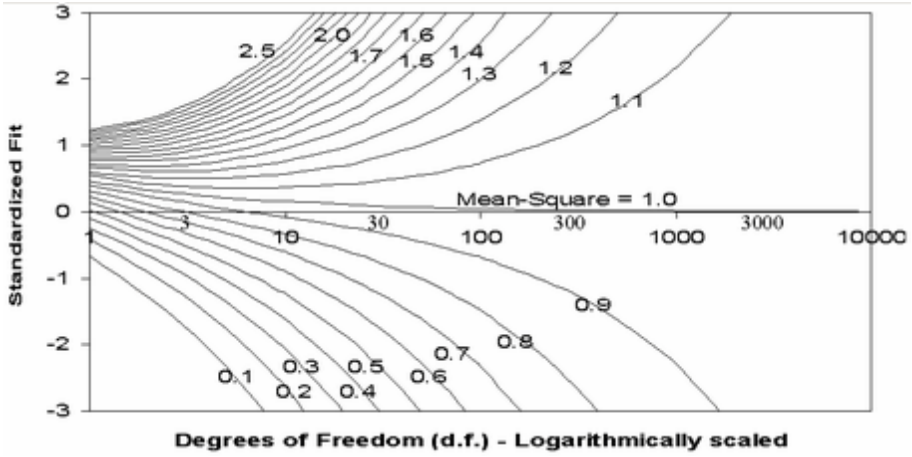
ZSTD INFIT is the "Standardized Weighted Mean Square" shown at the bottom of RSA p. 100.
ZSTD OUTFIT is the "Standardized Unweighted Mean Square" based on the terms on RSA p. 100.
The Wilson-Hilferty transformation converts mean-square values to their equivalent "standardized" normal deviates.
See RSA p. 101 and www.rasch.org/rmt/rmt162g.htm

$$t_i = (v_i^{1/3} - 1)(3/q_i) + q_i/3$$

In Facets, the chi-squared degrees of freedom (2/qi^2) are allowed to be less than 1.0.

Under certain circumstances, Wilson-Hilferty can correctly report the paradoxical finding that the mean-squares indicate overfit, but the normal deviate indicates underfit. To allow this possibility, specify WHEXACT=Y. To suppress it, specify WHEXACT=N. The final q/3 term is omitted from the transformation.

With WHEXACT= Yes, notice that the left side of this plot is non-centered vertically:




Example 1: You have obviously contradictory mean-square and standardized fit statistics. The mean-square is much less than 1, indicating considerable overfit, but the Zstd is much more than 0, indicating significant underfit. Here the Wilson-Hilferty approximation has failed. Specify WHEXACT=No

| WHEXACT=Yes | | | | WHEXACT=No | | | |
|---|---|---|---|---|---|---|---|
| Infit | | Outfit | | Infit | | Outfit | |
| MnSq | ZStd | MnSq | ZStd | MnSq | ZStd | MnSq | ZStd |
| .44 | -.9 | .13 | 2.1 | .44 | -1.1 | .13 | -.2 |
| .97 | .0 | .78 | .4 | .97 | -.1 | .78 | -.1 |
| 1.20 | .7 | .73 | .3 | 1.20 | .6 | .73 | -.1 |
| .43 | -.9 | .18 | .0 | .43 | -1.1 | .18 | -.3 |

Example 2: A person takes a test of 20 dichotomous items and obtains an unweighted chi-squared 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.

## 9.66 Write iteration details in output file = No

Progress toward convergence is always reported on the screen. When you also wish it written to your output file for later examination, specify:
Write = Yes

## 9.67 Xtreme (extreme score adjustment) = 0.3, 0.5

According to the Rasch model, extreme or "perfect" scores (i.e., 0 or minimum possible or maximum possible or perfect scores) imply an infinite estimate for the parameter of the corresponding element. Accordingly they are flagged as "Minimum" or "Maximum" on the output. A reasonable finite estimate is also supplied for the measure corresponding to each extreme score by adjusting extreme scores by a fraction of a score point. The fraction is specified by the first value following Xtreme=. The standard value of 0.3 tends to give a reasonable logit measure corresponding to extreme scores. A similar adjustment procedure can be applied to extreme scores in the bias estimation procedure using the second value following Xtreme=. Here, the standard adjustment is 0.5, because scores within 0.5 score points of their expected values are as unbiased as can be observed. 0.9 is about the highest reasonable value for the extreme score adjustment, giving the most central extreme measures.

If the observations are weighted in Models=, Labels= or Data=, then the extreme-score adjustment is multiplied by the smallest weight applied to an observation.

Example 1: Produce estimates for extreme scores based on increasing all 0 scores by 0.4 score points and decreasing all maximum scores by that same amount of 0.4 score points:
Xtreme = .4

Example 2: You want to produce more extreme measures for extreme scores by assigning a 0.2 score point correction to extreme scores in the main analysis, but a correction of .4 score points in the bias analysis:
Xtreme = .2, .4

Example 3: You want measures for extreme scores to produce extreme Fair Scores:
Xtreme = .01

```
Table 7.1.1  Students Measurement Report  (arranged by mN).
+-------------------------------------------------------------------
|  Total   Total   Obsvd  Fair(M)|        Model | Infit
|  Score   Count  Average Average|Measure  S.E. | MnSq ZStd
|--------------------------------+--------------+----------
|    60      10      6.00   6.00 |( 12.22  9.80)|Maximum
|    58      10      5.80   5.71 |   6.70   .80 |  .74  -.4
|    56      10      5.60   5.47 |   5.70   .65 |  .82  -.6
```

---

**Extreme scores for elements**

"Xtreme= fraction" controls the estimation of measures for extreme scores. *Facets* provides several options for processing extreme scores. These scores are minimum possible (usually 0) or maximum possible scores attained by some element. Extreme scores are problematic in measurement construction, because they imply out-of-bounds, infinite measures.

Facets first drops all responses for which two or more elements have extreme scores. Responses for which only one element has an extreme score are counted only for that element. Responses for which no element has an extreme score are counted for all elements. Facets reports elements with extreme scores as follows:

1) Ignore the fact that the measures are extreme.
This is done when analyzed responses are reported. In Table 7.1 etc, elements with extreme scores are included in the count, score, average score, and point-biserial bar-charts. In Table 8, elements with extreme scores are included in the count, score, average score, and point-biserial summary statistics.

2) Do not report (drop) elements with extreme scores.
In Table 7.1 etc, elements with extreme scores are omitted from the "Logit", "Logit S.E." and Fit bar-charts. In Table 8, elements with extreme scores are omitted from the "Logit", "Logit S.E." and Fit summary statistics.

3) Report elements with extreme scores at the extremes of the measurement continuum.
This is done for Table 7, the vertical rulers. Elements with extreme scores are shown at the extreme top, or bottom, of the column for their facets.

4) Introduce a fractional point score adjustment to the extreme score (of size specified by "Xtreme=fraction"). Then estimate a finite measure for the amended score.
For example, you may feel that even the worst performance would be worth 0.3 score points. This is done with "Xtreme = 0.3" (the standard). Extreme scores will be given measures, but these measures will not alter the measures of other elements, affect the summary statistics, or be used in centering or bias estimation. These extreme measures and their putative standard errors are listed in Table 8.

---

**Extreme scores for interaction terms.**

"Xtreme= fraction" controls bias measures. Measures for interaction bias on extreme scores are problematical. How much is Item 1 biased in favor of the boys, if all boys succeeded on it, but only half the girls?

Facets surmounts this hurdle by introducing a fractional point score adjustment to the off-target extreme score (of size specified by "Xtreme= ,fraction"). This enables the estimation of a finite measure for the amended score. The standard fraction is 0.5 score-points because observed scores are always integers, so that amounts less than this can not be observed.

Measures and standard errors for all bias terms (extreme or not) are included in the summary statistics for Table 13.

## 9.68 Yardstick (Table 6 columns lines low high extreme) =

Adjusts the layout of the vertical rulers in Table 6. This instruction helps make the "vertical rulers" more useful. Several ruler sizes can be specified in the same run. Use the Output Tables menu to experiment with different settings. Vertical= controls which facets are reported in Table 6.

| Yardstick = Horizontal columns , Vertical lines, Low range, High range, {"Measure" or "End"} | |
|---|---|
| Horizontal columns | maximum width of vertical ruler display in columns.<br>Default: 0 means as wide as needed. |
| Vertical lines | number of lines or rows per logit (or per Umean= user-scaled unit).<br>This can be decimal, e.g., 0.5 vertical lines for each logit (or user-scaled unit).<br>Default: 0 makes the rulers about 50 lines long, regardless of logit range. |

| | |
|---|---|
| Low range | lowest measure to show on rulers.<br>Default: 0 |
| High range | highest measure to show on rulers.<br>Default: 0 If low range = high range, then the range is set by the element measures. |
| Placement of Extreme scores | "Measure": position the elements with extreme scores according to their estimated measures<br>"Ends": position the elements with extreme scores at the ends of the rulers<br>Default: "Ends" |

Unspecified values take their standard values or are calculated from the estimated measures.

*Example 1*:Generate a large vertical ruler display, to be used for display:
Yardstick=0,10 ; 10 rows for each unit of measurement (logit)

Multiple copies of the vertical rulers, with different dimensions, can be generated using parentheses:
Yardstick = (Horizontal columns , Vertical lines, Low, High, Extreme),(H,V,L,H,E),(H,V,L,H,E),...
where each set (H,V,L,H,E) cause the generation of another set of rulers with the corresponding Horizontal column and Vertical lines and range specifications.

*Example 2:* Generate two versions of Table 6, a standard and a compact ruler display:
Yardstick = (80,5,-3,4),(40,4)
(80,5,-3,4) 80 columns wide with 5 rows per logit, for the range -3 to +4 logits.
(40,4) 40 columns wide with 4 rows per logit. The range is matched to the measures.
This produces two sets of "rulers". The second is:

```
Vertical = (1A,2A,3A,#S) Yardstick (columns,lines,low,high,extreme)= 40,4,-1,1,Ends
                      ----------------------------------------
                      |Measr|+Senior sci|+Junior |+Traits     |
                      ----------------------------------------
                      +   1 +           +        +            +
                      |     |           |        |            |
                      |     |           | Betty  |            |
                      |     | Cavendish | Edward | Attack     |
                      :     :           : George : Daring     :
                      *   0 * Avogadro  *        * Basis      *
                      :     : Brahe     :        :            :
                      |     | Davey     | Anne   | Clarity    |
                      :     :           : Chris  :            :
                      :     :           : David  :            :
                      |     |           | Fred   | Enthusiasm |
                      |     |           |        |            |
                      +  -1 +           +        +            +
                      ----------------------------------------
                      |Measr|+Senior sci|+Junior |+Traits     |
                      ----------------------------------------
```

## 9.69 Zscore minimum for reported bias terms (bias size/t-statistic) = 0, 0

Bias analysis calculates a bias measure, reported in Table 13 and Table 14, based on the data for which the elements include one from each of the facets specified. Most bias measures are small in logit size and statistically insignificant. This specification is a selection filter for which terms to report. The first control parameter places a

lower limit on the absolute logit (in logits or user-units if user-scaled) of the bias measures to report. Thus, to report only bias measures that are outside of or equal to ±1 logit,

Zscore = 1 ; 1 logit

The second control parameter places a lower limit on the absolute value of the bias t-statistics to report. The t-statistic is obtained by dividing the bias measure by its standard error. With more than 30 observations, a t-statistic is approximately normally distributed, i.e., a z-statistic. To report only bias measures which have t-statistic outside of or equal to ±2:

Zscore = , 2 ; Leading comma, then 2 t-statistic units for bias control

To report bias measures which are either large in logits or significant statistically or both, specify both control parameters:

Zscore = 1,2 ; 1 logit or 2 t-statistic units

To report all bias measures:

Zscore= 0,0 ; include everything

Historical note: *Facets* used to assume that bias terms have approximately infinite degrees of freedom. Consequently they were reported as t-statistics with infinite degrees of freedom. These have the form of a unit-normal distribution, which is the same distribution as z-scores. *Facets* now reports bias significance using t-statistics with more exact finite degrees of freedom. For reporting purposes, the absolute value of the t-statistic is compared with the specified second value in Zscore = bias-size, t-statistic value.

# 10 The Screen Output

## 10.1 Processing specifications

On your computer screen is displayed a report of progress through the analysis. This is the iteration report. This confirms that Facets has understood your specifications.

```
Facets for Windows Version No. 3.64.0 ....
Current folder: C:\FACETS\examples
Editor = notepad.exe
Use Files pull-down menu for Specification File Name, or Ctrl+O
Specification = C:\FACETS\examples\guilford.txt
Processing specifications from "C:\FACETS\examples\guilford.txt"
>.< There will be many ... when there are many specifications
Sorting element labels
>.<
>.<
Report Output = C:\FACETS\examples\guilford.out.txt
Writing specifications to output file: "C:\FACETS\examples\guilford.out.txt"
Table 1. Specifications from file "C:\FACETS\examples\guilford.txt"
Table 1. means the first output Table is being written to the Report file
Title = Ratings of Scientists (Psychometric Methods p.282 Guilford 1954) 10:11:33
Facets = 3
Labels =
 1,Senior scientists (elements = 3)
 2,Junior Scientists (elements = 7)
 3,Traits (elements = 5)          ^
```

^ Check that there are the expected number of elements in each facet.

## 10.2 Reading data

*Facets* reads your data, matches them with your model statements, and reports how much usable data was found.

```
Opening work files
 >.<
Table 2. Data Summary Report
```

```
 Assigning models to Data= "C:\Facets3.83.1\examples\Essays.txt"
 >.First active data line is: 05,1,1,11,4
 05 (5), 1, 1, 11,   4
  Processed as: 05 (5), 1, 1, 11,   4
..................................... 572  observations input so far (big files)
 <
```

*>=====< is the iteration bar. It is drawn across your screen to show progress through the data*

```
 Total lines in data file = 1154
 Total data lines = 1152
 Responses matched to model: ?,?B,?B,?,R9,1 = 1152
```

***Check that the responses matched your models in the way you expected.***

```
     Total non-blank responses found = 1152
 Responses with elements not specified in Labels= are ignored = 0 first at or near line 87
  Responses not matched to any model are ignored = 0
              Number of blank lines = 1
 Number of missing-code observations = 0
 Number of missing-null observations = 0
 Number of lines with R0 or less replications= 0
 Number of invalid observations treated as missing = 0
 Number of negative observations treated as missing = 0
 Number of observations matching invalid models treated as missing = 0
```

***If some observations did not match any of your models. You can use Query=Y to check what happened***

***Point-biserial correlations are calculated if "Pt-biserial=Y".***

```
Calculating point-biserial correlations
>=========================================================================<

 Valid responses used for estimation = 1152
  See Output Table 2 for list of 0 elements not specified in Labels=
```

## 10.3 Calculating estimates

Initial estimates are obtained by the PROX algorithm, and then more refined ones by the JMLE (UCON) algorithm (or by the *Facets* implementation of PMLE). Check that convergence is occurring by reviewing the "Residual" and "Change" columns for smaller values each iteration. The iterative process can be terminated by pressing Ctrl+S simultaneously.

With T3onscreen= No, so that detailed output is displayed:

```
   Table 3. Iteration Report
   Iteration 1   Largest Residuals  |----------Score-----------|   logit Measure
   Facet     Name          No./Cat   Raw  Expected Residual Resd%  Value    Change
   >=========================================================================<
   Junior Sc Betty              2     79                              .740     .740
   Iteration    1  PROX                                             max.=    .740
```

PROX - "normal approximation algorithm" is used to make initial estimates.
If "PROX Recount required" is displayed, some elements or categories had a 0 or perfect score. Measure for these can not be estimated directly, so their estimation is deferred and all marginal scores recounted without them. PROX ceases when the maximum change is less than 0.1 logits

```
   Iteration 2   Largest Residuals  |----------Score-----------|   logit Measure
   Facet     Name          No./Cat   Raw  Expected Residual Resd%  Value    Change
```

```
Checking subset connection..
>=============================================================================<
Scale:    ?B,?B,?,CREATIVI    4     31      6.3     24.7  79.8   -.948    -.948
Scale:    ?B,?B,?,CREATIVI    5      6      6.5      -.5  -8.7    .961     .961
Junior Sc Fred                6     39     18.0     21.0  17.5   -.409     .237
Iteration    2  JMLE                     max.=    21.0  17.5    max.=    .237
```

JMLE= "joint maximum likelihood estimation" makes more precise estimates. The least-converged elements or rating-scale categories are shown (largest residual, largest residual percent, largest logit change). We want max.= to approach 0.

```
Subset connection O.K.
```

Facets confirms that all elements can be measured unambiguously in one frame of reference.

```
 |
 more JMLE iterations
 |
Iteration 15  Largest Residuals  |----------Score-----------|   logit Measure
Facet     Name          No./Cat   Raw  Expected Residual Resd%  Value    Change
>=============================================================================<
Scale:    ?B,?B,?,CREATIVI    4     31     30.8       .2    .6 -1.477    -.001
Scale:    ?B,?B,?,CREATIVI    1      4      4.0       .0   -.7  -.644    -.009
Traits    Enthusiasm          5     58     58.4      -.4   -.2   .498     .004
Junior Sc Betty               2     79     78.7       .3    .2   .638     .005
Junior Sc Fred                6     39     39.3      -.3   -.2  -.562    -.004
Iteration   15  JMLE                     max.=     -.4   -.2    max.=    .005
```

JMLE ceases when the max.= values are small, and the convergence criteria are met. If the maximum residual does not decrease noticeably per iteration, then Estimation menu, "Bigger".. If the maximum residual oscillates between large positive and negative values, then Estimation menu "Reduce". See also "My analysis does not converge."

Column headings have the following meanings:
Facet = Name of the facet, as given in specifications or "Scale:" if details apply to a category of a rating scale (or partial credit).
Name = Name of element within Facet or model statement of a rating scale (or partial credit)
No./Cat = Number of element within facet or category number of a rating scale (or partial credit)

Score
Raw = Raw score (= sum of categories, after renumbering if necessary), or count of responses in category, an integer
Expected = Score or count expected based on current estimates, in tenths
Residual = Difference between observed Score and Expected score, in tenths
Resd% = Residual/(Maximum score - minimum score)%.

Logit Measure
Value = Revised estimate of parameter in log-odds units (logits) as a result of this iteration.
Change = Change in logit estimate from logit estimate of the previous iteration.

max.= is the maximum change or residual corresponding to any element, used to determine convergence.

Iteration = The number of times PROX or JMLE has been applied to the data.

---

**Analysis fails after recounting**

```
Table 3. Iteration Report
```

```
+-----------------------------------------------------------+
| Iteration      Max. Score Residual      Max. Logit Change |
|             Elements    %  Categories   Elements    Steps |
|-----------------------------------------------------------|
Validating subset connection... (Subsets=No to bypass)
>==========================================================<
| PROX   1     Recount required             3.6889          |
Consolidating 2 subsets
>.<
Subset connection O.K.
>==========================================================<
| PROX   2     Recount required             2.3615          |
>==========================================================<
| PROX   3     Recount required             3.6142          |
>==========================================================<
......
>==========================================================<
| PROX  20     Recount required           -20.4497          |
>==========================================================<
| PROX  21     Recount required            -5.2877          |

Error F36: All data eliminated as extreme.
```

This report suggests that Models= does not match the data or that there are not enough observations to estimate the rating-scale structure.
Suggestion: specify a predefined rating-scale structure, such as Binomial Trials. For instance:
Models = ?, ?, B6 ; instead of R6

## 10.4 Fit calculation and writing out results

In a further iteration through the data, the fit statistics are calculated.

```
Calculating fit statistics
>=============================================================================<
>.<
Computing inter-rater agreement: (when Inter-rater = facet-number)
>.<
Table 4. Unexpected Responses - appears after Table 8
Table 5. Measurable Data Summary
Computing fit summary statistics..
Writing score file for facet 1 to C:\FACETS\examples\GUILFSC.1.txt
Score file for facet 1 = C:\FACETS\examples\GUILFSC.1.txt
Writing score file for facet 2 to C:\FACETS\examples\GUILFSC.2.txt
Score file for facet 2 = C:\FACETS\examples\GUILFSC.2.txt
Writing score file for facet 3 to C:\FACETS\examples\GUILFSC.3.txt
Score file for facet 3 = C:\FACETS\examples\GUILFSC.3.txt
Table 6.0  All Facet Vertical "Rulers"
>.<
Table 6.1  Senior scientists Facet Summary
Table 6.2  Junior Scientists Facet Summary
Table 6.3  Traits Facet Summary
Table 7.1.1  Senior scientists Measurement Report  (arranged by mN)
>.<
Table 7.2.1  Junior Scientists Measurement Report  (arranged by mN)
>.<
Table 7.2.2  Junior Scientists Measurement Report  (arranged by 2N)
Table 7.3.1  Traits Measurement Report  (arranged by mN)
>.<
Table 8. Category Statistics
>=============================================================================<
Table 4.1 Unexpected Responses (4 residuals sorted by 1,2,3)
>.<
Table 4.2 Unexpected Responses (4 residuals sorted by 3,1,2)
```

```
>.<
Table 4.3 Unexpected Responses (4 residuals sorted by Z,3)
>.<
```

## 10.5 Performing Bias/Interaction analysis

When bias analysis has been requested in one or more Model= statements, then a secondary analysis is performed for each model with bias interactions ("B") specified.

Bias/Interaction analysis specified by Model: ?B,?B,?,CREATIVITY

*There are as many bias analyses as there are Model= statements with "B"s or that you request from the "Output Tables" menu.*

There are empirically 21 Bias terms

*There is one term for each element combination from the two facets.*

*Table 9.1 Bias Iteration Report.*

*Bias/Interaction: 1. Senior scientists, 2. Junior Scientists*

*There are empirically 14 Bias terms*

| Iteration | | Max. Score Residual Elements | % | Categories | Max. Logit Change Elements | Cells |
|---|---|---|---|---|---|---|
| BIAS | 1 | 9.8439 | 24.6 | -.8196 | | 14 |
| BIAS | 2 | -1.7060 | -4.3 | -.2122 | | 11 |
| BIAS | 3 | -.2214 | -.6 | -.0336 | | 7 |
| BIAS | 4 | -.0051 | .0 | -.0008 | | 0 |

```
Table 9.1  Bias Iteration Report
Bias/Interaction: 1. Senior scientists, 2. Junior Scientists
>=============================================================================<
There are empirically 21 Bias terms
Iteration 1   Largest Residuals  |----------Score-----------|   logit Measure
Model:1. Senior scientists, No.   Raw  Expected Residual Resd%  Value    Change
>=============================================================================<
2=Brahe 5=Edward             14      9     22.0    -13.0 -144.   1.000   -1.000
Iteration    1  BIAS                      max.=   -13.0 -144.     max.= -1.000

 |
 2 more Bias iterations
 |

Iteration 4   Largest Residuals  |----------Score-----------|   logit Measure
Model:1. Senior scientists, No.   Raw  Expected Residual Resd%  Value    Change
>=============================================================================<
2=Brahe 5=Edward             14      9      9.0       .0    .0   1.267     .000
Iteration    4  BIAS                      max.=      .0    .0     max.=   .000
```

Column headings have the following meanings:

Model:= Specification of bias model, reported over the number and label of each element interacting to estimate a bias term.
No.= Sequence number of bias term.

Score
Raw= Observed integer score for this interaction of elements.
Expected= Score expected based on current bias estimate.
Residual= Difference between observed Score and Expected score

Logit Measure
Value= Revised estimate of bias in log-odds units (logits).
Change= Change in logit bias estimate from logit estimate of the previous iteration.

Estimate mean =mean value of the bias estimates for this model, usually near 0.0.
S.D.=sample standard deviation of the estimates, which is not constrained to any value.
Count =number of reportable bias estimates
max.=Is the maximum change or residual for any element.
Iteration=The number of times BIAS has been applied to the data.
BIAS=A Newton-Raphson, residual-score to logit, estimation routine

Fit statistics are computed for the bias terms, and then results are reported:

```
Calculating Bias fit statistics
>=============================================================================<
Table 11.1  Bias/Interaction Measurement Summary
Table 12.1  Bias/Interaction Summary Report
Table 13.1.1  Bias/Interaction Report (arranged by mN)
>.<
Table 13.1.2  Bias/Interaction Report (arranged by 0fN)
>.<
Press Ctrl+F to stop Table 14 bias/interaction reporting.
Tables 14: There are 21 pairwise terms for 1. Senior scientists
Table 14.1.1.1  Bias/Interaction Pairwise Report (arranged by mN)
>=============================================================================<
>.<
Table 14.1.2.1  Bias/Interaction Pairwise Report (arranged by 0fN)
>=============================================================================<
>.<
Tables 14: There are 84 pairwise terms for 2. Junior Scientists
Table 14.1.1.2  Bias/Interaction Pairwise Report (arranged by mN)
>=============================================================================<
>.<
Table 14.1.2.2  Bias/Interaction Pairwise Report (arranged by 0fN)
>=============================================================================<
>.<
```

## 10.6 Completing analysis

At the end of the analysis, Facets reissues any advisory measures, and reports the duration of the analysis and the file holding the results:

```
Subset connection O.K.
Analyzed in time of 0: 0:14
Output to C:\FACETS\examples\guilford.out.txt
```

If there are multiple subsets, then the meaning of the measures is doubtful.

# 11 Tables Output from Main Analysis

## 11.1 Table 1 is a summary of your specifications

These results are written to your output file on disk. Table 1 summarizes your specifications, and enables you to confirm that they were processed as you expected. The specifications are grouped by function. See the specification instructions for information about each specification.

```
Facets for Windows Version No. 3.62.0  Copyright © 1987-2009, John M. Linacre. All rights
       reserved.

Ratings of Scientists (Psychometric Methods p.282 Guilford 1954)
Table 1. Specifications from file "C:\FACETS\examples\guilford.txt".

Title = Ratings of Scientists (Psychometric Methods p.282 Guilford 1954)
Data file = creativity.xls
Output file = C:\FACETS\examples\guilford.out.txt

; Data specification
Facets = 3
Non-centered =  1
Positive =  2
Labels =
 1,Senior scientists (elements = 3)
 2,Junior Scientists (elements = 7)
 3,Traits (elements = 5)
Model =?B,?B,?,CREATIVITY,1
Rating (or other) scale = CREATIVITY,R9,General,Ordinal
Rating (or other) scale = EXTRASCALE,R9,General,Ordinal
Warning (3)! This Rating (or partial credit) scale= is NOT referenced by a Model= specification

; Output description
Arrange tables in order = mN,2N,0fN
Bias/Interaction direction =difficulty ;severity: lower score = positive logit
Fair score = Mean
Heading lines in output data files = Y
Inter-rater coefficients reported for facet = 1
Scorefile = GUILFSC
Show unobserved elements = no
Barchart = Yes
T4MAX maximum number of unexpected observations reported in Table 4 = 100
Unexpected observations reported if standardized residual >= 2
Usort unexpected observations sort order = (1,2,3),(3,1,2),(Z,3)
Vertical ruler definitions = 2N,3A,2*,1L,1A
WHexact - Wilson-Hilferty standardization = Y
Zscore (Bias size/significance) not less than = 1, 2 ; bias size, t-score

; Convergence control
Convergence = .5, .01
Iterations (maximum) = 0 ; unlimited
Xtreme scores adjusted by = .3, .5  ;(estimation, bias)
```

## 11.2 Table 2 is the data summary report

Model=, Rating (or partial credit) scale= and Labels= control data selection for each analysis. This confirms that your data has been understood in the way you expected. If the counts in this table are unexpected, examine your data, Labels= and Model= specifications. If an element number is missing from the Labels= specifications, then all its data values may have been ignored.

```
Table 2. Data Summary Report.
```

```
Assigning models to "C:\Facets\Examples\Essays.txt"
Check (1)? Element not specified in Labels=: facet: 4 element: 16 in line 1198 - see Table 2
Check (2)? Invalid datum location: 1,1,b at or near line 52. Datum "b" is too big or not a
       positive integer, treated as missing.
Total lines in data file = 1154
Total data lines = 1154
Responses matched to model: ?,?B,?B,?,R9,1 = 1152
    Total non-blank responses found = 1153
Number of blank lines = 1 first at (line number)
Responses with elements not specified in Labels= = 1 first at (line number)
Number of missing-code observations = 1 (,x,y,) first at (line number)
Valid responses used for estimation = 1152

List of elements notspecified in Labels=. Please copy and paste into your specification file,
       where needed
Labels= ; Nobuild ; to suppress this list
 1, examinee,  ; facet 1
 35 =
 *
 4, Session, A ; facet 4
 16 =
 *
```

Total lines in data file =
is the number of lines in your data file, or following the Data= specification in your specification file. It includes blank lines, comments (;), and unmodeled data as well as modeled data.

Total non-blank responses found =
This is a count of all non-blank responses in your data file.

Responses with elements not specified in Labels= =
Elements referencing elements not specified after Labels= are dropped from the analysis, unless for facets marked "X" in Model= statements. To obtain information about the unspecified elements, specify Query=Y.

> Example: Check (1)? Unspecified element: facet: 1 element: 13 in line 2594
> In the data, line 2594, facet 1 has element number 13.
> In the Labels= list, under facet 1, there is no element 13. So the data is treated as unspecified.
> If you want to analyze it, add 13 to the element list for facet 1. Also any other similar elements.
> Facets only reports the first unspecified element found, so Facets may report another element as unspecified after you account for element 13.

Valid responses used for estimation =
This is a count of those responses in your data file which matched your model statements and were used in the analysis. It is 0 if no useable data was found.

Responses matched to model: (model description) =
This is a count of how many responses matched the particular model. This is useful for validation of model specifications. Invalid data values and models specifying data to be treated as missing are included here, but not in the "Valid responses found" count.

Example 1: I removed one element number in Guilford.txt:
*Labels=*
*1,Senior scientists*
*1=Avogadro*
*; element 2 is missing*
*3=Cavendish*
*
Then, in Table 2 in the Report Output file:
*Table 2. Data Summary Report.*

.....
*1, Senior scientists, ; facet 1*
*2 =*
*
So, I can copy-and-paste 2 = into Guilford.txt. Then that element number will not be missing.

Example 2: A facet does not apply to these data:
*Facets = 3*
*Keepasnull = 0 ; this element number is ignored*
*Data = 1,2,0, 4 ; facet 3 element 0 is because facet 3 does not apply to these data*

Table 2:
*Labels=Nobuild ; to suppress this list*
*3, Facet 3*
*0 = ; this element number is ignored, because Keepasnull=0 - this is only a warning*
*

## 11.3 Table 3 is the iteration report for the main analysis

Convergence= and Iterations= control the number of iterations performed. Write=Yes writes the details reported on screen into Table 2. The number of iterations required depends on how difficult it is to obtain good estimates from the data. Many iterations may be required if

1) there is a poor fit of the data to the Rasch model.
2) the element parameter distribution is badly skewed or multi-modal.
3) there are rarely observed response categories.
4) exceedingly precise Convergence= criteria have been specified.
5) the data matrix is composed of disjoint subsets of observations, e.g., boys rated by Judge A, but girls by Judge B. When this is detected, a warning message is displayed.

Table 3. Iteration Report.

| Iteration | | Max. Score Residual | | | Max. Logit Change | |
|---|---|---|---|---|---|---|
| | | Elements | % | Categories | Elements | Steps |
| PROX | 1 | | | | .7405 | |
| JMLE | 2 | 26.6978 | 22.2 | 29.9588 | .3374 | .9902 |
| JMLE | 3 | 22.4284 | 13.4 | 27.3399 | -.1155 | 1.0049 |
| JMLE | 4 | 11.0189 | 6.6 | 22.0935 | -.0514 | .9957 |
| JMLE | 5 | -3.5380 | -2.9 | 9.5224 | -.0304 | -.7481 |
| JMLE | 6 | -5.7141 | -4.8 | 2.6727 | -.0620 | -.2008 |
| JMLE | 7 | -4.5692 | -3.4 | 1.8210 | .0501 | -.0892 |
| JMLE | 8 | -3.5327 | -2.5 | 1.4746 | .0393 | -.0599 |
| JMLE | 9 | -2.7709 | -1.9 | 1.2091 | .0314 | -.0529 |
| JMLE | 10 | -2.1980 | -1.4 | .9963 | .0255 | -.0446 |
| JMLE | 11 | -1.7600 | -1.1 | .8245 | .0209 | -.0371 |
| JMLE | 12 | -1.4209 | -.9 | .6847 | .0172 | -.0309 |
| JMLE | 13 | -1.1553 | -.7 | .5703 | .0143 | -.0258 |
| JMLE | 14 | -.9451 | -.6 | .4761 | .0119 | -.0215 |
| JMLE | 15 | -.7771 | -.5 | .3982 | .0099 | -.0180 |
| JMLE | 16 | -.6418 | -.4 | .3336 | .0083 | -.0151 |
| JMLE | 17 | -.5319 | -.3 | .2798 | .0069 | -.0127 |
| JMLE | 18 | -.4422 | -.3 | .2349 | .0058 | -.0106 |
| JMLE | 19 | -.3685 | -.2 | .1974 | .0049 | -.0089 |

Subset connection O.K.

Iteration counts the number of times the data has been read.

PROX is the "normal approximation algorithm" to obtain approximate estimates speedily. Steps are not estimated during PROX.
JMLE is joint (unconditional) maximum likelihood estimation to obtain precise estimates.
Facets generally produces its results with high precision. This precision is rarely needed in practice before the final runs. There are several ways to lower the precision of the results. Most immediately, Ctrl+F forces Facets to move into the reporting phase at the end of the current iteration through the data. Other specifications include Iterations= and Convergence=. Inspection of the iterations, Table 3 of the output, indicates when the changes per iteration are too small to have any important meaning at the current stage of your analysis. Here this happens after just 4 iterations.

Max. Score Residual
Elements: the largest difference (residual), in score points, between the observed and expected score corresponding to any element's parameter estimate. 1.0 is the smallest observable (i.e., in the data) difference with the standard model weighting of 1.
%: the largest residual as a percent of the (maximum possible score - minimum possible score) for any element.
Categories: the largest difference between the observed and expected counts of occurrence corresponding to any category of a rating scale (or partial credit). 1.0 is the smallest observable difference with the standard model weighting of 1.

Recount required
when this appears, it means that the scores corresponding to some element parameters had extreme values (either 0 or the maximum possible). These parameters are dropped from estimation, forcing a recount of the marginal scores of the other elements.

Max. Logit Change
Elements: the largest change, in logits, between any element parameter estimate this iteration and its estimate the previous iteration. Starting estimates are either 0.0 logits, or the values given in the specification file.
Categories: the largest change, in logits, between any step parameter estimate this iteration and its estimate the previous iteration. Starting estimates are either 0.0 logits, or the values given in the specification file.

After the first few iterations, both "Max. Score Residual" and "Max. Logit Change" should steadily reduce in absolute size, i.e., draw closer to zero. There may be occasional perturbations due to unusual data. If the iterative procedure seems to have reached a plateau, you may force termination by pressing the Ctrl+"S" keys simultaneously.

The more detailed iteration report, which appears on your screen, can be recorded in your output file with a "Write=Yes" specification

Subset connection O.K.
Facets has verified that all measures can be estimated in one, unambiguous frame of reference. Warning messages here require investigation.

## 11.4 Table 4 reports Unexpected Responses

Juxtapose=, Left-hand= control the layout of Table 4. These are responses with standardized residuals equal to or exceeding the amount specified by Unexpected=. They are listed in the order specified by Usort=. Up to T4maximum= are reported.

Table 4 shows outlying (i.e., very unexpected) observations. Look for patterns. Is the same examinee, rater, item, .... appearing many times? Perhaps this is an indication of unusual performance or data entry error. If it is a rater, perhaps a misunderstanding of the rating criteria. The outlying observation may be irrelevant for my purposes and so can be bypassed as missing data using Models= ...,M. This happens, for instance, when our interest is in determining task difficulty, but some examinees misunderstand the task and so are rated low. Their low ratings are going to make the task appear more difficult, but the task will immediately return to its correct difficulty when the task instructions are revised to remove the misunderstanding.

Look carefully at these residuals, inspecting them for signs of:
1) Incorrectly formatted data. Were the original data correctly reformatted into Facets data?
2) Incorrect application of a scoring key. Data that was originally in multiple-choice question (MCQ) or some other coded form may not have been correctly scored.
3) Reversed rating scales (or partial credit items). Survey items with the word "not" in them or with a negative implication may need to have their scoring reversed in order to align with the hypothesis that "more of the variable implies a higher observed response". You can also use model statements with "-?" terms in order to reverse the direction of the variable for particular combinations of elements.
4) Idiosyncratic or "off-variable" behavior. Wild guessing, response sets, frequent selection of "don't know" are symptoms that the participants are not exhibiting the type of behavior that the instrument is intended to measure. For the construction of meaning, and from the measurement point of view, such data are better recoded "missing" (by means of "M" models, or ";" in their data lines).
5) Patterns of unexpected residuals loading onto certain elements. Facets reports the misfitting elements by their order in the data file. You may find it useful to transfer this Table into your word processor and sort it according to standardized residual size, facet etc. Systematic patterns of misfit may prompt a bias analysis or the reformulations of your Model= specifications.

There will always be some degree of misfit to the measurement model. You must decide when the fit of the data to the model is good enough for your purposes.

```
Table 4.1 Unexpected Responses (sorted by 1,2,3).
+-------------------------------------------------------------------+
| Cat  Score  Exp.  Resd StRes| Nu Senior sc Nu Junior Nu Traits     | Sequence
|-----------------------------+-------------------------------------|
|  2     2     5.2  -3.2 -2.0 |  2 Brahe      5 Edward  1 Attack     | 57
|  2     2     5.2  -3.2 -2.0 |  2 Brahe      5 Edward  4 Daring     | 60
|  6     6     3.3   2.7  2.2 |  2 Brahe      6 Fred    3 Clarity    | 64
|  9     9     5.3   3.7  2.3 |  2 Brahe      1 Anne    1 Attack     | 37
|-----------------------------+-------------------------------------|
| Cat  Score  Exp.  Resd StRes| Nu Senior sc Nu Junior Nu Traits     | Sequence
+-------------------------------------------------------------------+
```

Column headings have the following meanings:
Cat = Observed value of the category as entered in the data file.
Step = Value of category after it has been recounted cardinally commencing with "0" corresponding to the lowest observed category.
Exp. = Expected score based on current estimates
Resd = Residual, the score difference between Step and Exp.
StRes = The residual standardized by its standard error and truncated. A value of 1.6 is reported as 1. See Residuals file for a more exact value. StRes is expected to approximate a unit normal distribution.
Sequence = Sequence number of observation in the data file

For each modeled facet:
N is the number of the element in the specifications
Name of Facet is over the name of the element

## 11.5 Table 5 is the measurable data summary

Table 5 reports summary statistics about the data for the analysis.

```
+-------------------------------------------------+
| Cat  Score  Exp.  Resd StRes|                   |
|-----------------------------+-------------------|
| 4.80  4.80  4.80   .00  .00 | Mean (Count: 1152) |
| 1.63  1.63  1.03  1.27  .99 | S.D. (Population)  |
| 1.64  1.64  1.03  1.27  .99 | S.D. (Sample)      |
+-------------------------------------------------+
```

Column headings have the following meanings:
Cat = Observed value of the category as entered in the data file.
Score = Value of category after it has been recounted cardinally commencing with "0" corresponding to the lowest observed category.
Exp. = Expected score based on current estimates
Resd = Residual, the score difference between Step and Exp.
StRes = The residual standardized by its standard error. StRes is expected to approximate a unit normal distribution.

Mean = average of the observations
Count = number of observations
S.D. (Population) = standard deviation treating this sample as the entire population
S.D. (Sample) = standard deviation treating this sample as a sample from the population. It is larger than S.D. (Population).

The raw-score error variance % is 100*(Resd S.D./Cat S.D.)²

When the parameters are successfully estimated, the mean Resd is 0.0. If not, then there are estimation problems - usually due to too few iterations, or anchoring.
When the data fit the Rasch model, the mean of the "StRes" (Standardized Residuals) is expected to be near 0.0, and the "S.D." (sample standard deviation) is expected to be near 1.0. These depend on the distribution of the residuals.

Explained variance by each facet can be approximated by using the element S.D.^2 (^2 means "squared").

From Table 5:
Explained variance = Score Population S.D.^2 - Resd^2
Explained variance % = Explained variance * 100 / Score Population S.D.^2

From Table 7:

```
|------------------------------+--------------+----------------------+------+-------------+
+----------------------|
|   460.8    96.0     4.8   4.73|    .00   .08 | 1.00  -.1   .99  -.2|      |   .61       | |
Mean (Cnt: 12)       |
|    29.5     .0      .3    .32|    .19   .00 |  .23  1.8   .22  1.7|      |   .05       | |
S.D. (Population)    |
|    30.8     .0      .3    .33|    .20   .00 |  .24  1.9   .23  1.8|      |   .06       | |
S.D. (Sample)        |
+------------------------------------------------------------------------------------------
----------------------+
```

v1 = (measure Population S.D. facet 1)^2
v2 = (measure Population S.D. facet 2)^2
v3 = (measure Population S.D. facet 3)^2
vsum = v1 + v2 + v3 + .... (for all facets)

Compute Explained variance for each facet:
Explained variance % by facet 1 = (Explained variance %) * v1 /vsum
Explained variance % by facet 2 = (Explained variance %) * v2 /vsum
Explained variance % by facet 3 = (Explained variance %) * v3 /vsum

*Example:* Guilford.txt - a 3 facet analysis:
In Table 5:

```
Raw-score variance of observations   =  3.526 100.00%
Variance explained by Rasch measures =  1.446  41.02%
Variance of residuals                =  2.080  58.98%
```

In Table 7:

```
measure Population S.D. facet 1 = 0.13, variance = 0.02, % total variance = (0.02/0.29) * 41.02
= 2.83%
measure Population S.D. facet 2 = 0.42, variance = 0.18, % total variance = (0.18/0.29) * 41.02
= 25.47%
measure Population S.D. facet 3 = 0.30, variance = 0.09, % total variance = (0.09/0.29) * 41.02
= 12.73%
                                    variance sum = 0.29
```

*Confirmation of variance explained:* make all facets Positive=1,2,3 so that everything is addition. Then we want the S.D. of the sums of the 3 element measures. Easy! Output the Residualfile= to Excel and obtain the S.D. of the *Meas* column. We can then use www.rasch.org/rmt/rmt221j.htm with this S.D. as, say, the "person" S.D., and the item S.D. as zero.

For a Rasch-based Generalizability Coefficient:

G = (Explained variance% by target facet) / 100

A more specific Generalizability Coefficient can be formulated by selecting appropriate variance terms from Table 5, Table 7, and Table 13.

```
Estimable observations = 1152, Free parameters = 21, Degrees of freedom = 1129
Global Pearson chi-squared = 1137.55, probability = .4233

                                        Count   Mean   S.D.
Responses after end-of-file         =       0   0.00   0.00
Responses only in extreme scores    =       0   0.00   0.00
Responses in two extreme scores     =       0   0.00   0.00
Responses with invalid elements     =       0   0.00   0.00
Responses invalid after recounting  =       0   0.00   0.00
Responses non-extreme estimable     =    1152   4.80   1.63
Responses in one extreme score      =       0   0.00   0.00
All Responses                       =    1152   4.80   1.63
```

| Identification | Meaning |
|---|---|
| Estimable observations | (Weighted) count of all observations in non-extreme (minimum possible or maximum possible) response strings |
| Free parameters | Count of the minimum number of elements and thresholds which, when estimated, decide the estimates of all other elements.and thresholds |
| Degrees of freedom | The d.f. are the estimable observations less the free parameters |
| Global Pearson chi-squared | This summarizes the fit of all the data to the Rasch model. It is the sum of the squared standardized residuals (weighted if weights are specified). It assumes that the element measures are the maximum-likelihood estimates.<br>To produce this value from the Residual File:<br>Output Files menu<br>Residual/Response File<br>Select fields to output<br>Decimal places 4<br>OK<br>Output to Excel<br>Insert blank column<br>Divide residual^2 / variance into the blank column |

| | |
|---|---|
| | Sum the blank column (for more accuracy, only for Status 1 observations)<br>*Note:* there will be small rounding errors. |
| Probability | The probability of observing the chi-squared value (or larger) when the data fit the model |
| | |
| **Response Type** | **Responses *not* used for estimation: see Residual File** |
| Responses after end-of-file | A Facets internal work-file has too many responses. Please report this to Winsteps.com and rerun this analysis. |
| Responses only in extreme scores | The category of the rating scale cannot be estimated. |
| Responses in two extreme scores | These cannot be estimated nor used for estimating element measures. |
| Responses with invalid elements | Elements for these observations are not defined. See Table 2 with *Build* option. |
| Responses invalid after recounting | A dichotomy or rating scale has less than two categories, so it cannot be estimated. See Table 8 for missing or one-category rating scales. |
| **Response Type** | **Responses *used* for estimation: see Residual File** |
| Responses non-extreme estimable | This is the count of responses used in estimating non-extreme parameter values (element measures and rating scale structures). |
| Responses in one extreme score | These are only used for estimating the element with the extreme score |
| All Responses | Shown if there is more than one response type listed above |

```
Count of measurable responses           =      1152
Raw-score variance of observations      =   2.67 100.00%
Variance explained by Rasch measures    =   1.06  39.57%
Variance of residuals                   =   1.61  60.43%
Variance explained by bias/interactions =   0.14   5.24%
Variance remaining in residuals         =   1.47  55.06%
```

| An approximate Analysis of Variance (ANOVA) of the data | |
|---|---|
| **Identification** | **Meaning** |
| Count of measurable responses | All responses (including for extreme scores) with weighting (if any) |
| Raw-score variance of observations | Square of S.D. (Population) of Score |
| Variance explained by Rasch measures | Raw score variance - Variance of residuals. This is dominated by the spread of the elements. We usually want the raters to be equally lenient = explain no variance, etc.<br>The size of the expected variance for 2-facet models is shown in www.rasch.org/rmt/rmt221j.htm |
| Variance of residuals | Square of S.D. (Population) of Resd. |

| | |
|---|---|
| Variance explained by bias/interactions | The variance explained by the bias/interactions specified with "B" in your Models= statements |
| Variance remaining in residuals | Variance of residuals - Variance of interactions |

**Nested models:** Suppose we want to estimate the effect on fit of a facet.
Run twice:
First analysis: 3 facets
Models = ?,?,?, R
Second analysis: 2 facets:
Models = ?,?,X,R

We can obtain an estimate of the improvement of fit based on including the third facet:
Chi-squared of improvement = chi-squared (2 facets) - chi-squared (3 facets) with d.f. (count of elements in facet 3 - 1).

If global fit statistics are the decisive evidence for choice of analytical model, then *Facets* is not suitable. In the statistical philosophy underlying *Facets,* the decisive evidence for choice of model is "which set of measures is more useful" (a practical decision), not "which set of measures fit the model better" (a statistical decision). The global fit statistics obtained by analyzing your data with log-linear models (e.g., in SPSS) will be more exact than those produced by *Facets.*

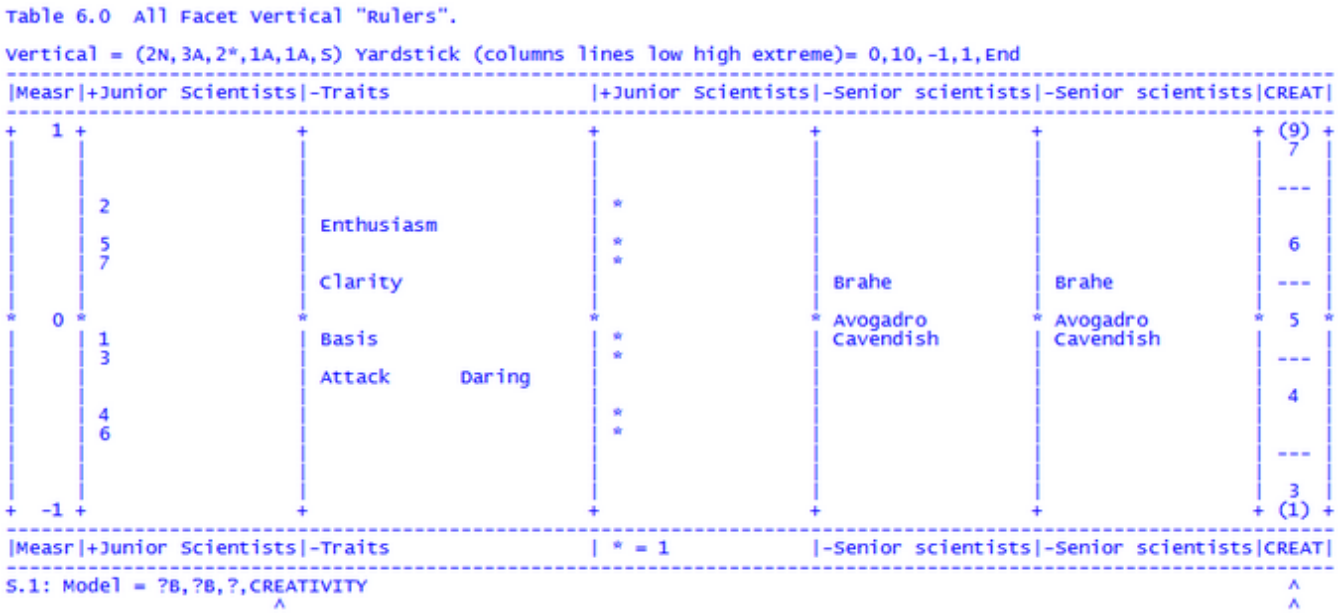
## 11.6 Table 6.0 is the All facet summary - Wright Map "rulers"

Table 6 can be invoked from the Output Tables menu. Vertical= and Yardstick= control this Table. This set of rulers provides a graphical description of the variable as a Wright Map. The facets are positioned horizontally. The "+" or "-" before the facet name indicates whether the facet measures are positively or negatively oriented (Positive= or Negative=). The vertical axis provides a linear definition of the variable. Each element name is positioned according to its measure, "Measr". Elements with extreme scores are not positioned by measure, but are placed at the extreme top or bottom of the column for their facet.

The average expected response on the scale is indicated on the right-hand side. S.1 refers to scale 1 in Table 8.1. The "scale" column in Table 6 is the x-axis of the Expected Score ICC, also called the IRF. "---" are position at average expected half-score points. The extreme values of the scale are indicated by (_), e.g., (9) and (1).

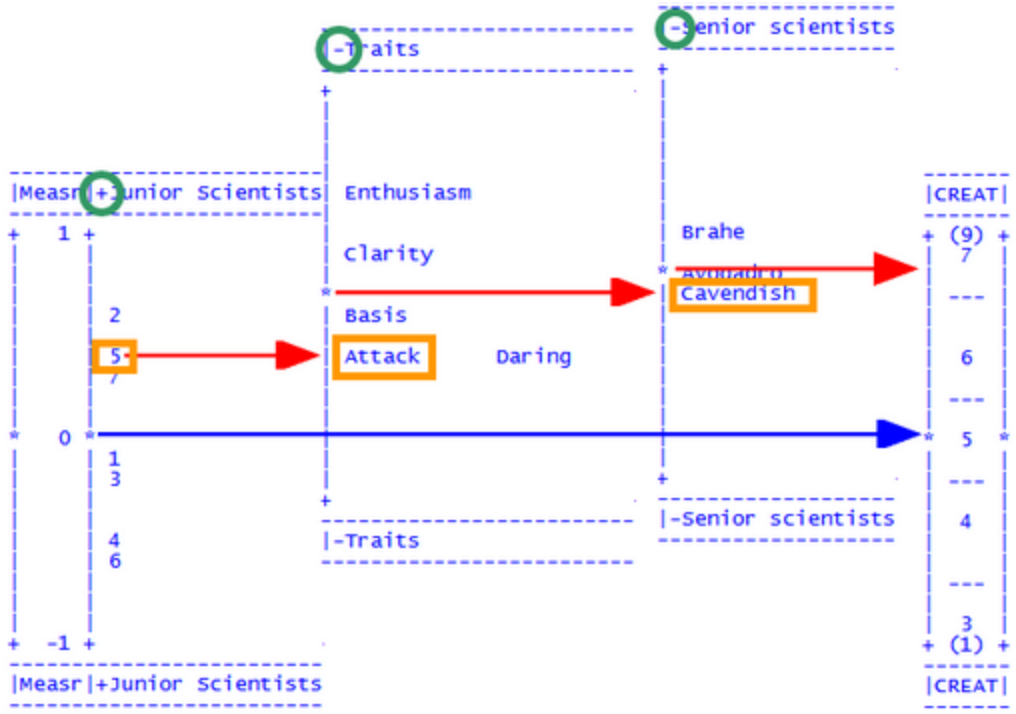
It is easier to compare elements with the rating scale (or partial credit) when all facets are positively oriented: Positive= 1,2,3,....

To choose what to show in Facets Table 6, see Vertical=
To set the width of Table 6, see Yardstick=



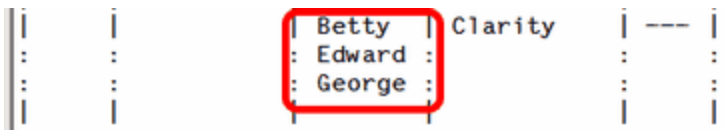

The rating scale in the right-hand column is identified underneath the "Rulers": CREATIVITY

---




Using the rulers: "+ Junior scientist "5" ability - Attach difficulty - Cavendish severity" = average expected rating of "7".

---




Lines marked with ":" (instead of "|") are continuations of the lines above, and have the same measure.

In this example, Betty, Edward and George have different measures, but display on the same line. They are listed alphabetically.

---

**Example 1. A published example**

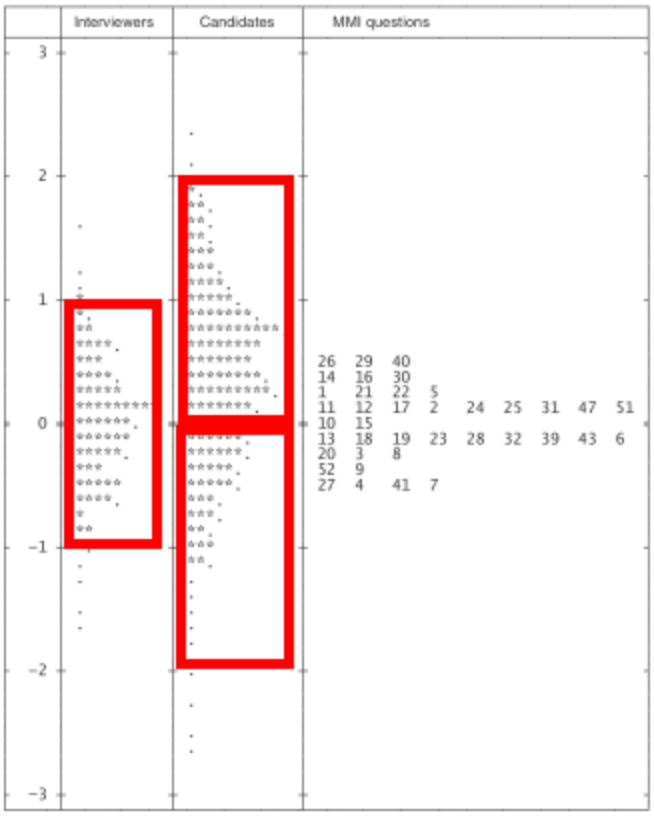



Table 6 as published in "Should candidate scores be adjusted for interviewer stringency or leniency in the multiple mini-interview?" by Chris Roberts, Imogene Rothnie, Nathan Zoanetti & Jim Crossley, Medical Education 2010: 44: 690–698

My red boxes: Notice that the logit range of the interviewers (raters) is about half that of the candidates. So that:

*smartest candidate + most severe rater ≈ least smart candidate + most lenient rater*

This was first noticed by Prof. Edgeworth around 1890, but 120 years later, Examination Boards continue to rely on the "luck of the draw" (as stated in Shavelson & Webb, "Generalizability Theory", 1991, p.8).

---

**Example 2. Prettify the Creativity "rulers"**

| Measr | -Senior scientists | +Junior Scientists | -Traits | CREAT |
|---|---|---|---|---|
| 1 | | | | (9) 7 |
| | | Betty | Enthusiasm | --- |
| | | Edward George | | 6 |
| | Brahe | | Clarity | --- |
| 0 | Avogadro Cavendish | Anne Chris | Basis | 5 |
| | | | Attack Daring | --- |
| | | David Fred | | 4 |
| | | | | --- |
| -1 | | | | 3 (1) |
| Measr | -Senior scientists | +Junior Scientists | -Traits | CREAT |

1. Do a standard Facets analysis of the Creativity data.
2. "Output Tables" menu
3. Table 6 (my setting: Yardstick = 1L, 2L, 3L)
4. Click on"Modify Specifications"
5. Click on "Ascii = Webpage"
6. click on OK
7. Temporary output file
8. Webpage displays
9. Reduce size until rulers are neatly visible (Ctrl -)
10. Copy and paste into a new Word document
11. Highlight all the Table with your mouse
12. Format font
13. 8pt

This should display neatly. If not,

14. 7pt (you can type in the 7 where the 8 was shown).

If there are spaces between the lines vertically, then

15. Format paragraph. Line spacing. Exactly. 6pt. (one point less than the font size)

## 11.7 Table 6.0.0 Subsets - disjoint element listing

When there are disjoint subsets, this Table reports them as a list, and in ruler format.

This is produced with the command Subsets= Report in the main output, or in Table 6 from the "Output Tables" pull-down menu.

Subset numbers are also shown in Table 7 and "Output Files" menu, "Subset group-anchor file".

```
Table 6.0.0  Disjoint Subset Element Listing.
Subset number: 1
Facet: 1. Senior scientists  2 Elements: 1 2
Facet: 2. Junior Scientists  3 Elements: 1-3
Subset number: 2
Facet: 1. Senior scientists  2 Elements: 3 4
Facet: 2. Junior Scientists  4 Elements: 4-7

Ratings of Scientists (edited to illustrate ambiguity in measurement)
Table 6.0.1  Disjoint Subset 1 Vertical Summary.
Vertical = (1A,2A,3A,S) Yardstick (columns lines low high extreme)= 0,1,45,56,End
-------------------------------------------------
```

```
|Measr|+Senior scientists|+Junior Scientists|Scale|
-------------------------------------------------
+  56 +                  + Betty            + (9) +
+  55 +                  +                  +  6  +
+  54 +                  +                  +     +
+  53 +                  +                  + --- +
+  52 +                  +                  +     +
+  51 + Brahe            +                  +  5  +
+  50 +                  +                  +     +
+  49 + Avogadro         + Anne             + --- +
+  48 +                  +                  +     +
+  47 +                  + Chris            +  4  +
+  46 +                  +                  +     +
+  45 +                  +                  + (1) +
-------------------------------------------------
|Measr|+Senior scientists|+Junior Scientists|Scale|
-------------------------------------------------
```

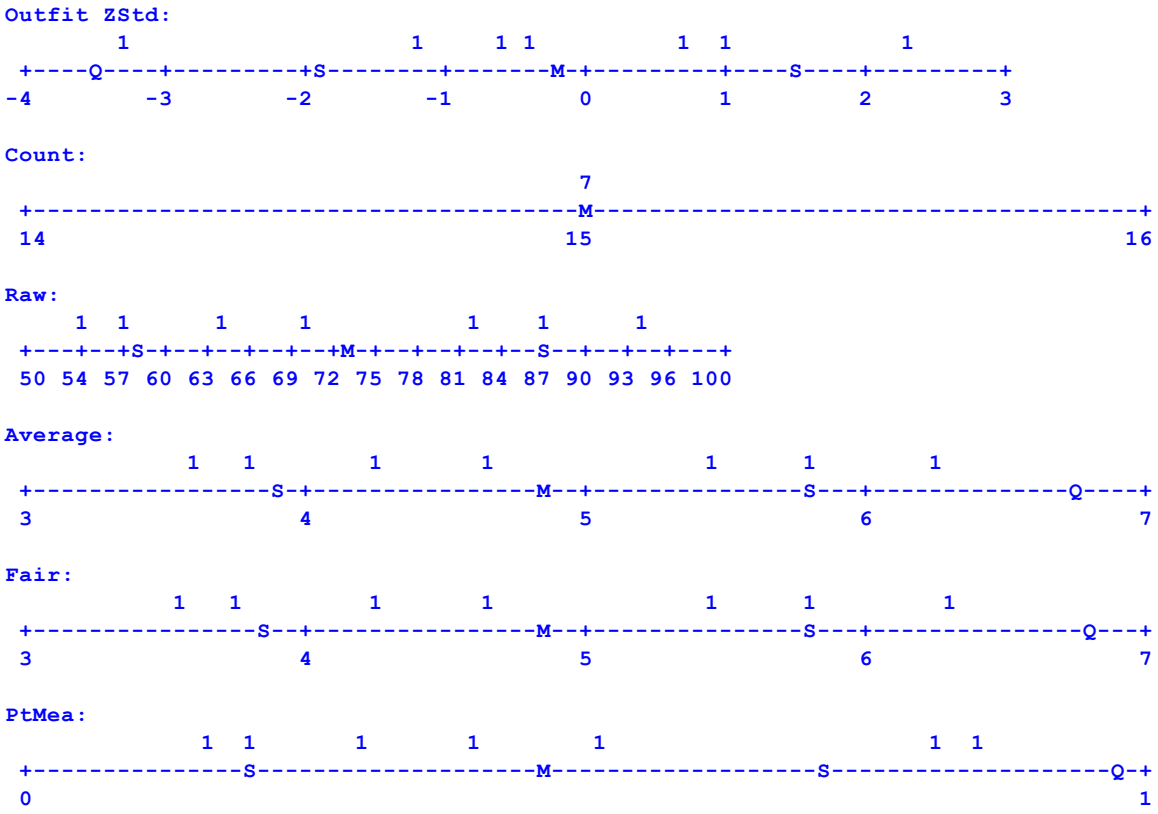
This shows which elements fall into subsets. Their measures can only be directly compared to measures in the same subset, or measures of elements which are common across all subsets. These are not shown in any of the subset lists.

## 11.8 Table 6.2 Graphical description of facet statistics

Controlled by Barchart=
One table is produced for each facet. Each entry in the table is a bar-chart reflecting the distribution of the reported statistics. The scale is from left to right, with range such that all facets have the same scale for each statistic. "M" represents the mean value of the statistic for the facet, "S" is one sample standard deviation each side of the mean, "Q" is two sample standard deviations, and "X" is three.

Frequencies above 9 are reported vertically. 10 is printed as
1
0

```
Ratings of Scientists (Psychometric Methods p.282 Guilford 1954)
Table 6.2  Junior Scientists Facet Summary.

Logit:
                  1   1       1      1             1     1       1
 +-----Q----------------S----------------M----------------S----------------Q-----+
-1                                      0                                       1

S.E.:
               61
 +------------QMQ-------------------------------------------------------------+
 0                                                                             1

Infit MnSq:
          1             1   1     1          1      1                         1
 +-----------------S--------------------M+-------------------S--------------------Q
 0                                      1                                       2

Outfit MnSq:
          1             1      1 1               1     1                       1
 +------------------S--------------------M---------------------S-----------------+
 0                                      1                                       2

Infit ZStd:
        1                    1   1    1       1   1             1
 +---Q-----+---------+S--------+------M--+---------+---S-----+---------Q
-4        -3        -2        -1         0         1         2         3
```




| **Definitions of numeric bar-charts:** | |
|---|---|
| Logit: | Logit (or user-scaled) measures of elements |
| S.E. | Standard errors of the measures (Barchart=All) |
| Infit MnSq: | Infit, information-weighted, inlier pattern sensitive, mean-square fit statistics |
| Outfit MnSq: | Outfit, outlier sensitive, mean-square fit statistics |
| Infit Zstd: | Standardized, approximately N(0,1), infit statistic |
| Outfit Zstd: | Standardized, approximately N(0,1), outfit statistic |
| Count: | Number of observations of the elements (Barchart=All). Total observations for all estimable elements. |
| Raw: | Raw scores of the elements (Barchart=All) |
| Average: | Average observations = Raw / Count (Barchart=All) |
| Fair: | Fair-values of the average observations (Barchart=All) |
| PtMea: | Point-measure or point-biserial correlations of the elements (Barchart=All, Ptbiserial=) |

## 11.9 Table 7 Facet elements measurement report

T7LGS= and Juxtapose= control this Table. One table is produced for each arrangement of each facet which lists its elements and their estimates. The arrangement is determined by the Arrange= specification, otherwise it is ascending sequence by element number.

```
Ratings of Scientists (Psychometric Methods p.282 Guilford 1954)
Table 7.1.1  Senior scientists Measurement Report  (arranged by mN).

+-------------------------------------------------------------------------------------------------------------------------------+
|  Total   Total   Obsvd  Fair(M)|   -    Model | Infit      Outfit    |Estim.| Corr. | Exact Agree. |                       |
|  Score   Count  Average Average|Measure  S.E. | MnSq ZStd  MnSq ZStd |Discrm| PtBis | Obs %  Exp % | Nu Senior scientists |
|-------------------------------+--------------+----------------------+------+-------+--------------+----------------------|
|    35      35     1.00    1.01 |( 115.6  18.1)|Maximum               |      |   .00 |   1.4    1.4 |  1 Avogadro          | in subset: 1
|   159      35     4.54    4.48 |   50.6   1.1 | 1.18   .8  1.19   .8 |  .38 |  -.02 |  11.4   12.0 |  2 Brahe             | in subset: 1
|   181      35     5.17    5.15 |   47.9   1.1 |  .78  -.9   .79  -.9 | 1.60 |   .27 |  12.9   11.5 |  3 Cavendish         | in subset: 1
|    10       1    10.00         |              |Unmeasurable          |      |   .00 |    .0     .0 | 10 Loner             | in subset: 2
|-------------------------------+--------------+----------------------+------+-------+--------------+----------------------|
|   125.0    35.0    3.57    3.54 |   71.3   6.8 |  .98  -.1   .99   .0 |      |   .08 |              | Mean (Count: 3)      |
|    64.3     .0     1.84    1.81 |   31.3   8.0 |  .20   .9   .20   .9 |      |   .13 |              | S.D. (Population)    |
|    78.7     .0     2.25    2.22 |   38.3   9.8 |  .28  1.3   .28  1.3 |      |   .16 |              | S.D. (Sample)        |
+-------------------------------------------------------------------------------------------------------------------------------+
   With extremes, Model, Popuin: RMSE 10.5  Adj (True) S.D. 29.5  Separation 2.82  Strata 4.09  Reliability (not inter-rater) .89
   With extremes, Model, Sample: RMSE 10.5  Adj (True) S.D. 36.9  Separation 3.52  Strata 5.03  Reliability (not inter-rater) .93
Without extremes, Model, Popuin: RMSE 1.1  Adj (True) S.D. .8  Separation .69  Strata 1.26  Reliability (not inter-rater) .32
Without extremes, Model, Sample: RMSE 1.1  Adj (True) S.D. 1.6  Separation 1.40  Strata 2.20  Reliability (not inter-rater) .66
With extremes, Model, Fixed (all same) chi-square:  16.4  d.f.: 2  significance (probability): .00
With extremes, Model,  Random (normal) chi-square:  4.3  d.f.: 2  significance (probability): .11
Inter-Rater agreement opportunities: 105  Exact agreements: 9 =  8.6%  Expected:  8.7 =  8.3%
```

when there are Groups, each Group has its own Table. such as Table 7.1.1.1. Controlled by T7LGS=

| **Column Heading** | **Meaning** (values that are not estimable are blank) |
|---|---|
| Total Score = (when Totalscore=Yes) | the observed raw score (including extreme elements) |
| Obsvd Score = Observed Score: (when Totalscore=No) | observed raw score, the sum of the active responses for that element, after removal of extreme elements and renumbering of rating scales or partial credit scales to eliminate unobserved intermediate categories, when Keep= is not specified, so that Ordinal is active. |
| Weightd Score = (when weighting is active) | weighted: the observed raw score (including extreme elements) multiplied by their weights |
| Total Count = (when Totalscore=Yes) | the observed response count (including extreme elements). This is the sum of the weighted when the data are weighted. |
| Obsvd Count = Observed Count: (when Totalscore=No) | the number of active responses observed for the element |
| Weightd Count = (when weighting is active) | weighted: the observed response count (including extreme elements) multiplied by their weights |
| *Fair(M) Average* = Fair Mean-based Average | Rasch measure to raw score conversion, producing an average rating for the element that is standardized so that it is fair to all elements in the facet, using the mean measures of the elements in each facet as the baseline. |
| *Fair(Z) Average* = Fair Zero-based Average | Rasch measure to raw score conversion, producing an average rating for the element that is standardized so that it is fair to all elements in the facet, using the facet local origins (zero points) as the baseline. |
| (Anchoring status Character =) A, G, X or none | none shown: the element is not anchored<br>A - the element measure is anchored (preset, fixed) in Labels=<br>G - the element measure is part of a group-anchor which includes extreme measures (if any).<br>X - the element measure is part of a group-anchor which excludes extreme measures (if any). |
| +- Measure | Rasch measure/calibration of ability/difficulty/leniency etc. in log-odds units (logits) or user-scaled units based on Umean= and Uscale=<br>(....) indicates that the measure is for an extreme score, estimated according to Xtreme=. |

| | |
|---|---|
| | + says this facet is positively oriented: higher score <-> higher measure<br>- says this facet is negatively oriented: higher score <-> lower measure |
| Model S.E. | the smallest-possible standard error assuming all randomness in the data accords with the Rasch model (best case) |
| Real S.E. | the larger standard error assuming excess randomness in the data contradicts the Rasch model (worst case) |
| (Element measurement status) | "Anchored" = this element is anchored, but does not participate in this analysis. Shown when Omit-unobserved=No.<br>"Minimum" = this element has the minimum possible extreme score. Fit statistics cannot be computed.<br>"Maximum" = this element has the maximum possible extreme score. Fit statistics cannot be computed.<br>"Unmeasurable" = this measure for this element cannot be estimated because its observations do not connect it with estimable elements<br>"One datum" = this element has only one observation<br>"Unknown" = this element is inestimable for unknown reasons. Please report this to www.winsteps.com |
| Infit MnSq = Infit Mean-Square | the information-weighted, inlier-pattern-sensitive, mean-square fit statistic, with expectation 1, and range 0 to infinity. Less than 1 indicates muting: too little variation, lack of independence. More than 1 indicates noise: unmodeled excess variation. A mean-square is a chi-squared fit statistic divided by its degrees of freedom.<br>For the chi-squared value, d.f., and probability, see Scorefile= |
| Infit Zstd = Infit Z-standardized *t*-statistic | the Infit MnSq statistic standardized toward a unit-normal distribution so effectively a t-statistic with infinite degrees of freedom, i.e., a z-score. The accuracy of this standardization is data dependent. This tests the statistical hypothesis: "Does the Infit Mean-Square indicate that apparent randomness in these data fit the Rasch model exactly?" |
| Outfit MnSq = Outfit Mean-Square | the unweighted, outlier-sensitive, mean-square fit statistic, with expectation 1, and range 0 to infinity. Less than 1 indicates muting: too little variation, lack of independence. More than 1 indicates noise: unmodeled excess variation. A mean-square is a chi-squared fit statistic divided by its degrees of freedom. For the chi-squared value, d.f., and probability, see Scorefile= |
| Outfit Zstd = Outfit Z-standardized *t*-statistic | the Outfit MnSq statistic standardized toward a unit-normal distribution so effectively a t-statistic with infinite degrees of freedom, i.e., a z-score. The accuracy of this standardization is data dependent. This tests the statistical hypothesis: "Does Outfit Mean-Square indicate that the apparent randomness in these data fit the Rasch model exactly?" |
| Estim. Discrm = Estimated Discrimination | an estimate of the item discrimination computed according to the "two-parameter logistic model" (2-PL) and "Generalized Partial Credit Model" approach www.rasch.org/rmt/rmt181b.htm, but without allowing the discrimination estimate to alter other estimates. 1.0 is the expected value. Values higher than 1.0 indicate a steeper than expected empirical ICC. Values less than 1.0 indicate a flatter |

| | |
|---|---|
| | empirical ICC. Negative values indicate reverse discrimination. Inspect these on the Graphs menu. According to www.rasch.org/rmt/rmt142a.htm discriminations in the range 0.5 to 1.5 provide reasonable fit to the Rasch model. |
| Displ. = Displacement measure | when a column with this heading is produced, it contains rough estimates of the displacements of the reported measures from those expected if:<br>i) the analysis had reached convergence, or<br>ii) the reported measure were not fixed at an anchored value.<br>Displacements less than the convergence criteria or the measure standard errors are not reported here. They can be seen in the Scorefile=. When there are no displacements to report, this column is not produced.<br>Displacement in logits = (observed raw score - expected raw score based on reported measure) / (model-derived raw-score variance)<br>The reported Displacement is (Displacement in logits) * (uscale user-scaling) * (orientation of facet: positive or negative). This is limited to the range -10 to +10 logits. |
| Correlation: PtMea = Point-measure correlation | the correlation between the element's observations and the measures modeled to generate them. Produced by Pt-biserial=Measure |
| Correlation: PtBis = Point-Biserial correlation (exclude) | a many-facet version of the point-biserial correlation between observations and their corresponding average observations (excluding the current observation). Produced by Pt-biserial=Yes |
| Correlation: PtBisI = Point-Biserial correlation (include) | a many-facet version of the point-biserial correlation between observations and their corresponding average observations (including the current observation). Produced by Pt-biserial=Include |
| Correlation: PtExp = Point-Correlation Expected value | the expected value of the reported adjacent point-correlation. Use this for deciding whether a correlation is too high or too low. www.rasch.org/rmt/rmt221e.htm |
| Exact Agreement: Obs % = Observed % | percent of exact agreements between raters on ratings made under identical conditions, see Interrater= |
| Exact Agreement: Exp % = Expected % | expected percent of exact agreements between raters on ratings made under identical conditions based on their Rasch measures. Compare the Observed % with this value. |
| Group | the group number to which the element is assigned in Labels=. This is used for group-anchoring and group reporting. |
| Weight | weighting applied to observations of this element (defaults to 1.00) in Labels=. This multiples the weight (if any) specified in Models=, and the R weight specified with the observations (if any) and the Labels= weights of the other elements in the Data= specification for each active observation. |
| N = number | the element number in Labels= |
| (Name of Facet in Labels=) | element label in Labels= |
| in subset: (numbers) | the disjoint subsets of which this element is a member |
| | |

| | |
|---|---|
| Mean (Count: ...) | Average of its column for active elements with count of relevant elements |
| S.D. (Population) | the (smaller) standard deviation of all values in its column (relevant to this analysis), computed on the basis that the elements are the entire population of elements |
| S.D. (Sample) | the (larger) standard deviation of all values in its column (relevant to this analysis), computed on the basis that the elements are a random sample from a population of elements |
| Reliabilities and Chi-squares | To Table 7 summary, reliability and chi-squared statistics |
| Inter-rater Agreement | To Table 7 agreement statistics |

**Table 7 and Excel**
The statistics in Table 7 are also in the Scorefile=, but it can be convenient to copy-and-paste line from Table 7 into Excel. These paste into one Excel column. The Excel "Data", "Text-to-Columns" splits the Table 7 lines into Excel columns.

## 11.10 Table 7 Reliability and Chi-square statistics

Table 7 also provides summary statistics by facet.

```
+-------------------------------------------------------------------------------------------------
       ----------------------+
|  Total    Total    Obsvd  Fair-M|        Model | Infit      Outfit   |Estim.| Correlation | |
                            |
|  Score    Count  Average Avrage|Measure  S.E. | MnSq Zstd  MnSq Zstd|Discrm| PtMea PtExp | |
       Nu Reader             |
|-------------------------------+--------------+---------------------+------+-------------+
       +---------------------|
.....
|-------------------------------+--------------+---------------------+------+-------------+
       +---------------------|
|   460.8     96.0      4.8    4.73|     .00    .08 | 1.00  -.1    .99  -.2|      |    .61       | |
       Mean (Cnt: 12)        |
|    29.5       .0       .3     .32|     .19    .00 |  .23  1.8    .22  1.7|      |    .05       | |
       S.D. (Population)     |
|    30.8       .0       .3     .33|     .20    .00 |  .24  1.9    .23  1.8|      |    .06       | |
       S.D. (Sample)         |
+-------------------------------------------------------------------------------------------------
       ----------------------+
Model, Populn: RMSE .08  Adj (True) S.D. .17  Separation 2.17  Strata 3.22  Reliability (not
       inter-rater) .82
Model, Sample: RMSE .08  Adj (True) S.D. .18  Separation 2.28  Strata 3.38  Reliability (not
       inter-rater) .84
Model, Fixed (all same) chi-squared: 66.3  d.f.: 11  significance (probability): .00
Model,  Random (normal) chi-squared: 9.4  d.f.: 10  significance (probability): .49
Inter-Rater agreement opportunities: 384  Exact agreements: 108 = 28.1%  Expected: 82.6 = 21.5%
```

or

```
With extremes, Model, Populn: RMSE 1.05  Adj (True) S.D. 1.98  Separation 1.88  Strata 2.84
       Reliability .78
With extremes, Model, Sample: RMSE 1.05  Adj (True) S.D. 2.01  Separation 1.91  Strata 2.89
       Reliability .79
Without extremes, Model, Populn: RMSE 1.02  Adj (True) S.D. 1.71  Separation 1.68  Strata 2.57
        Reliability .74
Without extremes, Model, Sample: RMSE 1.02  Adj (True) S.D. 1.75  Separation 1.71  Strata 2.62
        Reliability .75
```

```
With extremes, Model, Fixed (all same) chi-squared: 175.9  d.f.: 34  significance
       (probability): .00
With extremes, Model,  Random (normal) chi-squared: 33.8  d.f.: 33  significance (probability):
       .43
```

In summary:
"model" = "the unexpectedness in this facet is considered to be the randomness predicted by the Rasch model"
"population" = the elements in this facet are the entire population of possible elements
"sample" = the elements in this facet are a sample from the entire population of possible elements
"fixed" = we are testing the hypothesis "all the elements of this facet have statistically the same measure"
"random" = we are testing the hypothesis "all the elements of this facet are a random sample from a normally-distributed population"
RMSE = root-mean-square-error: it is the statistical average of the standard errors of the measures. It reports the overall precision of the measurement of the elements in the facet. RMSE is heavily influenced by the count of observations of each element.

| | |
|---|---|
| Mean = | arithmetic average |
| Count = | number of elements reported |
| S.D. (Populn) | is the standard deviation when this sample comprises the entire population.<br>If the element list includes every possible element for the facet: use the Population statistics, e.g., grade levels, genders (sexes), ... |
| S.D. (Sample) | is the standard deviation when this sample is a random sample from the population.<br>If there are "more like this" elements in addition to the current elements: use the Sample statistics, e.g., candidates, items (usually), tasks, .... |
| With extremes | including elements with extreme (zero and perfect, minimum possible and maximum possible) scores |
| Without extremes | excluding elements with extreme (zero and perfect, minimum possible and maximum possible) scores |
| Model | Estimated as though all noise in the data is due to model-predicted stochasticity (i.e., the best-case situation for randomness in the data) |
| Real | Estimated as though all unpredicted noise is contradicting model expectations (i.e., the worst-case situation |
| RMSE | root mean square standard error (i.e., the average S.E. statistically) for all non-extreme measures. |
| Adj (True) S.D. | "true" sample standard deviation of the estimates after adjusting for measurement error |
| Separation | Adj "true" S.D. / RMSE, a measure of the spread of the estimates relative to their precision. The signal-to-noise ratio is the "true" variance/error variance = Separation². See also Separation. |
| Strata | (4*Separation + 1)/3, a measure of the spread of the estimates relative to their precisions, when extreme measures are assumed to represent extreme "true" abilities. See also Strata |
| Reliability (not inter-rater) | Spearman reliability: Rasch-measure-based equivalent to the KR-20 or Cronbach Alpha raw-score-based statistic, i.e., the ratio of "True variance" to "Observed variance" (Spearman 1904, 1911). This shows how different the measures are, which may or may not indicate how "good" the test is. High (near 1.0) person and item reliabilities are preferred. This reliability is somewhat the opposite of an interrater reliability, so low (near 0.0) judge and rater reliabilities are preferred. See also Reliability. |

| | |
|---|---|
| Fixed (all same) chi-squared: | A test of the "fixed effect" hypothesis: "Can this set of elements be regarded as sharing the same measure after allowing for measurement error?" The chi-squared value and degrees of freedom (d.f.) are shown. The significance is the probability that this "fixed" hypothesis is the case. Depending on the sub-Table, this tests the hypothesis: "Can these items be thought of as equally difficult?" The precise statistical formulation is:<br>wi = 1/SE²i for i=1,L, where L is the number of items, and Di is the difficulty/easiness of item i.<br>**chi-squared = Sum(wi.D²i) - Sum( wi.Di)²/ Sum(wi) with d.f. = L-1**<br>Or this tests the hypothesis: "Can these raters be thought of as equally lenient?" Is there a statistically significant rater effect?<br>The precise statistical formulation is:<br>**wj = 1/SE²j for j=1,J**, where J is the number of raters, and Cj is the leniency/severity of rater j.<br>**chi-squared = Sum(wj.C²j) - Sum( wj.Cj)²/ Sum(wj) with d.f. = J-1**<br>And so on .... |
| Random (normal) chi-squared: | A test of the "random effects" hypothesis: "Can this set of elements be regarded as a random sample from a normal distribution?" The significance is the probability that this "random" hypothesis is the case. This tests the hypothesis: "Can these persons (items, raters, etc.) be thought of as sampled at random from a normally distributed population?" The precise statistical formulation is:<br>**var(D) = S(Di-Dmean)²/(L-1) - ( SSE²i)/L**<br>**wi = 1/(var(D)+SE²i)**<br>**chi-squared = S(wi.D²i) - ( Swi.Di)²/ Swi with d.f. = L-2** |
| Rater agreement opportunities | when Inter-rater= facet-number.<br>see Table 7 Agreement statistics |

## 11.11 Table 7 Agreement Statistics

When **inter-rater=** is used to specify a rater facet, then a count of the situations in which ratings are given in identical circumstances by different raters is made.

If exact inter-rater statistics are required, please do a special run of Facets in which all unwanted facets are Xed out, so that matching only occurs on facets relevant to agreement. For instance, if "rater gender" is irrelevant to agreement, then X out that facet in the Models= specifications.

The percent of times those ratings are identical is reported, along with its expected value. This supports an investigation as to whether raters are rating as "independent experts" or as "rating machines". The report is:

```
Table 7.3.1  Reader Measurement Report  (arranged by MN).
-------------------------------------------------------------------------------------------------
         -
| Obsvd   Obsvd  Obsvd  Fair-M|        Model | Infit      Outfit   | Exact Agree. |
         |
| Score   Count Average Avrage|Measure  S.E. |MnSq ZStd  MnSq ZStd | Obs %  Exp % | Nu Reader
         |
-------------------------------------------------------------------------------------------------
         -
|   1524    288     5.3   5.26|   -.30   .05 | 1.2   2    1.2   2  |  28.2   20.9 |  8 8
         |
|   1455    288     5.1   5.00|   -.16   .05 |  .5  -7     .5  -7  |  30.8   21.6 |  4 4
         |
....
-------------------------------------------------------------------------------------------------
         -
RMSE (Model)  .05 Adj S.D.  .19  Separation  4.02  Strata  5.69  Reliability  .94
......
```

```
Inter-Rater agreement opportunities: 60480  Exact agreements: 17838 = 29.5%  Expected: 13063.2
        = 21.6%
-------------------------------------------------------------------------------------------
       -
```

Exact Agree. is exact agreements under identical rating conditions. Agreement on qualitative levels relative to the lowest observed qualitative level.
So, imagine all your ratings are 4,5,6 and all my ratings are 1,2,3.
If we use the (shared) Rating Scale model. Then we will have no exact agreements.
But if we use the (individual) Partial Credit model, #, then we agree when you rate a 4 (your bottom observed category) and I rate a 1 (my bottom observed category). Similarly, your 5 agrees with my 2, and your 6 agrees with my 3.
If you want "exact agreement" to mean "exact agreement of data values", then please use the Rating Scale model statistics.

Obs % = Observed % of exact agreements between raters on ratings under identical conditions.
Exp % = Expected % of exact agreements between raters on ratings under identical conditions, based on Rasch measures.
If Obs % ≈ Exp % then the raters may be behaving like independent experts.
If Obs % » Exp % then the raters may be behaving like "rating machines".

Here is the computation for "Expected Agreement %". We pair up another rater with the target rater who rated the same ratee on the same item of the same task of the same ....., so the raters rated the same performance under identical circumstance.

Then, for each rater we have an observed rating. They agree or not. The percentage of times raters agree with the target rater is the "Observed Agreement%"

For each rater we also have an (average) expected rating based on the Rasch measures. The (average) expected ratings will not agree unless the raters have the same leniency/severity measure.

But we also have the Rasch-model-based probabilities for each category of the rating scale for each rater. Suppose this is a 1,2,3 (3-category) rating scale.

| Rater A | Rater B | Expected agreement between Raters A and B (assuming they are rating independently) |
|---|---|---|
| probability of category 1 = 10%<br>probability of category 2 = 40%<br>probability of category 3 = 50% | probability of category 1 = 20%<br>probability of category 2 = 60%<br>probability of category 3 = 20% | Category 1 10%*20% = 2%<br>Category 2 40%*60% = 24%<br>Category 3 50%*20% = 10%<br>Expected agreement in any category = 2+24+10% = 36% |

This expected-agreement computation is performed over all pairs of raters and averaged to obtain the reported "Expected Agreement %".

Higher than expected agreement indicates statistical local dependence among the raters. This biases all the standard errors towards zero. An approximate guideline is:
"True" Standard error = "Reported Standard Error" * Maximum( 1, sqrt (Exact agreements / Expected)) for all elements.
In this example, the inflator for the S.E.'s of all elements of all facets approximates sqrt( 17838/13063.2) = 1.17.

Alternatively, deflate the reported person-facet reliability, R, in accordance with the extent to which the raters are not independent. Based on the Spearman-Brown prophecy formula, an approximation is:

T = (100 - observed exact agreement%) / (100 - expected exact agreement%)
deflated reliability = T * R / ( (1-R) + T * R)

Example: 100 raters with a wide range of rater severity/leniency:

| | |
|---|---|
| Exact agreements | 781=18.8% |
| Expected | 577.5=13.9% |

With this large spread of rater severities, the prediction is that only 13.9% of the observations will show the raters giving the same rating under the same conditions. This accords with the wide range of severities.

There is somewhat more agreement than this in the data, 18.8%. This is typical of the psychology of rater behavior. We are conditioned from baby-hood to agree with what we conceive to be the expectations of others. This behavior continues even for expert raters. Subconsciously they continue to have a mental pressure to agree with the expectations of others. In this case, that pressure has increased observed agreement from 13.9% to 18.8%.

Whether you report this depends on the purpose for your paper. If it is an investigation into rater behavior, then this provides empirical evidence for a psychological conjecture. If your paper is a validity study of the instrument, then this aspect is probably too obscure to be meaningful for your audience.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/ "many texts recommend 80% agreement as the minimum acceptable interrater agreement."

See more at Inter-rater Reliability and Inter-rater correlations

## 11.12 Table 7 Group Summary

When the elements of a facet are assigned to more than one group in Labels=, then a group summary report is output. Controlled by T7LGS=

Table 7.5.2 is an example:

| Total Score | Total Count | Obsvd Average | Fair-M Avrage | Measure | Model S.E. | Infit MnSq | Infit ZStd | Outfit MnSq | Outfit ZStd | Estim. Discrm | Correlation PtMea | Correlation PtExp | Exact Agree. Obs % | Exact Agree. Exp % | Group | N Senior scientists |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 169.3 | 35.0 | 4.8 | 4.81 | .06 | .11 | .92 | -.4 | .94 | -.3 | | .59 | | | | | Mean (Count: 3) |
| 10.3 | .0 | .3 | .29 | .12 | .00 | .22 | 1.0 | .23 | 1.1 | | .30 | | | | | S.D. (Population) |
| 12.6 | .0 | .4 | .36 | .14 | .00 | .27 | 1.3 | .28 | 1.3 | | .36 | | | | | S.D. (Sample) |
| 176.0 | 35.0 | 5.0 | 5.00 | -.01 | .11 | .77 | -1.1 | .79 | -1.0 | | .80 | | | | 1 | Mean (Count: 2) |
| 5.0 | .0 | .1 | .14 | .06 | .00 | .09 | .5 | .11 | .6 | | .06 | | | | 1 | S.D. (Population) |
| 7.1 | .0 | .2 | .20 | .08 | .00 | .13 | .7 | .16 | .8 | | .08 | | | | 1 | S.D. (Sample) |
| 156.0 | 35.0 | 4.5 | 4.43 | .21 | .11 | 1.21 | .9 | 1.23 | 1.0 | | .18 | | | | 2 | Mean (Count: 1) |
| .0 | .0 | .0 | .00 | .00 | .00 | .00 | .0 | .00 | .0 | | .00 | | | | 2 | S.D. (Population) |

The columns are the same as Table 7 Facet elements report. This Table shows the summary statistics for the groups. They can be copied in Excel, and then organized into columns using Excel's "Data", "Text to columns".

For comparing the means of the groups, please use the Welch t-statistic.

## 11.13 Table 8.1 Dichotomy, binomial trial and Poisson statistics

Scale codes in Models= and Rating (or partial credit) scale= control this Table of statistics for scale structures. For each modeled scale found in the data, a table is produced.

To see the Average Measure for the categories of each item for the Rating Scale model: ?,?,?,...
1. do the rating scale analysis
2. output the anchor file: Anchorfile=
3. leave everything anchored, but change the Models= to put # for the facet you want the ability mean measures.
4. analyze the anchor file

---

**Dichotomies**
This is generated by
Models=?,?,D

```
Table 8.1 Category Statistics.

 Model = ?,?,D
+-------------------------------------------------+
|           DATA                |   QUALITY CONTROL |
|      Category Counts       Cum.|  Avge  Exp. OUTFIT|
|Score Total      Used    %    % |  Meas  Meas  MnSq |
|--------------------------------+-------------------|
|  0     289        240   50%  50%| -3.39  -3.38   .8 |
|  1     341        236   50% 100%|  3.06   3.05   .6 |
+-------------------------------------------------+
```

The column headings mean:

| | |
|---|---|
| DATA = | Information relating to the data |
| Score = | Cardinal value assigned to each category, i.e., its rating. |
| Category Counts<br>Total =<br>Used = | <br>Number of observations of this category in the analysis<br>Number of observations that participated in the estimation (excludes observations in extreme scores) |
| % = | Percent of the Used responses which are in this category. |
| Cum. % = | Cumulative percentage of responses in this category and lower. |
| FIT = | Information regarding validity of the data. |
| Avge Meas = | The average of the measures that are modeled to generate the observations in this category. If Avge Measure does not increase with category score, then the measure is flagged with a "*", and doubt is cast on the idea that larger response scores correspond to "more" of the variable. |
| Exp. Meas = | The expected value of the average measures. This provides guidance whether observations in a category are higher or lower than expected. |
| OUTFIT MnSq = | The unweighted mean-square for observations in this category. Mean-squares have expectation of 1.0. Values much larger than 1.0 indicate unexpected observations in this category. Central categories usually have smaller mean-squares than extreme categories. The INFIT MnSq is not reported because it approximates the OUTFIT MnSq when the data are stratified by category. |
| Obsd-Expd Diagnostic Residual = | score-point<br>---<br>difference between the observed count of responses and the expected count, based on the Rasch measures. This is shown only when it is greater than 0.5 score-points for some category. This can be due to |

| | |
|---|---|
| | i) lack of convergence<br>ii) anchor values incompatible with the data<br>iii) responses do not match the specified scale structure, e.g., Poisson counts. |
| Response Category Name =  | name of category from Rating (or partial credit) scale= specification |

**Dichotomies with anchored thresholds** are reported as Table 8.1 Rating Scales

```
+-------------------------------------------------------------------------------------------------------------+
|      DATA            |  QUALITY CONTROL  |RASCH-ANDRICH|  EXPECTATION  |  MOST  |  RASCH-  |Cat|Response|
| Category Counts  Cum.| Avge  Exp.  OUTFIT| Thresholds  |  Measure at   |PROBABLE| THURSTONE|PEAK|Category|
|Score   Used   %    % | Meas  Meas   MnSq |Measure  S.E.|Category  -0.5 |  from  |Thresholds|Prob|  Name  |
|---------------------+-------------------+-------------+---------------+--------+----------+----+--------|
|  0       27  79%  79%|  -.79   -.67   .6 |             |(   .87)       |   low  |   low    |100%| 0      |
|  1        7  21% 100%|  2.34   1.87   .3 |  2.00A      |(  3.07)   1.99|   2.00 |   1.99   |100%| 1      |
+-------------------------------------------------------------------------------------------------------------+
```

**Binomial Trials with Estimated Discrimination**

This is generated by
Models=?,?,B22 where 22 is the number of trials. This estimates the binomial discrimination.

or
Models=?,?,Trials
Rating (or partial credit) scale=Trials,B22
0=Lowest category
*

The Binomial trials discrimination parameterizes the binomial rating scale term in the model. The separate Rating Scale= specifies a scale discrimination. ai in the following model:

$$\log(Pnik/Pnik-1) = Bn - Di - a_i(\log(k/(m-k+1)))$$

Table 8.2 Category Statistics.

```
 Model = 22,?,-?,BB22
 Rating (or partial credit) scale = BB22,B22,G,O
-----------------------------------------------------
|      DATA            |  QUALITY CONTROL  | Obsd-Expd|
| Category Counts  Cum.| Avge  Exp.  OUTFIT|Diagnostic|
|Score   Used   %    % | Meas  Meas   MnSq | Residual |
-----------------------------------------------------
|  4       .5   2%   2%|  -.60 -1.32  1.3 |          |
|  5       .5   2%   4%|  -.93* -1.95   .2 |      -.7 |
|  6      3.5  14%  18%|  -.81   -.33  1.6 |      2.0 |
....
| 15      1.5   6%  82%|   .76    .77   .3 |          |
| 16      3.5  14%  96%|   .81    .33  1.6 |      2.0 |
| 17       .5   2%  98%|   .93   1.95   .2 |      -.7 |
| 18       .5   2% 100%|   .60*  1.32  1.3 |          |
-----------------------------------------------------
```

```
Binomial trials discrimination:  1.03  S.E. .07
```

**Binomial Trials with Fixed Discrimination**

This is generated by
Models=?,?,Trials
Rating (or partial credit) scale=Trials,B22
0=0,1.0,A ; 1.0 is the anchored (pre-set, fixed) discrimination of the binomial scale
*

```
Table 8.2  Category Statistics.

 Model = 22,?,-?,B22
-------------------------------------------------------
|      DATA            |  QUALITY CONTROL   | Obsd-Expd|
| Category Counts  Cum.| Avge  Exp.  OUTFIT|Diagnostic|
|Score   Used   %    % | Meas  Meas   MnSq | Residual |
-------------------------------------------------------
|  4       .5   2%   2%|  -.59 -1.31  1.2 |          |
|  5       .5   2%   4%|  -.91* -1.90   .2 |      -.7 |
....
| 17       .5   2%  98%|   .91   1.90   .2 |      -.7 |
| 18       .5   2% 100%|   .59*  1.31  1.2 |          |
-------------------------------------------------------
Binomial trials discrimination:  1.00  Anchored
```

Binomial trials discrimination: 1.00 Anchored
reports the discrimination of the numerical observations for binomial trials and Poisson counts, either pre-set (anchored) or with its S.E.
Discrimination is anchored, unless scale type is specified using Rating (or partial credit) scale=

**Poisson Counts with Estimated Discrimination**

Specify
Models=?,?,P
or
Models=?,?,Poisson
Rating (or partial credit) scale=Poisson,P
0 = Lowest
*

The separate Rating Scale= specifies a scale discrimination, ai, is to be estimated.
log(Pnik/Pnik-1) = Bn - Di - ai log(k)

```
Table 8.1  Category Statistics.

 Model = ?B,?B,?,CHOPS
 Rating (or partial credit) scale = CHOPS,P,G,O
-------------------------------------------------------
|      DATA            |  QUALITY CONTROL   | Obsd-Expd|
| Category Counts  Cum.| Avge  Exp.  OUTFIT|Diagnostic|
|Score   Used   %    % | Meas  Meas   MnSq | Residual |
-------------------------------------------------------
| 40       1   4%   4%|   .77    .14   .1 |       .8 |
| 41       0   0%   4%|                    |          |
| 42       0   0%   4%|                    |          |
| 43       1   4%   8%|   .79    .16   .0 |       .8 |
...
|133       0   0%  96%|                    |          |
```

```
|134        1    4% 100%|  1.02     .08   2.1 |        .9 |
-------------------------------------------------------------
Poisson discrimination:    .21   S.E.  .00
```

**Poisson Counts with Fixed Discrimination**

Specify
Models=?,?,Poisson
Rating (or partial credit) scale=Poisson,P
0=0,1.0,A ; 1.0 is the anchored (pre-set, fixed) discrimination of the Poisson scale
*

The separate Rating Scale= specifies a fixed Poisson scale discrimination.
log(Pnik/Pnik-1) = Bn - Di - log(k)

```
Table 8.1  Category Statistics.

 Model = ?B,?B,?,P
-------------------------------------------------------------
|      DATA             |  QUALITY CONTROL   | Obsd-Expd|
| Category Counts  Cum.| Avge   Exp.   OUTFIT|Diagnostic|
|Score   Used    %    % | Meas   Meas    MnSq | Residual |
-------------------------------------------------------------
| 40         1    4%    4%|  3.64     .63    .8 |        .8 |
| 41         0    0%    4%|                     |           |
| 42         0    0%    4%|                     |           |
| 43         1    4%    8%|  3.77     .72    .0 |        .8 |
|.....
|133         0    0%   96%|                     |           |
|134         1    4% 100%|  4.81     .37   9.9 |        .9 |
-------------------------------------------------------------
Poisson discrimination:  1.00  Anchored
```

## 11.14 Table 8.1 Rating scale statistics

For each modeled scale code in Models= and Rating (or partial credit) scale= with observations found in Data=, a table is produced. The heading describes to which model the scale applies. Only columns applicable to the type of scale are output.

```
+-------------------------------------------------------------------------------------------------------------------------------------------------------
------+
|          DATA                |  QUALITY CONTROL  |RASCH-ANDRICH|  EXPECTATION  |  MOST  |  RASCH-  | Cat| Obsd-Expd  ANDRICH|
Response   |
|      Category Counts       Cum.| Avge  Exp.  OUTFIT| Thresholds  |  Measure at   |PROBABLE| THURSTONE|PEAK|Diagnostic  Thresh.|
Category   |
|Score Total      Used    %    % | Meas  Meas   MnSq |Measure  S.E.|Category  -0.5 |  from  |Thresholds|Prob| Residual   Displac|
Name     |
|--------------------------------+------------------+-------------+---------------+--------+----------+----+-------------------
+----------|
|  0     452       378   20%  20%|  -.87  -1.03  1.2 |             |( -2.04)       |   low  |   low    |100%|      -.9          |
dislike   |
|  1     620       620   34%  54%|   .13    .33   .7 |  -.85    .07|    .00   -1.17|   -.85 |  -1.00   | 54%|                   |
don't know|
|  2     864       852   46% 100%|  2.23   2.15  1.5 |   .85    .06|(  2.05)   1.18|    .85 |    .99   |100%|                   | like
     |
+-------------------------------------------------------------------(Mean)---------(Modal)--
(Median)-------------------------------------+
```

The column headings mean:

| | |
|---|---|
| DATA | Information relating to the data |

| | |
|---|---|
| Score | Cardinal value assigned to each category, i.e., its rating.<br>If two Scores are shown, the second score is after structural zeroes have been removed |
| Category Counts<br>Total =<br>Used = | <br>Number of observations of this category in the analysis<br>Number of observations that participated in the estimation (excludes extreme scores) |
| % | Percent of the Used responses which are in this category.<br><br>The observed probability is the "Category Counts %" divided by 100.<br><br>The probability of paired agreement by chance is sum (probability of each category**2) across all the categories. In the Table above,<br>Probabilities = .20, .34, .46 (they sum to 1.0)<br>Probability**2 = Probability * Probability = .04, .13, .21<br>Sum (Probability**2) = .04 + .13 + .21 = .38 = Agreement by chance |
| Cum. % | Percent of the Used responses in or below this category. |
| QUALITY CONTROL | Information relating to the validity of the categorization. |
| Avge Meas | The average of the measures that are modeled to generate the observations in this category. If Average Measure does not increase with each higher category, then the category average measure is flagged with a "*", and doubt is cast on the idea that higher categories correspond to "more" of the variable. |
| Exp. Meas | The expected value of the average measure if these data fit the Rasch model. |
| OUTFIT MnSq | The unweighted mean-square for observations in this category.<br>Mean-squares have expectation of 1.0. Values much larger than 1.0 indicate unexpected observations in this category. Extreme categories have greater opportunity for large mean-squares than central categories.<br>The INFIT MnSq is not reported because it approximates the OUTFIT MnSq when the data are stratified by category. |
| RASCH-ANDRICH THRESHOLDS | Step calibrations, rating scale structure. If these are disordered, see Disordered Rasch-Andrich thresholds. |
| Measure | value of the Rasch-Andrich threshold, the location on the latent variable (relative to the center of the rating scale) where adjacent categories are equally probable. This is the Rasch model parameter. Use this for anchoring rating scale, or for estimation starting-values. |
| S.E. | standard error of the Rasch-Andrich threshold (step calibration).<br>For dichotomous items, no S.E. is reported because there is no estimable threshold parameter. |
| EXPECTATION Measure at | gives the details of the logit-to-expected-score ogive. This is the expected (mean) value of the observations for measures at this point on the latent variable relative to the rating scale. |
| at Category | logit measure for the expected score on the rating scale corresponding to the value in the category score column. Measures corresponding to extreme responses, e.g., (-2.70), correspond to expected responses 0.25 score points from the extreme response, i.e., half way between the extreme response and 0.5 score points. |
| at -0.5 | logit measure for the expected score corresponding to the value in the category score column less 0.5 score points. These can be thought of as the transition points into one expected score from the one below. |

| | |
|---|---|
| MOST PROBABLE from | lowest measure at which this category is the one most probable to be observed. It continues to be the most probable (modal) category until a numerically higher category becomes most probable. |
| "low" | indicates the most probable category at the low end of the scale. |
| "no" | indicates this category is never the most probable to be observed for any measure. |
| RASCH-THURSTONE Thresholds | measure at which the probability of being rated in this category or above equals that of being rated in any of the category below, i.e., is .5., i.e., the 50% (median) cumulative probability threshold. |
| Cat PEAK Prob | The largest percentage probability this category has of being observed at any measure. Extreme categories have a maximum probability of 100% at the extremes of the measurement continuum. Intermediate categories have their peak probabilities when the expected response value is numerically equal to the intermediate category's response value, the "at Category" value. |
| Obsd-Expd Diagnostic Residual | This column is produced only when the difference between the observed count of responses and the expected count, based on the Rasch measures, is greater than 0.5 for some category. This can be due to<br>i) lack of convergence: set smaller values in Convergence=<br>ii) anchor values incompatible with the data<br>iii) responses do not match the specified scale structure, e.g., Poisson counts.<br>iv) contradictory modeling, e.g., models = ?,?,#,#,R6 can imply contradictory estimates for elements. |
| ANDRICH Threshold Displacement | Add this to the Andrich threshold to obtain approximate value with no residual |
| Response Category Name = | name of category from Rating (or partial credit) scale= specification |

**Optimizing Rating-Scale Categorization: When/How to Collapse Categories for Better Measurement**

*Classical Test Theory* says "The categorization with the highest person "test" reliability is the best". We can evaluate this by looking at the Reliability in Facets Table 7 of the person facet. A similar investigation is done at www.rasch.org/rmt/rmt101k.htm

*Rasch Theory* says "Each advancing category of the rating scale corresponds to one higher qualitative level of performance." We can evaluate this by looking at the "Avge Meas" (Average Measure) for the rating scale in Facets Table 8.1. The average measures should advance and be close to their "Exp. Meas" (Expected Measures). Collapse together Average Measures that are disordered or very close together. Also, to avoid accidents in the data biasing results, we want no category to have less than 10 ratings. We also like to see that each category has reasonable fit statistics. If you intend to make inferences at the category level (as opposed to the overall score/measure level) then the Rasch-Andrich Thresholds should also advance.

---

***Unobserved Categories: Structural Zeroes or Incidental (Sampling) Zeroes***
**Structural zero:** A category cannot be observed, and is omitted from qualitative levels. (The default.)
"Category" shows the category number. "Score" shows the value used for analysis.

```
Model = ?,?,R3
+-------------------------------------------------------------------------------------------------------------------+
|         DATA            |  QUALITY CONTROL  |RASCH-ANDRICH|  EXPECTATION  |  MOST  |.5 Cumultv| Cat| Obsd-Expd|
| Category     Counts  Cum.| Avge  Exp.  OUTFIT| THRESHOLDS  | Measure at    |PROBABLE|Probabilty|PEAK|Diagnostic|
|   Score    Used   %    % | Meas  Meas   MnSq |Measure  S.E.|Category  -0.5 |  from  |    at    |Prob| Residual |
|--------------------------+-------------------+-------------+---------------+--------+----------+----+----------|
|  0    0      378  20%  20%|  -.87  -1.03  1.2 |             |( -2.04)       |   low  |   low    |100%|      -.9 |
|  1    1      620  34%  54%|   .13    .33   .7 |  -.85   .07|    .00  -1.17|   -.85 |  -1.00   | 54%|          |
|  2                         |                   |             |               |        |          |    |          |
|  3    2      852  46% 100%|  2.23   2.15  1.5 |   .85   .06|(  2.05)  1.18|    .85 |    .99   |100%|          |
```

```
+-------------------------------------------------------------------(Mean)---------(Modal)--
(Median)-----------------+
```

**Incidental (Sampling) zero**: A category can be observed (but not in this dataset). It is included in the qualitative levels. (Keep.) The "Category Score" is the value used for analysis.

```
Model = ?,?,R3K   <= K means "Keep unobserved intermediate categories"
+-------------------------------------------------------------------------------------------------------------+
|      DATA           |  QUALITY CONTROL  |RASCH-ANDRICH|  EXPECTATION  |  MOST  |.5 Cumultv| Cat| Obsd-Expd|
| Category Counts  Cum.| Avge  Exp.  OUTFIT| THRESHOLDS  | Measure at    |PROBABLE|Probabilty|PEAK|Diagnostic|
|Score   Used   %    % | Meas  Meas   MnSq |Measure  S.E.|Category  -0.5 |  from  |    at    |Prob| Residual |
|---------------------+-----------------+-------------+---------------+--------+----------+----+----------|
|  0      378  20%  20%|  -.68   -.74  1.2 |             |( -1.99)       |   low  |   low    |100%|      1.0 |
|  1      620  34%  54%|  -.11   -.06   .6 |  -.90    .07|   -.23   -1.09|   -.90 |   -.95   | 56%|     -.7  |
|  2        0   0%  54%|                   |             |    .63     .24|        |    .55   |  0%|          |
|  3      852  46% 100%|  1.35   1.34  1.7 |   .90    .07|(  1.50)   1.10|    .90 |    .55   |100%|          |
+-------------------------------------------------------------------(Mean)---------(Modal)--
(Median)-----------------+
```

---

**Equivalence of Facets Table 8 with Winsteps Table 3.2**

Facets Table 8

```
+-------------------------------------------------------------------------------------------------------------
--------------------------+
|      DATA            |  QUALITY CONTROL  |RASCH-ANDRICH|  EXPECTATION  |  MOST  |  RASCH-  |
Cat| Obsd-Expd|Response   |
| Category Counts  Cum.| Avge  Exp.  OUTFIT| Thresholds  |  Measure at   |PROBABLE| THURSTONE|
PEAK|Diagnostic|Category   |
|Score   Used   %    % | Meas  Meas   MnSq |Measure  S.E.|Category  -0.5 |  from  |Thresholds|
Prob| Residual |  Name     |
|---------------------+-----------------+-------------+---------------+--------+----------
+----+----------+-----------|
|  0      378  20%  20%|  -.87  -1.03  1.2 |             |( -2.04)       |   low  |   low    |
100%|      -.9 | dislike   |
|  1      620  34%  54%|   .13    .33   .7 |  -.85    .07|    .00   -1.17|   -.85 |  -1.00   |
54%|          | don't know|
|  2      852  46% 100%|  2.23   2.15  1.5 |   .85    .06|(  2.05)   1.18|    .85 |    .99   |
100%|          | like      |
+-------------------------------------------------------------------(Mean)---------(Modal)--
(Median)----------------------------+
```

Winsteps Table 3.2

```
-------------------------------------------------------------------------
|CATEGORY   OBSERVED|OBSVD SAMPLE|INFIT OUTFIT||STRUCTURE|CATEGORY|
|LABEL SCORE COUNT %|AVRGE EXPECT|  MNSQ  MNSQ||CALIBRATN| 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
-------------------------------------------------------------------------
-------------------------------------------------------------------------------
|CATEGORY    STRUCTURE   |  SCORE-TO-MEASURE   | 50% CUM.| COHERENCE|ESTIM|
| LABEL    MEASURE  S.E. | AT CAT. ----ZONE----|PROBABLTY| M->C C->M|DISCR|
|-----------------------+---------------------+---------+----------+-----|
|   0      NONE         |( -2.07) -INF   -1.19|         |  62%  42%|     | 0 Dislike
|   1        -.86    .07 |    .00  -1.19  1.19|   -1.00 |  54%  71%|  .73| 1 Neutral
|   2         .86    .06 |(  2.07)  1.19 +INF |    1.00 |  85%  78%| 1.19| 2 Like
-------------------------------------------------------------------------------
```

| Winsteps field: | Facets field: |
|---|---|
| CATEGORY LABEL | Category Score |

| | |
|---|---|
| CATEGORY SCORE | Category Score |
| COUNT | Used |
| % | % |
| ... | Cum. % |
| OBSVD AVRGE | Avge Meas |
| SAMPLE EXPECT | Exp. Meas |
| INFIT MNSQ | ... |
| OUTFIT MNSQ | OUTFIT MnSq |
| STRUCTURE CALIBRATN | RASCH-ANDRICH Thresholds Measure |
| *CATEGORY MEASURE | EXPECTATION Measure at Category |
| *STRUCTURE MEASURE | RASCH-ANDRICH Thresholds Measure |
| STRUCTURE S.E. | RASCH-ANDRICH Thresholds S.E. |
| *SCORE-TO-MEASURE AT CAT. | EXPECTATION Measure at Category |
| *SCORE-TO-MEASURE --ZONE-- | EXPECTATION Measure at -0.5 |
| *50% CUM. PROBALTY | RASCH-THURSTONE Thresholds |
| COHERENCE M->C | ... |
| COHERENCE C->M | ... |
| ESTIM DISCR | ... |
| ... | MOST PROBABLE from |
| ... | Cat PEAK Prob |
| OBSERVED-EXPECTED RESIDUAL DIFFERENCE | Obsd-Expd Diagnostic Residual |
| (text to right of table) | Response Category Name |
| * = In Winsteps only, includes item difficulty for Partial Credit model | |

## 11.15 Table 8 Scale structure bar-chart

This gives several pictorial representations of the rating scale (or partial credit).

```
Scale structure for "FACES"
Measr:-3.0       -2.0       -1.0        0.0        1.0        2.0        3.0
        +          +          +          +          +          +          +
  Mode:<0---------(^)----------01--------^-------12-----------(^)---------2>
```

```
Median:<0---------(^)--------01----------^---------12---------(^)---------2>

  Mean:<0---------(^)------01------------^-----------12--------(^)--------2>
       +          +          +          +          +          +          +
Measr:-3.0       -2.0       -1.0        0.0        1.0        2.0        3.0
```

| | |
|---|---|
| Mode: Most Probable Category<br><br>The values are relative to the relevant item difficulty. | The category labels are located at the measures where they commence to be the most probable observations. Omitted categories are never most probable. ^ indicates location of category peak probability. (^) indicates point at which extreme category becomes very probable. |
| Median: Rasch-Thurstone Threshold<br><br>The values are relative to the relevant item difficulty. | The category labels are located at the measures where there is the same probability (0.5) of being rated below that category as there is of being rated in that category or above. Conceptually, each category label is at the left-hand end of its "zone" on the variable. ^ indicate the mid-points of each zone. (^) indicates a conceptual mid-point of the infinitely long extreme zones. |
| Mean: Expected Score<br><br>The values are relative to the relevant item difficulty.<br><br>The mean values are usually more dispersed than the median, and the median than the mode. | The category labels are located at the measures where the expected values become closer to that score than the preceding score, i.e., at the half score-points. ^ indicates the locations where the category score is the expected score. (^) indicates the location at which the expected score is 0.25 score points away from the extreme category score.<br>the "mean" values below Table 8.2 (shown above) are the "scale" values on the right side of Table 6.0, and are the "Expectation" measures in Table 8.1. They answer the question: if we had 1,000 people at a particular measure relative to the item difficulty, what would we expect their average (mean) rating to be? |
| Measr: Measure relative to item difficulty | The scale of relative measures, e.g., persons relative to an item. The local origin is the mean of the Rasch-Andrich thresholds (step calibrations) or as determined by the anchored categories. |

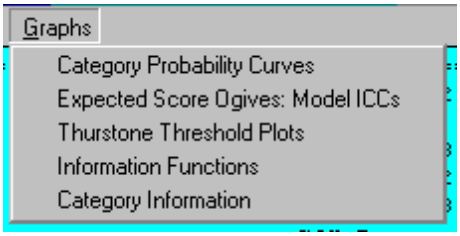
**Dichotomous scale structures are always the same:**

```
Scale structure for dichotomies

Measr:-2.0              -1.0              0.0              1.0              2.0
        +                 +                +                +                +
  Mode:<0---------------------------------(^)---------------------------------0>

Median:<0---------------------------------(^)---------------------------------0>

  Mean:<0---------------------------------(^)---------------------------------0>
       +                 +                +                +                +
Measr:-2.0              -1.0              0.0              1.0              2.0
```

```
Test
```

## 11.16 Table 8 Scale structure probability curves

Use the "Graphs" pull down menu to obtain these as bit-mapped plots.




Specify "Graphfile=" to obtain the numbers to plot these yourself.

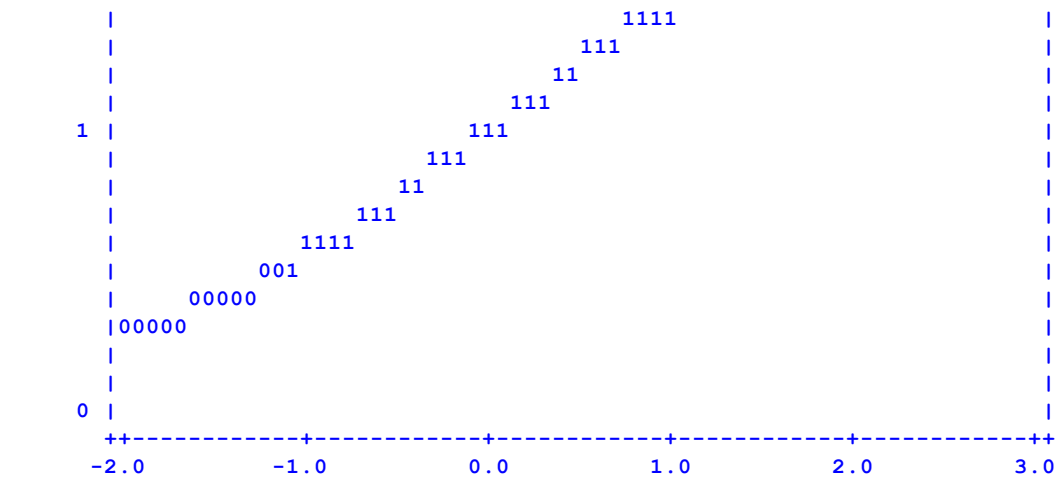
Table 8 includes a line-printer plot of the probability of occurrence for each category. The probability of the extreme categories always approaches 1.0 for corresponding extreme measures.

Most scale developers intend this to look like a series of hills, so that each category is in turn the most likely category along the latent variable.

```
Probability Curves
      -3.0       -2.0       -1.0        0.0        1.0        2.0        3.0
       ++----------+----------+----------+----------+----------+----------++
     1 |                                                                    |
       |                                                                    |
       |00                                                                22|
       |  00000                                                     22222   |
       |       000                                                222        |
     P |          000                                           222          |
     r |             00                                       22             |
     o |               00                                   22               |
     b |                 000                              222                 |
     a |                    00        1111111        22                     |
     b |                      0  11111       11111  2                       |
     i |                      1**                 **1                       |
     l |                   111   00             22   111                    |
     i |                 11        00         22        11                  |
     t |              111            00     22            111               |
     y |           111                 00*22                 111            |
       |       1111                    22 00                   1111        |
       |   1111                     222     000                    1111    |
       |111                    22222           00000                    111|
       |               22222222                     00000000               |
     0 |222222222222222                                   000000000000000|
       ++----------+----------+----------+----------+----------+----------++
      -3.0       -2.0       -1.0        0.0        1.0        2.0        3.0
```

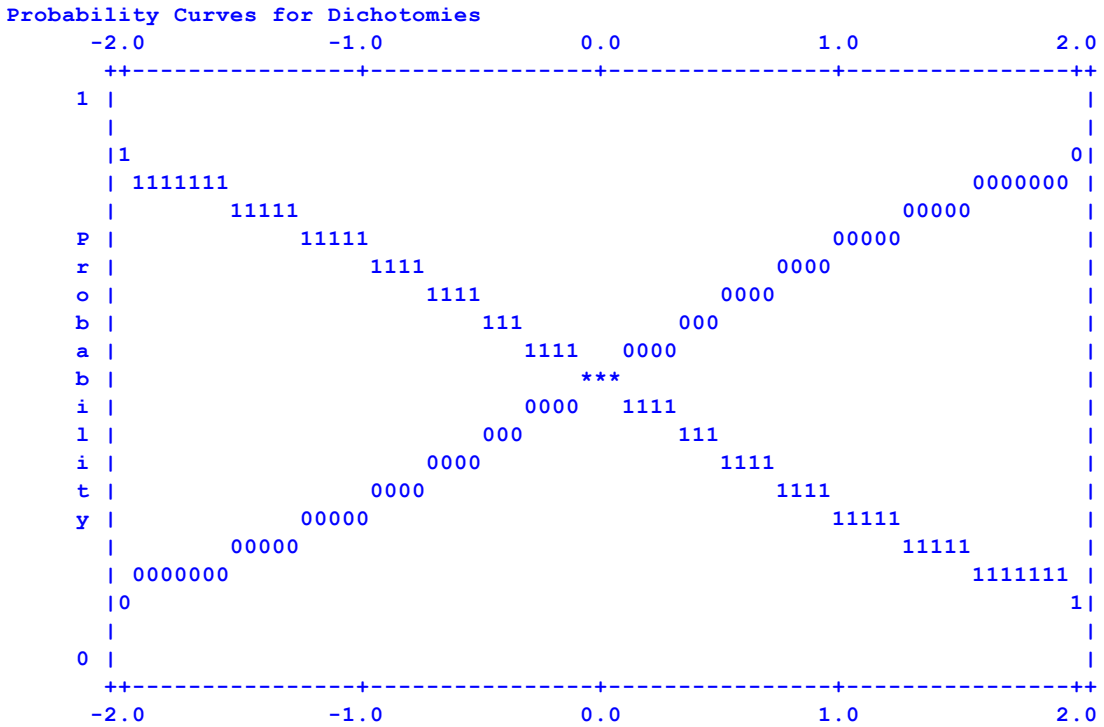
The expected score ogive shows the average rating value expected for any measure relative to the item, judge etc. The ogive also indicates the category "zone", e.g., the area between average ratings of 0.5 and 1.5 is marked by 1's plotted on the curve.

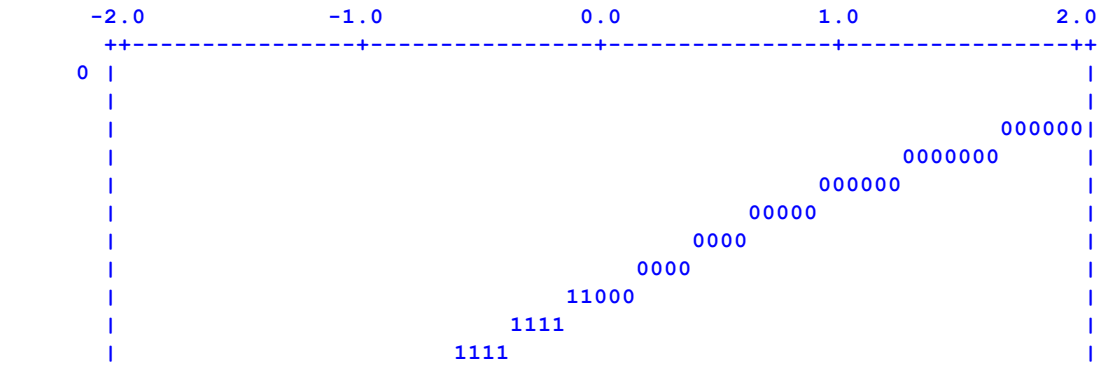
```
Expected Score Ogive
      -2.0         -1.0          0.0          1.0          2.0          3.0
       ++------------+------------+------------+------------+------------++
     2 |                                                                  |
       |                                                           222222|
       |                                                     2222222      |
       |                                                22222             |
       |                                           22222                  |
       |                                        122                       |
```

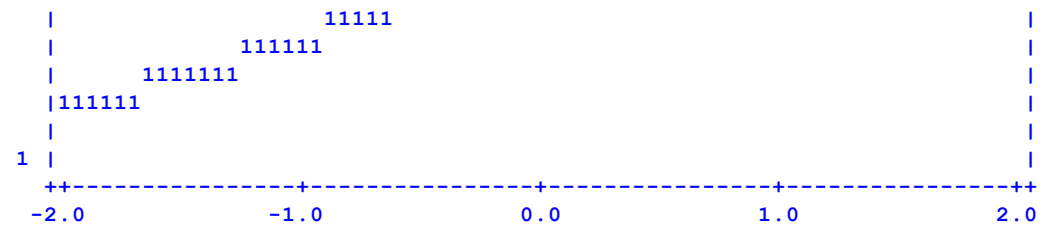



Dichotomous plots always look the same:

**Dichotomous scale structures are always the same:**






## 11.17 Table 9 is the iteration report for Bias estimation

Bias direction=, Write= and "B" in Models= statements control these Tables of bias interaction designators, "B", have been included in the measurement models, then bias estimates will be made. All estimates reported in the main analysis are fixed (anchored). Then Facets estimates the size of bias measures from the residuals of the main analysis. This process is repeated as many times as there are models with bias designators.

For each model specifying bias designators, a set of tables is produced, which estimate the designated bias as it is represented in all the modeled data (not just for the model containing the bias designators).

```
Table 9.1  Bias Iteration Report.

Bias/Interaction: 1. Senior scientists, 2. Junior Scientists

There are empirically 21 Bias terms
+-----------------------------------------------------------+
| Iteration      Max. Score Residual      Max. Logit Change |
|             Elements    %  Categories   Elements    Cells |
|-----------------------------------------------------------|
| BIAS   1    -12.9867 -32.5               -1.1329       21 |
| BIAS   2     -1.1019  -2.8                -.1322       18 |
| BIAS   3       .0164    .0                -.0021        4 |
| BIAS   4       .0001    .0                 .0000        0 |
+-----------------------------------------------------------+
```

| | |
|---|---|
| Table 9 | Report on the estimation of the size of the bias/interaction terms |
| 9.1 | the sequential number of the model specification that produced this set of Bias/Interaction Tables. |
| Bias/Interaction analysis specified by Model: | This shows the model statement. |
| There are empirically 21 Bias terms | the number of bias interaction terms for this model in the data. Suppose we have two facets selected for bias analysis, "B": items and ethnicities. Then *Facets* computes a bias size for every observed different combination of an item and an ethnicity. The maximum number of bias terms is (number of item elements) * (number of ethnicity elements). Empirically, not all these combinations may be observed, so the empirical number can be less. |
| Iteration = | how many times the data has been read to estimate bias terms. |
| Max. Score Residual: | the maximum (biggest) difference between an observed raw score and an expected score. 1.0 is the smallest observable difference. The convergence criterion is the minimum of Converge= and .01 score-points. |
| Elements | is the largest difference for any bias term. |
| Categories | (not relevant for bias terms) |

| | |
|---|---|
| Max. Logit Change: | is the largest change, in logits, between any bias estimate in this iteration and its estimate in the previous iteration. Starting estimates are 0.0 logits. The convergence criterion is the minimum of Converge= and .001 logits. |
| Elements | is the largest change for any bias term. |
| Cells | The number of Bias terms yet to be calculated to the specified convergence criteria |

Both Max. Score Residual and Max. Logit Change should steadily reduce in absolute size, i.e., draw closer to zero. You may force termination by pressing the Ctrl+"F" keys simultaneously.

The detailed iteration report on your screen can be recorded in your output file with the "Write=Yes" specification.

## 11.18 Table 10 lists responses, still unexpected after allowing for bias

This Table is not produced under usual circumstances. The measurement is over-parameterized at this point, so response residual computations can be misleading.

To examine response residuals after adjusting for bias, model the interaction term as the main effect and do not model the main effects, i.e.,
Usual model: $\log(P_{nik}/P_{ni(k-1)}) = B_n - D_i - F_k$
and then compute interaction $B_{ni}$
Instead model: $\log(P_{nik}/P_{ni(k-1)}) = B_{ni} - F_k$

## 11.19 Table 11 shows how many responses were used to calculate bias estimates

Table 11 summarizes the response modeling in the bias analysis

```
Bias/Interaction: 1. Student, 2. Format
+-------------------------------------------------+
| Cat  Score  Exp.  Resd StRes|                    |
|-----------------------------+--------------------|
| 9.60  9.60  9.60   .01  .01 | Mean (Cnt: 296)    |
| 1.34  1.34  1.30   .31  .53 | S.D. (Population)  |
| 1.34  1.34  1.30   .31  .53 | S.D. (Sample)      |
+-------------------------------------------------+
Count of measurable responses = 296.00
Raw-score variance of observations      = 1.80 100.00%
Variance explained by Rasch measures    = 1.53  85.00%
Variance of residuals                   = 0.27  15.00%
Variance explained by bias/interactions = 0.18  10.00%
Variance remaining in residuals         = 0.09   5.00%
```

| Identification | Meaning |
|---|---|
| Cat | Observed value of the category in the data file |
| Score | Value of category in this analysis. This is less than Cat. if there are structural zeroes. |
| Exp. | Expected score based on current measures |
| Resd | Residual, the score difference between Score and Exp. |
| StRes | The residual divided by its standard error. The Rasch model is now over-parameterized, so the standardized residuals are |

| | |
|---|---|
| | expected to have an S.D. less than 1.0. |
| Mean | average of the observations |
| Cnt: | Count is the number of observations |
| S.D. (Population) | standard deviation treating this sample as the entire population |
| S.D. (Sample) | standard deviation treating this sample as a sample from the population. It is larger than S.D. (Population). |
| Count of measurable responses | the number of ratings used in estimating the bias terms. This may be only a small fraction of the entire data set. |
| Raw-score variance of observations | square of S.D. (Population) of Score in Table 11 |
| Variance explained by Rasch measures | from Table 5 |
| Variance of residuals | from Table 5 |
| Variance explained by bias/interactions | Variance of Residuals - Variance remaining in residuals |
| Variance remaining in residuals | square of S.D. (Population) of Resd in Table 11 |

## 11.20 Table 12 is the Bias Summary Report

Zscore=, Bias= and Xtreme= control Table 12. One table is produced for each bias estimation. Each entry in the table is a bar-chart showing the distribution of reported statistics. The scale is from left to right. "M" represents the mean value of the statistic for the facet, "S" is one sample standard deviation each side of the mean, "Q" is two sample standard deviations, and "X" is three.

Table 12 reports the same statistics reported numerically in the later Table 13. So, -4, -3 etc are the numerical values of those statistics (compare with the numerical Table 13). M is the mean, S is one standard deviation away, Q is two standard deviations away.
Bias/Interaction Size = Bias Size (usually in logits) in Table 13
Bias/Interaction Significance = *t*-value in Table 13

Example:

```
Bias/Interaction Size:
                                1
                 12213457677888808866553422222
  2  21 1122472485329801372717797531096665273067453 2221111 1   1
 +---------+----Q----+-S-------M------S--+----Q----+---------+---------+
-3        -2        -1         0         1         2         3      4

Bias/Interaction Significance:
                             11111
                   1213358889100108775433222
     11 1 1 11276444773675833557684234395094893331222111  1     1
 +-----+-----+----Q+-----S-----M-----S-----+Q----+-----+-----+-----+
-5    -4    -3    -2    -1     0     1     2     3     4     5     6
```

The numbers are frequencies. They are printed vertically. The biggest number for "Bias/Interaction size" is 107. 107 interactions are about 0.1 logits in size. 115 interactions have a bias/interaction significance (t) of about -0.3

The frequencies summarize the full Table of interactions (Table 14) and are useful for seeing the distributional pattern.

In these pictures there are clearly a few outliers needing special attention, but the great majority of interactions are statistically and substantively close to 0.

## 11.21 Table 13 is the bias report

This Table compares the local (biased?) measure of one element with its measure from the main analysis. Zscore=, Bias=, Xtreme=, Arrange= and Juxtapose= control this Table.

```
+-------------------------------------------------------------------------------------------------------
        -------------------------+
|Observd  Expctd  Observd  Obs-Exp|  Bias  Model                     |Infit Outfit|    Senior
        scientists  Junior Scientis |
|  Score   Score    Count  Average|  Size   S.E.     t   d.f. Prob. | MnSq  MnSq | Sq N Senior
        sc  measr N Junior  measr |
|--------------------------------+--------------------------------+------------
        +--------------------------------------|
|    4.2     7.91     1.5    -2.47|  -1.00   .61  -1.63     1 .3496 |   .6    .6 | 14 2 Brahe
             .21 5 Edward    .34 |
|--------------------------------+--------------------------------+------------
        +--------------------------------------|
|    7.3     7.25     1.5      .01|    .01   .56    .03             |   .6    .6 | Mean (Count:
        21)                       |
|    2.1     1.25      .0      .98|    .42   .03    .72             |   .4    .4 | S.D.
        (Population)               |
|    2.1     1.28      .0     1.00|    .43   .03    .73             |   .4    .4 | S.D.
        (Sample)                   |
+-------------------------------------------------------------------------------------------------------
        -------------------------+
Fixed (all = 0) chi-squared: 10.8  d.f.: 21  significance (probability): .97
-------------------------------------------------------------------------------------------------------
        -------------------------
```

The column headings mean:

Observd Score = raw score of the estimable responses involving these elements simultaneously, as observed in the data file.

Expctd Score = expected score based on the measures from the main analysis.

Observd Count = number of estimable responses involving these elements simultaneously.

Obs-Exp Average = observed score less the expected score divided by observed count, the bias in terms of the response metric. For rater behavior, look at the "Obs-Exp Average". If this is positive, then the rater is more lenient than expected in this situation. If this is negative, then the rater is more severe than expected.

Bias Size = Size of bias measure in log-odds units, logits, relative to overall measures. Only large or significant biases are listed greater than Zscore=. For clarification, compare the ranking of the Obs-Exp Average with that of the Bias Size. In this case, larger observed scores correspond to higher Bias sizes, i.e., higher abilities, higher leniencies, higher easiness. The sign of the report bias is controlled by Bias=. ">" indicates an extreme low score, "<" an extreme high score

For (measure+bias), add Bias Size to the element "measr", or subtract Bias Size from "measr". Addition for persons that are locally more able. Subtraction for items that are locally more difficult. Look at the "Obs-Exp Average".

Model Error = standard error of the bias estimate.

t = Student's t-statistic testing the hypothesis "There is no bias apart from measurement error". The "Obsvd Count"-2 approximates the degrees of freedom of the t-statistic. With many observations, the t-statistic approximates a normal distribution with mean = 0, S.D. = 1, i.e., a z-score. The t-statistic is the report of a test of the statistical significance of the size of the bias. In this table, statistically significant bias indicates that the difference between the element measure for this interaction and the overall element measure is greater than the difference we would expect to see by chance.

Infit MnSq and Outfit MnSq = Does the bias explain all the misfit reported in Table 7 or is there also another source of misfit? Values are expected to be less than 1.0 because the bias is explaining some of the overall misfit. These statistics do **not** report the fit of the bias terms. In effect, we are deliberately over-parameterizing the

statistical model. Consequently we expect the mean-squares to be less than 1.0 (by an unknown amount). The reported mean-squares indicate how much of the misfit in Table 7 remains after the interactions are estimated. The reported mean-squares do not have the usual statistical properties of mean-squares (chi-squares) and so their statistical significance (Zstd) is unknown. The purpose of the mean-square fit statistics is to help you determine whether the misfit in the data is explained by the bias or is due to other causes. When there are choices for bias, such as raters and items, or raters and tasks, etc., this may help in choosing which bias to focus on: the one with the lowest mean-squares.

For each facet entering into the bias calculation:
Sq = a sequence number used to reference the bias term - useful for referring to a specific line in this Table.
N = element number with Facet. Only elements with non-extreme measures are included.
Senior Sc = Name of facet: elements listed below
measr = Measure of element from main analysis.

In the summary statistics,
Count = the number of modeled bias terms found in the data.
S.D. (Population) = the standard deviation if this sample is the whole population
S.D. (Sample) = the standard deviation if this sample is a random sample from the whole population
Fixed (all=0) chi-squared = A test of the "fixed effect" hypothesis: "Can this set of interactions be regarded as sharing the same measure of approximately 0.0 after allowing for measurement error?" The chi-squared value and degrees of freedom (d.f.) are shown. The significance is the probability that this "fixed" hypothesis is the case. This is not a test of "Can these interactions be disregarded?" Individual interactions may be large and significant. For instance, one bad tire in a production run of 1000 tires may not indicate a "statistically significant" problem in the production process, but I still don't want it on my car!

---

The chi-squares and t-tests evaluate hypothesis tests.

Think of driving along the road.

*A.) Is this road surface generally OK?*

Answer: `Fixed (all = 0) chi-squared: 20  d.f.: 21  significance (probability): .5`

So we can't reject the hypothesis that the road surface is generally OK.

*B.) Are there any pot-holes we should avoid (even if the road is generally OK)?*

Answer: the t-test Prob. for Brahe-David p=.02. This is a "pot-hole"!

---

Example using earlier Facets output:

Table 4. Bias interaction analysis by model.

Applying Rasch Measurement to Performance Assessment in Higher Education 10–11–2006 16:16:29
Table 13.1.1 Bias/Interaction Calibration Report (arranged by 0fN).
Bias/Interaction analysis specified by Model: ?B, ?B, ?, RATINGS

| Obsvd Score | Exp. Score | Obsvd Count | Obs-Exp Average | Bias+ Measure | Model S.E. | Z-Score | Infit MnSq | Outfit MnSq | Sq | N | Groups | measr | N | Jud | measr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 40 | 33.6 | 9 | .71 | −.70 | .34 | −2.04 | 1.2 | 1.2 | 21 | 3 | Group 3 | .49 | 4 | J4 | .02 |
| 28.8 | 28.8 | 9.0 | .00 | −.01 | .31 | .00 | 0.9 | 0.9 | Mean | (Count: 42) | | | | | |
| 5.8 | 4.9 | 0.0 | .35 | .31 | .02 | .97 | 0.9 | 0.3 | S.D. | | | | | | |

Fixed (all = 0) chi-square: 39.9 d.f.: 42 significance: .56

The observed score is 40. The expected score is 33.6. There are 9 observations. So, on average, Group 3 is being rated 0.71 score-points higher than expected by Judge J4. This corresponds to a change in judge severity of -.70 logits = less severe. The standard error is .34 logits. So the Z-score (t-test with infinite d.f.) is -2.04. The rater is significantly less severe (more lenient) at the .05 level (double-sided). If we want to report this as a change in group ability of +.71 logits, set Bias=Positive.

---

3-Way Interaction:

Model=
?,?B,?B,?,R9 ; 2-way interaction - also from Output Tables menu
?,?B,?B,?B,R9 ; 3-way interaction - only in Specification file: Models=
*

Table 13.2.1 Bias/Interaction Report (arranged by MN).
Bias/Interaction: 2. Essay, 3. Reader, 4. Session (higher score = higher bias measure)

```
+-------------------------------------------------------------------------------------------------------
-------------------------------------+
|Observd  Expctd  Observd  Obs-Exp| Bias+  Model                    |Infit Outfit|     Essay
   Reader       Session               |
|  Score   Score    Count  Average|  Size   S.E.     t   d.f. Prob. | MnSq  MnSq | Sq  N E
measr- Nu Re measr- Nu Session      measr- |
|---------------------------------+--------------------------------+------------
+-----------------------------------------------|
|   10      18.13     4      -2.03|  -1.29   .46  -2.79     3 .0687 |   .1    .1 |  70 1 A   -
.01 12 12   -.04 12 day 1 time 2    .00 |
|   10      17.52     4      -1.88|  -1.21   .46  -2.62     3 .0790 |   .4    .4 |  17 2 B
.08  6 6     .11 11 day 1 time 1    .00 |
```

---

**Measure of an Element of one Facet on an Element of Another Facet - such as Test-taker and Criterion**

Produce Facets Table 13 from the Facets Output Tables menu: Test-takers by Criteria. For each combination of test-taker and Criterion, this shows the observed average rating and the bias size. You may be reporting bias in a positive or negative direction:

When the observed score is higher than the expected score, is the bias size positive?
If yes, add the bias size to the test-taker measure (at the right of the table) for the test-taker's measure on the criterion.
If no, subtract Add the bias size from the test-taker measure (at the right of the table) for the test-taker's measure on the criterion.

You can split Table 13 into columns by pasting it into Excel, and using Excel's "Data", "Text to columns" feature.

## 11.22 Table 14 is the pairwise bias report

This Table compares the local (biased?) relative measure of one element with the local relative measure of another element. Table 14 can be output from your main analysis or from the "Output Tables" menu.

If no comparisons can be made, then "`Table 14: ... - No pairwise terms`" is displayed in the Analysis window and output to Table 14.

Zscore=, Xtreme= and Arrange= control this Table. This Table presents the bias/interaction information in a pairwise format. It has the same information as Table 13. Table 14 contains several subtables, each containing the same

information, but from different perspectives. Depending on how the Tables are conceptualized, they quantify item bias or differential item functioning, DIF, differential person functioning, DPF, differential rater functioning, DRF, differential task functioning, DTF, etc. Item bias and DIF are the same thing. The widespread use of the term "item bias" dates to the 1960's, the term "DIF" to the 1980's.

Here are two subtables. Brahe, a judge, has given Brahe rated Edward relatively low and Cavendish relatively high:

```
Table 14.3.1.1  Bias/Interaction Pairwise Report
-------------------------------------------------------------------------------------------------------
       ----------------
| Target    | Target     Obs-Exp Context     | Target     Obs-Exp Context     |  Target Joint
        Rasch-Welch    |
| N Junior | Measr S.E. Average N Senior sc | Measr S.E. Average N Senior sc |Contrast  S.E.
       t    d.f. Prob. |
-------------------------------------------------------------------------------------------------------
       ----------------
| 4 David  |   .25  .29    1.54 2 Brahe     | -1.05  .35    -.46 3 Cavendish |   1.30   .45
       2.86     8 .0211 |
```

David, the target examinee is rated 1.54 score points high (.25 logits) by Brahe, and .46 score points low (-1.05 logits) by Cavendish. So David's perceived difference in performance (.25 - -1.05) is 1.30 logits, which has a p=.02 probability of happening by chance.

```
-------------------------------------------------------------------------------------------------------
       ----------------
| Target    | Target     Obs-Exp Context     | Target     Obs-Exp Context     |  Target Joint
        Rasch-Welch    |
| N Junior | Measr S.E. Average N Senior sc | Measr S.E. Average N Senior sc |Contrast  S.E.
       t    d.f. Prob. |
-------------------------------------------------------------------------------------------------------
       ----------------
| 5 Edward |  -.84  .36   -2.60 2 Brahe     |  1.11  .36    2.00 3 Cavendish |  -1.96   .51 -
       3.85     8 .0049 |
```

Edward, the target examinee is rated 2.60 score points low (-.84 logits) by Brahe, and 2.00 score points high (1.11 logits) by Cavendish. So Edward's perceived difference in performance (-.84 - 1.11) is -1.96 logits, which has a p=.00 probability of happening by chance.

Target = Element for which the bias or interaction is to be compared in two contexts
 N = Element number. Only elements with non-extreme measures are included.
 Facet name heading with element name beneath
In one Context:
 Target Measr = local measure of element in this context (includes bias) = overall measure for target (Table 7 or Table 13) + bias (Table 13). ">" indicates an extreme low score, "<" an extreme high score
 Target S.E. = precision of local ability estimate
 Obs-Exp Average = the average difference between the observed and the expected (no bias) ratings for the Target element in this context.
 Context
  N = Element number. Only elements with non-extreme measures are included.
  Facet name heading with element name beneath
In the other Context:
 Target Measr = local measure of element in this context (includes bias) = overall measure for target (Table 7 or Table 13) + bias (Table 13). ">" indicates an extreme low score, "<" an extreme high score
 Target S.E. = precision of local ability estimate
 Obs-Exp Average = the average difference between the observed and the expected (no bias) ratings for the Target element in this context.
 Context
  N = Element number. Only elements with non-extreme measures are included.

Facet name heading with element name beneath
Target Contrast = difference between the Target Measures in the two Contexts
Joint S.E. = standard error of the difference
t = Welch's corrected version of Student's t-statistic of Contrast / S.E.
*Cohen's d* ≈ 2 * t / √ (d.f.)
d.f. = degrees of freedom of t-statistic (approximate)
Prob. = probability of t-statistic assuming t, d.f. are exact.

The individual interactions are summarized in Table 13, and the observations are summarized in Table 11.

Table 14 in Facets shows pairwise interactions. When there are 3 groups, A, B, C, then the pairings are reported as AB, AC, BC. But they could also be reported as AB, AC, CB or in many other ways. This makes any mean and S.D for the columns of Table 14 dependent on which of the two groups is placed in the left-hand column, and which is in the right-hand column.

Here, the judges are the Contexts. They compare their perceptions of the Targets, the examinees, i.e., the bias is interpreted as Differential Person Functioning, DPF. *The first line reads:* Brahe (a judge, the Context) perceives David (an examinee, the Target) to have a local ability measure of .25 logits (= -.46 David's overall measure from Table 7 + .71 David's local bias size from Table 13) with a precision of .29 logits, corresponding to David performing 1.54 score points per observation better than expected. Cavendish (a judge, the Context) perceives David (an examinee, the Target) to have an ability measure of -1.05 logits, performing .46 score points per observation worse than expected. The ability difference is 1.30 logits. Statistically, this difference has a *t* of 2.86 with 8 d.f., i.e., p=.02 for a two-sided t-test.

**Table 14.3.1.2 Bias/Interaction Pairwise Report**

| Target | Target | Obs-Exp | Context | Target | Obs-Exp | Context | Target | Joint | Rasch-Welch |
|---|---|---|---|---|---|---|---|---|---|
| N | Senior sc | Measr | S.E. | Average | N | Junior | Measr | S.E. | Average | N | Junior | Contrast | S.E. | t | d.f. | Prob. |
| 2 | Brahe | -.48 | .29 | 1.54 | 4 | David | 1.50 | .36 | -.66 | 5 | Edward | -1.98 | .46 | -4.28 | 8 | .0027 |
| 3 | Cavendish | .49 | .35 | -1.13 | 4 | David | -.78 | .36 | 3.27 | 5 | Edward | 1.27 | .50 | 2.55 | 8 | .0341 |

Here, the examinees compare their perceptions of the judges, i.e., the bias is interpreted as Differential Rater Functioning, DRF. *The first line reads:* David (an examinee, the Context) perceives Brahe (a judge, the Target) to have a severity of -.48 logits (= .24 Brahe's overall severity in Table 7, - .71 Brahe's local leniency bias size from Table 13) . Edward (an examinee, the Context) perceives Brahe (a judge, the Target) to have a severity measure of 1.50 logits. The difference is -1.98 logits. Statistically, this difference has a *t* of -4.28 with 8 d.f., i.e., p<.01 for a two-sided t-test.

**Table 13 Bias/Interaction Report**

| Obsvd | Exp. | Obsvd | Obs-Exp | Bias | Model | | Infit | Outfit | Sq | N | Senior sc | measr | N | Junior | measr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Score | Score | Count | Average | Size | S.E. | t | MnSq | MnSq | | | | | | | |
| 25 | 17.3 | 5 | 1.54 | .71 | .29 | 2.43 | .3 | .3 | 11 | 2 | Brahe | .24 | 4 | David | -.46 |
| 15 | 20.6 | 5 | -1.13 | -.58 | .35 | -1.69 | .5 | .5 | 12 | 3 | Cavendish | -.09 | 4 | David | -.46 |

Interpretation:
Observations by Brahe for David are 1.54 score points higher than expected = .71 logits more able
Observations by Cavendish for David are 1.13 score point slower than expected = -.58 logits less able.
Overall pairwise ability swing = .71 - -.58 = 1.29 logits.

```
Table 14.3.1.1  Bias/Interaction Pairwise Report
-------------------------------------------------------------------------------------------------------
       ----------------
| Target    | Target     Obs-Exp Context     | Target     Obs-Exp Context     |  Target Joint
        Rasch-Welch    |
| N Junior | Measr S.E. Average N Senior sc | Measr S.E. Average N Senior sc |Contrast  S.E.
       t   d.f. Prob. |
-------------------------------------------------------------------------------------------------------
       ----------------
| 4 David  |   .25  .29    1.54 2 Brahe     | -1.05  .35    -.46 3 Cavendish |   1.30   .45
       2.86     8 .0211 |
```

Interpretation:
For David, the observations by Brahe are 1.54 higher than expected. This corresponds to Brahe (context) perceiving David (target) to have an ability of "David + Bias" = -.46 + .71 = .25
For David, the observations by Cavendish are .46 lower than expected. This corresponds to Cavendish (context) perceiving David (target) to have an ability of "David + Bias" = -.46 + -.58 = -1.04 = -1.05 (due to rounding)
Overall pairwise ability swing = .25 - -1.05 = 1.30 logits

---

**Example:** When "higher score = higher measure" and Bias = Ability
Target means "apply the bias to the measure of this element"
Context means "the bias observed when the target element interacts with the context element"

```
Obsvd  Exp.  Bias    Model
Score   Score Measure S.E.  Context  measr  Items    measr
186       177     .19  .14   red      -.61  tulip     -.42
234       243    -.15  .13   green     .61  tulip     -.42

Bias                                  | target = tulip
Measure Context  measr  Items    measr | bias   overall        Context
0.19    red       -.61  tulip     -.42 |   .19 + -.42 = -.23 relative to red
-.15    green      .61  tulip     -.42 |  -.15 + -.42 = -.57 relative to green
```

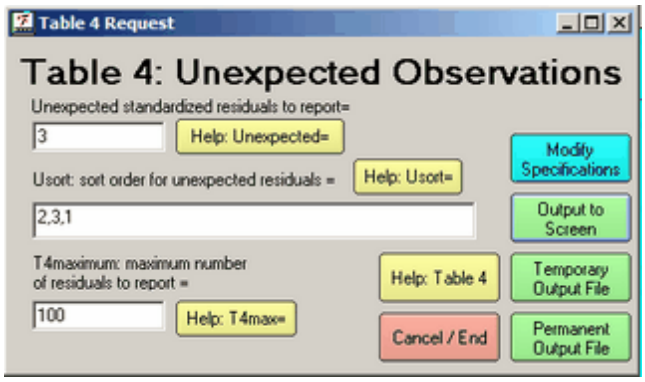
**Effect of imprecision in element estimates**
This computation treats the element measures as point estimates (i.e., exact). You can inflate the reported standard errors to allow for the imprecision in those measures. Formula 29 of Wright and Panchapakesan (1969), www.rasch.org/memo46.htm, applies. 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 S.E.s reported in this Table by sqrt(1.25) = 1.12 (and divide the t by 1.12), then you will be as conservative as possible in computing the bias significance.

**Multiple-comparisons in Table 14.** Here we have to be exact about the hypothesis we are testing. If the (global) hypothesis is "For every pair, the measures are the same", then we need a Bonferroni (or similar) correction. If the (local) hypothesis is "For this pair, the measures are the same", then no correction is needed.
But an easier global test is the "Fixed chi-squared" test at the bottom of Table 13. This tests the global hypothesis explicitly with one operation, rather than indirectly with many corrected t-tests.

# 12 Output Tables from menu

## 12.1 Table 4 from menu

From the Output Tables menu, select *Table 4* to output Table 4.
Table 4 reports Unexpected Responses




## 12.2 Table 5 from menu

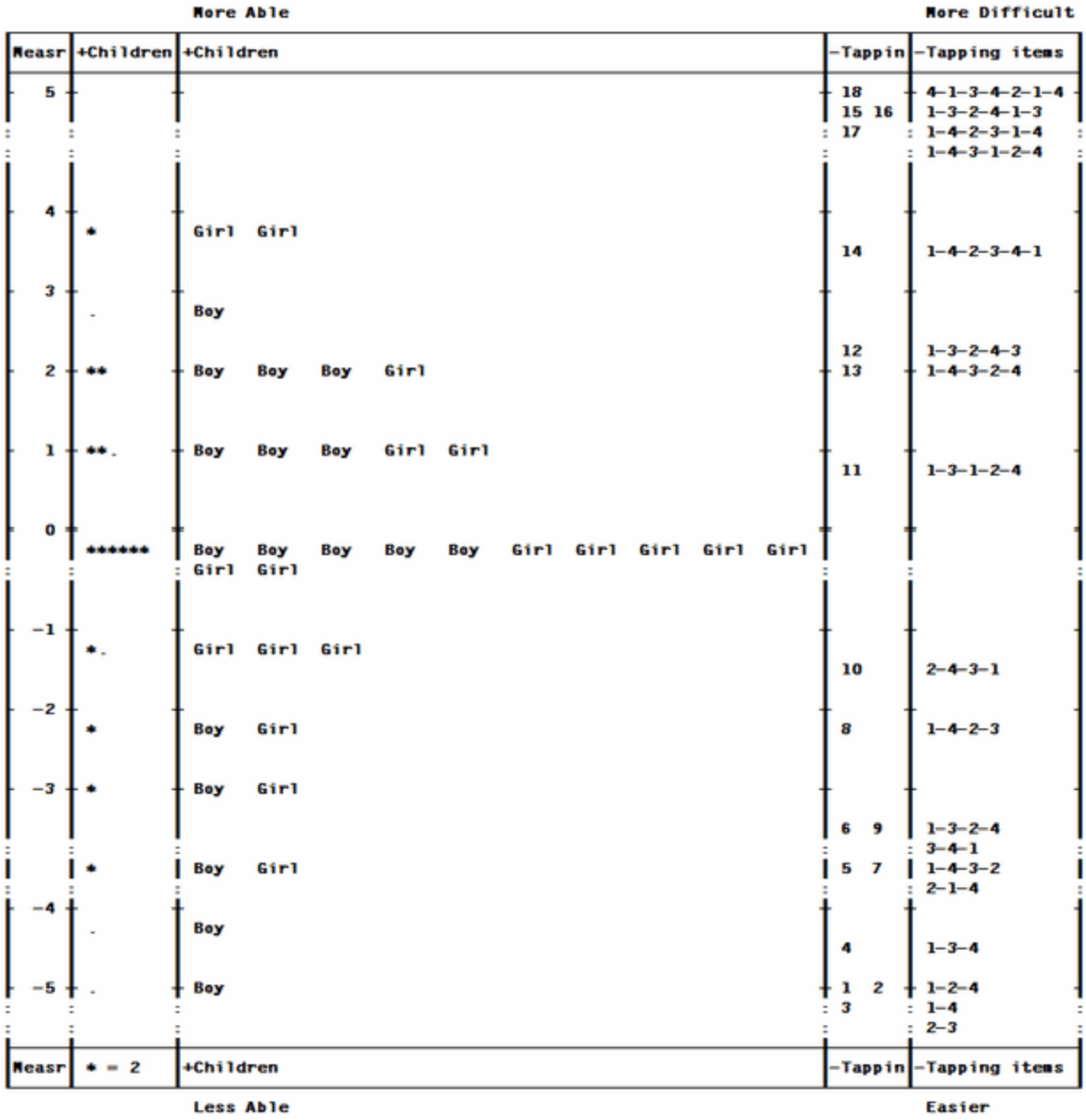
From the Output Tables menu, select *Table 5* to output Table 5.
Table 5 reports Measurable Data Summary

## 12.3 Table 6 from menu

From the Output Tables menu, select *Table 6* to output Table 6.
Table 6.0 is the All facet summary - "rulers"
Table 6.0.0 is the Disjoint element listing
Table 6.2 Graphical description of facet statistics

Table 6.0 All Facet Vertical "Rulers".

Vertical = (1*,1A,2N,2A,S) Yardstick (columns lines low high extreme)= 112,4,-5,5,End




For more details about the specifications: Vertical=, Yardstick=, Barcharts=, Show unobserved=, Subsets=

| Input Field: | Specifications: |
|---|---|
| Columns are:<br>Table 6 too wide? Omit columns, or replace labels with distributions. | Vertical= |
| Size is: (width in columns, height in lines per logit, lowest measure, highest measure)<br>Table 6 too long? Reduce lines per logit. | Yardstick= |
| Placement of extreme scores | Yardstick= |

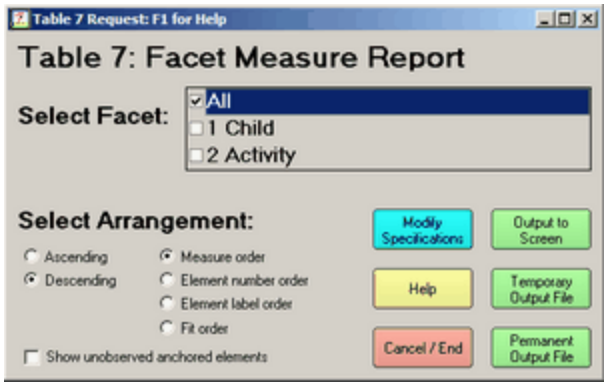
| | |
|---|---|
| Output summary barcharts | Barcharts= |
| All | Barcharts= for all barcharts |
| Show unobserved anchored elements | Omit-unobserved= |
| Subsets reported (shown if there are subsets) | Subsets= R for subset reporting |

Example of Table 6.0:

```
Vertical = (1A,2A,3A,S) Yardstick (columns lines low high extreme)= 0,10,-1,1,End
+-------------------------------------------------------------------------+
|Measr|-Senior scientists|+Junior Scientists|-Traits                |CREAT|
|-----+------------------+------------------+-----------------------+-----|
|   1 +                  +                  +                       + (9) |
|     |                  |                  |                       |  7  |
|     |                  |                  |                       |     |
|     |                  |                  |                       | --- |
|     |                  | Betty            |                       |     |
|     |                  |                  | Enthusiasm            |     |
|     |                  | Edward           |                       |  6  |
|     |                  | George           |                       |     |
|     | Brahe            |                  | Clarity               | --- |
|     |                  |                  |                       |     |
*   0 * Avogadro         *                  *                       *  5  *
|     | Cavendish        | Anne             | Basis                 |     |
|     |                  | Chris            |                       | --- |
|     |                  |                  | Attack      Daring    |     |
|     |                  |                  |                       |  4  |
|     |                  | David            |                       |     |
|     |                  | Fred             |                       |     |
|     |                  |                  |                       | --- |
|     |                  |                  |                       |     |
|     |                  |                  |                       |  3  |
|  -1 +                  +                  +                       + (1) |
|-----+------------------+------------------+-----------------------+-----|
|Measr|-Senior scientists|+Junior Scientists|-Traits                |CREAT|
+-------------------------------------------------------------------------+
```

## 12.4 Table 7 from menu

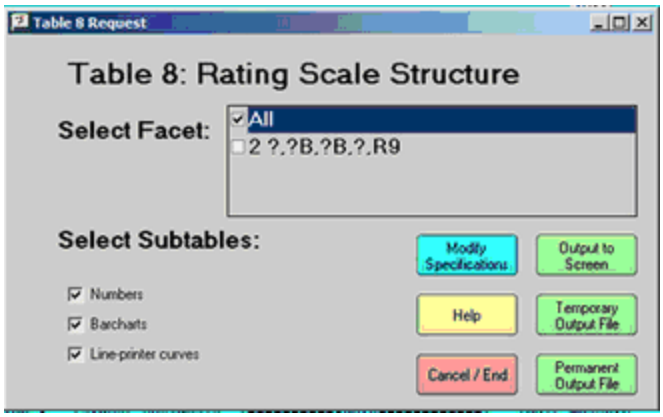
From the Output Tables menu, select *Table 7* to output Table 7.




For more details, Arrange=, Show unobserved=, Modify Specifications box

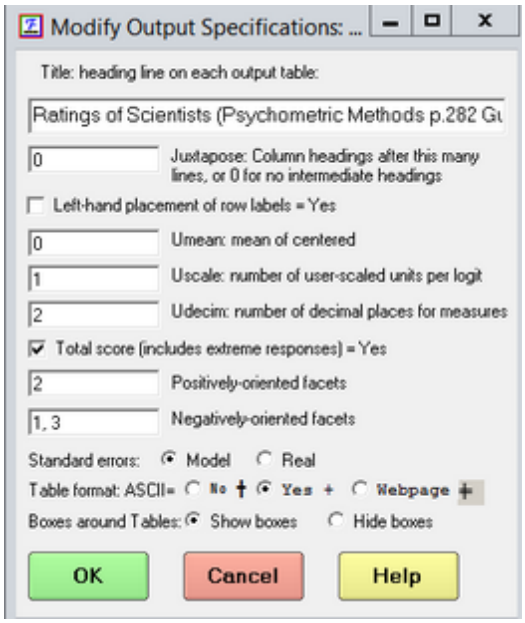
## 12.5 Table 8 from menu

From the Output Tables menu, select *Table 8* to output Table 8.
Table 8.1 Dichotomy, binomial trial and Poisson statistics
Table 8.1 Rating scale (or partial credit) statistics
Table 8 Scale structure bar-chart
Table 8 Scale structure probability curves




## 12.6 Modify output specifications

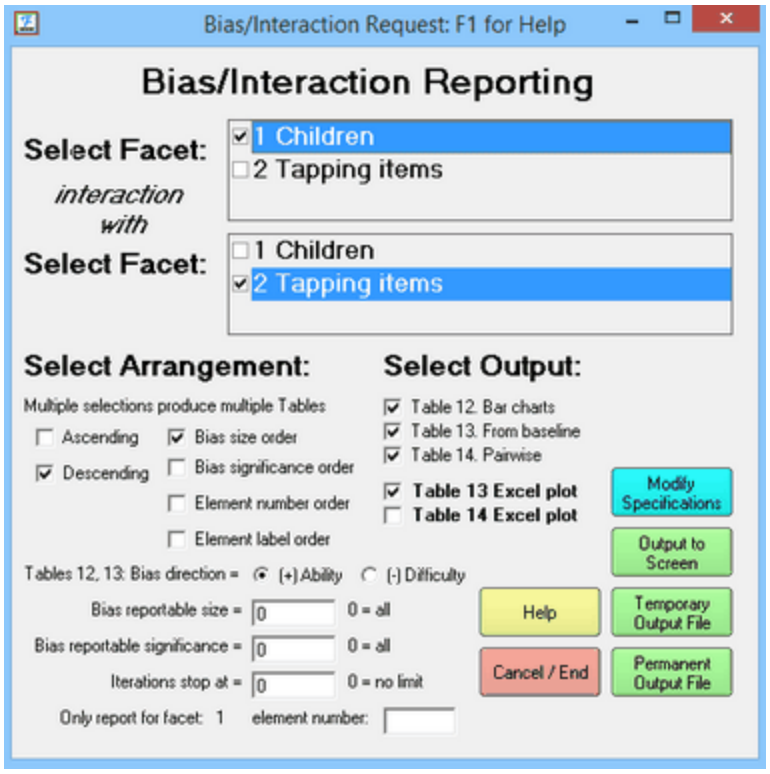
From the Output Tables menu, **Modify Specifications** allows modification of control specifications which affect general program output, but without requiring re-estimation.




**Title:** head line on each output table is Title=
**Juxtapose:** column heading after this many lines, or 0 for no intermediate headings is Juxtapose=,
**Left-hand** placement of row labels = Yes is Left-hand=
**Umean:** mean of centered measures is Umean= (first value)
**Uscale:** number of user-scaled units per logit is Umean= (second value)
**Udecim:** number of decimal places for measures is Umean= (third value)
**Total score** (includes extreme responses) = Yes is Totalscore=
**Positively-oriented** facets is Positive=
**Negatively-oriented** facets is Negative=
**Standard errors:** Model or Real is SE=
**Table format:** ASCII = No or Yes or Webpage is ASCII=
**Boxes around Tables:** Boxshow= Yes or No, useful when importing Tables into spreadsheets, etc.
**OK** actions the new values

## 12.7 Table 13-14 Excel bias plot

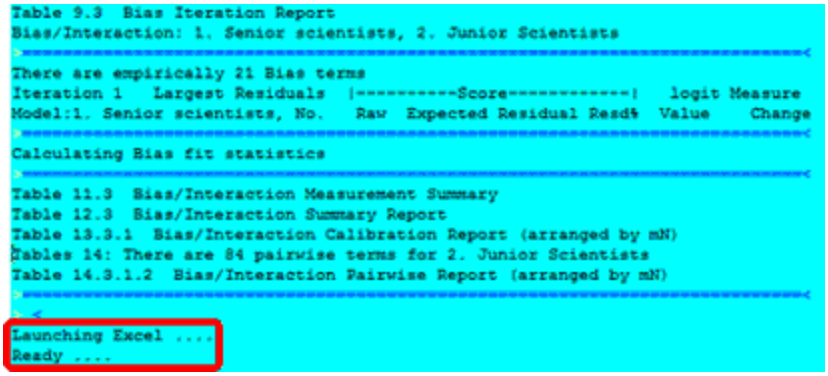
From the Output Tables menu, select *Table 13-14* to output Table 13. Excel plots can also be produced.
Table 13 Excel plots correspond to Table 13.
Table 14 Excel plots correspond to Table 14, the pairwise report. Excel may be slow to plot this, and may not display the legends correctly. See below how to identify points.




Bias reportable size and significance are the minimum absolute values to be reported. Only values more extreme are shown in the Tables. Typical significance sizes are 1.96 for 5% significance (double-sided) and 2.58 for 1% significance (double-sided).
For "Element number order", top facet is minor, bottom facet is major.

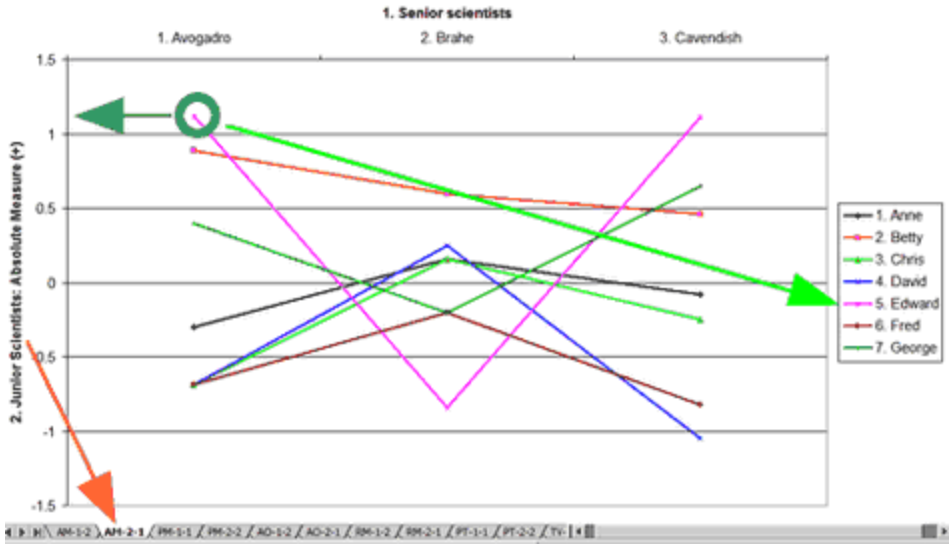
"Only report for facet: " - this is the facet number in the top Select Facet box
"element number:" - there is no number here, then all elements are reported. If there is an element number, then only this element of the selected facet is reported.

The bias/interaction terms are computed, and Excel is launched to plot them:



Warning! Plotted interactions pruned to first 250 on Excel plot
The number of series to be plotted has been pruned to fit into Excel limits.

When "Excel plots" is checked and Microsoft Excel is available, then Excel plots are produced of the numbers reported in Tables 13 14. The direction of the measures on the plots accords with the orientation of the facets (positive or negative). These plots can be edited using all the tools of Excel.




Here is plot AM-2-1. This shows the absolute measures (measures reported in Table 7 + bias/interaction measures reported in Table 13). The measures are for the elements of Facet 2, as perceived by the elements of Facet 1. The Green circle highlights the measure for Edward (facet 2, element 5) according to Avogadro (facet 1, element 1).

```
Table 13.4.1  Bias/Interaction Report (arranged by mN).
Bias/Interaction: 1. Senior scientists, 2. Junior Scientists (higher score = higher bias
       measure)
-------------------------------------------------------------------------------------------
       ------------
| Obsvd    Exp.  Obsvd  Obs-Exp|  Bias  Model        |Infit Outfit|    Senior scientists
       Junior Scientis |
| Score   Score  Count  Average|  Size   S.E.    t   | MnSq  MnSq | Sq N Senior sc  measr N
       Junior  measr |
-------------------------------------------------------------------------------------------
       ------------
|    36     29.2     5     1.36|    .69   .35   2.00 |   .5    .5 | 13 1 Avogadro    -.04 5
       Edward    .42 |
```

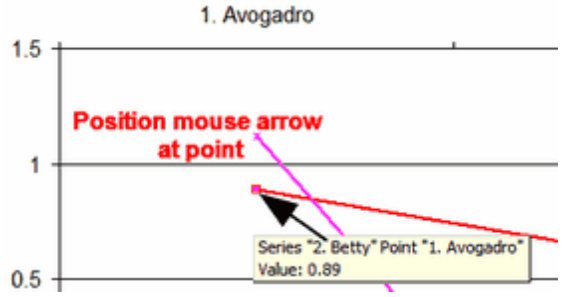
According to Table 13, when Avogadro rated Edward, the ratings were 1.36 higher than expected, from Avogadro's perspective, Edward was .69 logits more able than his overall measure. Edward's overall measure (at the extreme right) was .42 logits. So his measure from Avogadro's perspective was .42 + .69 = 1.11 logits, as indicated by the measure on the y-axis (heavy green arrow).

Each plot has a name like "AM-1-2". "AM" identifies the plot type. The first number, "1" is the facet number of the target facet. The second number "2" is the facet number of the context facet.

| Worksheet Code | *Facets* Table | Contents |
|---|---|---|
| AM | 13 | Absolute Measures: the absolute bias/interaction measures (overall measures + bias) are shown for each target element. |

| | | |
|---|---|---|
| PM | 14 | Paired Measures: the differences between the bias/interaction measures for each pair of target elements in a facet |
| AO | 13 | Average Observation for each target element |
| RM | 13 | Relative Measures: the bias/interaction measures relative to the overall measures for each target element |
| PT | 14 | Pairwise *t:* the t-statistic testing the hypothesis that there is no pairwise bias/interaction effect for each pair of target elements |
| TV | 13 | *t* oVerall: the t-statistic testing the hypothesis that there is no bias/interaction effect relative to its overall measure for each target element |
| *Worksheet* | | the numbers that Excel is plotting - these can be edited as desired. |

Position the mouse arrow over a point on the plot to see its details:




---

Excel Worksheet: This contains the plotted values. You can copy-and-paste for other uses.

| | A | B | C | D | E | F | |
|---|---|---|---|---|---|---|---|
| 1 | Bias/Intera | Ratings of Scientists (I | | ########## | | | |
| 2 | AM-1-2 | 1. Senior scientists: Al | | 2. Junior Scientists | | | |
| 3 | Number | 1. Senior scientists | | 1. Anne | 2. Betty | 3. Chris | 4. D |
| 4 | 1 | Avogadro | 1. Avogadr | 0.27 | -0.21 | 0.49 | |
| 5 | 2 | Brahe | 2. Brahe | 0 | 0.27 | -0.17 | |
| 6 | 3 | Cavendish | 3. Cavendi | -0.08 | 0.09 | -0.09 | |

# 13 Output Files from menu

## 13.1 Specification settings file

A list of the settings of all specifications

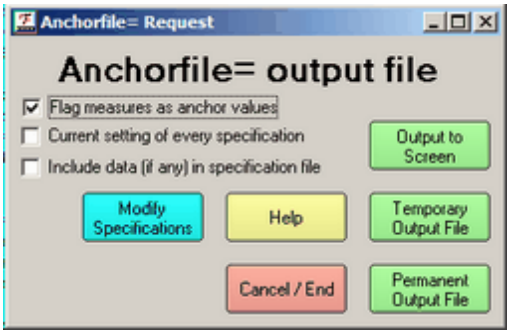
Example:

```
; Specifications for C:\Facets\examples\Guilford.txt
Title for each results table = Ratings of Scientists (Psychometric Methods p.282 Guilford 1954)
11/11/2019 3:16:22 PM
Facets in data = 3, 3 ; label facets, data facets
Anchor output file =
Arrangement by Number/Alpha/Measure/Fit/PBS (Asc./Desc.) = mN,2N,0fN
ASCII output table display format = Y
Barcharts output = Yes
```

```
Batch processing = N
Beep sound = No
Bias (direction: difficulty/ability/omit) = plus ; ability, leniency, easiness: higher score =
positive logit
Box show around Tables = Y
Center elements in facet =  2, 3
Convergence (criteria to end iteration: score residual + change + category residual + change) =
 .5, .01, 0, 0
CSV output format = Fixed
...
```

## 13.2 Anchor output file

Form the Output Files menu, click on Anchor Output File. Facets writes out a new specification file containing the input specifications for this analysis. This file omits any specifications entered as extra specifications, but includes the final estimated measures for all elements and scales. Also produced by Anchorfile=




**Flag measures as anchor values**: includes ",A" after the facet labels to indicate that all reported estimates are anchor values. These anchor values can be used as starting values for slightly different later runs. For this, uncheck "Flag measures as anchor values" or edit the ",A"'s out of the first line of each facet in the Labels= specifications and out of the category lines of Rating (or partial credit) scale=.

**Current setting of every specification:**
**checked** - this produces an anchorfile with every specification set to its current value. This is useful if you want a full record of the specifications for this analysis.
**unchecked** - this produces an anchorfile similar to the original specification file.

**Include data (if any) in the specification file:** Is the data in the specification file after "Data=" to be included in the anchorfile?

Example: Using an edited version of Guilford.txt. Anchorfile=Guilford.anc - "Current setting" is unchecked.

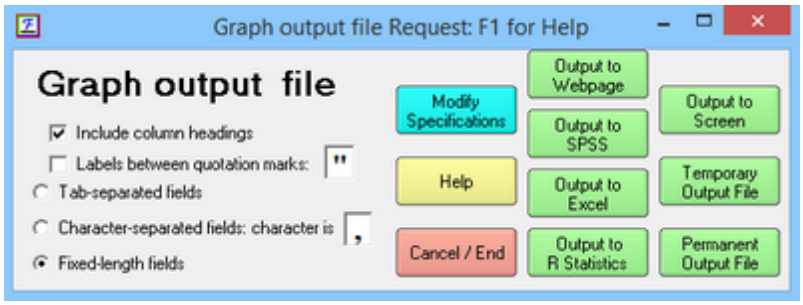
```
Title= Ratings of Scientists (Psychometric Methods p.282 Guilford 1954)
anchorfile=guilford.anc
Score file=GUILFORD.SC ; score files .SC1, .SC2, .SC3
Facets=3               ;three facets
Arrange=m,2N,0f        ;arrange tables in measure order, descending, etc
Positive=2             ;the scientists have greater ability with greater score
Non-centered=1         ;scientists and traits are centered, seniors are not
Unexpected = 0           ;report ratings if standardized residual >=|2|
Usort = (),(1,2,3),(3,1,2),(Z,3)   ;sort unexpected ratings various ways
Vertical=2N,3A,2*,1L,2A  ;define rulers
Zscore=1,2             ;report bias size than 1 logit or z>2
Models=
?,2,3,RS1,0, ; CREATIVITY
?,1,?,RS1,5, ; CREATIVITY
?,2,?,RS1,2, ; CREATIVITY
3,?B,?,RS2,1, ; CREATIVITY
```

```
#B,?B,?,CREATIVITY,1,
*
Rating (or partial credit) scale=RS1,R9,G,O ; CREATIVITY
 1=lowest,,,
 3=,0,A, ; A if "Flag measures" is checked
 5=middle,-3.823871,A,
 7=,1.151407,A,
 8=,.8946334,A,
 9=highest,1.777831,A,
*
Rating (or partial credit) scale=RS2,R9,G,O ; CREATIVITY
 1=lowest,,,
 5=middle,0,A,
 7=,.5815754,A,
 9=highest,,,
*
Labels=
1,Senior scientists,A
1,Avogadro,1
2,Brahe,
3,Cavendish,-1.990667
*
2,Junior Scientists,A
1,Anne,-2.213118 ; in subset: 2 (shown if the data exhibits subsetting)
2,Betty,1.296003
3,Chris,3
4,David,
5,Edward,
6,Fred,
7,George,.9171149
8,Extreme,
*
3,Traits,A
1,Attack,-2.379807
2,Basis,-.9213699
3,Clarity,2.452592
4,Daring,-1.218569
5,Enthusiasm,2.067154
*
Data= ; If "Include data" is checked
1,1,1-5,5,5,3,5,3
....
```

## 13.3 Graph output file

**A better plot may result from copying Facets graph coordinates, see "Copy Plot Data to Clipboard".**

From the Output Files menu, "Graph output file" provides the values plotted in Table 8 for the probability curves and expected score ogive. Also produced by Graphfile=.




For help with this dialog box, see Output Dialog Box.

The output is in fixed column format or, if comma-separated values are preferred, specify "filename, CSV" or CSV=Y. For Tab-separated values, specify "filename, Tab" or CSV=Tab

Output to SPSS, Excel and R Statistics are in their native-format. Excel and R Statistics are launched.

Heading lines= controls display of heading lines. QM quotation marks, controls whether labels are withing quotation marks.

Modify Specifications provides further options

The fixed column format is:

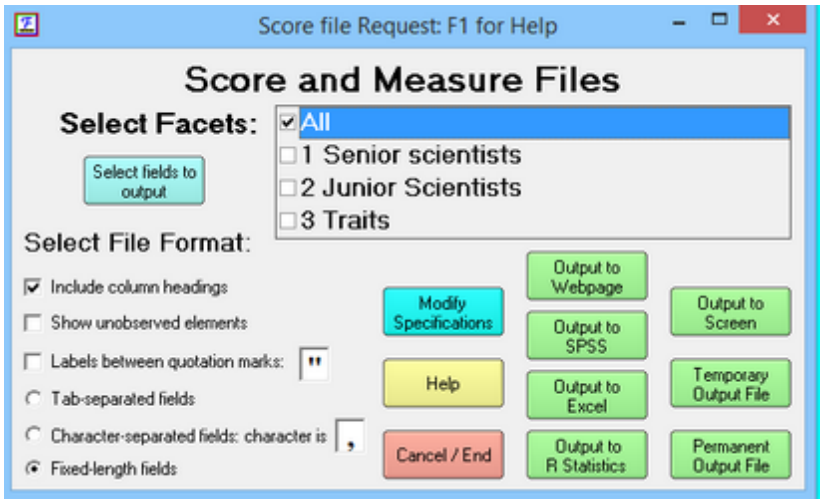
| Field | Columns | Contents |
|---|---|---|
| 1 | 1-10 | Scale number (matches Sub-Table number for Table 8) |
| 2 | 11-20 | Measure (X-Axis) |
| 3 | 21-30 | Expected Score |
| 4 | 31-40 | Category value corresponding to expected score |
| 5 | 41-50 | Probability for bottom category |
| 6 | 51-60 | Probability for second category |
| 7 ... | 61-70, ... | Probabilities for higher categories |

```
     Scale   Measure  Expected    ExpCat    Prob:1    Prob:2    Prob:3    Prob:4    Prob:5
Prob:6    Prob:7    Prob:8    Prob:9
         1   -6.0000    1.0307         1     .9697     .0299     .0004     .0000     .0000
.0000     .0000     .0000     .0000
         1   -5.9880    1.0311         1     .9693     .0303     .0004     .0000     .0000
.0000     .0000     .0000     .0000
         1   -5.9759    1.0315         1     .9689     .0307     .0004     .0000     .0000
.0000     .0000     .0000     .0000
```

Example:Provide coordinates so I can plot beautiful probability curves using my graphics package: Graphfile=plot.dat,CSV; the name of an output file with comma-separated values, produces, for a 3 category rating scale (or partial credit):

```
1,-2.00,.23,0,.79,.20,.01
1,-1.92,.24,0,.77,.21,.02
1,-1.85,.26,0,.76,.22,.02
```

## 13.4 Output Dialog Box



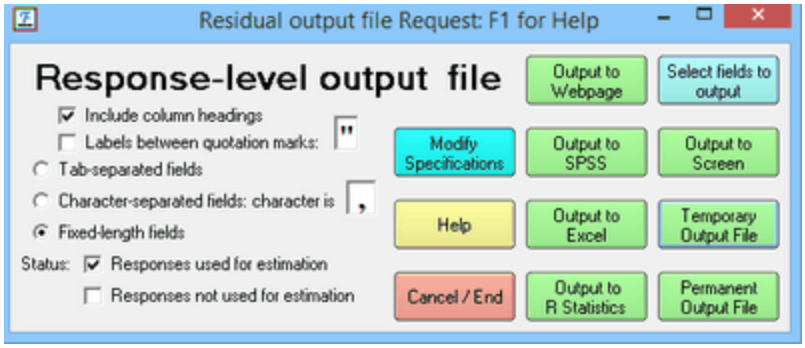

Here is an example of an Output Dialog box.

| Output Dialog Buttons | | | |
|---|---|---|---|
| **Select Facets** check the box for the facets you want in the output file(s). | **Modify Specifications** - provides more formatting options. Some choices can also be preset using Edit Initial Settings. | **Output to Webpage** uses your web browser to display a sortable table containing the output file | **Output to Screen** - writes the output file to the analysis window (generally slow) |
| **Select fields to output** lets you choose which fields you want to display | **Help** - displays the Help topic for this Dialog | **Output to SPSS** - writes the output as an SPSS .sav file to the specified file. SPSS-formatted files include the facet number, and also the facet label if element labels are specified. | **Temporary Output File** - writes the output files to temporary text files (generally quickly) |
| **Select File Format** to match the output you want and that works | **Cancel/End** - exits from this Dialog without any further action. | **Output to Excel** - writes the output files with one Excel temporary workbook for each facet. If Excel files do not display. | **Permanent Output File** - writes the output files to permanent text files. The file name builds on the specified name. |
| | | **Output to R Statistics** - writes the output file to a .rdata file and launches R. | |

**Output to Excel, SPSS or Word**
When Facets does not support direct output, then see export to Excel, SPSS and Word.

## 13.5 Residuals / Responses Output File

From the Output Files menu, clicking on Residuals/Response Output file. A file of residuals from the main analysis is produced. This file is designed for input into other programs with one line per measurable observation. It can be used for calculating other fit statistics and producing specialized diagnostic reports. Also produced by Residualfile=.




For help with this dialog box, see Output Dialog Box.

Heading lines= control the output of the heading line. CSV= allows tab-delimited and other formats. QM quotation marks, controls whether labels are within quotation marks.

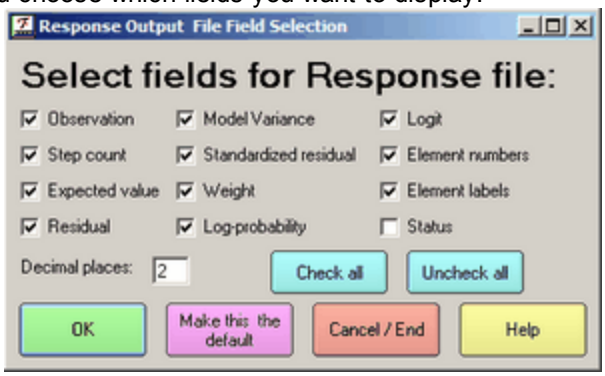
Output to SPSS, Excel and R Statistics are in their native-format. Excel and R Statistics are launched.

**Modify specifications** lets you change the values of some basic output settings

**Status: Responses used for estimation:** report responses used for estimation (Status code: 1, 2)
**Responses not used for estimation:** report responses not used for estimation (Status code: -5, -4 , -3, -2, -1)

Output directly to Excel and into an SPSS .sav file are supported.

**Select fields to output** lets you choose which fields you want to display:




If "Make this the default" is ignored, then you need to take ownership of file Facets.ini in your c:\Facets folder. See www.addictivetips.com/windows-tips/windows-7-access-denied-permission-ownership

Here is an example of the format with 4 decimal places in the "Select fields" dialog box. The precise format depends on the number of facets in your data:

```
Obs      Stp      Exp      Res      Var     StRes       Wt    LProb  Measure    Displ   Status    MPCat       El
     E2       M1       M2 Children  Tapping_it
  1       1      .97      .03      .02      .16     1.00    -.03     3.66     1.03        2        1        1
       1    -2.98    -6.64 Boy      1-4
  1       1      .97      .03      .02      .16     1.00    -.03     3.66     1.03        2        1        1
       2    -2.98    -6.64 Boy      2-3
```

The columns are:

| Fixed field columns | Abbreviation | Description |
|---|---|---|

<table>
<tr><td>1-10</td><td>Obs</td><td colspan="2">response as observed in the data file</td></tr>
<tr><td>11-20</td><td>Stp</td><td colspan="2">observed response as renumbered into a count of ordered steps</td></tr>
<tr><td>21-30</td><td>Exp</td><td colspan="2">expected score for this response (decimal places set in selection dialog box)</td></tr>
<tr><td>31-40</td><td>Res</td><td colspan="2">score residual: (observed Stp - expected Exp)</td></tr>
<tr><td>41-50</td><td>Var</td><td colspan="2">model variance of observed score around the expected score for this response, the statistical information in this response</td></tr>
<tr><td>51-60</td><td>StRes</td><td colspan="2">standardized residual: residual / sqrt (variance)</td></tr>
<tr><td>61-70</td><td>Wt</td><td colspan="2">weighting (model weight * observation weight * item weight)</td></tr>
<tr><td>71-80</td><td>LProb</td><td colspan="2">natural logarithm of the probability of the observation</td></tr>
<tr><td>81-90</td><td>Meas</td><td colspan="2">sum of the measures of the elements producing the observation. User-scaled: Meas = sum(element measures - umean) + umean</td></tr>
<tr><td>91-100</td><td>Disp</td><td colspan="2">displacement = measure residual = (score residual / variance)*(user-scaling). The measure of element 1 according to this observation is "element measure" for element + "displacement" * (orientation of facet 1). This is limited to the range -10 to +10 logits.</td></tr>
<tr><td rowspan="9">101-110</td><td rowspan="9">Status</td><td>Status Code</td><td>Meaning</td></tr>
<tr><td>-6 (not used for estimation)</td><td>Response in two multiple-observation ranges, such as 1-4, 2-6,...</td></tr>
<tr><td>-5 (not used)</td><td>Responses after end-of-file.</td></tr>
<tr><td>-4 (not used)</td><td>Responses only in extreme scores.</td></tr>
<tr><td>-3 (not used)</td><td>Responses with invalid elements. Elements for these observations are not defined. See Table 2.</td></tr>
<tr><td>-2 (not used)</td><td>Responses in two extreme scores</td></tr>
<tr><td>-1 (not used)</td><td>Responses invalid after recounting<br>A dichotomy or rating scale has less than two categories, so it cannot be estimated.</td></tr>
<tr><td>1 (used for estimation)</td><td>Responses used for estimation</td></tr>
<tr><td>2 (used)</td><td>Responses in one extreme</td></tr>
</table>

| | | | |
|---|---|---|---|
| | | | score |
| 111-120 | MPCat | most probable category to be observed. If two categories are equally probable, then the higher category is shown here | |
| 121-130 | E(facet number) | element number for facet 1 or null element, usually 0 | |
| \| | | | |
| \| | M(facet number) | element measure for facet 1 from Table 7 (user-scaled) | |
| \| | | | |
| \| | (facet label) | element label for facet 1 | |
| \| | | | |

For "Category implies Measure" (C->M) and "Measure implies Category" (M->C) statistics, for each observation in the Facets Residualfile=,
"expected score for this response" - round this to the nearest category number = expected average category
if "expected average category" = "observed response as renumbered into a count of ordered steps" 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

Example: The "Obs" (observed) is the original data. The "Stp" (step) is the ordinal version of the original data. This version is used for analysis, and is the version on which the "Exp" (expected) and the "Res" (residual) are based. This version may be the same as the original data, or the original data may be transformed either due to explicit instructions by the analyst, or by default operation of Facets.

For instance, suppose that the original data are observations of these three values: 10, 20 and 30. Then, by default, Facets will analyze these observations as the "steps": 10, 11, 12. If the original data are intended to be 10,11,12,13,14,....,28,29,30. Then please specify this is in your Models= statement:
Models=
?,?,..., R30K  ; where "K" means "Keep" the original numeration.

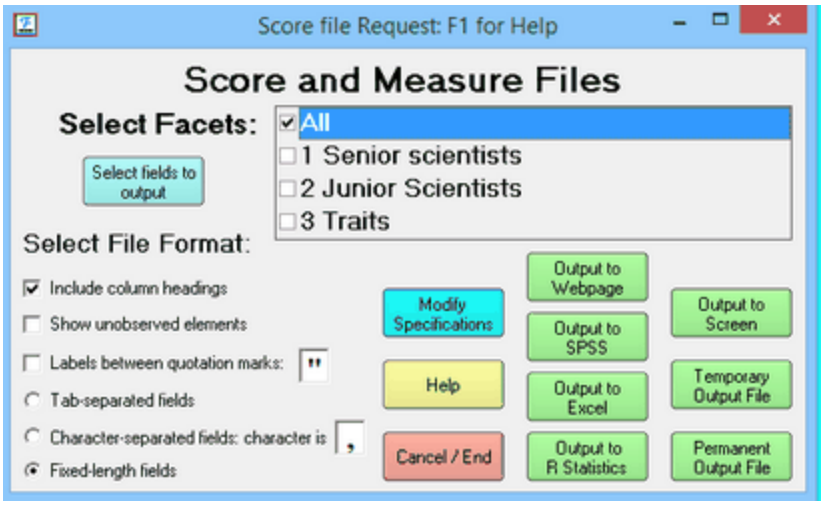
## 13.6 Score and measure file output

From the Output Files menu as Score and measure file output. This enables you to control the Score and measure file. Score and measure output files simplify the transfer of measures to other computer software. you may specify a file name that will be used to construct files to hold the numeric results presented in Facets measure tables. Also produced by Scorefile=.

Output to SPSS, Excel and R Statistics are in their native-format. Excel and R Statistics are launched.

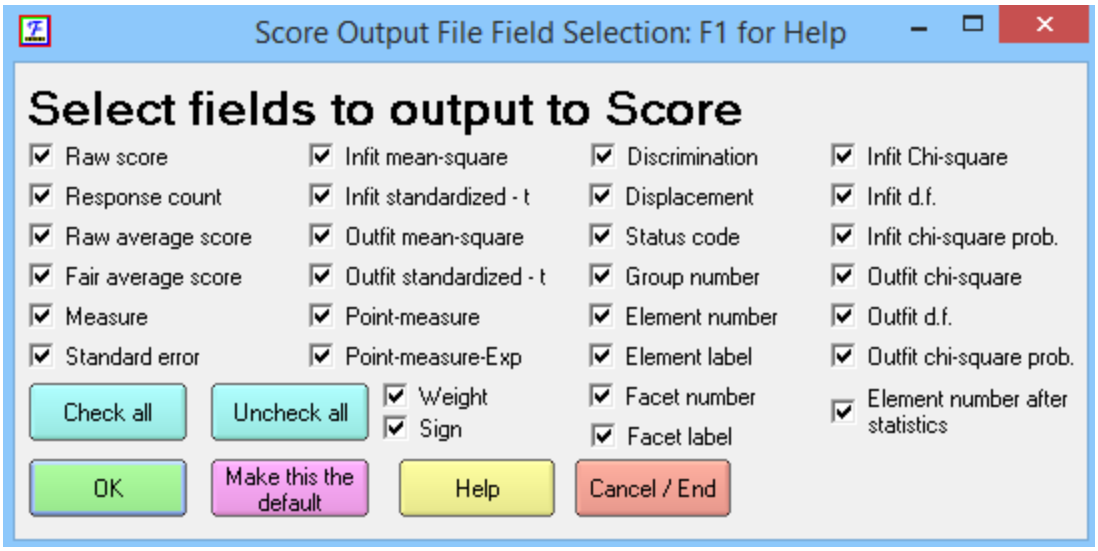
The measures for each facet are written to a separate file, whose name is that specified, with the facet number appended. E.g., if Scorefile=Results, then the measures for facet 1 are written to Results.1, for facet 2 to Results.2 etc.

For Excel-formatted output files, the scores and measures for the elements of each facet are written to separate workbooks.
For SPSS-formatted output files, the scores and measures for the elements of all facets are written to one SPSS .sav file, which includes facet number as a variable.
For R Statistics-formatted output files, the scores and measures for the elements of each facet are written to a separate Rdata file.




For help with this dialog box, see Output Dialog Box.

| | |
|---|---|
| **Select Facets:** | specifies the facets for which the elements are to be reported. |
| **Select file format:** | describes in what way you want the details of each element to be listed |
| **Include column headings** | Usually headings are not wanted when a score file is to be imported into a database. Heading lines= |
| **Show unobserved elements** | Reports all elements specified in Labels=. Otherwise only elements active in this analysis are reported. Omit-unobserved= |
| **Labels between quotation marks** | Useful if labels contain blanks. Required by some software. The quotation mark character can be specified |
| **Tab-separated fields** | best field format for Excel |
| **Character-separated fields** | best field format for some database programs. The separator character can be specified |
| **Fixed-length fields** | easiest field format to read by eye |

| | |
|---|---|
| **Button: Score fields to output** | to select which fields to include in the score file. Check the fields to be output. Then click on **OK**.<br>  <br><br>When Element number is checked, its place in the Score file is set by "Element number after statistics"<br>When "Element number after statistics" is *checked,* the Element number is placed after the numerical statistics and before the Element label in the output data line.<br>When "Element number after statistics" is *not checked,* the Element number is placed first in the output data line. |
| **(Other Buttons)** | see Output Dialog box. |

Example: Write out score files for each facet in fixed field format:
Score file= kct

This puts the measures for facet 1 in file "kct.1.txt", for facet 2 in "kct.2.txt", etc. The results depend on the score file field selection, but can be:

```
                                                
```

| | | |
|---|---|---|
| 0. Element number | 241-250 | Element number is the first field when "Element number after statistics" is *not checked* in the Score File Field Selection dialog box. Otherwise it is before the Element label. |
| 1. Raw score | 1-10 | T.Score includes extreme scores, when Totalscore=Yes<br>Score excludes extreme scores, when Totalscore=No |
| 2. Response count | 11-20 | T.Count includes extreme scores, when Totalscore=Yes<br>Count excludes extreme scores, when Totalscore=No<br>Elements with no observations are omitted unless Omit-unobserved=No |
| 3. Observed average score | 21-30 | (Raw Score)/(Raw Count) |
| 4. Fair average score | 31-40 | FairMAvge: Fair=Mean computes from the mean of the other facets<br>FairZAvge: Fair=Zero computes from the local zero of the other facets |
| 5. Measure | 41-50 | Decimal places set with Umean=x,x,x |
| 6. Standard error | 51-60 | Decimal places set with Umean=x,x,x |
| 7. Infit mean-square | 61-70 | |
| 8. Infit standardized | 71-80 | |
| 9. Outfit mean-square | 81-90 | |
| 10. Outfit standardized | 91-100 | |
| 11. Point-measure correlation | 101-110 | or Point-biserial correlation. See Pt-biserial= |
| 12. Point-measure expectation | 111-120 | .00 if not computed. |
| 13. Discrimination | 121-130 | |
| 14. Displacement | 131-140 | Decimal places set with Umean=x,x,x. This is limited to the range -10 to +10 logits. |
| 15. Status code | 141-150 | See **Status Code** list below |
| 16. Group number | 151-160 | |
| 17. Weight | 161-170 | |
| 18. Sign | 171-180 | |
| 19. Infit Chi-squared | 181-190 | = infit information-weighted mean-square * d.f. |
| 20. Infit Chi-squared d.f. | 191-200 | |

| | | |
|---|---|---|
| 21. Infit Chi-squared 2-sided probability | 201-210 | 2-sided probability of the chi-squared value. <.05 is significant underfit or overfit. |
| 22. Outfit Chi-squared | 211-220 | = outfit conventional mean-square * d.f. |
| 23. Outfit Chi-squared d.f. | 221-230 | |
| 24. Outfit Chi-squared 2-sided probability | 231-240 | 2-sided probability of the chi-squared value. <.05 is significant underfit or overfit. |
| 25. Element number | 241-250 | Element number is here when "Element number after statistics" is checked in the Score File Field Selection dialog box. Otherwise it is the first field. |
| 26. Element label | 251- ? | field length depends on the length of the element labels |
| 27. Facet number | 10 columns | field positions depends on the length of the element labels |
| 28. Facet label | ? | field length depends on the length of the facet label |

| Status code | Meaning for element measure |
|---|---|
| 1 | Anchored measure |
| 0 | No data available to estimate measure |
| -1 | Measure estimated from data |
| -2 | Measure estimated for extreme low score: XTREME= |
| -3 | Measure estimated for extreme high score: XTREME= |
| -4 | Unmeasurable, because all relevant data are generated by more than one extreme element |
| -5 | Only one active data point for this element |

## 13.7 Web output table

Output from Graphfile=, Residualfile= and Scorefile= can be to a sortable table in a webpage, displayed by your web browser.

A sortable webpage has this layout:

Scores: Ratings of Scientists (Psych

| T.Score | T.Count | Obs.Avge | FairMAvge ↓ | Measure ↔ | S.E. |
|---|---|---|---|---|---|
| 81.00 | 21.00 | 3.86 | 5.79 | -.05 | .10 |
| 81.00 | 21.00 | 3.86 | 5.79 | -.05 | .10 |
| 78.00 | 21.00 | 3.71 | 5.68 | -.03 | .10 |
| 72.00 | 21.00 | 3.43 | 5.43 | .03 | .10 |
| 63.00 | 21.00 | 3.00 | 5.01 | .11 | .10 |

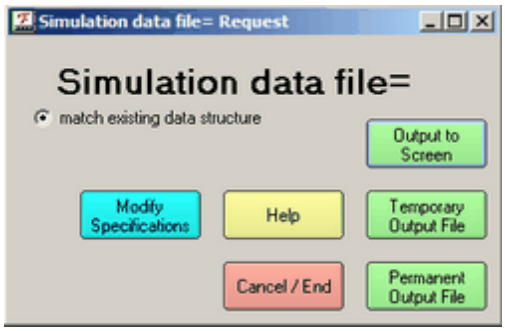
Position the mouse pointer to the right of the column title of the column to be sorted.
↔ this column is selected for sorting: click to sort descending
↑ sorted ascending: click to sort descending
↓ sorted descending: click to sort ascending
blank - not sorted

Save the webtable (HTML only) while Facets is running for use when Facets is not running.

Here is an example of an interactive sortable web table.

## 13.8 Simulated data file

A file of simulated data can be constructed from the measures estimated (or anchored) for the main analysis. It will have one simulated observation for each observation in the original data file. Each simulation is unique, so that multiple different simulations can be obtained with the Output Files pull-down menu. Also produced by Simulateddatafile=.



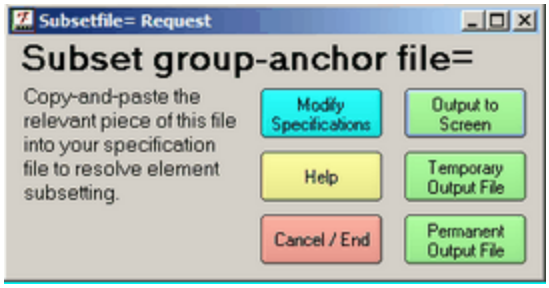

When written to a permanent output file, the simulated data can be analyzed using Data= (permanent file name) in the original specification file (or enter at Extra specifications? prompt). Comment out any Dvalues= specifications in the original specification file.

The simulated data file has the basic Facets data format:

```
; Simulated data matching the empirical data structure
; Ratings of Scientists (edited to illustrate ambiguity in measurement
; matching: C:\FACETS\examples\subsets.txt
1,2,1,7 ; 9 ; 1,2,1 are facet elements. 7 is simulated. 9 is the original data value.
1,2,2,7 ; 7
1,2,3,4 ; 5
1,2,4,9 ; 8
1,2,5,3 ; 5
```

## 13.9 Subset group-anchor file

When subsetting is reported, it can be difficult to eliminate. The Subset group-anchor file can help resolve this. From the Output files menu, click on Subset group-anchor file. It is only available when needed.




Modify Specifications allows changes to specifications.

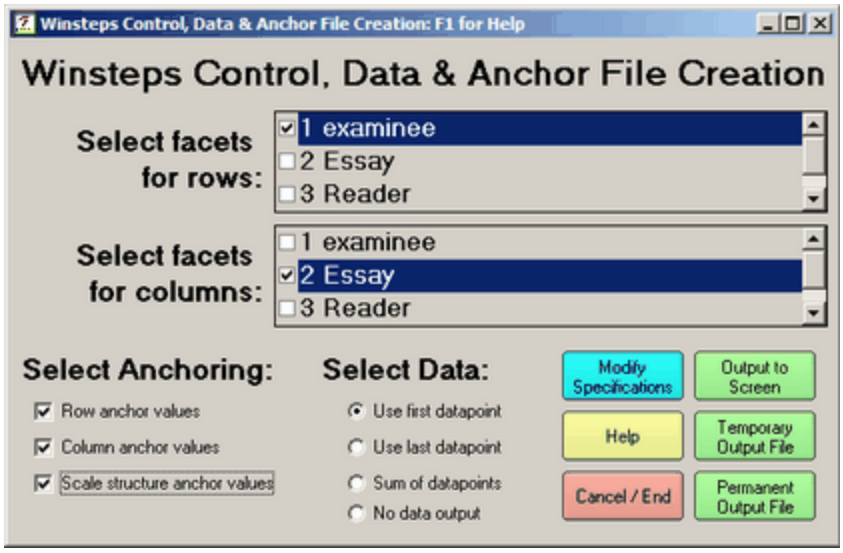
The output file suggests possible changes to your specification file to eliminate subsetting. Usually only changes to one facet are required, though the subset group-anchor file may suggest more than one.

Go to the subsets.txt analysis for an example of this. Its Subset group-anchor file is:

```
To resolve subset problems, copy-and-paste after Labels=
Non-center= must reference a facet that is not anchored or group-anchored.

Group anchor this facet:

1,Judges, G ; group-anchoring at Umean = 50
1,Avogadro,50, 1
2,Brahe,50, 1
3,Cavendish,50, 2
4,Davey,50, 2
*

And/or group anchor this facet:

2,Examinees, G ; group-anchoring at Umean = 50
1,Anne,50, 1
2,Betty,50, 1
3,Chris,50, 1
4,David,50, 2
5,Edward,50, 2
6,Fred,50, 2
7,George,50, 2
*
```

## 13.10 Winsteps control, data and anchor file

Winsteps provides analytical and output capabilities not implemented in Facets, such as the **principal components analysis PCA of residuals**. Winsteps requires a rectangular data layout. Select from the Output Files pull-down menu,




The Facets data are formatted into a rectangle. A row (or column) can comprise one facet. If one facet is checked, then each element of that facet becomes one row (or one column) in the data rectangle. If two or more facets are checked for rows (or columns), then each row (or column) comprises data corresponding to one combination of elements from each checked facet.

It is typical that formatting the data as a rectangle results in more than one observation for each row-column cell. If so, the "Select data" option specifies whether it is the first relevant datapoint in the Facets file, or the last, or the sum, which occupies the cell.

Row and column measure anchor values can also be written into the Winsteps file. These are obtained from the Facets estimates. If there is a rating-scale common to all the Facets data, this can also be used for Winsteps anchor values.

Example 1: The dialog box produces:

```
Title = AP English Essays (College Board/ETS)
item = Essay  ; column identification
person = examinee  ; row identification
xwide = 1  ; width of datapoint
codes = "12345678"  ; valid data codes
; first data point in cell used
ni = 3  ; number of data columns
item1 = 1  ; starting column of data
name1 = 5  ; start of row labels
@pf1=$S1W5 ; 1 examinee
namelen = 5 ; row label length
safile = *  ; rating structure anchor values
1 .00
...
9 2.75
*
iafile = *  ; item-column anchor values
1 -.02 ; 1 A
2 .13 ; 2 B
3 -.11 ; 3 C
*
pafile = *  ; person-row anchor values
1 -.82 ;  1 1
...
32 .03 ; 32 32
*
@if2=$S1W3 ; 2 Essay
```

```
&End
1 A  ; item-column labels
2 B
3 C
END LABELS
553  1 1   ; row data + label
454  2 2
....
644 32 32
```

Example 2: To verify that raters are following the judging plan, output a Winsteps file with the raters as rows, and the other ingredients of the judging plan as columns.

Example 3: We want a standard version of Winsteps Table 2.2 or other Winsteps Keyform-style output.
1. Do the Facets analysis, and output an Anchor file.
2. Output from the Facets "output files" menu a Winsteps control file. Specify which facet you want for the Winsteps columns (the rows on Table 2.2) and which facets you want for the Winsteps rows. Check "row anchor values" and "column anchor values". If "scale anchor values" are available, please check them.
3. Look at the Winsteps control file. If IAFILE= or SAFILE= anchor values are missing, please copy and reformat those values from the Facets anchorfile.
4. Run Winsteps and produce Table 2.2 or whatever Winsteps output you want.

Example 4. To investigate unidimensionality or multidimensionality and local independence, output a Winsteps control and data file with appropriate columns, analyze with Winsteps and output Winsteps Table 23.

Example 5. Checking for **unidimensionality** or **multidimensionality** using R Statistics (freeware):
We need complete data, so set up the Winsteps output with your "items" as columns and all the other facets as rows.

Delete all the rows in the Facets-produced Winsteps data file down to END LABELS. After that is data.
Your observations are probably one column wide.
Save the data as data.txt file.

install R (freeware) if you don't have it.
Launch R
Then construct an R dataset:

```
> mydataset <- read.fwf("data.txt", widths = c(1,1,1,1,1,1,1,1))
```
this has as many 1s as there are items (columns) in your data

add column names
```
> colnames(mydataset) <- c("item1","item2","item3",...)
```
replace item1 with the label for your first item, etc.

check that all is OK:
```
> options(max.print=100)
> mydataset
```

now do a PCA &CFA

```
> install.packages('psych',"FactoMineR')
> library (psych)
> library (FactoMineR)
> result <- fa.parallel (mydataset)
```
a Scree Plot displays of PCA and FA eigenvalues.
R Console window: "Parallel analysis suggests that the number of factors =  3"
Put 3 (or whatever) in the "factors" R instruction below

Principal Components Analysis:
> pca <- PCA(mydataset, graph=FALSE)
> pca$eig
> pca$var$coord
> head(pca$ind$coord)

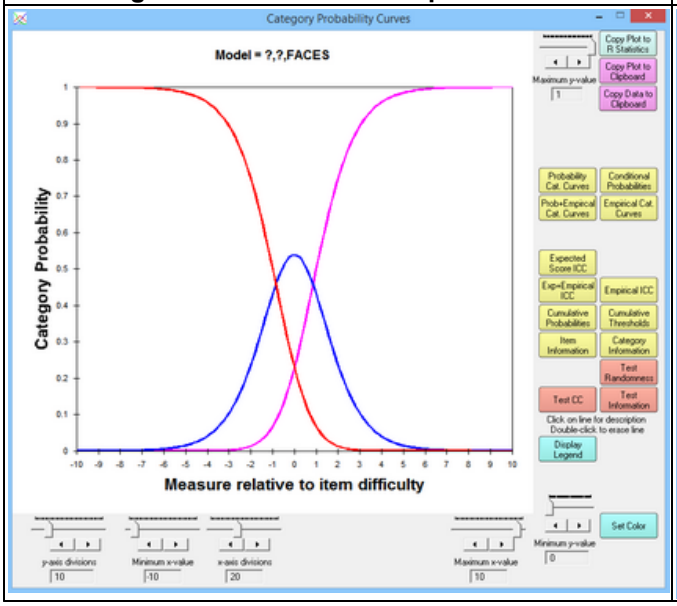
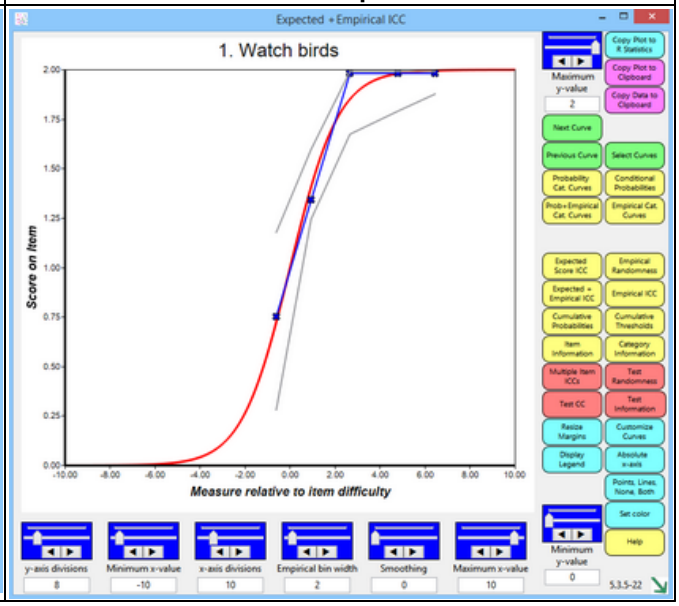
Factor analysis:
> factors <- fa(mydataset, 3)
> print (factors)
    the factor loadings are in columns MR1, MR2, ...
> plot(factors)

You are expecting to see that the first component is large, and the others are much smaller.

# 14 Graphs

## 14.1 Graph window

From the Graphs menu,

| Original and Standard Graph Windows | Enhanced Graph Window |
|---|---|
|  |  |



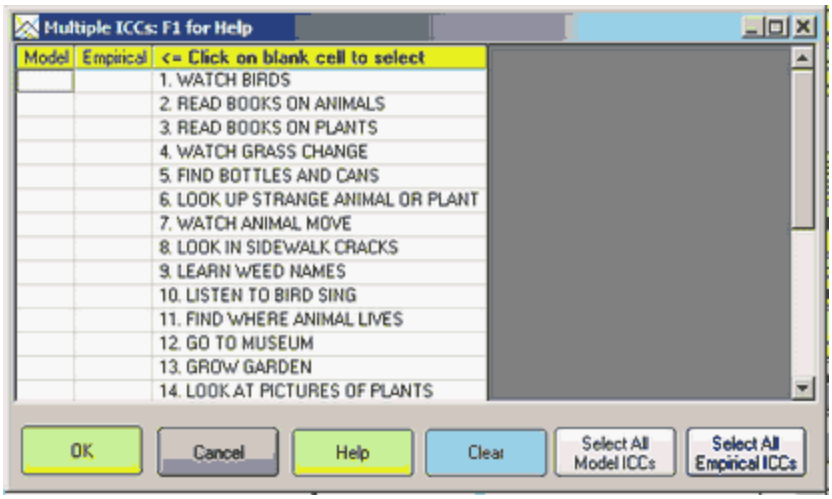
Category Probability Curves: (Probability Cat. Curves) model-based probability of observing each category of the response structure at each point on the latent variable (relative to the item difficulty)
Empirical Category Curves: (Empirical Cat. Curves) data-based relative frequencies of categories in each interval along the latent variable
Category Model and Empirical Curves: (Prob.+Empirical Cat. Curves) the model-based and data-based frequencies of each category.
Expected Score ICC shows the model-based Item Characteristic Curve (or Item Response Function IRF) for the response structure.
Empirical ICC shows the data-based empirical curve.
Cumulative Probabilities plot the model-based sum of category probabilities. The category median boundaries are the points at which the probability is .5. Click on a line to obtain the category accumulation.
Conditional Probability Curves show the model-based relationship between probabilities of adjacent categories. These follow dichotomous logistic ogives. Click on a line to obtain the category pairing.

Item Information Function plots the model-based Fisher statistical information for the item. This is also the model variance of the responses, see RSA p. 100.
Category Information plots the model-based item information partitioned according to the probability of observing the category. Click on a line to obtain the category number.
Test Information Function - this can be plotted with Excel.

| **Button or Slider** | **Function: for complete description, see Winsteps Graph Window** |
|---|---|
| Adjust maximum Y-value | Sets the maximum value on the vertical Y-axis |
| Adjust Y-axis divisions | Sets the number of tick-marks on the Y-axis |
| Copy Plot to Clipboard | 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. Open a graphics program, such as ***Paint,*** and **paste** in the image for editing. |
| Copy Data to Clipboard | 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.<br>Each row corresponds to a set of points to plot:<br>Column A - ignore (internal row identifier)<br>Column B - x-axis for point on first curve: measure relative to item difficulty or relative to latent variable<br>Column C - y-axis for point on first curve: probability, expected score, etc.<br>Column D - x-axis for point on second curve: measure relative to item difficulty or relative to latent variable<br>Column E - y-axis for point on second curve: probability, expected score, etc.<br>etc.<br>Scatterplot:<br>Column C (y-axis) against Column B (x-axis)<br>Column E (y-axis) against Column D (x-axis)<br>etc.<br>Combined plots: the curves on two or more Facets graphs can be combined into one Excel plot using "Copy Data" from the different graphs into one Excel worksheet. |
| Next Curve | Displays the same graph for the element with the next higher entry number is "#" is specified in the Models= specification or there are multiple Models= specifications. |
| Select Curves | Displays the graph for the selected item from the curve-selection-box: |
| Previous Curve | Displays the same graph for the item with the next lower entry number. |
| Probability Cat. Curves | Displays the Rasch-model probability-curve for each response-category |
| 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+Empircal 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. |
| 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. |
| Conditional Probabilities | Displays the Rasch-model probability-curve for the relative probability of each adjacent pair of categories. These are dichotomous ogives. |
| 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. |
| Test Information | Displays the test-information function according to the Rasch model. The inverse square-root of the test information is the standard error of the Rasch person measure. |
| *Original/Standard:* 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. |
| *Original/Standard:* 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. |
| *Enhanced: Resize Margins* | Adjust the left-margin width and the Legend box height below the graph |
| *Enhanced: Customize Curves* | Hide unwanted curves and change curve colors |
| Display Legend<br>Hide Legend | Click on "Display legend" to show descriptions of all the curves below the graph.<br>Click on "Hide legend" to remove the descriptions. |
| Absolute x-axis<br>Relative x-axis | Click on "Absolute x-axis" to show the graphs with measures on the latent variable along the x-axis.<br>Click on "Relative x-axis" to show the graphs with the x-axis plotted relative to the item difficulty. |
| 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. |

| | |
|---|---|
| Help | Click on Help to display this page. |
| Adjust minimum X-value | Sets the minimum value on the horizontal X-axis |
| Fine-tune minimum X-value | Makes small changes to the minimum value on the horizontal X-axis |
| Adjust X-axis divisions | Sets the number of tick-marks on the X-axis |
| Empirical divisions (slider) | Sets the number of empirical summary-intervals on the X-axis |
| Empirical divisions (arrows) | Makes small changes to the number of empirical summary-intervals on the X-axis |
| Fine-tune maximum X-value | Makes small changes to the maximum value on the horizontal X-axis |
| Adjust maximum X-value | Sets the maximum value on the horizontal X-axis |

| | |
|---|---|
| Curve selection Box | This box displays the item entry numbers and labels. Click on the item which you want displayed. |
| ICC Selection Box |  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. |
| Color Selection Box | Click on desired color, and then OK |

The probability curves will display. Graphs are plotted relative to the central difficulty of each item or scale group.

The buttons are:

**Copy Plot to Clipboard** places this plot on the clipboard as a graphic. Open a graphics program, such as ***Paint,*** and **paste** in the image for editing, or paste directly into a Word document.

**Copy Plot Data to Clipboard** places the plotted data on the clipboard. This can be pasted into an Excel worksheet so that you can use it to construct your own graphs.

**Next Curve** takes you to the curves for the next Group.

**Previous Curve** takes you to the curves for the previous Group.

**Select Curves** enables you to jump to the set of curves you want to see by clicking on the list that is displayed. Click on the desired curve.

**Probability Cat. Curves** plots the category probability curves according to the model. There is one curve for each category. They sum vertically to 1.0. Click on a line to obtain the category.

**Empirical Cat. Curves** plot the empirical category frequency points according to the data. The points are joined by lines for ease of viewing.

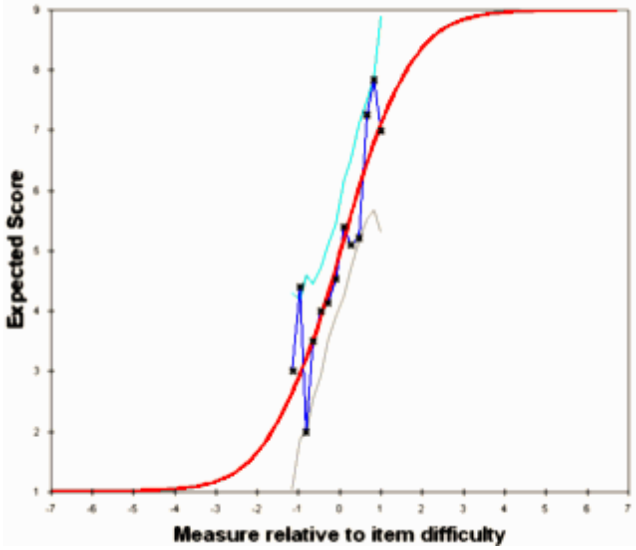
**Prob.+Empirical Cat. Curves** plots both the model and empirical category curves.

**Expected Score ICC** plots the model-expected item characteristic curve, and the empirical one, shown by **X**. **One curve is shown for each rating (or partial credit) scale.**

**Empirical ICC** shows only the empirical, data-derived, item characteristic curve.

---

**Exp+Empirical ICC** plots both the model-expected item characteristic curve, and the empirical, data-derived, curve, together with a confidence band for the empirical around the model curve.




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 with the same rating scale structure. The empirical blue line is constructed from the observed frequencies 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 upper green line (and the lower gray line) are at two model standard errors above (and below) the model "red line", i.e., form a two-sided 95% confidence band. The distance of these lines from the red line is determined by the number of observations in the interval, not by their fit.

---

**Cumulative Probabilities** plot the sum of category probabilities. The category median boundaries are the points at which the probability is .5. Click on a line to obtain the category identification.
**Conditional Probabilities** plots the model ogives for adjacent pairs of categories. Click on a line to obtain the category pairing.
**Information Function** plots the Fisher statistical information for the item. This is also the model variance of the responses, see RSA p. 100.
**Category Information** plots the item information partitioned according to the probability of observing the category. Click on a line to obtain the category number.

**Display Legend (Hide Legend)** displays identification for all lines on the plot
**Absolute x-axis** displays the curves positioned on the latent variable, rather than relative to the element difficulty. **Relative x-axis** positions the curves relative to the element difficulty.
**Points, Lines, None, Both** controls display of the empirical curves.
**Smoothing** smooths the empirical curves.
**Set color** changes the color of the currently selected part of the plot.
**Help** displays this page.

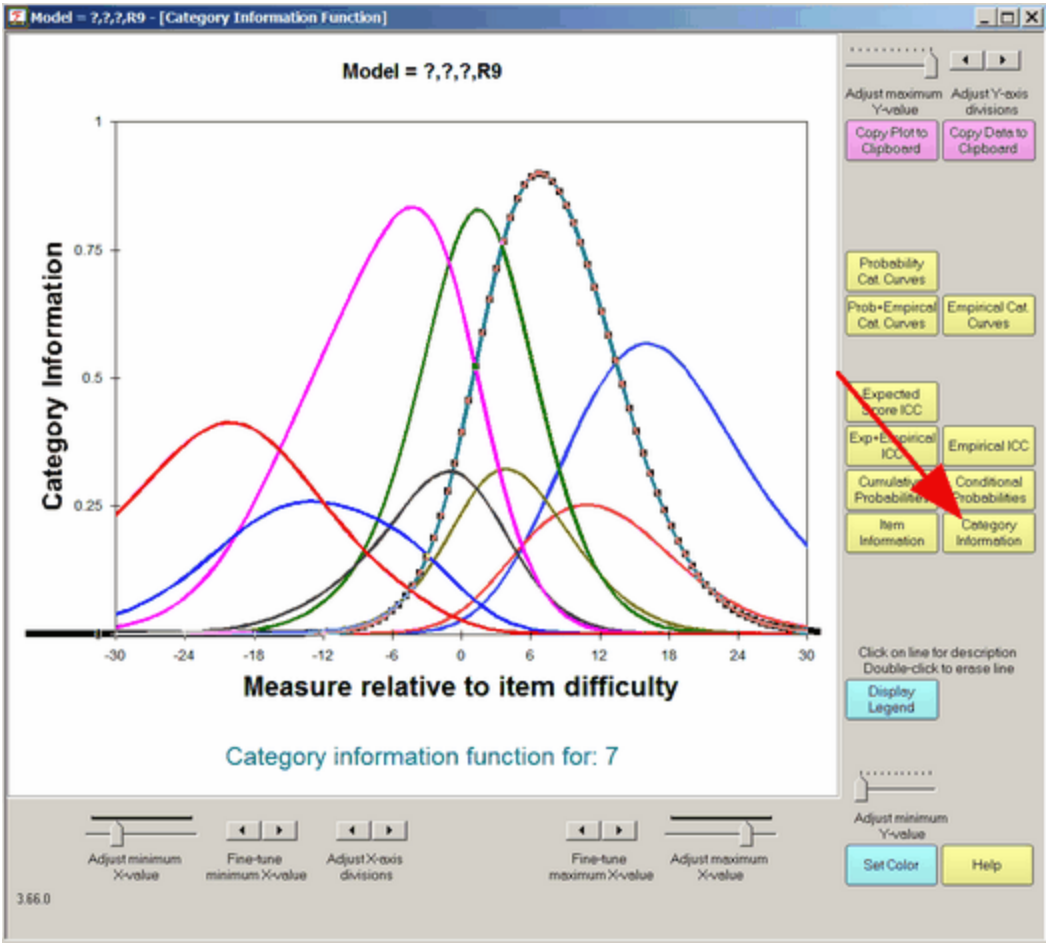
The graphical display is controlled by:
**Adjust** or **fine-tune minimum** or **maximum** enables you to change the x- and y- scale ranges.
**Adjust number of divisions** enables you to change the number of x- and y- tick marks and labels.
**Empirical interval** sets the width of the interval on the latent variable summarized by each **x** on the empirical lines. Displays only when empirical lines are displayed.

On the plot:
*Original/Standard:* **Traceline:** 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. 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.

*Suggestion:* If you need to copy-and-paste several graphs into a document, automate the process with a Windows short-cut macro, such as Autohotkey.

Click on **Close Box** in upper right corner to close the Graphs menu.

## 14.2 Category information function

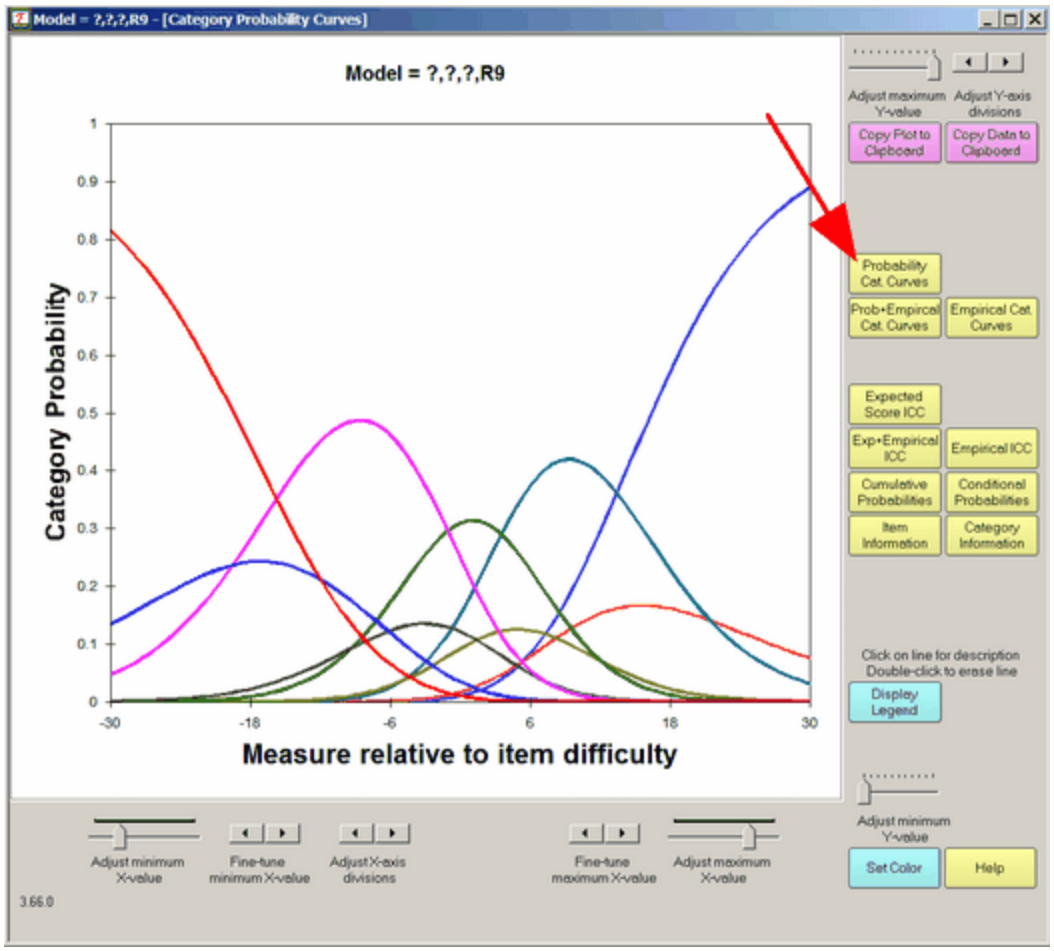



Select by clicking on "Category Information". This shows the item information partitioned according to the probability of observing each category.

Buttons are described in Graph window.

## 14.3 Category probability curves

Click on "Category Probability Curves", also called Item Response Curves, IRCs




Select by clicking on "Probability Cat. Curves". Buttons are described in Graph window.

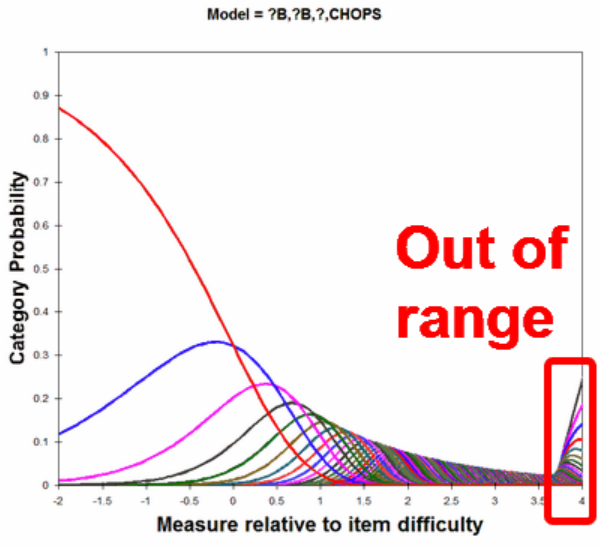
This shows the probability of observing each ordered category according to the Rasch model. To identify a category, click on it.

The caption can be clicked on and moved. "2" is the category score. "25% independent" is the category description from Rating Scale=. To clear the black highlighting, click somewhere else on the plot.

To delete the line corresponding to a category, double-click on it.

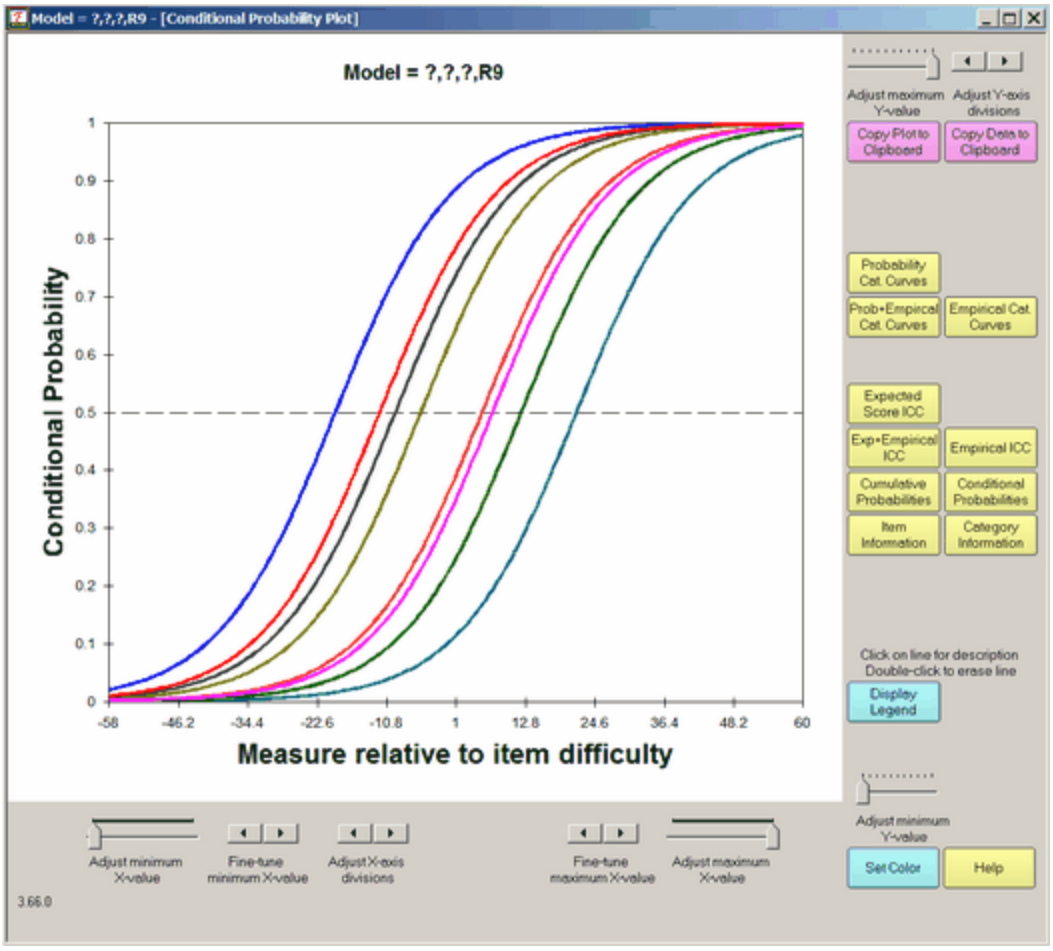
---

**Poisson Curves**




The Rasch-Poisson model has an infinity of categories. Facets computes 256 categories. Consequently measures are biased for observed categories above 200.

## 14.4 Conditional probability curves



Select by clicking on "Conditional Probabilities". 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.

You are looking at the conditional probability plot. Please ignore this plot unless you are concerned about this form of the Rasch model:

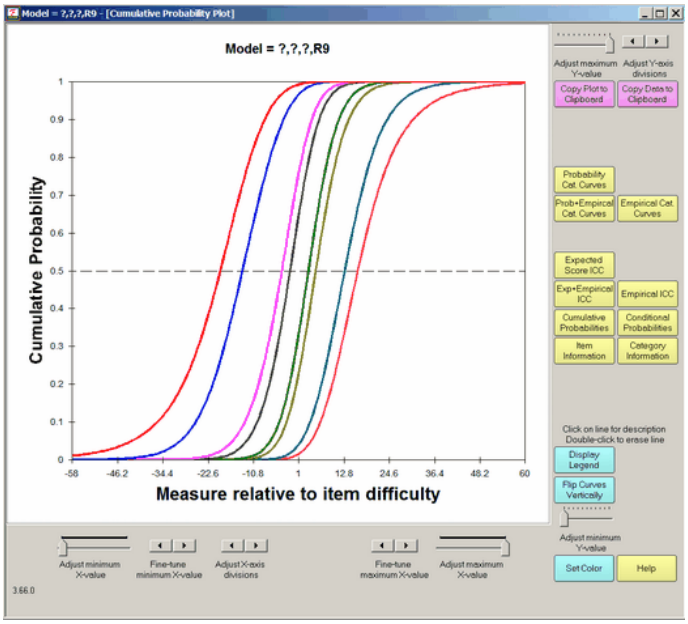
$\log (P_{nij} / P_{ni(j-1)}) = B_n - D_i - F_j$

If you are concerned about this form of the model, then the lines on the conditional probability plot are the parallel dichotomous ogives for the relationship:

Plotted line = $P_{nij} / ( P_{nij} + P_{ni(j-1)} )$

Buttons are described in Graph window.

## 14.5 Cumulative probabilities

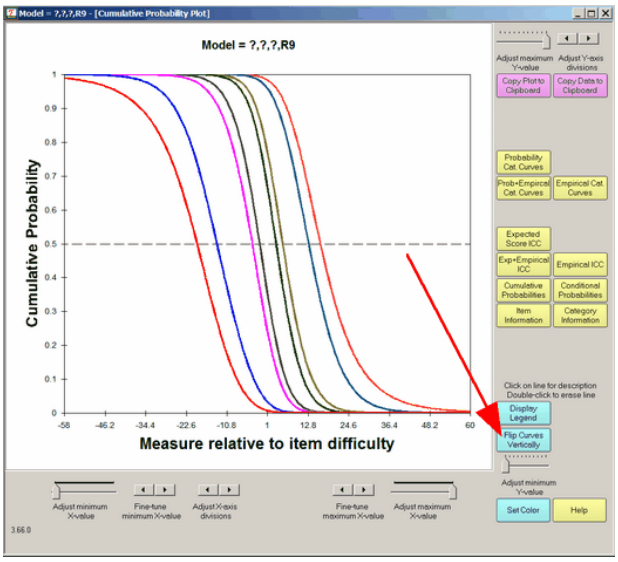



Select by clicking on "Cumulative Probabilities". Buttons are described in Graph window.

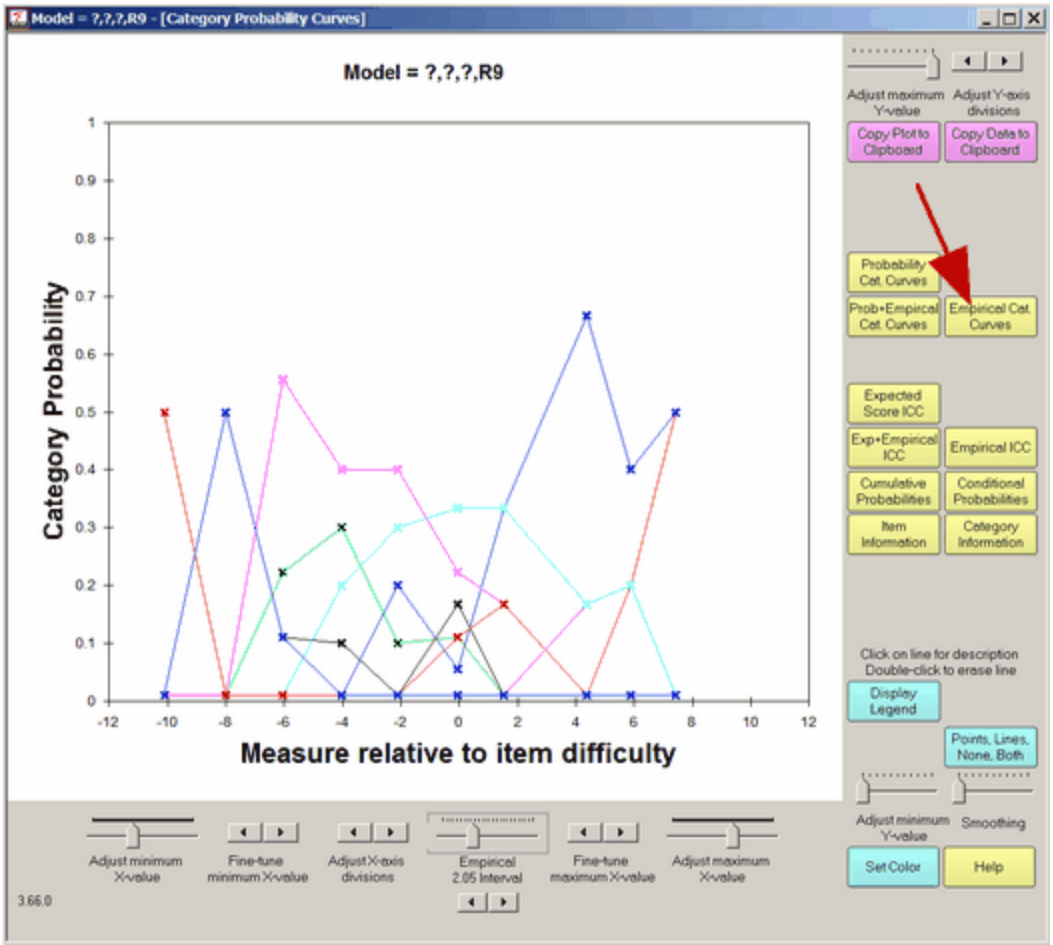
*Lines:* This shows the modeled category probability curves accumulated so that the left-hand curve (red) is the probability of being observed in the lowest category. The next curve to the right (blue) is the probability of being observed in the lowest or next lowest category. And so on further to the right.

*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.

Click on the "Flip Curves Vertically" button to change the vertical orientation of the curves.



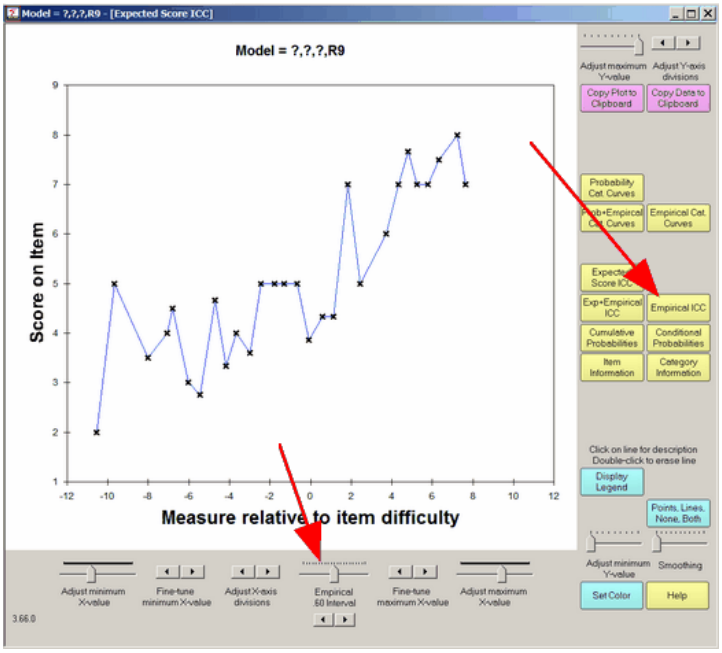

## 14.6 Empirical category curves - frequency polygons




These are the empirical (data-describing) category curves (frequency polygons). They are obtained by clicking on "Empirical Cat. Curves". The width of each empirical interval can be adjusted by the "Empirical Interval" slider. The smoothing of the empirical curves by means of cubic splines is adjusted by the smoothing control.

Buttons are described in Graph window.

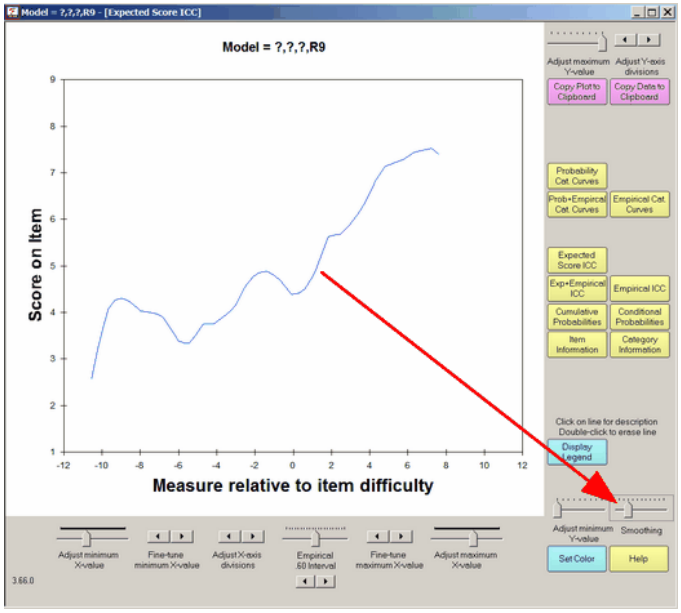
## 14.7 Empirical ICC or IRF



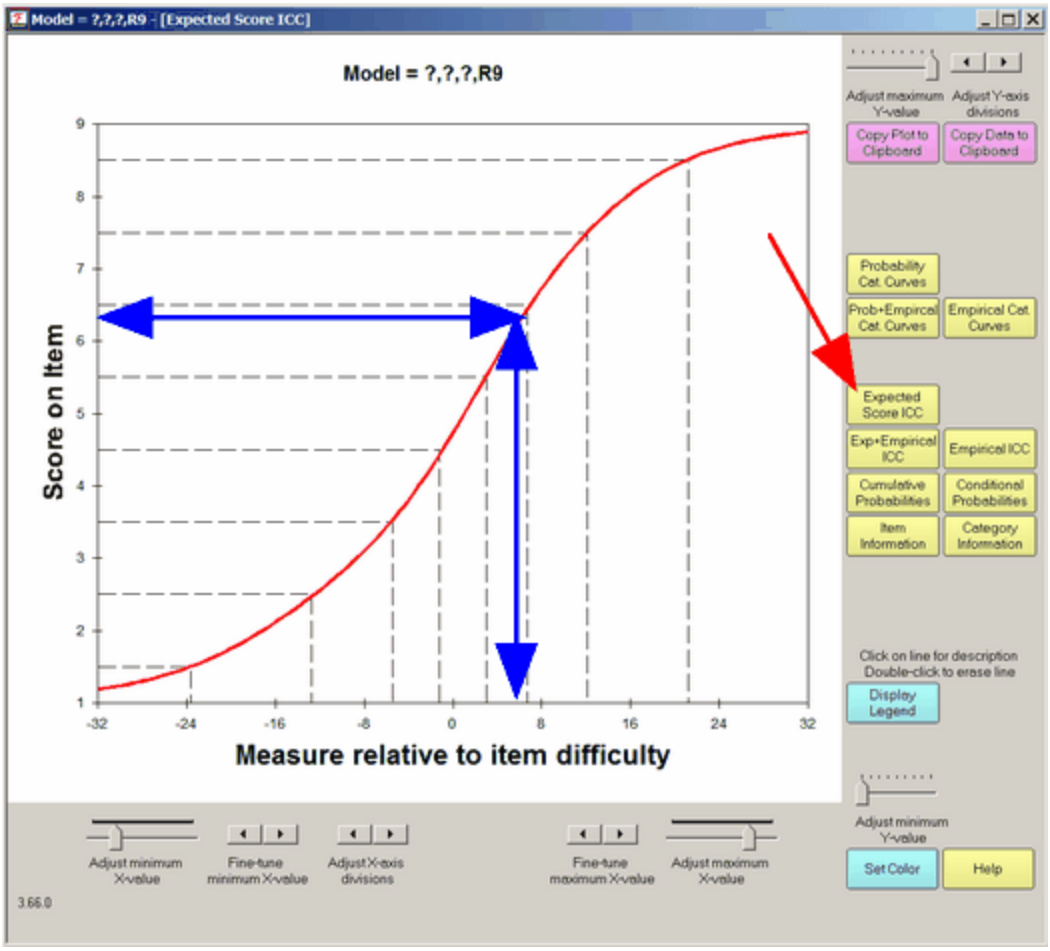

Empirical Item Characteristic Curve (ICC), also called the Empirical Item Response Function (IRF) is displayed. The control buttons are described in Graph window.

Change the Empirical interval (summarized to each "x") to make the plot the most meaningful.

Use the Smoothing slide to represent the points as a smooth curve:



## 14.8 Expected score ICC or IRF



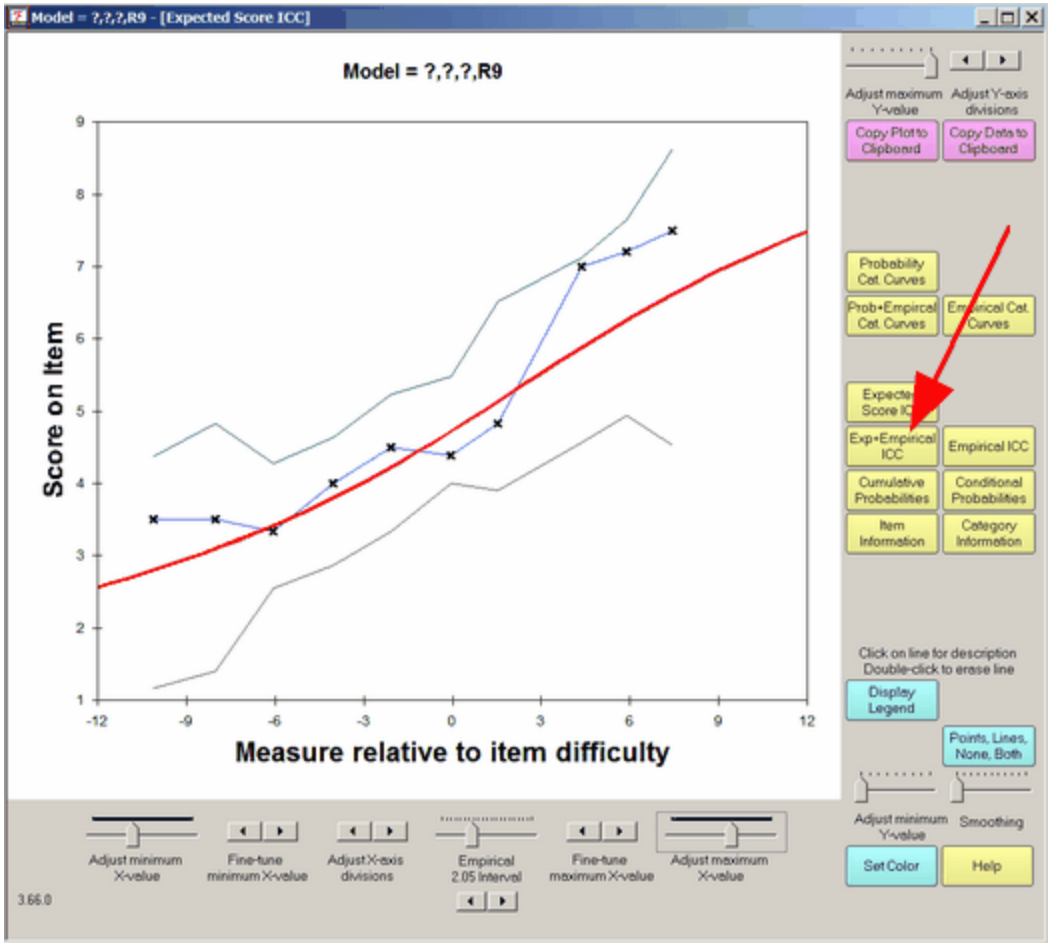

Select by clicking on "Expected Score ICC". The ICC is the Item Characteristic Curve, also called the IRF, the Item Response Function. **Expected Score ICC** plots the model-expected item characteristic curve. This shows the Rasch-model prediction for each measure relative to item difficulty Its shape 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 x-axis in this plot corresponds to the scales in the "vertical rulers", Table 6.0.0

Buttons are described in Graph window.

## 14.9 Expected+Empirical ICC or IRF

*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.




This shows the joint display of the expected and empirical Item Characteristic Curves, 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.

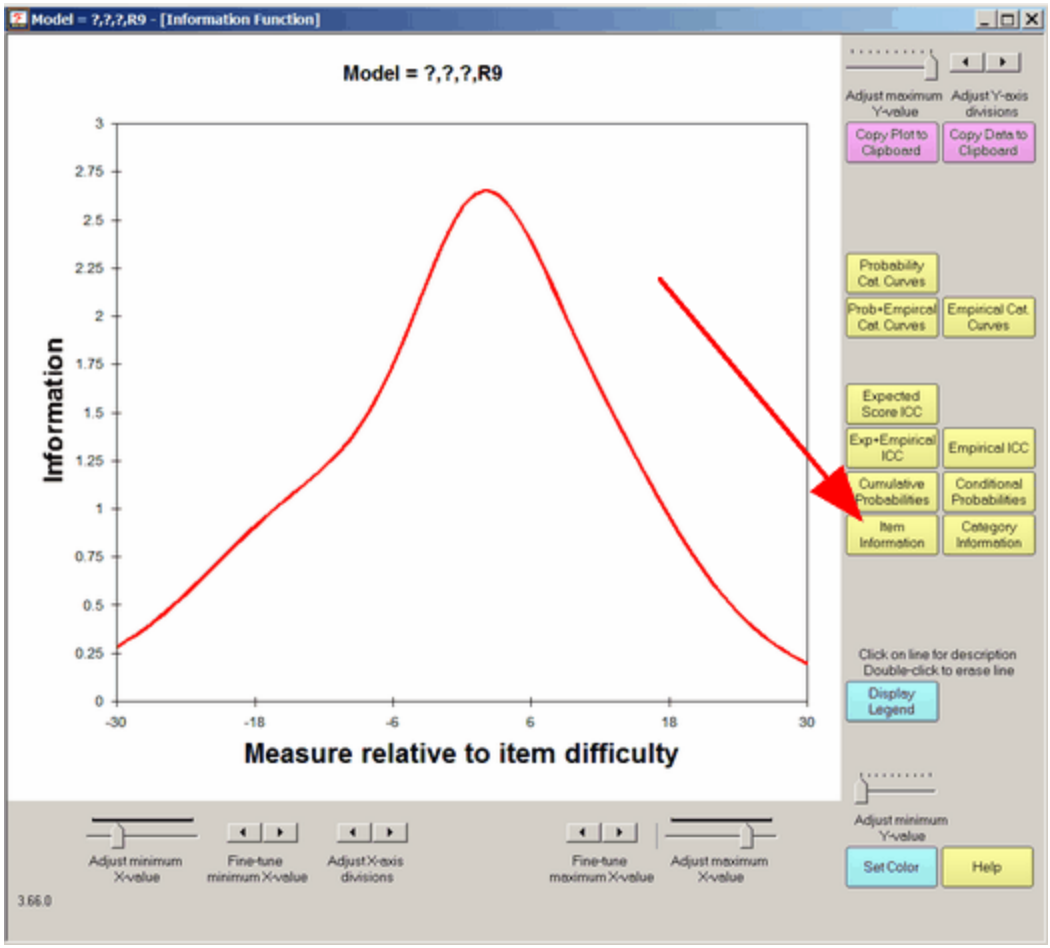
The solid red "model" line is generated by the relevant Rasch model. For a test of dichotomous items, these red curves will be the same for every item.

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.

## 14.10 Item information function

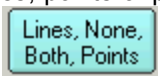



Select by clicking on "Item Information". This shows the (Ronald A.) Fisher measurement information in responses made to items. It is the same as the binomial variance (dichotomies) or the multinomial variance (polytomies).

Buttons are described in Graph window.

## 14.11 Points and lines

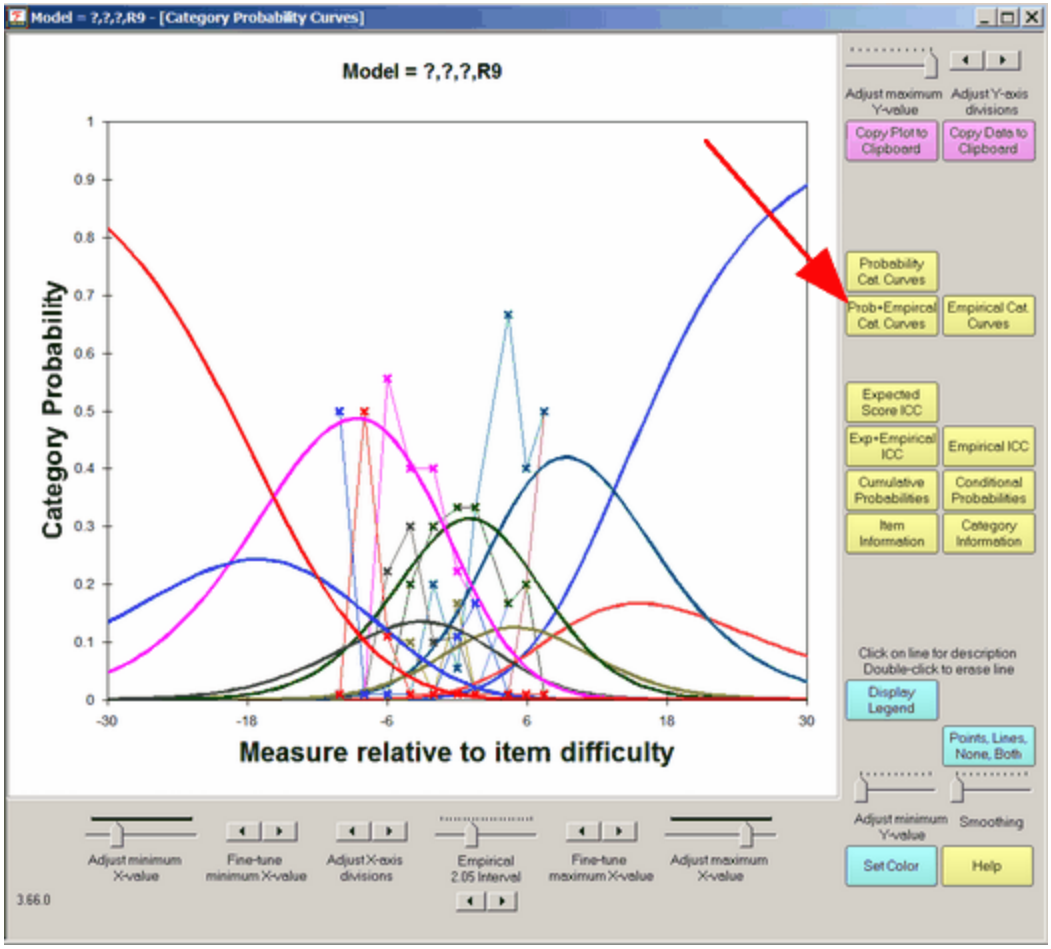
On the empirical curves you can control whether lines, points or points+lines are displayed.




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.

## 14.12 Prob.+Empirical Cat. Curves




These are the model category curves and the empirical (data-describing) category curves. The width of each empirical interval can be adjusted by the "Empirical Interval" slider. The smoothing of the empirical curves by means of cubic splines is adjusted by the smoothing control.

Buttons are described in Graph window.

## 14.13 Test information function

**Test information function:**
A test information function is not produced by Facets, because there is not a unique of a definition of a "test" in a many-facet situation.

You can construct your own test information function using an Excel spreadsheet.

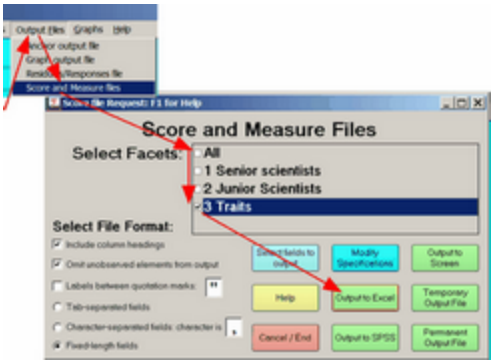
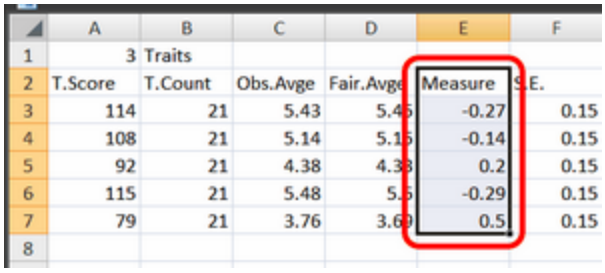
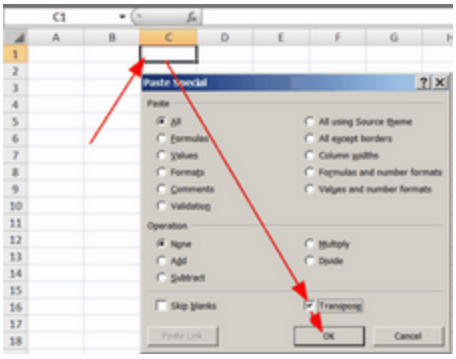
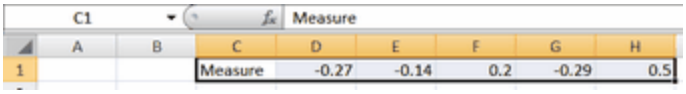
Open an Excel spreadsheet

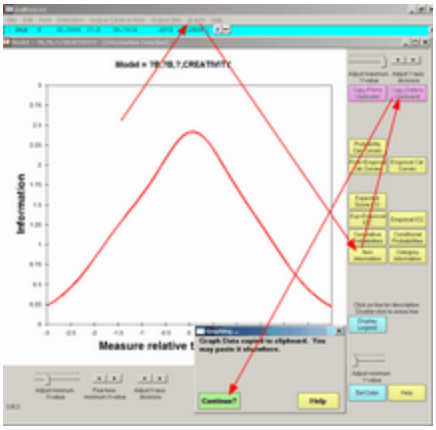
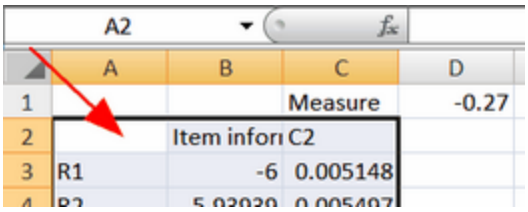
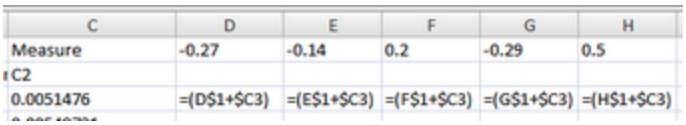
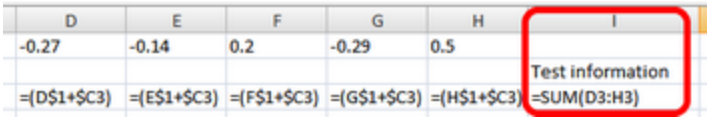
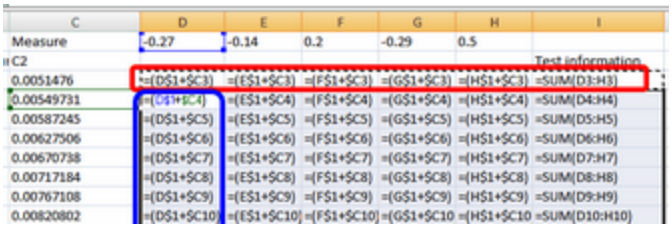
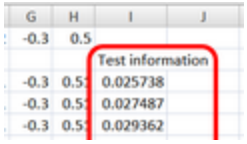
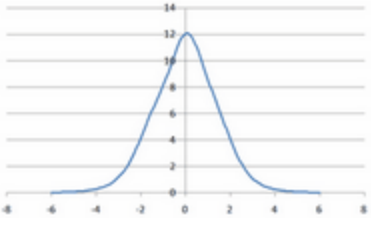
When the Facets information function display, "Copy data to clipboard"
Then "Paste" into an Excel worksheet.

In another the Excel worksheet
In column A, enter measures for a wide range, e.g., -10 logits to +10 logits every .1 logit
In column B, locate the item difficulty of your first item

copy the information function column from the first worksheet (which centers on 0), and paste into the second work sheet so that it centers on the item's difficulty.
In column C, locate the item difficulty of your second item
copy the information function column from the first worksheet (which centers on 0), and paste into the second work sheet so that it centers on the item's difficulty.
In column D, (same again), etc.

When all the item information functions are entered into the spreadsheet.
Sum the rows (except column A) to give the test information
Plot the sum against column A to see the test information function.

| | |
|---|---|
| 1. Suppose you want the test information function based on all the elements in facet 3 (the items) of Guilford.txt.<br><br>Perform your facets analysis of Guilford.txt | |
| 2. Facets menu bar<br>Click on "Output Files"<br>Click on "Score and measure files"<br>Click on "All" to uncheck it.<br>Click on your "test" facet<br>Click on "Output to Excel" |  |
| 3. The Excel spreadsheet displays with the measures in Column E.<br>Copy the measures. |  |
| 4. In a new Excel worksheet,<br>put cursor in cell C1<br>paste-and-transpose the measures into the top row. |  |
| 5. The first measure is in cell D1, going along the row |  |

| | |
|---|---|
| 6. Facets menu bar<br>Click on "Graphs"<br>Click on "Item information"<br>The item information function displays<br>Click on "copy data to clipboard"<br>"Graph data copied ..."<br>Click on "Continue?" |  |
| 7. Excel worksheet (with measures)<br>Paste into cell A2<br>Item information is in column C |  |
| 8. Excel<br>In cell D3, enter the formula<br>=(D$1+$C3)<br>Copy this cell along row 3 for all the measures |  <br>Press Ctrl+` (next to "1") to display the formulas |
| 9. At the end of row 3 "sum" cells D3 to the last measure column.<br>In the cell above the sum, enter "Test information" |  |
| 10. Copy the row of cells from D3 to the Sum.<br>Paste into all the other rows (down to about row 201). |  |
| 11. The test information function is in the right-hand "summed" column |  |
| 12. Scatter-plot the test information function.<br>x-axis is column B<br>y-axis is the "Test Information" column. |  |

# 15 Program Messages

## 15.1 Check! Warning! and other messages and their explanation

Message: *Cannot find the file '.... command.com'*

In your AUTOEXEC.BAT file or Windows Environment variables, check that COMSPEC= points to the correct location of *command.com*.
If so, please create the specified folder and copy *command.com* into it.
You can find *command.com* by using Find Files from the Start menu.

**Check (1)? Element not specified in Labels=: facet: ___ element: ___ in line ___ - see Table 2. Delements = N**

an observation has been matched to Model= statements, but then ignored because an element does not match a Labels= element. The element is listed Table 2 in the Output= file. Check for element numbers not listedafter Labels= or miscoded in the Data= file. Delements= specifies whether element numbers or element labels or both are expected in the data file.

Example: Element not specified in Labels=: facet: 1 element :13 in line 2594 - see Table 2. Delements = N
In the data, line 2594, facet 1 has element number 13.
In the Labels= list, under facet 1, there is no element 13. So the data is treated as unspecified. This element is shown in Table 2.
If you want to analyze it, add 13 to the element list for facet 1. Also any other similar elements.
Facets only reports the first unspecified element found, so Facets may report another element as unspecified after you account for element 13.

Example: R5 11001 106 1 1-3 4 3 4
Check (1)? Element not specified in Labels=: facet: 1 element: (unknown) in line 168 - see Table 2. Delements = N
In this line, blanks and tabs have both been used as separators. Use either blanks or tabs.

**Check (2)? Invalid datum location: "___" at or near line: ___. Datum "_" is too big or non-numeric, treated as missing.**
An observation has been treated as a missing response because it does not comply with the matching Model= statement. If the datum is meaningful, change the Model= statement, or recode the datum with a Rating (or partial credit) scale= specification. To avoid reporting missing-data codes, add them to Missing data codes=

**Warning (1)! Estimate out of range! Facet ___ Element: ___**
The current estimate of this element has exceeded the range of the mathematical routines. You may have elements with many responses, but almost extreme scores. You may have categories with very few responses. Contact us for assistance.

**Warning (2)! Subset checking bypassed.**
Specify "Subsets=Y" to verify that the data are connected, so that measures are in one frame of reference.

**Warning (3)! This is NOT referenced by a Model= specification**
Rating (or partial credit) scale= specifies a rating scale (or partial credit) that is not referenced in a Models=statement.

**Warning (4)! More than 32000 rating scales (or partial credit items) specified. Observation bypassed.**
The computational capacity of Facets has been exceeded, probably by using a "#" model for a facet with more than 32,000 elements.

**Warning (5)! Subset connection not verified.**
Facets did not run for enough iterations to check for subset connection. Specify "Subsets=Y" and "Iter=".

**Warning (6)! There may be ___ disjoint subsets**
See the explanation of connectedness. You need to redefine your facets, collect more data, or anchor some elements or groups.

**Warning (7)! Over-constrained? Noncenter= not in effect.**
There is no non-centered facet. The estimates may not converge. There may be column labeled "Displacement" in Table 7. This can occur when:
1) No non-centered facet is specified, e.g., Noncenter=0
2) The non-centered facet is not modeled, e.g.,
Noncenter=2
Models=?,,?,D; Facet 2 is not modeled
3) The non-centered facet is anchored, e.g.,
Noncenter=2
Labels=
2,Agents,A; Facets 2 is anchored or with ,D or ,G
1=Item 1,-2.3; Element 1 anchored at -2.3

**Warning (8)! Under-constrained? Noncenter= for two or more facets**
There may be ambiguity in the measures due to the facets moving relative to each other. Stability of measurement requires no more than one facet to be unconstrained for each data element. This problem often shows up in summary measures, such as by demographics. Only Noncenter= one demographic characteristic in each analysis.

**Warning (9)! Screen area is too small. Use Control Panel to adjust your Display.**
The bit-mapped graphics need high screen resolution. Please go to Control Panel - Display - Settings -Screen Area. And slide pointer to the right. 1024x768 or bigger are recommended.

**Warning (10)! No elements specified for facet ___**
A facet is defined in Labels= but immediately followed by *, another specification or end-of-file. Did you intend to define elements for the facet?

**Warning (11)! Plotted interactions pruned to first 250 per Excel plot**
An Excel file written by Facets would contain too many columns. It has been pruned to a valid number for most versions of Excel.

**Warning (12)! Duplicate ratingscale= ____**
Two rating scale= specifications define the same rating-scale name.

**Warning (13)! Measure values specified for elements in group-anchored facet ___ without a group number**
You may have omitted the group number after the measure value. Without a group number, the measure will be treated as a starting value.

**Warning (14)! Estimation diverged**
Facets was unable to obtain estimates for which the observed score = expected score. This may be because Noncenter= is not active or the estimates are anchored.

**Warning (15)! Duplicate element labels ignored for element matching: facet _ (label)**
Facets is trying to match an element identifier in the data with the element labels and has discovered that two elements in a facet have the same label or the same label abbreviation before :

**Warning (16)! Null element (Null=) is specified in Labels= for facet _ (facet number)**
If the Null= element is an active element in this facet, please specify a different element as the Null= element.

**Warning (17)! _ in "Delements=X" ignored, not N or L**
Delements= specifies whether element identifiers in the data are element numbers N, element labels L, or both NL LN

**Warning (18)! _ excess responses ignored**
There are too many responses in the data line. Extra responses are not processed.

**Warning (19)! at or near line ... * or number = value expected**
The list of elements in Labels= is expected to contain "facet number=facet label", then "element number = element label", then * at the end of the list for each facet. Something may be missing.

**Warning (20!) No labels found after Labels=**
A list of labels will be output in Table 2 for placement after Labels=
*Facets* builds a list to match the data file. Paste it into you specification file.

**Warning (21)! observation ...is weighted ... and is ignored**
If zero weighting is not deliberate, please check for extra commas in Models= and Labels= or R0 in the data.

**Warning: Active Excel columns to right of responses**
Please look at your Excel worksheet. There are non-blank columns in it on the right-hand-side of the columns of Facets data. If those columns are intentional, then please ignore the warning message. If those columns should not be there, then please delete those columns. If there are no non-blank columns, then please check that the Facets specifications are correct.

**Failure (___)! _____. Contact winsteps.com**
Internal program failure. Please email this information to winsteps.com along with your control and data files. These failures are rare, and we wish to remedy them immediately. You will be emailed how to download a corrected version of Facets.

## 15.2 Disk file problem

When Facets experiences difficulty with a disk file, it reports an error message.

**The problematic file is usually listed on the screen just before the error message.**

Can Facets read its input files or write to its output files and work files?

Work files are located in the working folder. Is the working folder write protected?

## 15.3 Disk space exhausted

1) Delete unneeded files to free up disk space.

2) Don't specify a *Residual output file=* or residual option on *Output file=.*

## 15.4 Memory problem

*Facets* dynamically assigns itself computer RAM.

(a) Terminate as many other programs as possible.

(b) Eliminate unused element numbers specified after Labels=

## 15.5 No valid matching responses found!

Facets stopped because did not find any useable observations in the data file:
no observations matched a Model= specification (other than M, missing)
or, matching the observations did not contain valid data values.

If the reason for failure to match is not obvious, the matching process may be diagnosed with the special *Query=Yes* specification.

## 15.6 Overflow problem

PC's come in many configurations, with and without math co-processors. Their mathematical capabilities vary considerably. If you experience an error message such as:

**Overflow** in module FAC265

then your PC does not have the mathematical range that Facets expects.

Try this solution:

*Facets* has a standard math range of "80". Try reducing this range in increments of 10 until the analysis runs. Use the Hardware= specification, e.g.,

```
Title=A Facets run
Hardware=70
Facets=3
```

## 15.7 Specification or data problem

When Facets cannot clearly understand your specifications or data, it stops with a diagnostic message such as:

```
Processing specifications from "SPEC.txt"
Specification is: $QUIT=YES
Error 1 in line 23: "Specification =" expected
Execution halted
Analyzed in time of 0: 0: 1 Output to spec.OUT.txt
```

Each such message has a number, "Error 1", a location "line 23", and a diagnosis, "Specification =" expected. In this case, the specification is "$QUIT=YES". Facets does not know this specification and so stops. Remove or alter the specification in your specification file, and start Facets again.

Control files with specifications like $QUIT are intended for use with Facform, those with &INST and &END are intended for use with Winsteps or Ministep.

---

**Facets expects its input files to be in MS-DOS text file format, i.e., readable by NotePad. Mac and UNIX end-of-line codes are allowed: CR, LF, CRLF and LFCR. CR = Carriage Return. LF = Line Feed.**

UTF-8 BOM (byte-order marks) are allowed, but not required.

Other file formats:

**{\rtf1\....**
Your control file is in RTF, not TXT format. Double-click on the file, and "Save as" "text with line breaks" or "MS-DOS format"

**ÿþT I T L E =**
Your control file is in Unicode, not ANSI coding. Double-click on the file, and "Save as" with "Encoding" "ANSI" or "MS-DOS format"

**þÿ T I T L E =**
Your control file is in Unicode Bigendian, not ANSI coding. Double-click on the file, and "Save as" with "Encoding" "ANSI" or "MS-DOS format"

## 15.8 Unexplained program failure

For Windows-related and other problems, see www.winsteps.com/problems.htm

If you experience any inexplicable difficulties, or you have any suggestions of your own, please contact us using the form at www.winsteps.com. We want to resolve any problems. We take great pains to bring you the most reliable Rasch analysis software. We also try to keep abreast of all the latest developments. Consequently Facets is constantly being revised and improved. If necessary, we will send you an updated version of Facets.

*Facets* is innovative software breaking new ground in statistical theory. A guiding design principle has been that a useful solution now is better than a perfect solution later. Consequently, many of the statistical niceties that are taken for granted with 70-year-old statistical techniques are not yet implemented.

In the Hegelian conflict between theory and data, *Facets* regards intentional measurement-oriented, prescriptive and deliberate theory as preeminent over inevitably ambiguous data. Conventional statistics regard deterministic data as dominant over arbitrary and accidental statistical models. Since Rasch theory embodies what we demand of our data in our endeavor to construct meaning, failure of the data to generate satisfactory measures implies gross failure in the data, not in the model. This way of thinking is quite unfamiliar to those accustomed to conventional statistical methodology.

**Facets on the Mac** - Facets is 32-bit software. Apple say: "Apple has been working with developers to transition their apps, and in 2018 Apple informed them that macOS Mojave would be the last version of macOS to run 32-bit apps." - *But they did not tell me! So, sorry, Facets does not run on the current macOS.*

# 16 Special Topics

## 16.1 Analysis - step-by-step

1). Conceptualize your analytical model. Facets is based on an additive model:

observation = person ability - item difficulty - rater severity - time-point-challenge - .....

What is your additive model?

2). "Dummy facets" which aren't part of the additive model, but are required in order to investigate fit and interactions, such as, "rater gender (sex)".

Do you have any dummy facets? These will be anchored at 0 logits in the analysis, so that they don't alter the additive model.

3). Number all the facets:

1. Persons
2. Items
3. Raters
4. Time-points
5. .....

6). Within each facet, there is a list of elements:
Persons: Fred, George, Mary, ...
Items: ...
Raters: ...

7). Construct the Excel spreadsheet with one column for each facet and one column for the observation.
One facet can have range, such as 1-4, so that there are 4 columns of observations.

8). Construct the Labels= list of facets and elements.

9). Put in anchor values of 0 for the elements of the dummy facet.

10). Write the Models= specifications that match your analysis.

11). Complete the Facets specifications.

12). Run the analysis.

## 16.2 Anchoring rating scale structures

Anchoring (fixing) the Rasch-Andrich thresholds (step calibrations) of a rating-scale (or partial-credit scale) enables rating scales in one analysis to be equated to rating scales in another analysis.

Anchoring also enables predetermined threshold values to be imposed on an analysis.

**Anchoring for equating**

Obtain the Rasch-Andrich thresholds (step calibrations) from Table 8 or from the Anchorfile=.

```
Table 8.1  Category Statistics.
Model = ?,?,?,?,R12
+-------------------------------------------------------------
|          DATA            |  QUALITY CONTROL  |RASCH-ANDRICH|
|Category      Counts  Cum.| Avge  Exp.  OUTFIT| THRESHOLDS  |
|   Score    Used   %    % | Meas  Meas   MnSq |Measure  S.E.|
|--------------------------+-------------------+-------------+-
|  5   0        1   0%   0%| -6.49 -6.84   .7 |             |
|  6   1        5   2%   2%| -5.89 -6.03   .5 | -7.59   1.03|
|  7   2       16   5%   7%| -3.60 -3.39   .7 | -5.82    .58|
|  8   3       33  11%  19%|  -.85  -.80   .8 | -2.63    .37|
|  9   4       72  24%  43%|  1.26  1.38   .7 |  -.47    .25|
| 10   5       93  31%  74%|  3.59  3.56   .8 |  2.20    .19|
| 11   6       59  20%  94%|  6.36  6.19  1.3 |  5.37    .22|
| 12   7       17   6% 100%|  9.08  9.08   .9 |  8.94    .36|
+-------------------------------------------------------------
```

Rating scale = Myscale, R12
5 = bottom, 0, A ; the bottom category has the dummy anchor-value of 0.
6 = , -7.59, A
.....
12 = , 8.94, A
*

**Pivot points**
The Rating scale= specification provides the capability to choose the pivot point in a dichotomy, rating scale, partial credit or other polytomous structure.

*Example: I am setting the standard for a "Competent" essay-writer. I need to anchor the mean for the essays at a particular point on the scale - "good" (6) and/or halfway between 5 and 6. How do I do this?*

Here is Table 8 from an unanchored run of Essays.txt. Looking at the "Rasch-Andrich Thresholds" (step calibrations), the standard 0 logit point for the scale is the point on the latent variable at which categories 1 and 9 (the lowest and highest categories) are equally probable.

```
-------------------------------------------------------------------------------------------
------------
```

```
|      DATA            |  QUALITY CONTROL  |RASCH-ANDRICH|  EXPECTATION  |  MOST  |.5 Cumul.|
Cat|Response|
| Category Counts  Cum.| Avge  Exp.  OUTFIT| THRESHOLDS  |  Measure at   |PROBABLE|Probabil.|
PEAK|Category|
|Score   Used   %    % | Meas  Meas   MnSq |Measure  S.E.|Category  -0.5 |  from  |    at   |
Prob|  Name  |
-------------------------------------------------------------------------------------------
-----------
|  1        4   4%   4%|  -.86   -.74   .8 |             |( -2.70)       |   low  |   low   |
100%| lowest |
|  2        4   4%   8%|  -.11   -.58  2.7 |  -.64    .53|  -1.65   -2.21|        |  -1.75  |
17%|        |
|  3       25  24%  31%|  -.36*  -.40   .9 | -2.32    .39|   -.93   -1.26|  -1.48 |  -1.39  |
48%|        |
|  4        8   8%  39%|  -.43*  -.22   .5 |   .83    .25|   -.41    -.66|        |   -.46  |
11%|        |
|  5       31  30%  69%|  -.04   -.03   .8 | -1.48    .24|    .02    -.19|   -.32 |   -.29  |
39%| middle |
|  6        6   6%  74%|  -.46*   .17  4.1 |  1.71    .25|    .44     .23|        |    .34  |
9%| six    |
|  7       21  20%  94%|   .45    .34   .6 | -1.00    .26|    .94     .68|    .35 |    .47  |
47%|        |
|  8        3   3%  97%|   .75    .50   .5 |  2.36    .44|   1.62    1.24|        |   1.37  |
16%|        |
|  9        3   3% 100%|   .77    .62   .8 |   .54    .60|(  2.69)   2.17|   1.45 |   1.70  |
100%| highest|
-----------------------------------------------------------(Mean)---------(Modal)--
(Median)---------------
```

If you want the 0 logit point to be that at which categories 5 and 6 are equally probable, then

Rating scale = Creativity,R9 ;Creativity is a rating scale with possible categories 0 to 9
1 = lowest ; name of lowest observed category
5 = middle ; no need to list unnamed categories
6 = six, 0, A ; anchor point at which 5 and 6 are equally probable at 0 logits
9 = highest ; name of highest observed category
*

If you want the 0 logit point to be that at which category 6 is most probable or at which 6 is the expected average rating (- these are the same point on the latent variable):

Rating scale = Creativity,R9 ;Creativity is a rating scale with possible categories 0 to 9
1 = lowest ; name of lowest observed category
5 = middle ; no need to list unnamed categories
6 = 1.27, 0, A ; anchor point to make expected measure of 6 at 0 logits
; 1.27 = "Rasch-Andrich threshold for 6" (1.71) - "Expectation" for 6 (.44)
9 = highest ; name of highest observed category
*

If you want the 0 logit point to be that at which 5.5 is the expected average rating:

Rating scale = Creativity,R9 ;Creativity is a rating scale with possible categories 0 to 9
1 = lowest ; name of lowest observed category
5 = middle ; no need to list unnamed categories
6 = 1.48, 0, A ; anchor point to make expected measure of 6 at 0 logits
; 1.27 = "Rasch-Andrich threshold for 6" (1.71) - "Expectation -0.5" for 6 (.23)
9 = highest ; name of highest observed category
*

Example 2: I want to set the person-item targeting on dichotomous items at 2 logits offset (= 88% probability of success) instead of the usual 0 logits offset (=50% chance of success). This will make all the persons appear to be 2 logits less able.

model = ?,#,MyDichotomy
rating scale = MyDichotomy, R2
0 = 0, 0, A ; place holder
1 = 1, 2, A ; anchored at 2 logits
2 = 2
*

```
+-------------------------------------------------------------------------------------------------
------------+
|      DATA            |  QUALITY CONTROL  |RASCH-ANDRICH|  EXPECTATION  |  MOST  |  RASCH-  |
Cat|Response|
| Category Counts  Cum.| Avge  Exp.  OUTFIT| Thresholds  |  Measure at   |PROBABLE| THURSTONE|
PEAK|Category|
|Score   Used   %    % | Meas  Meas   MnSq |Measure  S.E.|Category  -0.5 |  from  |Thresholds|
Prob|  Name  |
|----------------------+-------------------+-------------+---------------+--------+----------
+----+--------|
|  0       27  79%  79%|  -.79   -.67   .6 |             |(   .87)       |   low  |   low    |
100%| 0      |
|  1        7  21% 100%|  2.34   1.87   .3 |  2.00A      |(  3.07)   1.99|   2.00 |   1.99   |
100%| 1      |
+-------------------------------------------------------------------------------------------------
------------+
```

## 16.3 ANOVA-like Analysis

Statistical hypothesis tests can easily be performed on the resulting descriptive measures.

1) Do the genders (sexes) differ significantly in their ability?
The Fixed (all-same) chi-squared tests this.

2) Contrasts: If there were more than two races, any two could be compared by a t-test:
t = (Measure1 - Measure2)/ sqrt(SE1²+SE2²)

3) Interactions: Are there significant Race by Sex effects?
Use Model=?,?B,?B,,D
and the Bias Table to investigate these.

## 16.4 BASIC program to build Facets data and specification files

Here is an example program, which writes Facets data records and also builds a table of element labels for the specification file, using this example data file.

123 665 25467
452 339 34245
312 034 34452
columns 1-3, "123" in record 1, are examinee identifiers
columns 5-7, "665" in record 1, are judge identifiers
columns 9-13, "25467" in record 1, are the ratings on items 1 to 5.

```
100 DIM JUDGE$(100), examinee$(1000) 'to hold elements
110 OPEN "raw.dat" FOR INPUT AS #1 'input the raw data
120 OPEN "dat.txt" FOR OUTPUT AS #2 'output the Facets data file
130 OPEN "spec.txt" FOR OUTPUT AS #3 'output part of the specification file
140 LASTITEM% = 5 'there are 5 items
```

```
150 IF EOF(1) THEN 340 'are we at end-of-file ?
160 LINE INPUT #1, l$ 'read in a record from the raw data file
170 FOR J% = 1 TO LASTJUDGE% 'look down list of judges
180 IF MID$(l$, 5, 3) = JUDGE$(J%) THEN 220 'have we met the judge in cols 5-7 already?
190 NEXT J%
200 LASTJUDGE% = LASTJUDGE% + 1: J% = LASTJUDGE% 'no, add a new judge
210 JUDGE$(J%) = MID$(l$, 5, 3) 'add new judge to the array
220 FOR E% = 1 TO LASTEXAMINEE% 'look down list of examinees
230 IF MID$(l$, 1, 3) = EXAMINEE$(E%) THEN 270 'have we met the examinee in cols 1-3?
240 NEXT E%
250 LASTEXAMINEE% = LASTEXAMINEE% + 1: E% = LASTEXAMINEE% 'no, add a new examinee
260 EXAMINEE$(E%) = MID$(l$, 1, 3) 'add new examinee to the array
270 RATING$ = "" 'clear rating hold area
280 FOR I% = 1 TO LASTITEM% 'now append ratings, in cols 9-13, one for each item
290 RATING$ = RATING$ + "," + MID$(l$, 8 + I%, 1)
300 NEXT I%
310 PRINT #2, STR$(J%) + ",1-" + STR$(LASTITEM%) + "," + STR$(E%) + RATING$  'output data
320 GOTO 150 'go back for next record
330  ' now write out the element labels into the specification file
340 PRINT #3, "Title=Example conversion" ' Facets specifications
350 PRINT #3, "Facets=3"
360 PRINT #3, "Model=?,?,?,R"
370 PRINT #3, "Data=dat.txt"
380 PRINT #3, "Labels="
390 PRINT #3, "1,Judges"
400 FOR J% = 1 TO LASTJUDGE%
410 PRINT #3, STR$(J%) + "=" + JUDGE$(J%) 'the judge elements
420 NEXT J%
430 PRINT #3, "*"
440 PRINT #3, "2,Items"
450 FOR I% = 1 TO LASTITEM%
460 PRINT #3, STR$(I%) + "=" + STR$(I%) 'the item elements
470 NEXT I%
480 PRINT #3, "*"
490 PRINT #3, "3,Examinees"
500 FOR E% = 1 TO LASTEXAMINEE%
510 PRINT #3, STR$(E%) + "=" + EXAMINEE$(E%) 'the examinee elements
520 NEXT E%
530 PRINT #3, "*"
540 CLOSE
550 STOP
```

## 16.5 Bayesian Logistic Regression

The Observational Medical Outcomes Partnership (OMOP) of the Foundation for the National Institute of Health (FNIH) researches methods for analysis of healthcare databases. One of their methods is "Bayesian Logistic Regression"

Here is their description of the method:

## Statistical Model

Let's assume that $Y$ is a response variable that takes on two possible values, 0 and 1, and $x_1, x_2, ..., x_n$ are covariate variables. In our program $Y$ is an indicator of the outcome of interest (condition) and $x_1, x_2, ..., x_n$ are indicators of drugs and, possibly, of some other categorical variables (like age or sex groups). For $n$ explanatory variables the model is:

$$\log \frac{\Pr(Y=1)}{1-\Pr(Y=1)} = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n.$$

Estimation of the $i$-th coefficient, $\hat{\beta}_i$, can be interpreted as an approximation to the log-odds ratio, $\log\left(\frac{\Pr(Y=1|x_i=1)}{1-\Pr(Y=1|x_i=1)} / \frac{\Pr(Y=1|x_i=0)}{1-\Pr(Y=1|x_i=0)}\right)$, that characterizes the influence of the $i$-th covariate adjusted for all other explanatory variables in the model.

In the Bayesian interpretation one puts priors on the regression coefficients. Such an approach allows to avoid overfitting by assuming that each regression coefficient $\beta_i$ is likely to be close to 0; see, for example, [2, 5] for more details.

The blue box is a dichotomous Facets-type Rasch model with n+1 facets. n facets have two elements 0 and 1. The B0 facet has one element, non-centered.

The Red box is an artificial centralizing bias on the element measures. They could obtain a similar effect by using dummy weighted data in Facets.

Genkin, A., Lewis, D., Madigan, D., (2007). Large-Scale Bayesian Logistic Regression for Text Categorization, *Technometrics, 49,* 3, 291–304.

---

Example: with 2 covariates, $x_1$, $x_2$

Data from "Bayesian Logistic Regression for Medical Claims Data", Ivan Zorych, Patrick Ryan, David Madigan, 5-24-2010, The 33rd Annual Midwest Biopharmaceutical Statistics Workshop, Muncie, Indiana. www.mbswonline.com/upload/presentation_6-5-2010-6-23-40.pdf

```
Title=BLR
Newton = 0.5 ; use Newton-Raphson iteration
Converge = .000001, .0000001 ; with very tight convergence
Facets = 3
Noncenter = 1
Arrange = N
Models=
?,?,?,D        ; Nausea (Y) =1, No nausea = 0
*
Keepzero=999
Labels=
1, Noncenter
1, B0
*
2, Rosinex (x1), A ; published size: B1 = 4.4
0, 0, 0    ; No Rosinex
1, B1      ; Rosinex
```

```
*
3, Ganclex (x2) , A  ; published size: B2 = 0.0
0, 0, 0   ; No Ganclex
1, B2    ; Ganclex
*
Data=
R81, 1, 1, 1, 1 ; non-center, Rosinex, Ganclex, Nausea
R9,  1, 1, 1, 0
R1,  1, 0, 1, 1
R9,  1, 0, 1, 0
R9,  1, 1, 0, 1
R1,  1, 1, 0, 0
R90, 1, 0, 0, 1
R810,1, 0, 0, 0

; dummy data to centralize the estimates
; R3.5, 1, 1, 1, 1, 1, 1 ; weighting of 3.5 observations of Y=1
; R3.5, 1, 1, 1, 1, 1, 0 ; weighting of 3.5 observations of Y=0
```

## 16.6 Bias interaction DIF DPF DRF estimation

After estimating the measures, Facets checks to see if any Model= specifications include Bias specifiers, "B". If so, for each such model, the specified Bias interaction is estimated for all the data (not just the data matching that particular model). Bias can be due to any type of interaction including Differential Item Functioning DIF, Differential Person Functioning DPF, Differential Rater Functioning DRF.
This is done by iterating through the data again and, after convergence, doing one further iteration to calculate statistics.

Computation of interactions is a two-stage process.
1. The measures for the elements, and the structure of the rating scale, are estimated. Then those values are anchored (fixed, held constant).
2. The expected values of the observations are subtracted from the observed values of the observations, producing the residuals.
3. The residuals corresponding to each interaction term (e.g., examinees rated by judge 4) are summed. If this sum is not zero, then there is an interaction.
4. The size of the interaction is estimated. A first approximation is:
Interaction (logits) = (sum of residuals) / (sum of the statistical-information in the observations).

Algebraically, first the Bn, Di, Cj, Fk are estimated using a Rasch model such as:

$$\log ( P_{nijk} / P_{nij(k-1)}) = B_n - D_i - C_j - F_k$$

Then the Bn, Di, Cj, Fk are anchored, and the bias/interaction terms, e.g., Cij, are estimated:

$$\log ( P_{nijk} / P_{nij(k-1)}) = ( B_n - D_i - C_j - F_k ) - C_{ij}$$

Thus the Cij are estimated from the residuals left over from the main analysis. The conversion from residual score to bias interaction size is non-linear. Bias sizes may not sum to zero.

Bias, (also called interaction, differential item function, differential person function, etc.,) estimation serves several purposes:

1) in diagnosing misfit:
The response residuals are partitioned by element, e.g., by judge-item pairs, and converted into a logit measure. Estimates of unexpected size and statistical significance flag systematic misfit, focusing the misfit investigation.

2) in investigating validity:

A systematic, but small, bias in an item or a judge, for or against any group of persons, may be overwhelmed by the general stochastic component in the responses. Consequently it may not be detected by the usual summary fit statistics. Specifying a bias analysis between elements of facets of particular importance provides a powerful means of investigating and verifying the fairness and functioning of a test.

3) in assessing the effect of bias:
Since bias terms have a measure and a standard error (precision), their size and significance (t-statistic) are reported. This permits the effect of bias to be expressed in the same frame of reference as the element measures. Thus each element measure can be adjusted for any bias which has affected its estimation, e.g., by adding the estimate of bias, which has adversely affected an element, to that element's logit measure. Then the practical implications of removing bias can be determined. Does adjustment for bias alter the pass-fail decision? Does adjustment for bias affect the relative performance of two groups in a meaningful way?

4) in partitioning unexplained "error" variance:
The bias logit sample standard deviation corrected for its measurement error, can be an estimate of the amount of systematic error in the error variance (RMSE).
e.g., for a bias analysis of judges,
Bias logit S.D. = 0.47, mean bias S.E. = 0.32 (Table 13),
so "true" bias S.D. = $(0.47^2 - 0.32^2) = 0.35$ logits,
but, this exceeds the RMSE for judges = 0.12 (Table 7).
Here, locally extreme judge-person scores cause an overestimation of systematic bias.

---

**Adjusting for bias:**

A straight-forward approach is to define the biased element as two elements: one element for one subset of judges (examinees, etc.) and a different items for the other subset of judges (examinees, etc.). This can done by defining an extra item element, and then adjusting item references in the data file accordingly.

Example:

```
Facets = 4 ; Items, candidates, examiners, bias adjustment
Non-center = 2 ; candidates float

Models =
?, 28, 17, 1, myscale ; allow for bias adjustment between candidate 28 and examiner 17
?,  ?, ?, 2, myscale
*

Rating scale = myscale, R9

Labels=
1, Items
...
2, Candidates
...
3, Examiners
....
4, Bias adjustment, A
1, 28-17 adjustment ; the bias will be absorbed by this element, relative to element 2.
2, Everyone else, 0
*

Data=
1-5, 28, 17, 1, 1,2,3,4,5
1-5, 29, 23, 2, 5,4,3,2,1
.....
```

---

**Non-Uniform DIF**

Create a dummy facet for ability levels: then a three-way interaction between person group dummy facet, ability level dummy facet, and item (or whatever).
Models=?,?B,?B,?,?B,R3  (or whatever)

---

**Paired-Comparison Bias/Interaction Analysis:**

Facets produces meaningful numbers in the Bias/interaction analysis when:
1) Use mirrored data, but set the weight = 1.0, instead of 0.5. For the main analysis, use weight 0.5.
2) Arrange the data so that the Models= is  ..., -?,?,... instead of ...,?,-?,...

## 16.7 Convergence problems

**"*Facets* runs and runs, but does not converge! What has gone wrong?"**

Are the current set of estimates good enough for your purposes? Estimates with extremely high precision are rarely needed, and seldom improve overall fit of the data to a Rasch model. Set Convergence= and Iterations= to reasonable values.

If the changes per iteration are very small or are oscillating up and down about equally, then the estimates are as good as they can be. Press Ctrl+F to stop iteration and move on to fit computation. You can also use the Estimation menu to make the changes per iteration smaller, or to change the estimation method.

If Facets runs for more than 100 iterations, and the residuals are not converging (approaching zero), then please force Facets to stop iterating by Ctrl+F (or use the Estimation pull-down menu) - then look at the screen and the output file. Investigate elements with large displacements and rating-scale categories with large "Observed - Expected Diagnostic Residuals".

There are numerous possibilities. Here are some to check.

1) **Noncenter=**
   The analysis may be
   i) over-constrained because all facets are centered, anchored or group-anchored,
   ii) under-constrained because more than one facet is non-centered.
   For most analyses, noncenter= must specify one active facet.
   Make sure that facet is specified is not marked X or omitted from the model statement.
   Makes sure that the non-centered facet does not include anchor or group-anchor values.

2) **Null element= (Keepasnull=)**
   Element 0 in the data is usually a dummy element that means "this facet does not apply to this observation". If element 0 is an active element in your analysis, then please assign a different unused element number as the dummy element number, e.g., *Null element = 999*

3) **Your facets are nested**.
   When one facet is nested within another facet (without anchoring or group-anchoring), Facets makes an arbitrary allocation of statistical information between the facets, this can lead to unstable estimation. For instance, if there is a "student" facet and also a "student gender" (sex) facet, then nesting needs to be resolved by anchoring or group-anchoring either the "student" facet or the "student gender" facet.

4) **Low category frequencies**
   Rating (or partial credit) scale categories with very low frequencies (less than 10 observations) are difficult to use

as the basis of estimation. Convergence may take a long time.

5) **Rating (or partial credit) scales with many categories.**
Estimating the parameters of long rating (or partial credit) scales (more than 10 categories) is difficult. Adjustments tend to ripple up and down the scale, like a caterpillar moving. Convergence may take a long time.

6) **Category frequencies are lumpy.**
When some categories of a rating scale have high frequencies and other categories have low frequencies, convergence may take a long time.

7) **Clumps of data or stringy data due to missing data. Disconnected or weakly connected subsets.**
Data with disjoint subsets of data, or with only thin connections between the subsets, may not converge, or may converge to non-repeatable values. For instance, your judging design may be very sparse.

8) **The Maximum Likelihood curve is almost flat.**
The data are such that there are a range of estimated measures that match the data equally well. All are equally valid.

9) **The Maximum Likelihood curve has two peaks very close together.**
The data are such that there are two almost equally good sets of estimates that match the data. Facets cannot choose between them.

## 16.8 Cut-off points

Setting pass-fail cut-off points requires a precise conceptualization of the cut-off point. There are many possible conceptualizations. Here is one that is usually easy to explain:

In the "Essays" example, we want a cut-off point for the examinees between "5" and "6" on the rating scale.

So let's conceptualize this as:
"1,000 examinees whose performance is at the cut-off point between 5 and 6 would have an average rating of 5.5 on a "standard" (average) essay when rated by a standard (average) reader."

```
Table 7.2.1  Essay Measurement Report  (arranged by MN).
-------------------------------------------------------------------------------------------------
         --
|  Obsvd  Obsvd  Obsvd  Fair-M|        Model | Infit      Outfit   |Estim.|
         |
|  Score  Count Average Avrage|Measure  S.E. | MnSq ZStd  MnSq ZStd|Discrm| N Essay
         |
-------------------------------------------------------------------------------------------------
         --
|  1843.0   384.0   4.8   4.73|    .00   .04 | 1.00  -.2   .99  -.3|      | Mean (Count: 3)
         |

Table 7.3.1  Reader Measurement Report  (arranged by MN).
-------------------------------------------------------------------------------------------------
         -----------------
|  Obsvd  Obsvd  Obsvd  Fair-M|        Model | Infit      Outfit   |Estim.| Exact Agree. |
                        |
|  Score  Count Average Avrage|Measure  S.E. | MnSq ZStd  MnSq ZStd|Discrm| Obs %  Exp % | Nu
        Reader           |
-------------------------------------------------------------------------------------------------
         -----------------
|   460.8    96.0   4.8   4.73|    .00   .08 | 1.00  -.1   .99  -.2|      |              | Mean
        (Count: 12)    |
```

According to Table 7:
The average difficulty of the essays is 0.00 - this would be the "standard" essay difficulty

The average reader has a severity/leniency of 0.00

According to Table 8:

```
Table 8.1  Category Statistics.
-------------------------------------------------------------------------------------------------
---
|      DATA            |  QUALITY CONTROL  |RASCH-ANDRICH|  EXPECTATION  |  MOST  |.5 Cumul.|
Cat|
| Category Counts  Cum.| Avge  Exp.  OUTFIT| THRESHOLDS  |  Measure at   |PROBABLE|Probabil.|
PEAK|
|Score   Used   %    % | Meas  Meas   MnSq |Measure  S.E.|Category  -0.5 |  from  |    at   |
Prob|
-------------------------------------------------------------------------------------------------
---
|  5      232  20%  66%|  -.01    .00  1.0 |   .04    .07|    .05    -.24|    .04 |   -.17  |
27%|
|  6      207  18%  84%|   .24    .22  1.0 |   .23    .07|    .61     .33|    .23 |    .32  |
29%|
```

A rating of 5.5 on the rating scale has a Rasch-Andrich threshold (step calibration) of 0.33 [see "EXPECTATION at -0.5" for category 6 in Table 8]
So the cut-off level for 7 to 8 is (0.0 + 0.0 + 0.33) = 0.33 logits.

You can check this by looking at the "Fair Average" column in Table 7 for the examinees. Look for a candidate with a fair average close to 5.5. The candidate measure should be close to 0.33.

```
Table 7.1.1  examinee Measurement Report  (arranged by MN).
-------------------------------------------------------------------------------------------------
--
|  Obsvd  Obsvd  Obsvd  Fair-M|        Model | Infit      Outfit   |Estim.|
 |
|  Score  Count Average Avrage|Measure  S.E. | MnSq ZStd  MnSq ZStd|Discrm| Nu examinee
 |
-------------------------------------------------------------------------------------------------
--
|    198     36     5.5   5.50|    .33   .12 | 1.17   .8  1.17   .8|  .70 |  8 8
 |
```

## 16.9 Decimal and continuous data

Facets expects the data to be in the form of integers. Ascending integers are taken to represent a hierarchical order of qualitatively more of whatever is being observed. Each higher integer means one more level of the latent variable.

**Averaged data:** Example: For one element, the data looks like 2.7, 5.6, 8.9, ranging from 0 to 10. but other items are integers.
In Facets, every advancing score must mean one higher level of the latent variable. In your data,
0,1,2,3,4,5,6,7,8,9,10  are 11 levels of performance.
What does 2.7 mean?  From the viewpoint of Facets it means:
an observation of 2 weighted 0.3 (= 0.6) + an observation of 3 weighted 0.7 (= 2.1 so that 0.6+2.1=2.7)
so, if the original Facets observation was:
    facet1, facet2, facet3, 2.7
then the Facets data is:
    Data=
    R0.3, facet1, facet2, facet3, 2
    R0.7, facet1, facet2, facet3, 3

**Decimal data due to intermediate levels:** Example: performances are scored in increments of 0.5, e.g., 2.0, 2.5, 3.0.

Multiply the observations by whatever is required to put them into integer form, or use the Rating scale= specification to recode the data. For performances scored in increments of 0.5, e.g., 2.0, 2.5, 3.0. Multiply the observations by 2 for analysis: 4, 5, 6. To maintain the original raw scores, weight these by 0.5:

Model = ?,?,?,R6, 0.5

Example: The performances have ratings like 3.5, but Facets expects integers.

A. Convert the observation from 3.5 to 7:

```
Rating scale = Wscale, R10
7 = ,,, 3.5 ; convert 3.5 into 7
..... ; and all the other possible observations
 *
```

B. Weight the observation by a half

```
Models=?,?,?,Wscale,0.5 ; Weight the model 0.5
```

**Continuous data:** chunk this up into qualitatively advancing pieces and number the chunks in ascending order. For instance, a timed task in which shorter time is better. 0.0-0.5 seconds = 6; 0.5-1.5 seconds = 5, 1.5-3.0 seconds = 4, 3.0-6.0 seconds = 3, etc. This recoding can be done with Rating scale=.

```
Rating scale = Tscale, R6
6 = ,,,  0.0+0.1+0.2+0.3+0.4+0.5 ; convert 0.0-0.5 into 6
56 = ,,,  0.6+0.7+0.8+0.9+1.0+1.1+1.2+1.3+1.4+1.5 ; convert 0.5-1.5 into 5
```

## 16.10 Decision Consistency

Assess.com discusses What is Decision Consistency? They write: "Decision consistency [DC] is an estimate of how consistent the pass/fail decision is on your test. That is, if someone took your test today, had their brain wiped of that memory, and took the test again next week, what is the probability that they would obtain the same classification both times?"

Facets does not report a DC statistic directly, but Wan *et al.* (2007) describe how DC statistics are verified by means of data simulation. Accordingly an MFRM DC statistic can be computed directly by simulation. The procedure is

1.) Rasch-analyze all the data. Identify candidates above and below the cut-score.

2.) Use the Rasch results to simulate and analyze 100 hypothetical datasets matching the observed dataset.

3.) From analyses of the simulated data, compute the average % of candidates for whom the pass-fail decision is the same as for the observed data. This is the proposed Decision Consistency statistic. We expect Decision Consistency to be near 100% - that's the idea!

Details of the Decision Consistency procedure:

1. Use a dataset you already have to try this out.
2. Start with 2 simulated data files to verify this procedure works for you.
3. Do it with a loop in a batch file to perform 2 data simulations and analyses. Output 2 Scorefiles.
4. Use NotePad++ or similar to rectangular copy the candidate scores from the original analysis score file into Excel column 1. This is the master list. Followed by columns of scores for each of the simulations.
5. Whatever the cutscore, use an Excel =if formula to put 1 for pass, 0 for fail in column 1 of Worksheet 2 for the candidates.
6. Do the same thing for the simulated scorefiles in columns 2 and 3. So you have 3 columns of candidate scores in Worksheet 1. 3 columns of 1/0 in worksheet 2.
7. In Worksheet 3, column 2 put 1 if worksheet 2 column 1 =if with worksheet 2 column 2. otherwise 0 for each row
8. In Worksheet 3, column 3 put 1 if worksheet 2 column 1 =if with worksheet 2 column 3 otherwise 0 for each row
9. Decision consistency % is the average of the numbers in Worksheet 3 * 100%
10. When this procedure works for you, repeat with 100 simulated data files

Note: This can be streamlined using a VBA macro so that we don't need Worksheets 2 and 3, but to start off, it is easier to check that all is correct if we can see the intermediate results.

CASMA Research Report Number 22, Estimating Classification Consistency for Complex Assessments, Lei Wan, Robert L. Brennan, Won-Chan Lee, January 2007

## 16.11 Deleting or Ignoring data values, observations or elements

There are several ways to drop, delete, remove or ignore data values, observations or elements.

1) Ignore elements or element combinations by modeling with an "M", Missing, model.
This method is most flexible, self-documenting, and easiest to undo. Before:

```
Models=
?,2,R ; Item 2 is on a rating scale (or partial credit)
?,?,D ; All other items are dichotomies
*
```

After:

```
1,3,M ; Ignore response by person 1 to item 3
4,?,M ; Ignore all responses by person 4
?,2,R ; Item 2 is on a rating scale (or partial credit)
?,?,D ; All other items are dichotomies
*
```

2) Delete the datum from the Data= file.

For single value references:
Before:

```
Data=
1,2,1
1,3,2 ; Datum to be deleted
1,4,3
```

After:

```
Data=
1,2,1
1,4,3
```

For multiple-value references:
Before:

```
Data=
1,2-4,1,2,3 ; Datum values of 2 is to be deleted
```

After:

```
Data=
1,2-4,1,,3 ; Blank data values are ignored
```

3) Comment out the datum value

For single value references:
Before:

```
Data=
1,2,1
1,3,2 ; Datum to be deleted
1,4,3
```

After:

Data=
1,2,1
; 1,3,2 ; this datum is ignored
1,4,3

For multiple-value references:
Before:
Data=
1,2-4,1,2,3 ; all 3 data values are to be deleted

After:
Data=
; 1,2-4,1,2,3 ; all 3 data values are ignored

4) To ignore all data values that include a specific element, delete or comment out that element number in the Labels= specification, e.g.,

Labels=
1, Students
1=George
; 2=Mary ; all elements referencing element 2 in facet 1 will be ignored
3=John

4B) You can keep highly misfitting elements (persons, items, raters, etc.) in the analysis, but minimize their impact on other elements by giving them a very small weight:

Labels=
1, Students
1-4000 ; 4,000 students
37 = , , , 0.001 ; student 37 has a very small weight, so is effectively deleted

5) To ignore facet element references, BUT STILL VALIDATE ELEMENT REFERENCES, leave the facet control character blank.

Model=?,,?,D ; ignore facet element references for facet 2
Data=
4,5,6,1 ; analyzed as 4,0,6,1

6) To ignore an entire facet in the data, you can use the Entered= specification.

Entry = 1,0,2 ; the second facet in the data is to be ignored completely. No element number validation occurs.
Model=?,?,D ; model only the two active facets (1st and 3rd)
Data=
4,5,6,1 ; analyzed as 4,6,1

or use the "X" facet control character:
Model=?,X,?,D

7) To ignore every occurrence of particular response values in the data, you can use Missing=, e.g., data values "9" and "." are to be treated as missing data:

Missing = 9,. ; all 9's and periods are treated as missing.

8) To ignore every occurrence of particular response values for a particular Model= specification, e.g, value "3" is missing for rating scale (or partial credit) "Speed".

Model = ?, ?, Speed
*
Rating (or partial credit) scale = Speed, R6
-1=missing,,,3 ; recode 3 as -1, the missing value code.

9) To remove all misfitting observations, create a new data file:
1. In your current Facets analysis, output the residual file to Excel from the Facets "Output Files" menu.
2. Sort the Excel file by residual or standardized residual.
3. Delete the residuals you don't want at the top and bottom of the Excel file.
4. Delete all the columns except Observation and the Element numbers
5. Move the observation column after the element number columns
6. Export the Excel file as a text file: gooddata.txt
7. Make a copy of your Facets specification file as backup.
8. Change Data= to Data=gooddata.txt
9. If there is any data in the specification file, delete it.
10. If there are Dvalues= or Delements= in the specification file, delete them. Other changes may be needed.
11. Run Facets with the revised specification file.

## 16.12 Diagnosing Misfit

What do Infit Mean-square, Outfit Mean-square, Infit Zstd (z-standardized), Outfit Zstd (z-standardized) mean?

General rules:
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.

If the mean-squares average much below 1.0, then the data may have an almost Guttman-pattern. Please use much tighter convergence criteria.

Zstd are t-tests of the hypotheses "do the data fit the model (perfectly)?". These are reported as z-scores, i.e., unit normal deviates. 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. See Score files for more exact values.

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".

**Guidelines:**
(a) Look for negative bi-serial correlations and large response residuals. Explain or eliminate these first.
(b) If Zstd is acceptable, usually <|2| or <|3|, then there may not be much need to look further.
(c) If mean-squares indicate only small departures from model-conditions, then the data are probably useful for measurement.
(d) If there are only small proportion of misfitting elements, including or omitting them will make no substantive difference. If in doubt, do analyses with and without them and compare results.
(e) If measurement improves without misfitting elements, then
(i) omit misfitting elements
(ii) do an analysis without them and produce an anchorfile=
(iii) edit the anchorfile= to reinstate misfitting elements.
(iv) do an analysis with the anchorfile.
The misfitting elements will now be placed in the measurement framework, but without degrading the measures of the other elements.

Anchored runs:
Anchor values may not exactly accord with the current data. To the extent that they don't, they 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.

Mean-square interpretation:
>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.

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.

| Diagnosing Misfit | | | | |
|---|---|---|---|---|
| **Classification** | **INFIT** | **OUTFIT** | **Explanation** | **Investigation** |
| | Noisy | Noisy | Lack of convergence<br>Loss of precision<br>Anchoring | Final values in Table 0 large?<br>Many categories? Large logit range?<br>Displacements reported? |
| Hard Item | Noisy | Noisy | Bad item | Ambiguous or negative wording?<br>Debatable or misleading options? |
| | Muted | Muted | Only answered by top people | At end of test? |
| Item | Noisy | Noisy | Qualitatively different item<br>Incompatible anchor value | Different process or content?<br>Anchor value incorrectly applied? |
| | | ? | Biased (DIF) item | Stratify residuals by person group? |
| | | Muted | Curriculum interaction | Are there alternative curricula? |
| | Muted | ? | Redundant item | Similar items?<br>One item answers another?<br>Item correlated with other variable? |
| Rating scale | Noisy | Noisy | Extreme category overuse | Poor category wording?<br>Combine or omit categories?<br>Wrong model for scale? |
| | Muted | Muted | Middle category overuse | |
| Person | Noisy | ? | Processing error<br>Clerical error<br>Idiosyncratic person | Scanner failure?<br>Form markings misaligned?<br>Qualitatively different person? |
| High Person | ? | Noisy | Careless<br>Sleeping<br>Rushing | Unexpected wrong answers?<br>Unexpected errors at start?<br>Unexpected errors at end? |
| Low Person | ? | Noisy | Guessing | Unexpected right answers? |

| | | | | |
|---|---|---|---|---|
| | | | Response set<br>"Special" knowledge | Systematic response pattern?<br>Content of unexpected answers? |
| | Muted | ? | Plodding<br>Caution | Did not reach end of test?<br>Only answered easy items? |
| Person/Judge Rating | Noisy | Noisy | Extreme category overuse | Extremism? Defiance? |
| | Muted | Muted | Middle category overuse | Conservatism? Resistance? |
| Judge Rating | | | Apparent unanimity | Collusion? |
| INFIT:<br>OUTFIT:<br>Muted:<br>Noisy: | information-weighted mean-square, sensitive to irregular inlying patterns<br>usual unweighted mean-square, sensitive to unexpected rare extremes<br>unmodeled dependence, redundancy, error trends<br>unexpected unrelated irregularities | | | |

## 16.13 Dichotomous Misfit Statistics

```
  Responses:          Diagnosis             INFIT OUTFIT
  Easy--Items--Hard   Pattern               MnSq    MnSq
-------------------------------------------------------
  111|0110110100|000  Modelled/Ideal         1.1     1.0
  000|0000011111|111  Miscode                4.3    12.6
  011|1111110000|000  Carelessness/Sleeping  1.0     3.8
  111|1111000000|001  Lucky Guessing         1.0     3.8
  101|0101010101|010  Response set/Miskey    2.3     4.0
  111|1000011110|000  Special knowledge      1.3     0.9
  111|1111100000|000  Guttman/Deterministic  0.5     0.3
  111|1010110010|000  Imputed outliers *     1.0     0.6
-------------------------------------------------------
Right|Transition|Wrong
  high - low - high   OUTFIT sensitive to outlying observations
                      >>1.0 unexpected outliers
                      <<1.0 overly predictable outliers

   low - high - low   INFIT sensitive to pattern of inlying observations
                      >>1.0 disturbed pattern
                      <<1.0 Guttman  pattern
```

* as when a tailored test is filled out by imputing all "right" response to easier items and all "wrong" to harder items.

The exact details of these computations have been lost, but the items appear to be uniformly distributed about 0.4 logits apart, from Linacre, Wright (1994) Rasch Measurement Transactions 8:2 p. 360

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.

## 16.14 Differential performance on items

It is desired to obtain an ability measure for each person on each item, but in a common frame of reference.

A) Format the control and data file.
Include facets for persons, items, etc. Non-center the persons.

B) Perform a conventional analysis.
Each person has one ability. Each item has one difficulty. Produce an Anchorfile=

C) Edit the anchor file.

Remove the ",A" for the person facet. Comment out all except one item with ";" in the items part of the Labels= section.

D) Perform an analysis.
This will produce measures for the persons on the specified item.

E) Repeat (C) and (D) for all other items to produce person measures on each of the items in turn.

## 16.15 Disordered Rasch-Andrich thresholds

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 move 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(frequency of category below/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".

## 16.16 Double ratings

If a rater rates the same performance twice, we may want to look at the relationship between the first and second rating. Here is a start:

1) Each rater has two element numbers, e.g.,
101 and 102 ; this is rater 10 first rating and second rating
111 and 112
121 and 122
.....

2) Give the raters group numbers:
Labels=
...
*
2, Raters
101=10 first, , 1 ; first ratings are in group 1
102=10 second, , 2 ; second ratings are in group 2
....
*

3) There is a "double rating" facet:
Labels=
....
*
4, Rating sequence, D ; dummy facet
1=First rating
2=Second rating
*

You could also model the second rating with an additional rating scale. Give the Models= specification for this rating a very small weight:
Rating scale=R2

0=second rating lower than first
1=second and first rating the same
2=second rating higher than first
*

Enter 0,1 or 2 into the data for each double-rating, in addition to the two ratings.

## 16.17 Dummy facets and group-anchoring for interactions

Dummy facets are facets intended only for investigating interactions, not for measuring main effects. All the elements of a dummy facet are anchored at 0.

This feature is useful for dummy facets removing classification elements from the measurement. These may be used to partition fit, detect bias or select rating scale (or partial credit) models.

**Specifying dummy facets:**

```
 Labels=
 2=Classifier,A ; anchoring wanted
 1=Type A, 0 ; anchor at 0, so doesn't affect measurement
 2=Type B, 0
```

or, and easier,

```
 Labels=
 2=Classifier, D ; dummy-facet anchoring wanted
 1=Type A ; no anchor value required
 2=Type B
```

When Umean= is used, use the Umean= value for these "dummy" elements

```
 Umean=50,10 ; User mean is 50, user scaling is 10 per logit.
 Labels=
 2=Classifier,A ; anchoring wanted
 1=Type A, 50 ; anchor at umean value, so doesn't affect measurement
 2=Type B, 50
```

Or, easier

```
 Labels=
 2=Classifier,D ; dummy-facet anchoring wanted
 1=Type A ; no anchor value required; automatically set at umean value.
 2=Type B
```

**Modeling with dummy facets:**

```
 Facets = 3 ; persons, classifiers, items
 Models = ?, ?B, ?B, R ; interactions between classifiers and items
```

Example: Let's suppose we have three facets, "Rater leniency", "Rater experience", and "Examinees"

Scenario 1: measure the raters, report the groups

```
Facets=3
Models=?,?,?,R
Noncenter =1 ; noncentered facet cannot be anchored or dummy
Labels=
1, Rater leniency
1=Jose , , 1     ; novice
2=Maria , , 2    ; expert
```

....
*
2, Rater experience, D  ; dummy facet, all elements automatically anchored at 0
1=Novice
2=Expert
*
3, Examinees
...
*

Scenario 2: Measure the experience. Rater measures are relative to experience.

Facets=2
Models=?,?,R
Noncenter=2
Labels=
1, Rater leniency, A  ; group anchored by group
1=Jose, 0 , 1
2=Maria, 0 , 2
....
*
2, Rater experience  ; standard facet
1=Novice
2=Expert
*
3, Examinees
...
*

## 16.18 Effect size

There are several definitions of "effect size" - https://en.wikipedia.org/wiki/Effect_size

A common one is:
Effect size = (treatment B mean effect - treatment A mean effect) / (treatment A sample S.D.).

In *Facets* this would be:
Effect size = (Task B measure - Task A measure) / (S.D. of person measures)

Also:

Cramer's V converts a chi-squared into on effect size.

For an individual rater:
Here is one similar to Cohen's D:
Effect size = (rater leniency - mean rater leniency)/rater leniency S.D. of sample

One similar to Glass Delta $\Delta$ is
Effect size = (rater leniency - mean rater leniency)/rater latency S.E. of this rater

**Derivation of Cohen's d with joint (pooled) S.E.s**

Cohen's d: quoting Wikipedia: "difference between two means divided by a pooled standard deviation for the data"

So, in a simple case, let's assume the two sample sizes are the same, n, so that the combined sample is 2*n = N

The sample S.D. is used in Cohen's d, e.g., https://scientificallysound.org/2017/06/29/cohens-d-and-effect-size/

Sample S.D. for one sample = √(sample variance / (n-1)).
S.E. of the sample mean = √(sample variance / (n*(n-1))) = S.D. / √(n)
Rearranging, SD = √ (n) * SE

Let's add up the variances for two samples: pooled variance = (n-1)*SD1² + (n-1)*SD2²

pooled S.D. = √ ( (n-1)*SD1² + (n-1)*SD2²) / (n+n-2))
= √ (( SD1² + SD2²) / 2 )
= √ ((n*SE1² + n*SE2²) / 2)
pooled S.D.= √ (SE1² + SE2²) * √(n/2)

but n = N/2 where N is the combined sample
= √ (SE1² + SE2²) * √(N/4)

Facets: t = (M1-M2) / √ (SE1² + SE2²)
Cohen's d = (M1-M2) / pooled S.D.
Cohen's d = Facets t * √ (SE1²+ SE2²) / pooled S.D.
= t * √ (SE1²+ SE2²) / ( √ (SE1² + SE2²) * √(N / 4)
= t * 2 / √(N)

So, for Facets, when there is a joint S.E., its *t* and its d.f., the approximation becomes:
Cohen's *d* = 2 * t /√(d.f.)

## 16.19 Equating tests

Equating in many-facet situations is more complex than equating two facets, e.g., multiple-choice tests.

The general rules are:

(a) All facets except one must be equated. That one is usually persons, except in studies of rater behavior.

(b) There need to be at least 5 common elements for each facet. This is a rule of thumb. In principle, one element will work. But it gives no quality-control.

(c) First do independent analyses of each data set.

(d) Then cross-plot the measures for common elements in each facet.

(e) Disregard as "common" those elements whose measures are clearly different across datasets.

(f) If the remaining common elements for a facet are close to a line parallel to the identity line, than anchor those elements, or combine the data sets.

(g) If the remaining common elements for a facets are still somewhat of a cloud, then use group-anchoring of that facet to equate it.

(h) Continually monitor criterion-level-based success-rates and other indicators to check that the equating is making sense.

## 16.20 Estimates do not converge

**"*Facets* runs and runs, but does not converge! What has gone wrong?"**

Are the current set of estimates good enough for your purposes? Estimates with extremely high precision are rarely needed, and seldom improve overall fit of the data to a Rasch model. Set Convergence= and Iterations= to reasonable values.

If the changes per iteration are very small or are oscillating up and down about equally, then the estimates are as good as they can be. Press Ctrl+F to stop iteration and move on to fit computation. You can also use the Estimation menu to make the changes per iteration smaller, or to change the estimation method.

If Facets runs for more than 100 iterations, and the residuals are not converging (approaching zero), then please force Facets to stop iterating by Ctrl+F (or use the Estimation pull-down menu) - then look at the screen and the output file. Investigate elements with large displacements and rating-scale categories with large "Observed - Expected Diagnostic Residuals".

There are numerous possibilities. Here are some to check.

1) **Noncenter=**
   The analysis may be
   i) over-constrained because all facets are centered, anchored or group-anchored,
   ii) under-constrained because more than one facet is non-centered.
   For most analyses, noncenter= must specify one active facet.
   Make sure that facet is specified is not marked X or omitted from the model statement.
   Makes sure that the non-centered facet does not include anchor or group-anchor values.

2) **Null element= (Keepasnull=)**
   Element 0 in the data is usually a dummy element that means "this facet does not apply to this observation". If element 0 is an active element in your analysis, then please assign a different unused element number as the dummy element number, e.g., *Null element = 999*

3) **Your facets are nested**.
   When one facet is nested within another facet (without anchoring or group-anchoring), Facets makes an arbitrary allocation of statistical information between the facets, this can lead to unstable estimation. For instance, if there is a "student" facet and also a "student gender" (sex) facet, then nesting needs to be resolved by anchoring or group-anchoring either the "student" facet or the "student gender" facet.

4) **Low category frequencies**
   Rating (or partial credit) scale categories with very low frequencies (less than 10 observations) are difficult to use as the basis of estimation. Convergence may take a long time.

5) **Rating (or partial credit) scales with many categories.**
   Estimating the parameters of long rating (or partial credit) scales (more than 10 categories) is difficult. Adjustments tend to ripple up and down the scale, like a caterpillar moving. Convergence may take a long time.

6) **Category frequencies are lumpy.**
   When some categories of a rating scale have high frequencies and other categories have low frequencies, convergence may take a long time.

7) **Clumps of data or stringy data due to missing data. Disconnected or weakly connected subsets.**
   Data with disjoint subsets of data, or with only thin connections between the subsets, may not converge, or may converge to non-repeatable values. For instance, your judging design may be very sparse.

8) **The Maximum Likelihood curve is almost flat.**
The data are such that there are a range of estimated measures that match the data equally well. All are equally valid.

9) **The Maximum Likelihood curve has two peaks very close together.**
The data are such that there are two almost equally good sets of estimates that match the data. Facets cannot choose between them.

## 16.21 Estimation considerations: JMLE estimation bias

**Sample size**
Facets produces estimates for any data set in which there is overlapping randomness in the observations across elements. It cannot estimate measures when the data have a Guttman pattern or there is only one observation element. But even in these cases, anchoring and other analytical technique may make the element measures estimable. See also "My analysis does not converge."

Estimates which are likely to have some degree of stability across samples require at least 30 observations per element, and at least 10 observations per rating-scale category.

There is no maximum sample size apart from those imposed by computer constraints. Large sample sizes tend to make convergence slower and fit statistics overly sensitive.

**Estimation techniques: JMLE and PMLE**
*Facets* generally uses JMLE (= Joint Maximum-Likelihood Estimation), but for Models=?,-? Facets uses a customized implementation of PMLE (= Pairwise Maximum-Likelihood Estimation). RSA and MFRM derive Newton-Raphson iteration equations for the estimation of Rasch measures using the JMLE and PMLE approach. In *Facets*, this approach is not robust enough against spiky data. Consequently, since the basic functional shape of all the estimation equations is the logistic ogive, Facets estimates measures by means of iterative curve-fitting to that shape. The resulting measures also accord with the JMLE and PMLE approach.

1. All the parameters are given reasonable starting values.
2. The estimate of each parameter (element measure or Andrich threshold) is updated as though the values of all the other parameters are known.
3. Then all the parameters are reestimated using the updated estimates of all the other parameters.
4. (3.) is repeated until no parameter value changes by more than the convergence limit.
5. At convergence, for every element or rating-scale category, the expected score = the observed score.

For a two-facet analysis, this process is the same as shown in the Excel spreadsheets linked from www.rasch.org/moulton.htm

This is explained in Linacre "Many-Facet Rasch Measurement"

**Choice of estimation algorithm**
JMLE: Every estimation method has strengths and weaknesses. The advantages of JMLE far outweigh its disadvantages. JMLE is estimable under almost all conditions including idiosyncratic data designs, arbitrary and accidental patterns of missing data, arbitrary anchoring (fixing) of parameter estimates, unobserved intermediate categories in rating scales, flexible equating designs and multiple different Rasch models in the same analysis. All elements are treated as equal. Each element of every facet has a parameter estimate, standard error and fit statistics obtained in exactly the same way.

The primary weakness of JMLE is that estimates can have estimation bias. In MFRM analyses, estimation bias is usually of minor concern because either the dataset is large or the structure of the data negates the importance of estimation bias. The size of the estimation bias can be discovered using simulated datasets.

In a typical two-facet analysis, JMLE estimation bias is most obvious in a test of two dichotomous items (Andersen, 1973) with a large sample of persons, a two-facet analysis. In such an analysis, JMLE estimates the difference between the item difficulties of the two items to be twice its true value. In Winsteps, the estimation bias for two-response situations is corrected by specifying PAIRED=Yes. The estimation bias for longer tests is corrected by STBIAS=Yes, but there are drawbacks to bias correction as discussed in Winsteps Help. In many practical situations, the statistical bias is less than the standard errors of the estimates, and becomes negligible as test length and sample size increase. See Wright (1988).

PMLE: When Models=?,-? (or similar) are specified, Facets uses a customized version of PMLE. PMLE is also implemented in RUMM2030. PMLE produces unbiased estimates under suitable conditions.

MFRM analysis can be performed with other estimation methods including CMLE (Conditional Maximum Likelihood Estimation), MMLE (Marginal Maximum Likelihood Estimation) which can all produce better estimates than JMLE under *ideal* conditions. In practice, they all have the same crucial drawback. They model MFRM data as two-facet data. One of the MFRM facets is modeled to be the "item" facet. All the other MFRM facets are compressed into the second, "person", facet. After the "person" estimates are obtained from a two-facet analysis, those estimates are linearly decomposed into the estimates for all the elements of all the other facets. This approach limits admissible data designs, restricts element-anchoring and introduces distortions into the estimates of most elements. Changing the facet modeled to be the "item" facet results in different estimates of all elements of all facets.

Andersen, E. B. Conditional inference for multiple-choice questionnaires. British Journal of Mathematical and Statistical Psychology, 1973, 26, 31-44.

Wright, B.D. The Efficacy of Unconditional Maximum Likelihood (JMLE) Bias Correction. 1988.
www.rasch.org/memo45.htm

---

**JMLE estimation bias**

*Facets* makes no correction for estimation bias in JMLE. The size of the bias adjustment (if any) depends on exact details of the data design. Estimation bias is almost always less than the standard error of measurement. Estimation bias is approximately a linear transformation of the unbiased estimates. This bias slightly expands the logit range of the elements. If you encounter a situation in which this slight expansion is crucial for decision-making, please do tell Winsteps.com.

To estimate the size of the estimation bias with *Facets:*
1. analyze your data with Facets. Output the element measures to Excel.
2. simulate many datasets (at least 10, but 100 would be better) using the *Facets* "Simulate" function
3. analyze all the simulated datasets. Use *Facets* in batch mode to do this.
4. average each element measure across all the simulated datasets.
5. crossplot the averaged measures (y-axis) against the original measures from 1 (x-axis).
6. the slope of the trend line in the plot indicates the size of the estimation bias. We expect the trend line to have a slope slightly above 1.0.
7. divide the original element measures by the slope of the trend line to obtain unbiased estimates.
8. the bias in the original standard errors approximates the square-root of the slope of the trend-line.

## 16.22 Excel files do not display

Windows may activate at button on the Windows task bar instead of displaying the Excel window. If you want the Excel window to display, then please change the Windows Registry (at your own risk):

*a.* Windows: Start >> Run >> *regedit*

*b.* Navigate to the following path: *HKEY_CURRENT_USER\Control Panel\Desktop*

*c.* Modify the value named *ForegroundLockTimeout* to 0
If *ForegroundLockTimeout* does not exist, then create *ForegroundLockTimeout* with DWORD value 0

## 16.23 Expected score

The expected score for an element is the sum of the expected scores for each observation in which the element participates.

For the observation, M is the sum of the measures of the elements for the observation. Then, from www.rasch.org/rmt/rmt122q.htm, the probability of observing a category for observation ni is:

$$P_{nij} = \frac{e^{j(M \quad) - \sum_{k=b}^{j} F_k}}{\sum_{h=b}^{t} e^{h(M \quad) - \sum_{k=b}^{h} F_k}}$$

and, and the score for an element is the sum of the expected scores for the L observations of the element:

$$\sum_{i=1}^{L} \sum_{j=b}^{t} j P_{nij}$$

For the rating-scale categories for the observation i are numbered b,..,t. $F_b$=0, and the other $F_k$ are the Rasch-Andrich step thresholds. The denominator is the sum of the numerators for all categories, so that the sum of the probabilities across all categories is 1:

---

**Expected score for a standard (zero) examiner/judge/rater**

"Expected" means what we expect the observation or score to be based on the examiner, task, candidate, etc.

If you need the value of the expected observation when the examiner severity is zero, then do this:

1. Do your standard Facets analysis
2. Output an Anchorfile= with anchor values
3. In the Anchorfile=, "X" out the examiner facet in the Models= specification. Don't change anything else.
4. Use the anchor file as your Facets specification file
5. The "expected" values in the output will be for examiners at 0.

## 16.24 Export Facets Tables and other files to Excel, SPSS or Word

**Output Facets Table to Excel**

1. In the Facets output report, locate the Facets Table that contains the numbers you want for your Excel analysis
2. Copy the rows containing the numbers you want from Facets Output Tables in NotePad into an Excel worksheet
3. In Excel, "Data", "Text to Columns" to put the numbers into columns
4. Save the Excel file

**Output Facets Table to SPSS**

1. In the Facets output report, locate the Facets Table that contains the numbers you want for your SPSS analysis
2. Copy the rows containing the numbers you want from Facets Output Tables in NotePad into an Excel worksheet
3. In Excel, "Data", "Text to Columns" to put the numbers into columns (SPSS variables)
4. Save the Excel file

5. SPSS can open the EXCEL file.

**Output Facets Table to Word**
1. In the Facets output report, locate the Facets Table that contains the numbers you want for your Word document
2. Highlight its columns by dragging your Mouse
3. Copy (ctrl+c)
4. In your Word document, paste (ctrl+v) the Facets Table.
5. Highlight the Facets Table
6. Change the font to Courier New (or another fixed-space font).

## 16.25 External measures as anchor values or ratings

**External measures as anchor values**

"In analyzing clinical or educational data, persons (or items) may already be characterized with "numbers" that are asserted to be linear measures. But it is not known what is the linear transformation between these measures and the local logits of the current data set.

One simple approach to this problem is to analyze the data, and plot the resulting person measures (or item difficulties) against their "numbers" to identify the transformation. This technique, though generally successful, overlooks misfit and sample-item targeting.

Another approach is to optimize the fit statistics. In this technique, one chooses an initial number-to-logit conversion, uscale, that spreads the sample (or items) conspicuously wider along the measurement variable than is reasonable. Perform a data analysis with the persons (or items) anchored at their converted numbers. All summary mean-squares (average infit and outfit mean-squares summarizing all persons and items) are expected to be noticeably above 1.0. If not, double the number-to-logit conversion factor, uscale, and repeat this step of the procedure.

Once all four summary mean-squares are greater than 1.0, a useful number-to-logit conversion factor has proved to be uscale = uscale divided by the average of the four summary mean- squares. Reanalyze the data, and, if the average of the four summary mean-squares is still noticeably greater than 1.0, repeat this step of the procedure.

In preliminary investigations with clinical indicators, this has proved to be a fast and easy way to build useful measurement systems.

Example: When the persons were anchored at their clinical indicator values, the mean person infit statistic was 1.93, outfit 2.19. The mean item infit statistic was 1.59, outfit 2.27. The average of these four numbers is 2.0. On reanchoring the persons at their clinical indicator values, divided by 2.0, the mean-squares became, respectively, .98, 1.08, 1.02, 1.08. This suggests that a probabilistically interpretable measurement system has been constructed." Linacre J.M. (2000) at www.rasch.org/rmt/rmt142n.htm

**External measures as observations on a rating scale**

When each participant has external measures for each task, these can be included in a Facets analysis as ratings with examiner "0", which tells Facets that there is no examiner for these ratings. The external measures must be stratified into rating scales, e.g., 3-categories, "low, medium, high", or 5 categories, probably not more than 7 categories. The number of categories cannot be more than can be discriminated by an informed person. Over-categorization can be identified by examining the fit statistics from your Facets analysis. If the "external measures" item noticeably underfits (mean-square >> 1.0) then there are too many categories.

## 16.26 Extreme scores

See Xtreme= for how Facets estimates extreme scores (minimum possible scores, and maximum possible scores).

## 16.27 FACDOS

Valid specifications for Facets 3.22, the 1999 DOS version of Facets:

Barcharts output = Yes (the defa
Bias (direction: difficulty/abil
Convergence (criteria to end ite
CSV = (the defau
Data file name = facets.dat (the
Entry order of facets in data =
Facets in data = 2 (the default)
General statistics including poi
Graph file name =
Hardware = 80 (the default)
Hsize (size of heap) = 2700 (the
Iterations (stop after) = 100 (t
Juxtapose (column headings after
Keep as null element zero = 0 (t
Labels of facets and elements =
Left-hand placement of row label
Missing data codes = (the defaul
Model to be used in the analysis
Noncenter facet = 1 (the default
Output results/anchors/residuals
Positively-oriented facet = 1 (t
Query to monitor matching of dat
Rating scale (or Response model)
Score files name and CSV status
Standard (error is Asymptotic/Re
Subset detection = Y (the defaul
Title for each results table = (
Umean (user mean and scale) = 0,
Unexpected (standardized residua
Usort sort order of unexpected r
Vertical rulers in results file
Write iteration details in resul
Xtreme (score adjustment) = 0.3,
Yardstick (columns/lines) = 80,
Zscore minimum for reported bias

## 16.28 Fair average

A "Fair Average" for each element of the target facet is computed as though every observation of that element in the data was for an average element of every other facet. Then these artificial observations are averaged by the number of observations of the target element. This is "fair" in the sense that every element of the target facet encounters elements of the same logit values in the other facets. The Fair Average is computed using an average rater, essay, etc., omitting elements with extrme scores. Since the mathematics is non-linear (logistic), a rating by an average rater is not the same as the average of the ratings by all the raters, but is usually close to it.

We can check this by outputting an Anchorfile=. Keep everything anchored. Then, for every facet except one (the target), change the anchor values of the elements to the average logit value of the elements in that facet. Then the expected average for each element in the target facet should be the same as the fair average.

**Meaning of the Fair Average:**

The "Fair Average" is what the raters would have had if they had all rated the same "average" elements under the same "average" conditions. For instance, suppose one rater rated all the easy tasks (= high observed average rating) but another rater rated all the hard tasks (= low observed average rating). Then the "Fair Average" says "What if both those raters had rated the same average task, what would the rating have been?". We can then use the Fair Average to compare the severity/leniency of the raters as though they had rated under the same conditions.

Example: What's the relationship between the raw score and Fair Averages?
They are both in the original rating-scale metric. The raw score on an item is the original observation. The Fair Average is the original observation adjusted for its context. Suppose that my performance receives a rating of 3 from a lenient rater. My Fair Average is 2.5. Your performance receives a rating of 3 from a severe rater. Your Fair Average is 3.5. Comparing your "3" with my "3" is unfair, because you had a severe rater (who generally gives low ratings) and I had a lenient rater (who generally gives high ratings). After adjusting for rater severity/leniency, our "Fair Average" ratings are 2.5 and 3.5. These give a fair comparison.

The "fair average" transforms the Rasch measure back into an expected average raw response value. This value is in a standardized environment in which all other elements interacting with this element have a zero measure or the mean measure of all elements in their facet. This is "fair" to all elements in the facet, e.g., this adjusts raw ratings for severe and lenient raters. This enables a "fair" comparison to be made in the raw score metric, in the same way that the measure does on the linear latent variable. *Fair-M* uses the facet element means as the baseline. *Fair-Z* uses the facet local origins (zero points) as the baseline. These are set by Fair average=.

The original purpose of Facets (in 1986) was to construct software that would automatically adjust for differences in rater severity/leniency. So this has been done for your data. No adjustment or trimming of usefully-fitting raters is necessary, regardless of their severity/leniency.

But Facets does assume that the average leniency of the raters is at the required standard of severity/leniency (usually by centering the rater facet at zero logits). If you see, from evidence external to the data, that the average leniency of the of the raters is too high or too low, then please
1. include an additional "adjustment" facet in your analysis with a ? in Models=
2. this facet has one element.
3. Anchor this element at the adjustment value
4. Include the element in all the observations in the dataset by using dvalues=
5. perform an analysis of all the data including the adjustment element The rater leniencies are now correct, but the fair averages are not correct for the adjusted leniencies.
6. output an anchorfile
7. replace the ? in the model specification with an X
8. analyze the anchorfile with the dataset
9. there will be displacements of the size of the adjustment value. The fair averages are correct for the adjusted leniencies.

This procedure will require several attempts before it produces the correct results.

**Standard Error of the Fair Average:**

The S.E. of the Fair Average is about the same as the S.E. of the Observed Average.

Output the Residualfile= to Excel. Sort on the facet of interest.
The S.E. of the observed average is =STDEV(cells of observations for element)/SQRT(COUNT(observations for element)

**Calculation of the Fair Average Score**
The observed average score is the average rating received by the element. The logit measure is the linear measure implied by the observations. This is adjusted for the measures of whatever other elements of other facets participated in producing the observed data. It is often useful to transform these measures back into the original raw score metric to communicate their substantive meaning. Fair Average does this. It is the observed average adjusted for the measures of the other elements encountered. It is the observed average that would have been received if all

the measures of the other elements had been located at the average measure of the elements in each of their facets.

Fr the Fair Average computation, elements with extreme scores are omitted. So the observed average when Totalscore=No matches the Fair Average.

A basic many-facet Rasch model for observation Xnmij is:

```
log ( Pnmijk / Pnmij(k-1)) = Bn - Am - Di - Cj - Fk
```

where
Bn is the ability of person n, e.g., examinee: Mary,
Am is the challenge of task m, e.g., an essay: "My day at the zoo".
Di is the difficulty of item i, e.g., punctuation,
Cj is the severity of judge j, e.g., the grader: Dr. Smith,
Fk is the barrier to being observed in category k relative to category k-1,
where k=0 to t, and F0=0.

To compute the fair average for person n (or task m, item i, judge j), set all element parameters except Bn (or Am, Di, Cj) to their mean (or zero) values. Thus, the model underlying a fair rating, when Fair=Mean, is:

```
log ( Pnmijk / Pnmij(k-1)) = Bn - Amean - Dmean - Cmean - Fk
```

or, when Fair=Zero, it becomes:

```
log ( Pnmijk / Pnmij(k-1)) = Bn - Fk
```

and the Fair average is sum(k Pnmijk) across categories k=0 to t.

---

***What Facets does ...***

1. compute mean logit value of all the elements in each facet: MFlogit(f) for facet f.

2. for each observation, sum the mean logit values of the relevant facets that are modeled to generate the observation: sum(MFlogit(f)) = MFtotal

3. for each observation: subtract the mean logit from the element measure for each facet: flogit(f) = element logit(f) - MFlogit(f)
*example:* element 3 of facet 4 has a measure of 3.2 logits
flogit(4) = 3.2 - MFlogit(4)

4. for each observation, for each element and facet generating it, compute the "fair" expected score for a logit measure of MFtotal+flogit(f) on the rating scale relevant to the observation

5. accumulate the "fair" expected scores for each element across all the observations in which it participates

6. divide each element's accumulated "fair" expected scores by the number of observations for that element

7. the result is the "fair average" for the element

---

*Example:* Students are rated on 6 items. I want a fair-average for each student on each item.
1. Perform the analysis of all the data.
2. Output an Anchorfile=

3. Unanchor the students (remove the ,A for students in Labels=)
4. Analyze the anchorfile one item at a time by commenting out all items except 1 in Labels=
5. The reported measures and fair-averages for the students will be the fair averages for each item.
6. To assemble these fair-averages, output a Scorefile= from each one-item analysis to, say, Excel, selecting student identification and fair average fields.

---

**Verifying the Fair Average**

Analyze your data in Facets. Choose a Facet and an element whose Fair Average in Table 7 you want to verify.

In Facets,
"Output Files Menu"
"Residuals File"
"Select fields to output"
"Uncheck all"
Check - observation, expected value, element numbers
OK
Output to Excel.

In Excel, sort on the element numbers for your Facet.
Delete all rows except those for your element of your Facet.

Count and Sum the observations for your element of your Facet.
The count and sum should agree with Facets Table 7.
Average the observations. This should agree with the Table 7 Observed Average.

Now for the Fair Average: This is the expected value for your element when it encounters elements of mean difficulty (usually 0) in all the other facets. Extreme-score elements are excluded from computing the mean.

From Facets Table 7, choose elements near the mean of the other facets.

In the Excel table, discover the expected values for your element combined with the mean elements of the other facets. The expected values should agree with the Table 7 Fair Average.

---

**Problems with the Fair Average**

If the "Fair averages" do not monotonically increase with the element measures, then
the misalignment of fair score with Rasch measures can occur when some items/task/raters etc. have different rating scales to other items/task/raters and all candidates are not rated on all rating scales.

For instance,
you do task 1 which has a rating scale from 0-10
but I do task 2 which has a rating scale from 0-5

We both have a measure of 2.00 logits.

Then your "fair score" will be 8.2 on the 0-10 item
But my "fair score" will be 4.1 on the 0-5 item

To get around this,
1. Write out an Anchorfile= from the Facets analysis.

2. Construct dummy data in which every candidate has a rating on every task (it doesn't matter what the value of the rating is).
3. Analyze the dummy data
4. The "fair score" for each candidate will be averaged across all the tasks.

## 16.29 Fractional dichotomous and polytomous models

Rasch models are designed to estimate measures from ordinal, integer, data. There are situations in which the data are better expressed in fractions, such as the Netflix Prize Challenge. For instance, instead of 0-1, analyzing the data as 0.1-0.9 might give better measures. Data like this can be implemented in Facets using response-level weighting, R.

So, instead of,
Data=
234, 54, 1
234, 55, 0
the data becomes
R0.9, 234,54,1 ; weighted observations have a combined weight of 1, but a combined score of 0.9
R0.1, 234,54,0
R0.9, 234,54,0 ; weighted observations have a combined weight of 1, but a combined score of 0.1
R0.1, 234,54,1

## 16.30 Frame of reference

The observations are understood to be the result of stochastic interactions between elements of facets. Each element is to be measured or calibrated within a common frame of reference. This frame of reference must be defined. There are three decisions:

a) Where is the scale origin is to be established for each facet? Center= or Noncenter=
Usually the origin of the scale, the "zero" point for each facet, is set at the mean value of the measures for that facet, i.e., centered. When this has been done for all facets except one, the placement of the origin for that last facet is then determined by the data. That facet will be non-centered=.
Usually the noncenter= facet is the examinees.

b) In which direction is each facet to be oriented? Positive= or Negative=
Does a greater raw score correspond to a greater measure? This positive orientation is usually how person abilities and judge leniencies are conceptualized.
Does a greater raw score correspond to a lesser measure? This negative orientation is usually how item and task difficulties and judge severities are conceptualized.
In some analyses, there is only one positively oriented facet, the measured people. All other facets are negatively oriented: item difficulty, judge severity, task challenge etc.
In other analysis, all facets are positively oriented. person ability, item easiness, judge leniency, task facility.
Careful choice of facet orientation avoids confusion in interpreting Facets output.

c) What is the unit of measurement? Umean=
The conventional Rasch unit of measurement is the "Logit", and this the standard unit. It is often convenient, however, to linearly rescale the logit into RITS, WITS, or to give the measures a useful range, e.g., 0 - 100 (see BTD chapter 8). Umean= accomplishes this.

## 16.31 Generalizability Theory

For G-Theory computations using Facets, see Table 5

The aims of a Facets analysis and G-Theory are different. G-Theory operationalizes a "raters are rating machines" philosophy, so we want the "machines" to operate in exactly the same way. Facets operationalizes a "raters are

independent experts" philosophy, so we want the "experts" to concur on their overall verdicts, but we expect them to disagree on the details.

In G-Theory, the aim is to reduce the unwanted variance so that the observed scores become the "truth". Of course, there is always measurement error, and some raters are always more lenient or more severe than other raters, Consequently the observed scores are always influenced by the "luck of the draw" (Shavelson & Webb, 1991). In Facets, the aim is to adjust for unwanted variance so that the estimated measures (after adjustment) become the "truth".

Usable judging plans (experimental designs) for Facets and G-Theory differ. G-Theory accepts nesting, such as raters within test center. Facets demands partial crossing, such as by having raters at all test centers rate the same videotaped performances. G-Theory demands judging designs for which the variance terms are estimable. Facets accepts deliberate and accidental missing data.

*Shavelson, R. J., & Webb, N. M. (1991). Generalizability Theory: A Primer. Thousand Oaks, CA: Sage.*

## 16.32 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 on the variable. |
| Agent of Measurement | the tool (items, questions, etc.) used to define a variable and position objects or persons 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 (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 along the variable. Also "step calibration" |

| | |
|---|---|
| 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<br>CATS | levels of performance on an observational or response format. |
| Cell | Location of data in the spreadsheet, given by a column letter designation and row number designation e.g. B7 |
| Classical Test Theory<br>CTT | Item analysis in which the raw scores are treated as linear numbers. |
| PMLE | Pairwise Maximum Likelihood Estimation. This was devised by Bruce Choppin and is used in RUMM2030. A customized version is implemented in *Facets* for Models = ?,-? |
| 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. |
| 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 PCA 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 are responding to the item, also called "item bias" |
| Difficulty | the level of resistance to successful performance of the agents of measurement on the variable. |
| 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 measures 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<br>END NAMES | Winsteps: the end of the list of item identifying labels. This is usually followed by the 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<br>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()<br>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. |
| 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 |
| 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", "these is no DIF". The null hypothesis is rejected if the results of the fit test are significant ($p<.05$) or highly significant ($p<.01$). 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 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. |
| Item | agent of measurement, not necessarily a test question, e.g., a product rating. |
| 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()<br>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. 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. A log odds transformation of the probability of a correct response. |
| 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 linear 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 variable. |
| Matrix | a rectangle of responses with rows (or columns) defined by objects and columns (or rows) defined by agents. |
| MCQ<br>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<br>MnSq | Also called the relative chi-squared and the normed chi-squared. A mean-square fit statistic is a chi-squared 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<br>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 |
| 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 | people, products, sites, to be measured or positioned along the variable. |
| OBS<br>Observed | Value derived from the data |
| Observation<br>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 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<br>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. |
| 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 formula, used by some Rasch programs for the first part of the iteration 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 equal-interval 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. |
| Ratio scale | Scale with a defined origin (reference point) so we can say that measure X is twice measure Y. See Ratio scale |
| 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 | 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 indicating degree of success by an object on an agent, and entered into the appropriate cell of an Input Grid. |
| 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 may be 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 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.<br>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 | Winsteps and Facets: A 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 | the root mean square of the differences between the calculated logits and their mean. |
| 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<br>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 ordered by the definition of the variable. |
| Strata | = (4*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 tales 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 he 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) ordering (hierarchy) according to 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 | the idea of what we want to measure A variable is defined by the items or agents of measurement used to elicit its manifestations or responses. |
| 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. |
| &END | The end of the list of Winsteps control variables |
| &INST | The beginning of the list of Winsteps control variables. This is not necessary. |

## 16.33 Group comparisons

Comparing groups of elements: It always helps in these situations to consider the equivalent situation in physical measurement. If Group B were boys and Group G were girls, and you were measuring their heights, how would you compare them?

You would get a height measure for every boy, and a height measure for every girl and then do the standard comparisons of means and standard deviations. This would be easiest for your audience to understand. So let's do the same thing with the Rasch measures.

Use group number "1" in the person labels of all boys, and group number "2" in the person labels of all girls.

*Facets* will then give a report by group for the person facet. Here it is for the "kct.txt" example.

Knox Cube Test (Best Test Design p.31)
Table 7.1.4 Children Measurement Report (arranged by N).

| Total Score | Total Count | Obsvd Average | Fair-M Avrage | Model Measure | Model S.E. | Infit MnSq | Infit ZStd | Outfit MnSq | Outfit ZStd | Estim. Discrm | Group | Nu Children |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 9.7 | 18.0 | .5 | .46 | 1.62 | 1.03 | 1.00 | -.2 | .69 | .9 | Boys+Girls | | Mean (Count: 35) |
| 2.4 | .0 | .2 | .31 | 2.23 | .17 | .95 | 1.2 | 1.30 | 1.1 | | | S.D. (Population) |
| 2.4 | .0 | .2 | .31 | 2.26 | .18 | .96 | 1.2 | 1.32 | 1.1 | | | S.D. (Sample) |
| 9.5 | 18.0 | .5 | .47 | 1.42 | 1.05 | 1.11 | .0 | .72 | 1.0 | Boys | 1 | Mean (Count: 17) |
| 2.7 | .0 | .2 | .33 | 2.50 | .23 | .91 | 1.1 | 1.41 | 1.2 | | 1 | S.D. (Population) |
| 2.8 | .0 | .2 | .34 | 2.58 | .23 | .94 | 1.2 | 1.46 | 1.2 | | 1 | S.D. (Sample) |
| 10.0 | 18.0 | .5 | .45 | 1.80 | 1.02 | .89 | -.3 | .66 | .8 | Girls | 2 | Mean (Count: 18) |
| 2.0 | .0 | .1 | .29 | 1.91 | .10 | .96 | 1.2 | 1.19 | 1.1 | | 2 | S.D. (Population) |
| 2.0 | .0 | .1 | .30 | 1.97 | .10 | .99 | 1.2 | 1.23 | 1.1 | | 2 | S.D. (Sample) |

The standard error of the mean = Population S.D. / √(Count-1)
The "Model S.E." in the Mean rows is the average of the Model S.E.s for the sample, not the S.E. of the Mean Measure.

So the significance of (Boys-Girls) is *t* = (Boy mean - Girl mean) / √(S.E. (mean Boys)² + S.E. (mean Girls)²) = (1.42 - 1.80) / √( 2.50²/(17-1) + 1.91²/(18-1)) = -0.49 which is not significant with about 32 d.f.

## 16.34 Guessing elimination

Facets does not have a mechanism for compensating for guessing.

The nearest thing would be to trim the data to remove all responses to items that are too difficult for the person.

1. Do the standard Facets analysis

2. Output the Residual file= to Excel

3. Sort the Excel worksheet on the column Meas.

4. Delete all rows when the Meas is too negative = the challenge is too high, so provoking guessing

5. Construct the trimmed data file, by reaarranging the Excel columns:
Element number for Facet 1 Element number for Facet 2 .... Obs (delete the other columns)

6. Rerun the Facets analysis with Data= the Excel file from 5.

## 16.35 Hierarchical Rater Models HRM

The Hierarchical Rater Model (HRM) of Patz et al. (2002) and the Hierarchical Rater Model based on Signal Detection Theory (HRM-SDT) by DeCarlo et al. (2011) are effective for investigating rater effects such as leniency differences, halo effects and central tendency, also item effects such as difficulty differences. HRM is a diagnostic tool. Unlike MFRM (*Facets),* it does not construct measures on a linear latent variable.

R Statistics packages for HRM include immer and sirt.

## 16.36 Inestimable elements or excessive logit ranges

Facets reports elements are inestimable when the data cannot produce an estimate.

1. Inestimable due to all observations being in the same category of a partial-credit item.
Solution: Use Rating Scale= to anchor the thresholds of the item at reasonable values.

2. A frequently-encountered problem in the analysis of paired-comparison data is an almost Guttman ordering of the pairings. This can lead to unrealistically huge logit ranges for the estimates of the elements or inestimable elements.

To solve this problem, we apply a little Bayesian logic. We know that the range of paired performances is not exceedingly wide, and we can also easily imagine a performance better than any of those being paired, and also a performance worse than any of the those being paired. Let's hypothesize that a reasonable logit distance between those two hypothetical performances is, say, 20 logits.

https://www.rasch.org/rmt/rmt151w.htm is a parallel situation for sports teams.

a. Hypothesize a "best"performance against which every other performance is worse. Anchor it at 10 logits.
b. Hypothesize a "worst" performance against which every other performance is better. Anchor it at -10 logits
c. Hypothesize a dummy judge who compares the best and worst performances against all the other performances.
d. Include these dummy observations in the analysis.

e. Analyze the actual observations + the dummy observations. The analysis should make sense, and the logit range of the performances will be about 20 logits. For reporting, we don't want the dummy material, so we write an Anchorfile= from this analysis.

f. We then use the Anchorfile as the Facets specification file, commenting out the "best" and "worst" performance elements and the dummy judge. We analyze only the actual observations. All the elements are anchored at their estimates from the actual+dummy analysis. In this anchored analysis, the "displacements" indicate the impact of the dummy data on the estimates.

g. If you perceive that the logit range of 20 logits is too big or too small, please adjust the "best" and "worst" anchored values.

3. Inestimable because elements only have one observation.

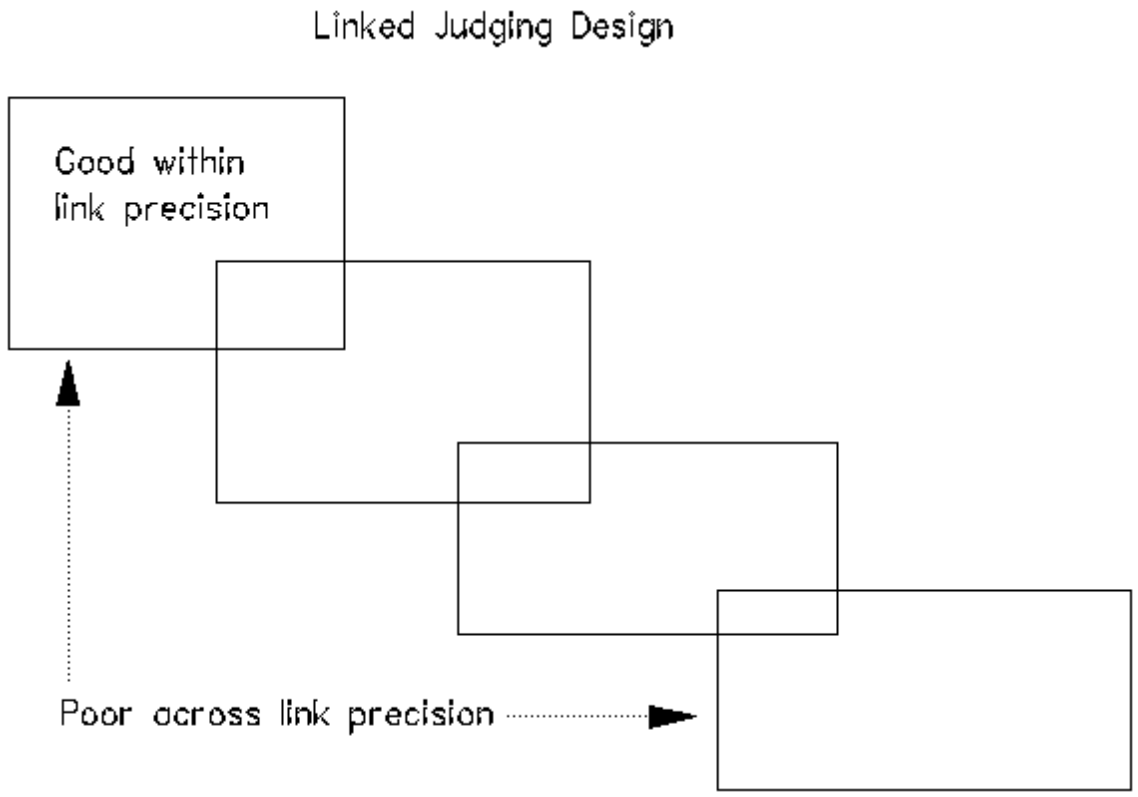
Solution: add a dummy person who interacts with the other elements in a central category. Give the data for this person a very small weight. Example:

R0.01, 101, 1-9, 2,2,2,2,2,2,2,2,2 ; dummy observations weighted low to make the data estimable

Those ratings are arbitrary additions to your data to make your data estimable. Your data are too thin for an ordinary analysis. We need at least two observations of each element. Accordingly I added an extra person, 101, who interacted with every item 1-9 in the same category 2 on the 1-5 rating scale. I gave this person a very small weight R0.01 in the analysis. Person 101 ties all the items together, which makes all the items, and so all the persons, estimable. However this additional person is arbitrary. *Facets* produces estimates, but different arbitrary data would produce different estimates.

If you are sharing your results with a non-technical audience, remove person 101 from the output reports.

## 16.37 Instability of Long Links




Facets can certainly analyze this type of design, but it does not include across-link imprecision in its estimates of standard errors. See BTD, p. 98 ff. for some guidance on test linking networks.

## 16.38 Inter-rater and intra-rater Reliability

```
Table 7.3.1  Reader Measurement Report  (arranged by MN).
-------------------------------------------------------------------------------------------------
        -
| Obsvd   Obsvd  Obsvd  Fair-M|        Model | Infit      Outfit   | Exact Agree. |
        |
| Score   Count Average Avrage|Measure  S.E. |MnSq ZStd  MnSq ZStd | Obs %  Exp % | Nu Reader
        |
-------------------------------------------------------------------------------------------------
        -
|   1524    288     5.3   5.26|   -.30   .05 | 1.2   2    1.2   2  |  28.2   20.9 |  8 8
        |
|   1455    288     5.1   5.00|   -.16   .05 |  .5  -7     .5  -7  |  30.8   21.6 |  4 4
        |
....
-------------------------------------------------------------------------------------------------
        -
RMSE (Model)  .05 Adj S.D.  .19  Separation  4.02  Strata  5.69  Reliability  .94 <-- Spearman
......
Inter-Rater agreement opportunities: 60480  Exact agreements: 17838 = 29.5%  Expected: 13063.2
        = 21.6%
-------------------------------------------------------------------------------------------------
       -
```

Inter-rater= facet-number to report inter-rater agreement statistics.

In Facets Table 7, there are "Reliability" indexes reported for every facet. This reliability is the Spearman Reliability. Cronbach Alpha is also an estimate of Spearman Reliability. This reliability "distinguishes between different levels of severity among" the elements of the facet. Higher reliability = more levels.

Inter-rater reliability is not the same as Spearman Reliability. For inter-rater reliability, higher reliability = more similarity. There are three families of inter-rater reliability statistics. (i) Do the raters agree with each other about the examinee's rating? (ii) Do the raters agree with each other about which examinees are better and which examinees are worse? (iii) Do the raters give the correct rating to the performance?

Inter-rater reliability (i)is used for pass-fail decisions about the examinees. Inter-rater reliability (ii) is used when the rank-order of the examinees is crucial. Inter-rater reliability (iii) is used when certifying raters.

---

**Intra-rater reliability** can be deduced from the rater's fit statistics. The lower the mean-square fit, the higher the intra-rater reliability. This is because high intra-rater reliability implies that the ratings given by the rater can be accurately predicted from each other

**Intra-rater consistency** would be maximized when a rates rates every performance in the same category. This would be equivalent to the "attenuation paradox" in Classical Test Theory. There is an optimal consistency beyond which validity drops.

So here is an approach:
1. Model each rater to have a unique rating scale. e.g., Models=?,#,?,?,R5
2. We want the category frequencies to be a smooth unimodal distribution with all categories well represented. This will also force the Andrich thresholds to be ordered.
3. We want the average measures for each category to be close to their expectations.
3. We want the mean-square fit statistics for each category to be close to 1.0

A rater who meets these requirements is consistent from a Rasch perspective.

---

There is no generally-agreed index of **inter-rater reliability** (IRR). The choice of IRR depends on the purpose for which the ratings are being collected, and the philosophy underlying the rating process.

For raters, there are a number of steps in deciding what quality-indexes to report:
1. Are the raters intended to act as independent experts or as "rating machines"?
2. Are the ratings reflective of criterion-levels or of relative performance?
3. How are differences in rater leniency to be managed?
4. How are rater disagreements to be managed?

First, you have to decide what type of rater agreement you want.

Do you want the raters to agree exactly with each other on the ratings awarded? The "rater agreement %".

Do you want the raters to agree about which performances are better and which are worse? Correlations

Do you want the raters to have the same leniency/severity? "1 - Separation Reliability" or "Fixed Chi-squared"

Do you want the raters to behave like independent experts? Rasch fit statistics

Typical indexes include: proportion of exact agreements (Cohen's kappa), correlations, variances (G-Theory).

In the literature there is no clear definition of this, so you must decide what the term means for your situation.

A. It can mean "to what extent to do pairs of raters agree on the same rating?". This is the "exact observed agreement" statistic. If you want your raters to act like "rating machines" (human optical scanners), then you expect to see agreement of 90%+. Raters are often trained to act like this.
B. It can mean "are the ratings of pairs of raters highly correlated?". Facets does not report this directly.
C. Are pairs of raters acting like independent experts (the ideal for Facets)? If so the "observed agreements" will be close to the "expected agreements".
D. Do raters have same level of leniency/severity? This is reported by the "Reliability (not inter-rater)" statistic. For raters we like to see this close to 0, so that the rater measures are not reliably different. We also like to see the "Fixed all-same" chi-squared test not be rejected.

| Rasch Agreement | Interpretation |
|---|---|
| Observed < Expected | Indicates disagreement, normally happens with untrained raters. |
| Observed ≈ Expected | Raters act independently. Need verification with fit statistics. |
| Observed somewhat > expected | Normal for trained raters. Training emphasizes agreement with others, but rating process requires raters to rate independently. |
| Observed >> expected | Raters do not rate independently. There may be pressure to agree with other raters. |
| Observed > 90% | Raters behave like a rating machine. Seriously consider excluding from the measurement model. Specify the rater facet as a demographic one ,D |

There is not a generally agreed definition of "inter-rater reliability". Do you intend your raters to act like "rating machines" or as "independent experts"? "Rating machines" are expected to give identical ratings under identical conditions. "Independent experts" are expected to show variation under identical conditions. *Facets* models raters to be "independent experts". An interrater reliability coefficient, IRR, is not computed. But, from one perspective, it is the reverse of the Separation Reliability, i.e., 1 - Separation Reliability.

For "rating machines", there are several inter-rater approaches. For these you need to use other software:

1. Raters must agree on the exact value of the ratings: use a Cohen's-Kappa type of inter-rater reliability index. Cohen's Kappa is (Observed Agreement% - Chance Agreement%)/(100-Chance Agreement%) where chance is determined by the marginal category frequencies. A Rasch version of this would use the "Expected Agreement%" for an adjustment based on "chance + rater leniency + rating scale structure". Then the Rasch-Cohen's Kappa would be: (Observed%-Expected%)/(100-Expected%). Under Rasch-model conditions this would be close to 0.

For Rasch-Kappa, Mojtaba Taghvafard's investigations suggest that:

| Rasch-Kappa Value | Meaning |
|---|---|
| -0.2 to +0.2 | model-expected level of agreement |
| -0.2 to -0.4<br>+0.2 to +0.4 | little more agreement than modeled |
| => +0.5 or =<-0.5 | high agreement |

2. Raters must agree on higher and lower performance ratings: use a correlation-based inter-rater reliability index.

3. Inter-rater variance must be much less than inter-examinee variance: compare the Rater facet S.D. with the Examinee facet S.D.

When the raters are behaving like rating machines, alternative analytical approaches should be considered, such as Wilson M. & Hoskens M. (2001) The Rater Bundle Model, *Journal of Educational and Behavioral Statistics, 26, 3,* 283-306, or consider specifying the rater facet in your analysis as a demographic facet ,D which has all elements anchored at zero.

**The computation**
How many ratings are made under identical conditions (usually by different raters) and how often are those ratings in exact agreement? This investigation is done pairwise across all raters. All facets except the Inter-rater= facet participate in the matching. If the inter-rater facet is Entered= more than once, only the first entry is active for this comparison.

To exclude dummy facets and irrelevant ones, do a special run with those marked by X in the model statements. For example, facet 1 is persons, facet 2 is gender (sex) (dummy, anchored at zero), facet 3 is rater, facet 4 is item, facet 5 is rating day (dummy, anchored at 0). Then Gender and Rating Day are irrelevant to the pairing of raters:
Inter-rater=3
Model = ?,X,?,?,X, R6

**1. Inter-Rater Agreement Opportunities**

| Raters: Senior scientists | Junior Scientists | Traits | Observation | Inter-Rater Agreement Opportunities | Observed Exact Agreement |
|---|---|---|---|---|---|
| Avogadro | Anne | Attack | 5 | 1 | 0.5 (agrees with Cavendish but not Brahe) |
| Cavendish | Anne | Attack | 5 | 1 | 0.5 (agrees with Avogadro but not Brahe) |
| Brahe | Anne | Attack | 6 | 1 | 0 (disagrees) |
| | | | | | |
| Avogadro | Anne | Basis | 5 | 1 | 1 (agrees) |
| Brahe | Anne | Basis | 5 | 1 | 1 |
| Cavendish | Anne | Basis | 5 | 1 | 1 |
| | | | | | |
| Avogadro | Anne | Clarity | 3 | 1 | 0 (disagrees) |
| Brahe | Anne | Clarity | 4 | 1 | 0 |
| Cavendish | Anne | Clarity | 5 | 1 | 0 |

In the Table above, "Inter-Rater Agreement Opportunities" are computed for each rater. There is one opportunity for each observation awarded by a rater under the same circumstances (i.e., same person, same item, same task, ....)

as another observation. In the Guilford.txt example, there are 105 observations, all in situations where there are multiple raters, so there are 105 agreement opportunities.

"Observed Exact Agreement" is the proportion of times one observation is exactly the same as one of the other observations for which there are the same circumstances. If, under the same circumstances, the raters all agree, then the Exact Agreement is 1 for each observation. If the raters all disagree, then the Exact Agreement is 0 for each observation. If some raters agree, then the Exact agreement for each observation is the fraction of opportunities to agree with other raters. In the Guilford data, there are 35 sets of 3 ratings: 5 sets of complete agreement = 5 *3 =15. There are 18 sets of partial agreement = 18 * 2 * 0.5 = 18. There are 12 sets of no agreement = 12 * 0 = 0. The agreements sum to 33.

**2. Fleiss' Kappa**

By contrast, Fleiss' kappa has the formula: Kappa = (Pobserved - Pchance) / (1 - Pchance) computed across all raters and rating-scale categories.

Proportion of observations in each category j is reported in Table 8 as the "Count %".

| DATA | | | |
|---|---|---|---|
| Category Score | Counts Used | % | Cum. % |
| 1 | 4 | 4% | 4% |
| 2 | 4 | 4% | 8% |
| 3 | 25 | 24% | 31% |
| 4 | 8 | 8% | 39% |
| 5 | 31 | 30% | 69% |
| 6 | 6 | 6% | 74% |
| 7 | 21 | 20% | 94% |
| 8 | 3 | 3% | 97% |
| 9 | 3 | 3% | 100% |

Pchance = $\Sigma(\text{Count \%}/100)^2$ = $.04^2 + .04^2 + .24^2 + .08^2 + .30^2 + .06^2 + .20^2 + .03^2 + .03^2 = 0.20$

Then, considering "7 Junior scientists + 5 Traits" as 35 "subjects", so that there are three raters, and n = 3 observations for each subject.

**Table of Inter-rater Agreements**

| Subject | | Senior Scientist = Rater | | | |
|---|---|---|---|---|---|
| Junior Scientist | Trait | 1 | 2 | 3 | Agreements |
| 1 | 2 | 5 | 5 | 5 | Complete |
| 1 | 5 | 3 | 3 | 3 | Complete |
| 2 | 2 | 7 | 7 | 7 | Complete |
| 2 | 3 | 5 | 5 | 5 | Complete |
| 7 | 4 | 5 | 5 | 5 | Complete |
| 1 | 3 | 3 | 4 | 5 | None |
| 1 | 4 | 5 | 6 | 7 | None |
| 2 | 1 | 9 | 8 | 7 | None |
| 3 | 4 | 7 | 6 | 5 | None |

| Subject | | Senior Scientist = Rater | | | |
|---|---|---|---|---|---|
| Junior Scientist | Trait | 1 | 2 | 3 | Agreements |
| 3 | 5 | 1 | 6 | 5 | None |
| 4 | 3 | 1 | 4 | 3 | None |
| 4 | 5 | 3 | 5 | 1 | None |
| 5 | 4 | 8 | 2 | 7 | None |
| 5 | 5 | 5 | 3 | 7 | None |
| 6 | 2 | 5 | 4 | 3 | None |
| 6 | 5 | 1 | 2 | 3 | None |
| 7 | 5 | 5 | 4 | 7 | None |
| 1 | 1 | 5 | 6 | 5 | Partial |
| 2 | 4 | 8 | 7 | 7 | Partial |
| 2 | 5 | 5 | 2 | 5 | Partial |
| 3 | 1 | 3 | 4 | 3 | Partial |
| 3 | 2 | 3 | 5 | 5 | Partial |
| 3 | 3 | 3 | 3 | 5 | Partial |
| 4 | 1 | 7 | 5 | 5 | Partial |
| 4 | 2 | 3 | 6 | 3 | Partial |
| 4 | 4 | 3 | 5 | 3 | Partial |
| 5 | 1 | 9 | 2 | 9 | Partial |
| 5 | 2 | 7 | 4 | 7 | Partial |
| 5 | 3 | 7 | 3 | 7 | Partial |
| 6 | 1 | 3 | 4 | 3 | Partial |
| 6 | 3 | 3 | 6 | 3 | Partial |
| 6 | 4 | 5 | 4 | 5 | Partial |
| 7 | 1 | 7 | 3 | 7 | Partial |
| 7 | 2 | 7 | 3 | 7 | Partial |
| 7 | 3 | 5 | 5 | 7 | Partial |

Pi = extent of agreement for subject i = ($\Sigma$(count of observations for subject i in category j)$^2$ - n)/(n(n-1))

For 5 subjects where all 3 raters rate in the same category, Pi = ($3^2$ - 3) / (3*2) = 1
For 18 subjects where 2 raters rate in the same category, and 1 rater in a different category, Pi = ($2^2$ + 1 - 3)/(3*2) = 0.33
For 12 subjects where all 3 raters rate in different categories, Pi = ($1^2$ + $1^2$ + $1^2$ - 3)/(3*2) = 0

Pobserved = Mean (Pi) = (1*5 + 18*0.33 + 12*0)/35 = 0.31

Fleiss kappa = (0.31 - 0.20) / (1 - 0.20) = 0.11 / 0.80 = 0.14, which is considered "slight agreement" in Wikipedia - Fleiss Kappa

Here are the results from Guilford's data. Note that Avogadro and Cavendish show much higher agreement rates than the model predicts. It seems that they share something which contrasts with Brahe:

| Obsvd Score | Obsvd Count | Obsvd Average | Exact Agree. Obs % | Exact Agree. Exp % | N Senior scientists |
|---|---|---|---|---|---|
| 156 | 35 | 4.5 | 21.4 | 25.2 | 2 Brahe |
| 171 | 35 | 4.9 | 35.7 | 25.8 | 1 Avogadro |
| 181 | 35 | 5.2 | 37.1 | 25.3 | 3 Cavendish |

Rater agreement opportunities: 105 Exact agreements: 33 = 31.4% Expected: 26.7 = 25.4%

When an anchorfile= is produced, and used for a subsequent analysis with Brahe commented out, then the agreement between Avogadro and Cavendish is twice what is expected!

| Obsvd Score | Obsvd Count | Obsvd Average | Exact Agree. Obs % | Exact Agree. Exp % | N Senior scientists |
|---|---|---|---|---|---|
| 171 | 35 | 4.9 | 51.4 | 25.9 | 1 Avogadro |
| 181 | 35 | 5.2 | 51.4 | 25.9 | 3 Cavendish |

**3. Krippendorff's Alpha**

"Krippendorff's alpha (Kα) is a reliability coefficient developed to measure the agreement among observers, coders, judges, raters, or measuring instruments drawing distinctions among typically unstructured phenomena or assign computable values to them. Kα emerged in content analysis but is widely applicable wherever two or more methods of generating data are applied to the same set of objects, units of analysis, or items and the question is how much the resulting data can be trusted to represent something real." *K. Krippendorff (2011) Computing Krippendorff's Alpha-Reliability*.

Kα = 1 - Observed Disagreement / Expected Disagreement
where Kα =1 is perfect agreement, Kα=0 is agreement by chance or worse.

1,m = rating-scale categories
A = Agreement opportunities = count of situations where a pair of raters have the same elements for all other facets
nj = count of observations of category j across all raters
N = sum (nj) for j=1,m
C = Agreement by chance = sum (nj*nj)/N for j=1,m - N

a) assuming all raters have the same leniency/severity!
O = Observed agreement = count of situations where pairs of raters have given the same rating
E = Expected agreement = sum(pj*pj) for j=1 to m across A, with measures of each situation (except raters)

b) assuming all raters have their own severity
O = Observed agreement = count of situations where pairs of raters have rating residuals that differ by 0.5 or less.
E = Expected agreement = sum(pj*pk) for j=1 to m, k=1 to m, and residual difference<=0.5, across A, with measures of each situation (with raters)

## 16.39 Inter-rater Correlations

Inter-rater consistency: In the Table below, from a Facets analysis of the example "Essays.txt" data, the Correlation columns give the observed and expected correlations between the ratings given by each reader and the element measures. The element measures are based on the ratings given by all the readers, so this Point-Measure (PT-Biserial=Measure) correlation summarizes the agreement between this Reader and the consensus of all the readers. If you want the same correlation, but using the unadjusted ratings (not measures), then Pt-Biserial=Yes

produces the column at the right of this table. The correlations are lower because they are not adjusted for rater leniency. BTW, this data is high quality, produced by ETS using their best raters for a special study.

AP English Essays (College Board/ETS) 8/24/2023 11:07:23 AM
Table 7.3.1 Reader Measurement Report (arranged by MN).

| Total Score | Total Count | Obsvd Average | Fair(M) Average | Measure | S.E. | Model MnSq | ZStd | Infit MnSq | ZStd | Outfit MnSq | ZStd | Estim. Discrm | Correlation PtMea | PtExp | Exact Agree. Obs % | Exp % | Nu | Reader | Corr. PtBis |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 508 | 96 | 5.29 | 5.26 | -.30 | .08 | 1.23 | 1.6 | 1.21 | 1.4 | .75 | | .60 | .62 | | 20.8 | 20.4 | 8 | 8 | .32 |
| 485 | 96 | 5.05 | 5.00 | -.16 | .08 | .52 | -4.2 | .53 | -4.1 | 1.48 | | .67 | .62 | | 21.2 | 21.7 | 4 | 4 | .39 |
| 484 | 96 | 5.04 | 4.99 | -.15 | .08 | 1.02 | .1 | 1.01 | .0 | .97 | | .64 | .62 | | 24.1 | 21.6 | 9 | 9 | .36 |
| 479 | 96 | 4.99 | 4.93 | -.12 | .08 | 1.13 | .9 | 1.13 | .9 | .83 | | .55 | .62 | | 28.8 | 21.7 | 7 | 7 | .29 |
| 473 | 96 | 4.93 | 4.86 | -.08 | .08 | 1.06 | .5 | 1.06 | .4 | .93 | | .56 | .62 | | 20.8 | 21.7 | 2 | 2 | .29 |
| 470 | 96 | 4.90 | 4.83 | -.06 | .08 | 1.40 | 2.6 | 1.37 | 2.4 | .63 | | .65 | .62 | | 27.8 | 22.0 | 12 | 12 | .33 |
| 466 | 96 | 4.85 | 4.79 | -.04 | .08 | 1.14 | .9 | 1.11 | .8 | .81 | | .62 | .61 | | 30.6 | 21.9 | 11 | 11 | .33 |
| 461 | 96 | 4.80 | 4.73 | .00 | .08 | .71 | -2.3 | .71 | -2.2 | 1.31 | | .68 | .61 | | 42.4 | 21.9 | 10 | 10 | .36 |
| 444 | 96 | 4.63 | 4.55 | .11 | .08 | .85 | -1.1 | .84 | -1.1 | 1.14 | | .60 | .61 | | 36.1 | 22.0 | 5 | 5 | .35 |
| 434 | 96 | 4.52 | 4.44 | .17 | .08 | 1.04 | .3 | 1.06 | .4 | .93 | | .65 | .60 | | 38.9 | 21.9 | 6 | 6 | .37 |
| 433 | 96 | 4.51 | 4.43 | .18 | .08 | 1.06 | .4 | 1.03 | .2 | .99 | | .64 | .60 | | 27.8 | 21.8 | 3 | 3 | .35 |
| 392 | 96 | 4.08 | 4.00 | .45 | .08 | .79 | -1.5 | .79 | -1.5 | 1.23 | | .48 | .59 | | 19.7 | 19.8 | 1 | 1 | .27 |
| 460.8 | 96.0 | 4.80 | 4.73 | .00 | .08 | 1.00 | -.1 | .99 | -.2 | | | .61 | | | | | Mean (Count: 12) | | .33 |
| 29.5 | .0 | .31 | .32 | .19 | .00 | .23 | 1.8 | .22 | 1.7 | | | .05 | | | | | S.D. (Population) | | .03 |
| 30.8 | .0 | .32 | .33 | .20 | .00 | .24 | 1.9 | .23 | 1.8 | | | .06 | | | | | S.D. (Sample) | | .04 |

Model, Populn: RMSE .08 Adj (True) S.D. .17 Separation 2.17 Strata 3.22 Reliability (not inter-rater) .82
Model, Sample: RMSE .08 Adj (True) S.D. .18 Separation 2.28 Strata 3.38 Reliability (not inter-rater) .84
Model, Fixed (all same) chi-squared: 66.3 d.f.: 11 significance (probability): .00
Model, Random (normal) chi-squared: 9.4 d.f.: 10 significance (probability): .49
Inter-Rater agreement opportunities: 384 Exact agreements: 108 = 28.1% Expected: 82.6 = 21.5%

---

*Amother approach:*

Use the Facets "Output Files" option to produce a Winsteps file.
Select the raters columns (items), and the relevant combinations of facets (e.g., examinees and tasks) as the rows.
This will produce a data file which can be used to produce inter-rater correlations in Excel.

If the inter-rater correlation computation is done in Winsteps,
In the Winsteps control file,
PRCOMP=Observation
ICORFILE=*inter-rater-correlations.txt*
Rasch estimates are not needed, so Ctrl+F soon after iteration starts.
*inter-rater-correlations.txt* contains a list of the inter-rater correlations based on the observations.

Otherwise, see Table 7 Agreement Statistics

## 16.40 Investigating differences between rater groups

**The Problem:**
Videotapes of 2 different flight crews, consisting of 2 crew members each, the pilot-in-command (PIC) and the second-in-command (SIC), are rated on 10 items that measure technical and teamwork skills. 2 overall items assess crew technical and teamwork performance. There is one item for overall evaluation of the PIC, and one item for overall evaluation of the SIC. There are 3 groups of judges, each group consisting of 7 to 10 judges. Each judge rates both crews on all items (4-point Likert scale). I want to investigate the differences between the groups of judges.

**Solution:**
Here is the skeleton of a suggested Facets control file:

```
Title="Air Crew Videotapes"

; facet 1 is judges
; facet 2 is judge group (a dummy facet)
; facet 3 is crew number
; facet 4 is crew component: crew, PIC or SIC
; facet 5 is items
Investigating differences between rater groups
facets=5

models=
?,#,?,?,?,Likert     ; model each group to have its own rating scale (or partial credit)
?,?B,?,?,?B,R        ; look for group by item interactions
?,?B,?B,?,?,R        ; look for group by crew number interactions
*

rating scale = Likert,R4,Specific   ;we model each group to have its own scale
1=SD                                ; compare the rating scale (or partial credit) structures
2=D                                 ; reported for the 3 groups
3=A
4=SA
*

labels=
1,Judges
1=Fred,,1          ;Fred is in group 1: Note the ,, to skip over the pre-set logit value field.
2=Mary,,3          ;Mary in group 3
3=Alfonso,,2       ;Alfonson in group 2
....               ;there will be subtotals by group
*
2,Group,A   ;dummy facet - used for classification
1=Group A,0 ;anchored at 0 to remove for measurement,
2=Group B,0 ;group facet used for classification and summary fit statistics
3=Group C,0
*

3,Crew number ; the videotapes
1=Crew X
2=Crew Y
*
4,Component
1=Crew      ;used for items 1-12
2=PIC       ;used for item 13
3=SIC       ;used for item 14
*
5,Items
```

```
1=technical
....
10=teamwork
11=technical performance
12=teamwork performance
13=PIC overall
14=SIC overall
*
data=      ;here are some example data lines:
2,3,1,1,1-12, 2,1,3,2,4,1,2,1,3,2,1,2 ;Mary in group 3 rates crew 1's crew on items 1-12
2,3,1,2,13,   4                   ;Mary in group 3 rates crew 1's PIC on item 13
2,3,1,2,14,   3                   ;Mary in group 3 rates crew 1's SIC on item 14
```

**Here is the skeleton of a likely Facform control file:**
Facform can be downloaded from www.winsteps.com/facets.htm

Original data format (say) in file rawdata.txt:
judge id  cols 1-3
judge group  cols 4
Crew number  col  5
Item ratings cols 6-19  (14 items)

```
$Title = "Format the Air Crew judge data"
$Input = "rawdata.txt"
$Output= "facdata.txt"

$Facets=5                    ;define number of facets
$Flabel=1,"Judges"           ;define facet names
$Flabel=2,"Group,A"
$Flabel=3,"Crew number"
$Flabel=4,"Component"
$Flabel=5,"Items"

$Label=1,$S1W3              ;identification for facet 1 judge
$Label=2,$S4W1              ;identification for facet 2 group
$Label=3,$S5W1              ;identification for facet 3 crew number

$label=4,1,"Crew"           ; identification for facet 4 - the crew
item = 1
$Do=12
    $Label=5,item           ; identification for facet 5 - item number
    $Rating=$S(item+5)W1    ; response to item 1 etc.
    item = item +1          ; next item
$Again

$label=4,2,"PIC"           ; identification for facet 4 - PIC
$Label=5,13,"PIC overall"  ; identification for facet 5 - item 13
$Rating=$S18W1             ; response to item 13

$label=4,3,"SIC"           ; identification for facet 4 - SIC
$Label=5,14,"SIC overall"  ; identification for facet 5 - item 14
$Rating=$S19W1             ; response to item 14
```

## 16.41 Iterations do not stop

**"*Facets* runs and runs, but does not converge! What has gone wrong?"**

Are the current set of estimates good enough for your purposes? Estimates with extremely high precision are rarely needed, and seldom improve overall fit of the data to a Rasch model. Set Convergence= and Iterations= to reasonable values.

If the changes per iteration are very small or are oscillating up and down about equally, then the estimates are as good as they can be. Press Ctrl+F to stop iteration and move on to fit computation. You can also use the Estimation menu to make the changes per iteration smaller, or to change the estimation method.

If Facets runs for more than 100 iterations, and the residuals are not converging (approaching zero), then please force Facets to stop iterating by Ctrl+F (or use the Estimation pull-down menu) - then look at the screen and the output file. Investigate elements with large displacements and rating-scale categories with large "Observed - Expected Diagnostic Residuals".

There are numerous possibilities. Here are some to check.

1) **Noncenter=**
   The analysis may be
   i) over-constrained because all facets are centered, anchored or group-anchored,
   ii) under-constrained because more than one facet is non-centered.
   For most analyses, noncenter= must specify one active facet.
   Make sure that facet is specified is not marked X or omitted from the model statement.
   Makes sure that the non-centered facet does not include anchor or group-anchor values.

2) **Null element= (Keepasnull=)**
   Element 0 in the data is usually a dummy element that means "this facet does not apply to this observation". If element 0 is an active element in your analysis, then please assign a different unused element number as the dummy element number, e.g., *Null element = 999*

3) **Your facets are nested**.
   When one facet is nested within another facet (without anchoring or group-anchoring), Facets makes an arbitrary allocation of statistical information between the facets, this can lead to unstable estimation. For instance, if there is a "student" facet and also a "student gender" (sex) facet, then nesting needs to be resolved by anchoring or group-anchoring either the "student" facet or the "student gender" facet.

4) **Low category frequencies**
   Rating (or partial credit) scale categories with very low frequencies (less than 10 observations) are difficult to use as the basis of estimation. Convergence may take a long time.

5) **Rating (or partial credit) scales with many categories.**
   Estimating the parameters of long rating (or partial credit) scales (more than 10 categories) is difficult. Adjustments tend to ripple up and down the scale, like a caterpillar moving. Convergence may take a long time.

6) **Category frequencies are lumpy.**
   When some categories of a rating scale have high frequencies and other categories have low frequencies, convergence may take a long time.

7) **Clumps of data or stringy data due to missing data. Disconnected or weakly connected subsets.**
   Data with disjoint subsets of data, or with only thin connections between the subsets, may not converge, or may converge to non-repeatable values. For instance, your judging design may be very sparse.

8) **The Maximum Likelihood curve is almost flat.**
   The data are such that there are a range of estimated measures that match the data equally well. All are equally valid.

9) **The Maximum Likelihood curve has two peaks very close together.**
   The data are such that there are two almost equally good sets of estimates that match the data. Facets cannot choose between them.

## 16.42 Judging Plans and Experimental Designs

Suggestion: "Output Files", "Winsteps control file" is a convenient way of producing a rectangular picture of your judging plan from your data.

The only requirement on the judging plan is that there be enough linkage between all elements of all facets that all parameters can be estimated without indeterminacy within one frame of reference. Fig 1 illustrates an ideal judging plan for both conventional and Rasch analysis. The 1152 ratings shown are a set of essay ratings from the Advanced Placement Program of the College Board. These are also discussed in Braun (1988). This judging plan meets the linkage requirement because every element can be compared directly and unambiguously with every other element. Thus it provides precise and accurate measures of all parameters in a shared frame of reference. For robust estimation of measures, we need 30 observations of each element, and at least 10 observations of each rating-scale category.

Less data intensive, but also less precise, Rasch estimates can be obtained so long as overlap is maintained. Fig. 2 illustrates such a reduced network of observations which still connects examinees, judges and items. The parameters are linked into one frame of reference through 180 ratings which share pairs of parameters (common essays, common examinees or common judges). Accidental omissions or unintended ratings would alter the judging plan, but would not threaten the analysis. Measures are less precise than with complete data because 83% less observations are made.

Judging is time-consuming and expensive. Under extreme circumstances, judging plans can be devised so that each performance is judged only once. Even then the statistical requirement for overlap can usually be met rather easily. Fig. 3 is a simulation of such a minimal judging plan. Each of the 32 examinees' three essays is rated by only one judge. Each of the 12 judges rates 8 essays, including 2 or 3 of each essay type. Nevertheless the examinee-judge-essay overlap of these 96 ratings enables all parameters to be estimated unambiguously in one frame of reference. The constraints used in the assignment of essays to judges were that (1) each essay be rated only once; (2) each judge rate an examinee once at most; and (3) each judge avoid rating any one type of essay too frequently. The statistical cost of this minimal data collection is low measurement precision, but this plan requires only 96 ratings, 8% of the data in fig. 1. A practical refinement of this minimal plan would allow each judge to work at his own pace until all essays were graded, so that faster judges would rate more essays. A minimal judging plan of this type has been successfully implemented (Lunz *et al.,* 1990).

See also "Does Sparseness Matter? Examining the Use of Generalizability Theory and Many-Facet Rasch Measurement in Sparse Rating Designs", S.A. Wind et al., https://doi.org/10.1177/01466216231182148

| Judge<br>Essay | 1<br>ABC | 2<br>ABC | 3<br>ABC | 4<br>ABC | 5<br>ABC | 6<br>ABC | 7<br>ABC | 8<br>ABC | 9<br>ABC | 10<br>ABC | 11<br>ABC | 12<br>ABC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Personl | 553 | 686 | 877 | 687 | 777 | 685 | 565 | 667 | 586 | 567 | 776 | 696 |
| 2 | 454 | 542 | 445 | 534 | 334 | 344 | 433 | 526 | 444 | 445 | 533 | 534 |
| 3 | 434 | 544 | 343 | 555 | 433 | 544 | 563 | 443 | 554 | 454 | 443 | 343 |
| 4 | 345 | 426 | 232 | 545 | 445 | 225 | 464 | 456 | 642 | 446 | 445 | 335 |
| 5 | 443 | 548 | 656 | 545 | 657 | 448 | 558 | 466 | 464 | 448 | 547 | 348 |
| 6 | 544 | 846 | 843 | 565 | 633 | 367 | 788 | 673 | 666 | 566 | 564 | 454 |
| 7 | 545 | 665 | 454 | 667 | 755 | 646 | 773 | 785 | 874 | 565 | 745 | 447 |
| 8 | 553 | 763 | 655 | 675 | 775 | 653 | 773 | 656 | 784 | 576 | 573 | 574 |
| 9 | 343 | 643 | 643 | 645 | 534 | 523 | 665 | 674 | 753 | 546 | 545 | 765 |
| 10 | 564 | 766 | 884 | 776 | 655 | 667 | 875 | 778 | 778 | 667 | 649 | 888 |
| 11 | 535 | 524 | 537 | 544 | 545 | 435 | 546 | 557 | 326 | 446 | 456 | 334 |
| 12 | 436 | 644 | 444 | 546 | 666 | 555 | 574 | 445 | 745 | 356 | 763 | 676 |
| 13 | 445 | 486 | 657 | 566 | 246 | 366 | 368 | 448 | 467 | 348 | 569 | 349 |
| 14 | 446 | 533 | 333 | 344 | 545 | 343 | 463 | 353 | 354 | 346 | 462 | 363 |
| 15 | 548 | 855 | 743 | 746 | 766 | 656 | 665 | 765 | 854 | 666 | 862 | 844 |
| 16 | 644 | 653 | 547 | 545 | 643 | 454 | 556 | 467 | 666 | 447 | 558 | 667 |
| 17 | 414 | 817 | 625 | 628 | 536 | 518 | 425 | 618 | 717 | 627 | 639 | 436 |
| 18 | 334 | 655 | 443 | 445 | 243 | 473 | 445 | 747 | 654 | 445 | 435 | 334 |
| 19 | 747 | 745 | 837 | 756 | 755 | 847 | 664 | 688 | 737 | 656 | 847 | 938 |
| 20 | 443 | 666 | 735 | 556 | 557 | 557 | 588 | 667 | 666 | 557 | 476 | 488 |
| 21 | 242 | 443 | 336 | 465 | 245 | 243 | 263 | 245 | 441 | 253 | 342 | 254 |
| 22 | 564 | 765 | 747 | 666 | 864 | 577 | 667 | 576 | 667 | 557 | 667 | 785 |
| 23 | 446 | 566 | 753 | 646 | 444 | 565 | 475 | 388 | 576 | 557 | 557 | 776 |
| 24 | 332 | 422 | 334 | 433 | 322 | 214 | 423 | 223 | 323 | 313 | 233 | 223 |
| 25 | 543 | 664 | 544 | 657 | 646 | 544 | 454 | 448 | 547 | 545 | 456 | 464 |
| 26 | 644 | 764 | 955 | 756 | 545 | 658 | 655 | 867 | 776 | 646 | 756 | 885 |
| 27 | 342 | 346 | 334 | 344 | 346 | 234 | 256 | 256 | 345 | 345 | 256 | 253 |
| 28 | 343 | 463 | 335 | 334 | 465 | 573 | 341 | 475 | 442 | 243 | 462 | 272 |
| 29 | 433 | 444 | 323 | 446 | 334 | 333 | 235 | 336 | 423 | 336 | 323 | 343 |
| 30 | 542 | 564 | 244 | 655 | 445 | 224 | 546 | 575 | 645 | 446 | 432 | 555 |
| 31 | 325 | 514 | 313 | 425 | 315 | 314 | 334 | 225 | 525 | 314 | 324 | 314 |
| 32 | 644 | 744 | 445 | 545 | 533 | 553 | 567 | 584 | 664 | 447 | 556 | 364 |

Figure 1. Complete judging plan for the Essay data.
(Courtesy: Robert G. Cameron of the College Board).

| Judge<br>Essay | 1<br>ABC | 2<br>ABC | 3<br>ABC | 4<br>ABC | 5<br>ABC | 6<br>ABC | 7<br>ABC | 8<br>ABC | 9<br>ABC | 10<br>ABC | 11<br>ABC | 12<br>ABC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Personl | 553 | 686 | | | | | | | | | | |
| 2 | | 542 | 445 | | | | | | | | | |
| 3 | | | 343 | 555 | | | | | | | | |
| 4 | | | | 545 | 445 | | | | | | | |
| 5 | | | | | 657 | 448 | | | | | | |
| 6 | | | | | | 367 | 788 | | | | | |
| 7 | | | | | | | 773 | 785 | | | | |
| 8 | | | | | | | | 656 | 784 | | | |
| 9 | | | | | | | | | 753 | 546 | | |
| 10 | | | | | | | | | | 667 | 649 | |
| 11 | | | | | | | | | | | 456 | 334 |
| 12 | 436 | | | | | | | | | | | 676 |
| 13 | 445 | | | | | | 368 | | | | | |
| 14 | | 533 | | | | | | 353 | | | | |
| 15 | | | 743 | | | | | | 854 | | | |
| 16 | | | | 545 | | | | | | 447 | | |
| 17 | | | | | 536 | | | | | | 639 | |
| 18 | | | | | | 473 | | | | | | 334 |
| 19 | 747 | | | 756 | | | | | | | | |
| 20 | | 666 | | | 557 | | | | | | | |
| 21 | | | 336 | | | 243 | | | | | | |
| 22 | | | | 666 | | | 667 | | | | | |
| 23 | | | | | 444 | | | 388 | | | | |
| 24 | | | | | | 214 | | | 323 | | | |
| 25 | | | | | | | 454 | | | 545 | | |
| 26 | | | | | | | | 867 | | | 756 | |
| 27 | | | | | | | | | 345 | | | 253 |
| 28 | 343 | | | | | | | | | 243 | | |
| 29 | | 444 | | | | | | | | | 323 | |
| 30 | | | 244 | | | | | | | | | 555 |
| 31 | rating performed by any available judges | | | | | | | | | | | |
| 32 | rating performed by any available judges | | | | | | | | | | | |

Figure 2. Rotating test book judging plan.

| Judge<br>Essay | 1<br>ABC | 2<br>ABC | 3<br>ABC | 4<br>ABC | 5<br>ABC | 6<br>ABC | 7<br>ABC | 8<br>ABC | 9<br>ABC | 10<br>ABC | 11<br>ABC | 12<br>ABC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Person 1 | ... | ... | ... | ... | ... | ... | .6. | ..7 | 5.. | ... | ... | ... |
| 2 | ... | ... | ... | 5.. | .3. | ... | ... | ... | ... | ..5 | ... | ... |
| 3 | 4.. | ... | ... | ... | ... | ... | ... | ... | ... | .5. | ..3 | ... |
| 4 | ... | .2. | ... | ... | ... | ..5 | ... | ... | ... | 4.. | ... | ... |
| 5 | ... | ... | ... | ... | ..7 | ... | ... | ... | ... | .4. | 5.. | ... |
| 6 | 5.. | ..6 | ... | ... | ... | ... | ... | ... | ... | ... | .6. | ... |
| 7 | ... | ... | ... | ... | .5. | ... | ..3 | ... | ... | ... | ... | 4.. |
| 8 | ... | .6. | ... | ... | ... | ... | ... | ..6 | ... | 5.. | ... | ... |
| 9 | ... | ... | ... | ... | .3. | ... | 6.. | ..4 | ... | ... | ... | ... |
| 10 | ... | ... | ... | .7. | ..5 | ... | ... | ... | 7.. | ... | ... | ... |
| 11 | ... | ... | ... | ... | ... | ... | ... | ... | ..6 | ... | 4.. | .3. |
| 12 | 4.. | ... | ... | ..6 | ... | ... | ... | ... | ... | ... | ... | .7. |
| 13 | .4. | ... | ... | ..6 | ... | ... | ... | ... | 4.. | ... | ... | ... |
| 14 | ..6 | .3. | ... | ... | ... | ... | 4.. | ... | ... | ... | ... | ... |
| 15 | .4. | ... | ... | ... | ... | ... | ... | ... | ..4 | 6.. | ... | ... |
| 16 | ... | 6.. | .4. | ... | ... | ... | ... | ... | ... | ... | ..8 | ... |
| 17 | ... | ... | .2. | 6.. | ..6 | ... | ... | ... | ... | ... | ... | ... |
| 18 | ... | ... | ... | ... | ... | 4.. | .4. | ... | ... | ... | ... | ..4 |
| 19 | ... | 7.. | ... | ..6 | ... | .4. | ... | ... | ... | ... | ... | ... |
| 20 | ... | ... | 7.. | .5. | ... | ... | ... | ... | ..6 | ... | ... | ... |
| 21 | ... | ... | ... | ... | ... | 2.. | .6. | ... | ... | ..3 | ... | ... |
| 22 | ... | ... | ..7 | ... | 8.. | ... | ... | ... | .6. | ... | ... | ... |
| 23 | ... | ..6 | 7.. | ... | ... | ... | ... | .8. | ... | ... | ... | ... |
| 24 | ... | ... | ... | ... | ... | 2.. | ..3 | ... | ... | ... | ... | .2. |
| 25 | ... | ... | ... | ... | ... | .4. | ..4 | ... | ... | ... | 4.. | ... |
| 26 | ... | ... | ... | ... | ... | ... | ... | ... | .7. | ..6 | ... | 8.. |
| 27 | ... | ... | .3. | ... | ... | ... | ... | 2.. | ... | ... | ..6 | ... |
| 28 | .4. | ..3 | ... | ... | ... | ... | ... | ... | ... | ... | ... | 2.. |
| 29 | ... | ... | 3.. | ... | ... | ... | ... | .3. | ... | ... | ... | ..3 |
| 30 | ..2 | ... | ... | ... | ... | ... | ... | 5.. | ... | ... | .3. | ... |
| 31 | ... | ... | ... | .2. | ... | ..4 | ... | 2.. | ... | ... | ... | ... |
| 32 | ... | ... | ..5 | ... | 5.. | .5. | ... | ... | ... | ... | ... | ... |

Figure 3. Minimal effort judging plan.

In the Minimal Plan, Person 3 and 6 are connected by Rater 1 and Essay A. This starts a chain of connections which connects all the elements.

```
Title = "Minimal Judging Plan"
Facets = 3
Models = ?,?,?,R9
Labels =
1, Persons
1-32
*
2, Raters
1-12
*
3, Essays
1=A
2=B
3=C
*

data=
;Person Rater   Essay   Rating
3       1       1       4
6       1       1       5
12      1       1       4
13      1       2       4
15      1       2       4
28      1       2       4
14      1       3       6
30      1       3       2
16      2       1       6
19      2       1       7
4       2       2       2
8       2       2       6
14      2       2       3
6       2       3       6
```

| | | | |
|---|---|---|---|
| 23 | 2 | 3 | 6 |
| 28 | 2 | 3 | 3 |
| 20 | 3 | 1 | 7 |
| 23 | 3 | 1 | 7 |
| 29 | 3 | 1 | 3 |
| 16 | 3 | 2 | 4 |
| 17 | 3 | 2 | 2 |
| 27 | 3 | 2 | 3 |
| 22 | 3 | 3 | 7 |
| 32 | 3 | 3 | 5 |
| 2 | 4 | 1 | 5 |
| 17 | 4 | 1 | 6 |
| 10 | 4 | 2 | 7 |
| 20 | 4 | 2 | 5 |
| 31 | 4 | 2 | 2 |
| 12 | 4 | 3 | 6 |
| 13 | 4 | 3 | 6 |
| 19 | 4 | 3 | 6 |
| 22 | 5 | 1 | 8 |
| 32 | 5 | 1 | 5 |
| 2 | 5 | 2 | 3 |
| 7 | 5 | 2 | 5 |
| 9 | 5 | 2 | 3 |
| 5 | 5 | 3 | 7 |
| 10 | 5 | 3 | 5 |
| 17 | 5 | 3 | 6 |
| 18 | 6 | 1 | 4 |
| 21 | 6 | 1 | 2 |
| 24 | 6 | 1 | 2 |
| 19 | 6 | 2 | 4 |
| 25 | 6 | 2 | 4 |
| 32 | 6 | 2 | 5 |
| 4 | 6 | 3 | 5 |
| 31 | 6 | 3 | 4 |
| 9 | 7 | 1 | 6 |
| 14 | 7 | 1 | 4 |
| 1 | 7 | 2 | 6 |
| 18 | 7 | 2 | 4 |
| 21 | 7 | 2 | 6 |
| 7 | 7 | 3 | 3 |
| 24 | 7 | 3 | 3 |
| 25 | 7 | 3 | 4 |
| 27 | 8 | 1 | 2 |
| 30 | 8 | 1 | 5 |
| 31 | 8 | 1 | 2 |
| 23 | 8 | 2 | 8 |
| 29 | 8 | 2 | 3 |
| 1 | 8 | 3 | 7 |
| 8 | 8 | 3 | 6 |
| 9 | 8 | 3 | 4 |
| 1 | 9 | 1 | 5 |
| 10 | 9 | 1 | 7 |
| 13 | 9 | 1 | 4 |
| 22 | 9 | 2 | 6 |
| 26 | 9 | 2 | 7 |
| 11 | 9 | 3 | 6 |

```
15      9       3       4
20      9       3       6
4       10      1       4
8       10      1       5
15      10      1       6
3       10      2       5
5       10      2       4
2       10      3       5
21      10      3       3
26      10      3       6
5       11      1       5
11      11      1       4
25      11      1       4
6       11      2       6
30      11      2       3
3       11      3       3
16      11      3       8
27      11      3       6
7       12      1       4
26      12      1       8
28      12      1       2
11      12      2       3
12      12      2       7
24      12      2       2
18      12      3       4
29      12      3       3
; end of minimal data
```

## 16.43 LLTM - Linear Logistic Test Model

*Recommendation:* In *Facets,* all facets are treated the same. The only difference between facets is that one facet can be specified as the inter-rater facet which reports a few more statistics. For LLTM for any facet, my recommendation is that you do a standard MFRM analysis. This estimates measures for the elements of the facet. Then decompose the element measures for that facet using linear regression in a standard statistics program.

---

LLTM is designed to deconstruct and then predict item difficulties and is based on measurement. Here is how LLTM works:
Imagine a complex arithmetic item. Solving it correctly requires numerical operations, such as addition, subtraction, multiplication, division. Perhaps several of each of these operations.
We model the difficulty of this item to be a linear combination of the difficulties of these operations.
We repeat this for all our items, producing a design matrix.
We then analyze our response data using the design matrix.
LLTM tells us the difficulty of each operation.
As we construct further items, we can use the difficulties of these operations to predict the item difficulties.
The most successful application of this approach is the Lexile system for readability. This can predict the difficulty of English prose with great accuracy.

---

The Linear Logistic Test Model, LLTM, with integer "basic parameter" weights can be modeled with *Facets* easily.

For non-integer weights, multiply all the weights for all the items by the same value until all the weights approximate integers. Then use the integer weights.

A standard LLTM model for two "basic parameters" (components of item difficulty) looks like
Bn - W1i * P1 - W2i *P2 -> Xni
where Bn is the ability of person n
P1 and P2 are the difficulties of basic parameters 1 and 2.
W1i and W2i are the pre-assigned weights of P1 and P2 in item i.

The same approach applies to any facet such as rater leniency, person ability, etc.

The standard LLTM can be estimated with Facets if W1i and W2i are integers. The Facets model is
Bn - P1 - P1 - ... - P2 - P2 - .... -> Xni
with W1i "P1" facets and W2i "P2" facets

Here is the procedure with *Facets:*

1. The "basic parameters", BP, the components of the items, are elements of a "basic parameter" facet.

2. The items are a dummy facet. The item labels include the element numbers of the BP elements.

3. There are as many BP facets in the data as the largest sum of all the integer BP weights for any item = BPL. For example, BPL= 7

4. The Facets specifications:

```
Facets = 9  ; Person+item+BPL facets in the data
Entered-in-data = 1, 2, 3, 3, 3, 3, 3, 3, 3  ; BPL=7 facets in the data for facet 3
Models = ?, ?, ?,?,?,?,?,?,?, D ; 2+BPL=9 facets
Labels=
1, Persons
1= ...
...
*
2, Items, D ; a dummy facet
1=1124000 ; the BP elements for item 1 are 2*BP1 + BP2 + BP4. 0 = no BP element.
2=1222344 ; the BP elements for item 2 are BP1+3*BP2+BP3+2*BP4
....
10=3440000
*
3, Basic Parameters (BP)
1 = basic parameter one
....
4 = basic parameter four
*
Dvalues=
3,2,1,1 ; the element number for facet 3 in the data (BP) is the first character in the label for the element in facet 2
(item)
4,2,2,1
5,2,3,1
6,2,4,1
7,2,5,1
8,2,6,1
9,2,7,1
*
; only facets 1 and 2 are needed in the data, Facets 3-9 are imputed by Dvalues=
Data=
1,1-10, 0,0,1,1,1,0,1,1,0,1 ; person 1, items 1-10, dichotomous scored responses
.....
```

5. In the output,
The BP estimates are the estimates for the "basic parameters".
The item displacement is the difference between the LLTM estimate and the item estimate from the data.
The sum of the weighted BP for an item is -"sum of the measures" in the Residualfile= for a Person with measure 0.00, or add in the Person measure to the -"sum of the measures".

Fischer, G. H., 'The linear logistic test model as an instrument in educational research', Acta Psychologica 37 (1973), 359-74.

---

**Simple LLTM example:**

LLTM is similar to a Facets-style model. It models the item difficulty as a combination of components. Each component has a difficulty and components are weighted to reflect their contribution to the item. In the simplest case, the weighting is 0 = does not contribute, 1 = does contribute.

For a Facets analysis, let's say there are 4 components. In the Facets "Labels=" specification, we set up a "component" facet with 4 elements. In the data, we have 4 component facets, all referencing the one Labels= component facet.

The Facets specification file:

```
Facets = 6  ; person, item, 4 components facets in data
Entered-in-data = 1, 2, 3, 3, 3, 3 ; all 4 components facets in the data reference the same "component" facet
Models= ?,?,?,?,?,?, D ; dichotomous right/wrong items
Labels=
1, Persons
1 = Jose
....
23 = Maria
...
*
2, items, D ; this is a dummy facet for specifying the components in each item
1 = 1300 item1 with active components 1, 3. Inactive components specified with 0.
...
20 = 2340 item 20 with active components 2,3,4
*
3, components of item difficulty
1= recognition
2= comprehension
3= computation
4= realization
*
Dvalues=
2, 1-20 ; facet 2 in the data: all persons respond to a 20 item test
3, 2, 1 ; facet 3 in the data: decode the components from the item labels
4, 2, 2
5, 2, 3
6, 2, 4
*
data=
23,  1, 0, 1, ... ; success by person 23, Maria, on item 1, failure on item 2, etc. of the 20 item test
...
```

## 16.44 Maximum size limitations on specifications and data

The *Facets* software program will operate with a minimum of 1 facet, 1 element and two observations per element. For statistically stable measures to be estimated, 30 observations per element are needed: www.rasch.org/rmt/rmt74m.htm

| Size Limitations on Specifications and Data | | | |
|---|---|---|---|
| **Component** | **Maximum number** | **Limit** | **Notes** |
| Responses in file | 2,147,438,646<br>tested to 250,000,000 | D | Data= |
| Ordinal response categories<br>in a rating scale | 255 | F | Models= or Rating scale=, per rating scale |
| Facets | 255<br>tested to 60 | F | Facets= |
| Elements | Element numbers must be in range<br>0 - 2,147,438,646<br>Maximum number of elements<br>across all facets is 4,000,000 | D | Labels=, numbered up to 2,147,483,646 |
| Element label length | 32,000 | F | Labels= |
| Element groups | 32,766 | F | Labels=, per facet |
| Model statements | 32,767 | F | Models= |
| Rating scales<br>or other ordinal response scales | 32,767 | F | Models= or Rating scale=, includes scales generated by models with "#" |
| Bias terms | 2,147,438,646 | D | Models=, per model statement |
| Table arrangements | 255 | F | Arrange= |
| Element measure range | 16 logits | P | Estimation can go higher, useful up to 44 logits range |
| Line length in specification or data file | 64,000<br>or<br>32,000 | F<br>D | |
| D = limited by available disk and memory space<br>F = Fixed upper limit<br>P = limited by arithmetic precision. | | | |

Most Facets computations are at the level of single-precision floating point arithmetic (32-bits). Some crucial computations are at the level of double-precision floating point (64-bits). See https://en.wikipedia.org/wiki/Single-precision_floating-point_format and https://en.wikipedia.org/wiki/Double-precision_floating-point_format

For a few computations which can exceed the integer-precision capabilities of double-precision floating point, "currency" computations are used (which are those used for exact computations of huge sums of money.)

## 16.45 Measures and calibrations

Various terms are used for the Rasch-model parameter estimates associated with each element.

Measures: this is the general term used to indicate the values of the estimates in a linear, interval, additive, frame of reference.

Logit values: the natural values for Rasch estimates are in "log-odds units" or logits. Logit differences have a direct relationship to probabilities for dichotomous items. For polytomous items, the probabilistic relationship is conditional.

Rescaled or user-scaled values: logit values can be rescaled linearly, i.e., by adding a constant amount and multiplying by a constant amount. The rescaled values maintain their linear quality. These are set with Umean=

Calibrations or difficulties: these are measures for items.

Abilities: these are measures for persons.

In dichotomous situations, the measure is the point on the latent variable at which there is a 50% probability of success or failure when encountering a person, item or other element known to be at that point on the variable.

In polytomous situations, the measure is the point on the latent variable at which there is an equal probability (however small) of being observed in the top or bottom polytomous category when encountering a person, item or other element known to be at that point on the variable.

The standard reference point for a dichotomy or polytomy can be altered by using structure anchoring, Rating scale=.

## 16.46 Measuring the elements

Facets commences by reading and checking the specifications. If it finds obviously faulty specifications, omissions or contradictions it will stop and display a diagnostic message on the screen.

When the specifications are successfully processed, it writes a summary of them to the output file. The less commonly used specifications are written to the file only if a non-standard option has been selected.

Facets then locates the data, which may be at the end of the specification file following the Data= line, or in a separate file (either the standard "FACETS.DAT" or one that you have named). As each datum is read from the data file, a matching Model= specification is sought. If one is found, the combined information is written to a work file. After this input phase is completed, a report is written detailing how many data points have been matched to each model statement.
Examine the Data Summary Report carefully. Unexpected counts are diagnostic of problems in the Model= specifications or data file. Precise details of thee matching process can be obtained by specifying "Query=Y".
If Pt-biserial=Y has been specified, a pass is made through the data calculating a point-biserial correlation corresponding to each element.

Initial estimates are obtained for each element using the PROX (normal approximation algorithm) method. Simultaneously, extreme scores are flagged, and the connectedness of the element structure is examined.

Facets then attempts joint unconditional maximum likelihood estimation for all unanchored element parameters of all facets. This is an iterative procedure, and progress towards convergence is shown on the screen.

The screen displays, for each iteration, the largest marginal residual score and largest logit change for any estimate encountered so far. Only amounts larger than convergence values are displayed.

The active horizontal bar-chart (>=======<) shows how far through the data Facets has progressed in that iteration. To stop iteration prematurely, press the Ctrl+"S" keys simultaneously. The bar-chart changes symbol to (>######<) to acknowledge that you have forced termination at the end of this iteration. Complete output will be produced, as though convergence has been reached.

To cancel the program immediately, press Ctrl+C keys simultaneously. Output will consist only of what Facets has already produced. This can be useful for diagnosing problems with control specifications or data file contents. Immediate cancellation can cause Facets to leave its work files behind. Files with names starting "FACETS.$$_" are not needed and can be deleted.

After convergence is achieved or termination forced, a further iteration through the data is made to calculate fit statistics and write out residual information for the responses. The element parameter estimates and summary statistics are now written to the output file.

After estimating bias interactions, if specified, Facets concludes by reporting how long it took to complete the entire analysis.

## 16.47 Missing and sparse data considerations: Sample size and Category counts

Here are remarks from the Abstract of an instructive paper::

"Previous research includes frequent admonitions regarding the importance of establishing connectivity in data collection designs prior to the application of Rasch models. ... A simulation design whose characteristics were intended to reflect practical large-scale assessment networks with sparse connections were used to consider the influence of locations on the latent variable, model–data fit, and sample size within linking sets on the stability and model–data fit of estimates. Results suggested that parameter estimates for examinee and task facets are quite robust to modifications in the size, model–data fit, and latent-variable location of the link. Parameter estimates for the rater, while still quite robust, are more sensitive to reductions in link size."

The Stabilizing Influences of Linking Set Size and Model–Data Fit in Sparse Rater-Mediated Assessment Networks. Stefanie A. Wind, Eli Jones. Educational and Psychological Measurement. 2018. Volume: 78 issue: 4, page(s): 679-707.

For codes for missing: Missing score=

Sample size and Category Counts: please see "Sample Size and Item Calibration [or Person Measure] Stability"

For statistically robust estimates (= estimates likely to be observed next time, at least approximately), we usually need at least 10 observations of each category, but this is not required, and often not achieved, even in large datasets.

If your purpose is exploratory or to report on a specific dataset, e.g., to give feedback to raters, then any number of observations in a category is good. If your purpose is to establish norms for a population, i.e., to make generalizations about a population from this dataset, then at least 10 observations in each category is recommended.

## 16.48 Mixed rating scale forms and MCQ scoring keys

*Facets* handles mixtures of rating scales and partial credit items within the same test easily using Models=. Datasets containing only MCQ items can be analyzed.

Items 1-10 are dichotomies, already scored 0,1
Items 11-14 are 3-category rating scales 1,2,3 - each with its own scale definition, i.e., partial credit items.
Items 15 is a binomial count 0-10 successes

Facet 1 is person, Facet 2 is judge (only for items 11-14), Facet 3 is item:

Models=
?, X, 1-10, D  ; items 1-10 -dichotomies - no judge
?, ?, 11-14#,R3 ; items 11-14 - each with own scale
?, X, 15, B10 ; item 15 - 10 binomial trials
*

Here are typical lines of data for person 23:
data=
23,0,1-10,0,1,1,1,1,0,0,0,1,1 ; person 23, with no judge, responds to items 1-10, scored 0 or 1.
23,3,11-14,3,2,3,1  ; person 23, judge 3, rated on items 11-14 on rating scales with highest category 3.
23,6,12-14,1,,3  ; person 23, judge 6, rates only items 12 and 14
23,0,15,7  ; person 23, no judge, 7 successes out of 10 binomial trials.

---

**To score multiple-choice MCQ dichotomies:**

The scoring key for the 10 items is: CADBDCDABB

Also shown are items 11-14 which are partial credit or rating scale items.
Item 15 is a binomial trial item giving successes out of 10.

```
Models=
?,X,1,C  ; using scoring key C as the model code
?,X,2,A
?,X,3,D
?,X,4,B
?,X,5,D
?,X,6,C
?,X,7,D
?,X,8,A
?,X,9,B
?,X,10,B
?, ?, 11-14#,R3 ; items 11-14 - each with own scale
?, X, 15, B10 ; item 15 - 10 binomial trials
*
Rating (or partial credit) scale=A,D ; Scale A - dichotomous
0=Wrong,,,B+C+D+ ; add other possible wrong responses here
1=Right,,,A  ; ",,," is to skip over anchor value positions
*
Rating (or partial credit) scale=B,D
0=Wrong,,,A+C+D+.    ; add other possible wrong responses here
1=Right,,,B
*
Rating (or partial credit) scale=C,D
0=Wrong,,,A+B+D+.    ; add other possible wrong responses here
1=Right,,,C
*
Rating (or partial credit) scale=D,D
0=Wrong,,,A+B+C+.    ; add other possible wrong responses here
1=Right,,,D
*
; Here are typical lines of data for person 23:
data=
23,0,1-10,D,B,C,A,.,D,A,B,D,C ; person 23, with no judge, responds to items 1-10
23,3,11-14,3,2,3,1  ; person 23, judge 3, rated on items 11-14
23,6,12-14,1,,3   ; person 23, judge 6, rates only items 12 and 14
23,0,15,7   ; person 23, no judge, 7 successes out of 15.
```

# 16.49 Model - matching data with measurement models

If your models don't seem to match correctly, use **Query=Y**.Facets allows more than one model statement to be specified for an analysis after a Model= specification. The sequence of the model statements is important. Facets attempts to match each line of data to each model statement in turn, beginning with the first statement. Each datum is analyzed only once in accordance with first model it actually matches. The first "match" selects that datum into the analysis according to that model and that model only.

If a datum does **not match** any model, then it is ignored.

More information: details of Models= and Model statements.

Example 1: Multiple models used in one analysis. Here, each item has a different model, and only items 1,2,3,4,5,6 and 23 are explicitly listed. Other items will be ignored.

```
Model=
?,23,?,M ; All responses on Item 23 in Facet 2 are specified to be "missing"
?,1,?,D,2 ; Item 1 is a dichotomy, to be weighted 2
?,2,?,R ; Item 2 has a rating scale (or partial credit)
?,3,#,R ; Each judge on Item 3 has his own rating scale, i.e., a partial credit scale
?,4, ,B2 ; Item 4 is two binomial trials, with no judge
?B,5,?B,R15 ; Item 5 has, at most, a 0-15, 16 category, rating scale (or partial credit)
; a Person-Judge Bias analysis is requested for all data
?,6B,?B,D ; Item 6 is a dichotomy
; an Item-Judge Bias analysis is requested for all data
*
```

Here is a sample line from the data file:

```
2,4,1,0 ; Person 2 rated on item 4 by judge 1 with 0.
```

**Which model does it match?**

```
Model=
?,23,?,M ; no match:
        the matching process compares
        model elements = ?,23, ?
        data elements = 2, 4, 1
        Facets 1 and 3 of the model are "?" (=every element), and so match the corresponding data elements,
        but Facet 2 of the model is element 23, which does not match element 4 in the data, so there is no match.
?,1,?,D,2 ;no match: facet 2, "1" in model, "4" in data
?,2,?,R ;no match: facet 2, "2" in model, "4" in data
?,3,#,R ;no match: facet 2, "3" in model, "4" in data
?,4, ,B2 ;matches, because
        facet 1, "?" (any person) in model, "2" in data
        facet 2, "4" in model, "4" in data
        facet 2, " " (ignore judge) in model, "1" in data. "1" is checked against the Labels= list.
        so this datum is modeled as two binomial trials ("B2")

?B,5,?B,R15 ; datum has already matched and been used. No further matching is done.
?,6B,?B,D
*
```

Example 2: The next data line, with the same model statements as Example 1:

```
1,23,1,4 ; person 1 rated on item 23 by judge 1 with 4.
```

Which model does it match?

```
Model=
```

?,23,?,M ;matches, so this data is treated as "M", missing.

?,1,?,D,2 ; datum already matched and used. No further matching is done
*

Example 3: The next data line:
102,3,6,2 ; person 102 rated on item 3 by judge 6 with 2

Which model does it match?

Model=
?,23,?,M ;no match: facet 2
?,1,?,D,2 ;no match: facet 2
?,2,?,R ;no match: facet 2
?,3,#,R ;matches
facet 1, "?" (any person) in model, "102" in data
facet 2, "3" in model, "3" in data
facet 2, "#" (any judge) in model, "6" in data
so this datum is modeled on a rating scale (or partial credit) ("R") as used by the particular judge ("#"), which in this case is judge 6, "6" in the data. This rating sale will have Rasch-Andrich thresholds (step calibrations) based solely on judge 6's ratings.

?,4, ,B2 ; not used
?B,5,?B,R15
?,6B,?B,D
*

Example 4: The next data line:
79,3,3,1 ; person 79 rated on item 3 by judge 3 with 1

Which model does it match?

Model=
?,23,?,M ;no match: facet 2
?,1,?,D,2 ;no match: facet 2
?,2,?,R ;no match: facet 2
?,3,#,R ;matches
facet 1, "?" (any person) in model, "79" in data
facet 2, "3" in model, "3" in data
facet 2, "#" (any judge) in model, "3" in data
so this datum is modeled on a rating scale (or partial credit) ("R") as used by the particular judge ("#"), which in this case is judge 3, "3" in the data. This rating scale (or partial credit) will be similar to that for judge 6 in Example 3 above, but will be calibrated on judge 3's ratings alone.

?,4, ,B2 ; not used
?B,5,?B,R15
?,6B,?B,D
*

Example 5: The next data line:
7,5,6,1 ; Person 7 is rated on item 5 by judge 6 with 1

Which model does it match?

Model=
?,23,?,M ;no match: facet 2

?,1,?,D,2 ;no match: facet 2
?,2,?,R ;no match: facet 2
?,3,#,R ;no match: facet 2
?,4, ,B2 ;no match: facet 2
?B,5,?B,R15 ;matches
facet 1, "?B" (any person) in model, "7" in data
facet 2, "5" in model, "5" in data
facet 2, "?B" (any judge) in model, "6" in data
so this datum is modeled on a rating scale (or partial credit) ("R") with a valid range of 0 to 15.
The "B" characters in the model are ignored for matching purposes.

?,6B,?B,D ;ignored
*

Example 6: The next data line:
67,8,32,5 ; Person 67 rated on item 8 by judge 32 with 5

Which model does it match?

Model=
?,23,?,M ;no match: facet 2
?,1,?,D,2 ;no match: facet 2
?,2,?,R ;no match: facet 2
?,3,#,R ;no match: facet 2
?,4, ,B2 ;no match: facet 2
?B,5,?B,R15 ;no match: facet 2
?,6B,?B,D ;no match: facet 2
*
There is no match, so this datum is ignored for this analysis.

Example 7: The next data line:
8,5,4,12 ; Person 8 is rated on item 5 by judge 4 with 12

Which model does it match?

Model=
?,23,?,M ;no match: facet 2
?,1,?,D,2 ;no match: facet 2
?,2,?,R ;no match: facet 2
?,3,#,R ;no match: facet 2
?,4, ,B2 ;no match: facet 2
?B,5,?B,R15 ;matches
facet 1, "?B" (any person) in model, "8" in data
facet 2, "5" in model, "5" in data
facet 3, "?B" (any judge) in model, "5" in data
The "B" characters in the model are ignored for matching purposes.
?,6B,?B,D ;ignored
*

## 16.50 Model statement examples

Model statements are best understood through examples. Here are a number of model statements which could be used in an analysis where
Persons comprise facet 1,
Items comprise facet 2,
Judges comprise facet 3.

More information: details of Models= and Matching data with measurement models.

Model= 23,?,?,M
When person 23 ("23" for facet 1) is rated on any item ("?" for facet 2) by any judge ("?" for facet 3), treat the datum as missing ("M"). This has the effect of deleting person 23.

Model= ?,1,?,D
For any person ("?") rated on item 1 ("1") by any judge ("?"), treat the "0" and "1" data as dichotomous ("D"), i.e.,
log() = Bn - D1 - Cj for item i=1

Model= ?,2,?,D3
For any person ("?") rated on item 2 ("2") by any judge ("?") dichotomize the data ("D3"), treating 0,1,2 as 0, and 3 and above as 1, i.e.,
log() = Bn - D2 - Cj with data recoding for item i=2

Model= ?,2,?,R
For any person ("?") rated on item 2 ("2) by any judge ("?") use a common rating scale (or partial credit) ("R"). Valid ratings are in the range 0 through 9, i.e.,
log() = Bn - D2 - Cj - Fk for i=2, k=1,9

Model= ?,2,?,R2
For any person ("?") rated on item 2 ("2) by any judge ("?") use a common rating scale (or partial credit) ("R"). Valid ratings are in the range 0 through 2, i.e.,
log() = Bn - D2 - Cj - Fk for i=2, k=1,2

Model= ?,3,#,R
Let each judge ("#") apply his own version of the rating scale ("R"), i.e., a partial credit scale, to every person ("?") on item 3 ("3"), i.e.,
log() = Bn - D3 - Cj - Fjk for i=3, k=1,9

Model= ?, ,?,B2
For each person ("?"), ignore the item number (", ,") and let every rating by each judge ("?") be considered two binomial trials ("B2") scored 0 or 1 or 2, i.e.,
log() = Bn - Di - log(k/(3-k)) for k=1,2

Model= ?,?,0,P
For each person ("?") observed on each item ("?") which is not judged ("0"), the data are Poisson counts of successes. These are in the theoretical range of 0 to infinity, but in the empirically observed range of 0 to 255, i.e.,
log() = Bn - Di - log(k) for k=1,...

Model= ?,?,?,R,2
For any person ("?") rated on any ("?") by any judge ("?") use a common rating scale ("R"), but give each datum a double weight in the estimation, i.e.,
log() = Bn - Di - Cj - Fk for k=1,9

Model= ?B,?,?B,D
For any person ("?B") rated on any item ("?") by any judge ("?B"), the data are on a dichotomous scale ("D"), i.e.,
log() = Bn - Di - Cj

Then, after that estimation has been completed and all measures and rating-scale structures have been anchored, estimate bias measures for the bias interactions between each person ("?B") and each judge ("?B") across the whole data set for all models specified, i.e.,

log() = {Bn,Di,Cj,Fk,...} + Cnj
where {...} are the final estimates of the previous stage used as anchors and only the Cnj bias terms are now estimated. Cnj terms are appended to all model statements. The modeled expectation of Cnj is zero, but the mean of all estimated Cnj will not be zero due to the non-linear conversions between accumulated raw score residuals and bias measures in logits. Each bias term is a diagnostic specialization which turns a systematic misfit into a measure.

Model= ?,-?,?,D
For any person ("?") rated on any item ("-?") by any judge ("?"), the outcome is a dichotomy ("D"). The orientation of the second, item facet is reversed ("-") for data matching this model only, i.e.,
log() = Bn - (-Di) - Cj = Bn + Di - Cj

Model= ?, ,?,R
For any person ("?"), irrespective of the item (" "), rated by any judge ("?"), the outcome is a rating ("R"). The item facet is ignored, except that, if the item element number for a matching datum is not specified after Labels=, the datum is treated as missing.

Model= ?,X,?,R
For any person ("?"), irrespective of the item ("X"), rated by any judge ("?"), the outcome is a rating ("R"). The item facet is entirely ignored, so that, even if the item element number for a matching datum is not specified after Labels=, the datum is still treated as valid. If a facet is never referenced anywhere, then it may be more convenient to use Entry= rather than "X".

## 16.51 Multiple choice test

Facets is not designed for the analysis of multiple-choice (MCQ) tests, but can do it. Datasets containing both rated and MCQ items can be analyzed.

```
Facets=2 ; 1= persons, 2= items
labels=
1, persons
1-1000   ; 1000 persons
*
2, items
1-40   ; 40 items
*
; assume the responses are ABCD
; and the scoring key BCDBA ..... BA

model=
?, 1, KB ; item 1. Correct answer B.
         ; KB is the "rating scale" for an MCQ item with "B" as the correct answer
?, 2, KC
....
?,39, KB
?,40, KA
*

rating scale= KA, D, Specific
            ; KA is the "rating scale" for an MCQ item with "A" as the correct answer
0 = Wrong,,,B+C+D
1 = Right,,,A
*

rating scale= KB, D, Specific
0 = Wrong,,,A+C+D
1 = Right,,,B
*

rating scale= KC, D, Specific
```

```
0 = Wrong,,,A+B+D
1 = Right,,,C
*

rating scale= KD, D, Specific
0 = Wrong,,,A+B+C
1 = Right,,,D
*
data=
1, 1-40, A, C, D, ........ C, D ; 40 responses by person 1
```

## 16.52 Nested facet designs: remedying ambiguities

Here is a typical nested design. There are 7 prompts, but candidates only respond to one. From the measurement perspective, we cannot tell whether Mary scored higher than George was because Mary is the more able or because she answered the easier prompt.

There are several reasonable measurements solutions to this problem.

**(A) Random Equivalence of Nested Subsets.**

If the candidates are assigned to each prompt randomly, then the candidate-prompt subsets can be thought of as randomly equivalent. This implies that we can assert that the subsets have the same mean. This is implemented with group-anchoring.

---

A little word-processing of Table 7 can turn the "disjoint subset number" into the "group number":

```
Table 7.3.1  rater Measurement Report  (arranged by MN).

-------------------------------------------------------------------------------
| Obsvd   Obsvd  Obsvd  Fair-M|        Model | Infit      Outfit   |           |
| Score   Count Average Avrage|Measure  S.E. |MnSq ZStd  MnSq ZStd | Num rater |
-------------------------------------------------------------------------------
|    124     18     6.9   6.64|  -2.49   .38 |  .4  -1     .4  -1  |  74 Mary  | in subset: 12
|    123     18     6.8   6.58|  -2.35   .31 |  .4  -1     .4  -1  |  83 Fred  | in subset: 14
|    124     18     6.9   6.40|  -2.05   .39 |  .4  -1     .4  -1  | 430 Jose  | in subset: 29
|    121     18     6.7   6.31|  -1.93   .27 |  .4  -2     .4  -2  | 180 Chin  | in subset: 27
```

becomes

```
Labels=
.....
3,Raters,G
 74, Mary, 0, 12
 83, Fred, 0, 14
430, Jose, 0, 29
180, Chin, 0, 27
```

This can be done with the Output files menu, Subset file.

---

**Example of a specification file for group-anchoring:**

```
Facets=3  ; Prompt, Candidate, Rater
Positive=2  ; Candidates are oriented: larger score=>greater message
Noncenter=1  ; Prompts float
Labels=
1,Prompts
1=Prompt 1
```

```
2=Prompt 2
...
7=Prompt 7
*
2,Candidates,G  ; group-anchoring is requested: label groups by their prompt number.
5433=George,0,3 ; George, candidate 5433, is anchored at 0, as part of group 3
7629=Mary,0,2   ; Mary, candidate 7629, is anchored at 0, as part of group 2
3245=Anne,0,4   ; Anne responded to prompt 4
6751=Jose,0,2   ; Jose responded to prompt 2
.....
4765=Zoe,0,5   ; Zoe responded to prompt 7
*
3,Rater    ; The raters rated all prompts. More than one rater per candidate response.
1,Dr. Jones
2,Ms. Smith
3,Sra. Lopez
....
*
Data=
2,3,2,4    Prompt 2, Mary: Ms. Smith gives rating of 4.
......
```

Facet 2, Candidates, is specified with ",G". This implements group-anchoring. The members of each group are allowed to float relative to each other, but their mean measure is constrained to be the mean of their anchor values. In this case, the group means of candidates are set to zero. The differences between the groups are forced into the prompt difficulties. High scoring subsets answered easy prompts. Low scoring groups answered more challenging prompts.

**(B) Equally difficult or known measures for the prompts.**
group-anchoring is not used, but the prompts are anchored.

```
Facets=3   ; Prompt, Candidate, Rater
Positive=2   ; Candidates are oriented: larger score=>greater message
Noncenter=2   ; Candidates float
Labels=
1,Prompts,A
1=Prompt 1,0   ; Prompts all anchored to be the same difficulty
2=Prompt 2,0
...
7=Prompt 7,0
*
2,Candidates
5433=George
7629=Mary
3245=Anne
6751=Jose
.....
4765=Zoe
*
3,Rater    ; The raters rated all prompts. More than one rater per candidate response.
1,Dr. Jones
2,Ms. Smith
3,Sra. Lopez
....
*
Data=
2,3,2,4    Prompt 2, Mary: Ms. Smith gives rating of 4.
......
```

The prompts will all have the same measure of zero. This is equivalent to ignoring the prompt facet for measurement, but each prompt is reported with fit statistics commenting on the consistency with which it is rated.

**(C) Candidates choose prompt.**
In this situation, less able candidates may choose one prompt, "My day at the zoo", but more able ones may choose another, "Einstein's theory of relativity." This requires further information beyond the data.

"Virtual equating" is a useful post-hoc technique if the candidate essays and content experts are available. Each candidate-prompt subset is analyzed separately. For each prompt, a set of consistently-rated pieces of work is chosen, located about half-logit apart along the variable. These form substantive "prompt rulers". The experts then slide these rulers relative to each other to align pieces of work representing the same latent ability.
This provides the anchor values for the prompts.

**(D) Equivalent elements.**
If two elements are functionally the same, then the principle of "Virtual equating" applies. A convenient way of assigning two different elements the same element number is the "Target element number" in Labels=.

**(D) Equal measures for equal candidate scores.**
A specific candidate raw score is chosen as the banchmark value. All candidates with this score are anchored at the same measure, All other measures (raters, items, other candidates, etc.) align with this measure.

## 16.53 Omitting elements without editing your data

You want to drop elements 8 and 9 of facet 4 without editing the data. Here are two ways:

(a) comment out the item from the Labels= list

```
Facets=4
Labels=
1, Facet 1
...*
4, Facet 4
1= ...
...
; 8= Element 8  ; commented out
; 9= Element 9
....
*
```

(b) make a "missing data" model with Models=

```
Facets=4
Models=
?,?,?,8-9,M  ; observations matching items 8 or 9 treated as missing
?,?B,?B,?,R  ; put the model statement for your Facets here.
*
```

## 16.54 One observation for an element

If there is only one observation for an element, that element may be dropped from the analysis, or the analysis may not run.

Here is a way to get your analysis to run ....
1. Put all the data in twice copy-and-paste the current data after itself.

2. Weight the data 0.5
Models= ?, ?, ?, R, **0.5**

## 16.55 Paired comparisons of objects

Paired comparisons are simple in Facets. With paired-comparison data, the measurement model is:
measure of object A - measure of object B = log-odds(choice of A over B)

The person doing the measuring is not part of the measurement model. We want to include that person so that we can identify who has unusual characteristics, or where there is miscoding in the data. Since the person measure is irrelevant, it can be anything. Anchoring all persons at 0, as a dummy facet, is convenient. If any persons have displacements, then there is an error in the data.

*Example 1:* Paired comparison of objects by persons. In each pairing, one object wins.

Facet 1 is the objects to be compared. Each object is an element in the facet
Facet 2 is the persons doing the comparing. Each person is an element in the facet. This facet is a "dummy" facet. It is not used for measurement. It is used for fit analysis and interactions only.
So, here is what the Facets specification and data file look like:

Facets= 3 ; each observation has 3 elements in the data, 2 objects + 1 person
Entered= 1, 1, 2 ; the first two elements are for facet 1, the third element is for facet 2
Models= ?, -?,?,D ; the observation is "element of facet 1 - element of facet 1 + element of facet 2" produces a dichotomous 0/1 observation
Labels=
1, Objects ; the object facet
1=A
2=B
3=C
....
*
2, Persons, D ; the person facet: this is a Dummy facet. It is ignored for estimation
4=Mary
5=George
.....
*
Data=
1,2,4,1 ; Object A is compared with Object B by Mary. Object A wins
2,3,5,0 ; Object B is compared with Object C by George. Object B loses

or, better because it is more stable computationally,

Models= ?, -?,?,D, **0.5** ; weight each observation by 0.5

Data= ; every observation in data file twice
1,2,4,1 ; Object A is compared with Object B by Mary. Object A wins
2,1,4,1 ; ObjectB is compared with Object A by Mary. Object B Loses

2,3,5,0 ; Object B is compared with Object C by George. Object B loses
3,2,5,1 ; Object C is compared with Object B by George. Object C wins

*Example 2:* Paired comparison of objects by persons. In each pairing, one object wins and one object loses, or they are tied, draw, equal.

Score: 2=Win 1=Tie 0=Loss.

Facet 1 is the objects to be compared. Each object is an element in the facet
Facet 2 is the persons doing the comparing. Each person is an element in the facet. This facet is a "dummy" facet. It is not used for measurement. It is used for fit analysis and interactions only.

So, here is what the Facets specification and data file look like:

Facets= 3 ; each observation has 3 elements in the data, 2 objects + 1 person
Entered= 1, 1, 2 ; the first two elements are for facet 1, the third element is for facet 2
Models= ?, -?,?,R2, 0.5 ; the observation is "element of facet 1 - element of facet 1 + element of facet 2" produces a polymous 0/1/2 observation that is weighted 0.5 because each observation is twice in the data file.
Labels=
1, Objects ; the object facet
1=A
2=B
3=C
....
*
2, Persons, D ; the person facet: this is a Dummy facet. It is ignored for estimation
4=Mary
5=George
.....
*
Data=
; each observation twice (mirrored):
1,2,4,2 ; Object A is compared with Object B by Mary. Object A wins
2,1,4,0 ; Object B is compared with Object A by Mary. Object B loses

2,3,5,0 ; Object B is compared with Object C by George. Object B loses
3,2,5,2 ; Object C is compared with Object B by George. Object C wins

1,2,5,1 ; Object A is compared with Object B by George. Object A ties
2,1,5,1 ; Object B is compared with Object A by Mary. Object A ties

*Example 3.* Baseball.
*Example 4.* Flavor Strength of Gels

**Bayesian imputation for unrealistically huge logit ranges or inestimable elements**

A frequently-encountered problem in the analysis of paired-comparison data is an almost Guttman ordering of the pairings. This can lead to unrealistically huge logit ranges for the estimates of the elements or inestimable elements.

To solve this problem, we apply a little Bayesian logic. We know that the range of paired performances is not exceedingly wide, and we can also easily imagine a performance better than any of those being paired, and also a performance worse than any of the those being paired. Let's hypothesize that a reasonable logit distance between those two hypothetical performances is, say, 20 logits.

https://www.rasch.org/rmt/rmt151w.htm is a parallel situation for sports teams.

1. Hypothesize a "best"performance against which every other performance is worse. Anchor it at 10 logits.
2. Hypothesize a "worst" performance against which every other performance is better. Anchor it at -10 logits
3. Hypothesize a dummy judge who compares the best and worst performances against all the other performances.
4. Include these dummy observations in the analysis.

5. Analyze the actual observations + the dummy observations. The analysis should make sense, and the logit range of the performances will be about 20 logits. For reporting, we don't want the dummy material, so we write an Anchorfile= from this analysis.

6. We then use the Anchorfile as the Facets specification file, commenting out the "best" and "worst" performance elements and the dummy judge. We analyze only the actual observations. All the elements are anchored at their

estimates from the actual+dummy analysis. In this anchored analysis, the "displacements" indicate the impact of the dummy data on the estimates.

7. If you perceive that the logit range of 20 logits is too big or too small, please adjust the "best" and "worst" anchored values.

---

**Paired-Comparison Bias/Interaction Analysis:**

Facets produces meaningful numbers in the Bias/interaction analysis when:
1) Use mirrored data, but set the weight = 1.0, instead of 0.5. For the main analysis, use weight 0.5.
2) Arrange the data so that the Models= is ..., -?,?,... instead of ...,?,-?,...

## 16.56 Pairwise Maximum Likelihood Estimation PMLE with Facets

Dichotomous pairwise maximum likelihood estimation PMLE can be implemented directly in Facets using the paired comparison technique. The estimate match those using the pairwise maximum likelihood estimation method in Bruce Choppin (1983, 1985)."A Fully Conditional Estimation Procedure for Rasch Model Parameters."

Procedure:
1. from a dataset, such as the Knox Cube Test, use a utility program to loop through all the items, output two data line for each pair of observations where there is a '1" for one item and a "0" for another item. The data looks like this
first item, second item, 1
second item, first item, 0

2. Then construct a Facets specification and data file looking like this:

```
title = "facets pmle kct.txt"
facets=2
models=?,-?,D
entered=1,1
centered=1
negative=1
labels=
1, items
1-14
*
2,dummy
*

data=
1          6          1
6          1          0
1          7          1
7          1          0
.......
```

## 16.57 Polytomous Misfit Statistics

| Response String<br>Easy.....Hard | INFIT<br>MnSq | OUTFIT<br>MnSq | Point-Measure<br>Correlation RPM | 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 |
| 3333331110010200001 | **1.03** | **1.00** | 0.81 | in meaning |
| | | | | |
| II. overfitting (muted): | | | | |
| 33222222221111111100 | 0.18 | 0.22 | 0.92 | Guttman pattern |
| 33333222221111100000 | 0.31 | 0.35 | 0.97 | high discrimination |
| 32222222221111111110 | 0.21 | 0.26 | 0.89 | low discrimination |
| 32323232121212101010 | 0.52 | 0.54 | 0.82 | tight progression |
| | | | | |
| III. limited categories: | | | | |
| 33333333332222222222 | 0.24 | 0.24 | 0.87 | high (low) categories |
| 22222222221111111111 | 0.24 | 0.34 | 0.87 | central categories |
| 33333222222222211111 | 0.16 | 0.2 | 0.93 | only 3 categories |
| | | | | |
| IV. informative-noisy: | | | | |
| 32222222201111111130 | 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: | | | | |
| 22222222222222222222 | 0.85 | 1.21 | **0.00** | 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 | **0.00** | 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

In this Table, INFIT MnSq and OUTFIT MnSq usually point to the same diagnosis. If so, report only the OUTFIT MnSq, because that is a conventional chi-squared statistic divided by its degrees of freedom. INFIT uses less familiar information-weighting.

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.

* "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".)

## 16.58 Prettifying (beautifying) Facets output

Facets output can look plain, even ugly. Suggestions how to make Facets output look more attractive are in ASCII=, T8NBC=, Table 6.0 and Table 8. There are detailed procedures in Appendix 1 of Facets Tutorial 4.

## 16.59 R Statistics

Facets 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: https://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 |
| q() | (R terminates) | quit (exit from) R Statistics |

**Import R statistics data** in a .rdata or .rda file. See R Data file

**Export Facets output files as R statistics files**: .rdata or ,rda. Graph file, Residual file or Score file from the Output Files menu

## 16.60 Rasch models: implement in Facets

Facets can estimate all commonly-used Rasch unidimensional models, and many less familiar ones.

1. Rasch dichotomous model (usually equivalent to 1-PL model).

Example: True-false test. Multiple-choice test

```
Models = ?,?,?,D
Data =
1, 2, 3, 0 ; for failure
1, 2, 3, 1 ; for success
```

2. Andrich rating-scale (restricted) model
Example: survey using Likert scale

```
Models = ?,?,?,Rn  ; where n is highest category of the rating scale
Data =
1, 2, 3, 1 ; for strongly disagree
1, 2, 3, 2 ; for disagree
1, 2, 3, 3 ; for neutral
1, 2, 3, 4 ; for agree
1, 2, 3, 5 ; for disagree
```

3. Masters partial-credit (unrestricted) model
Example: Multiple-choice test with distractors scored by knowledge level.

```
Models = ?,?,#,Rn   ; where n is highest category of the partial-credit scale
Data =
1, 2, 3, 0 ; for wrong
1, 2, 3, 1 ; for partially correct
1, 2, 3, 2 ; for correct
```

4. Bernoulli binomial-trials model (with item discrimination), see https://www.rasch.org/rmt/rmt132a.htm
Example: hits on a target

```
Models = ?,?,?,Bn ; where n is the number of trials (attempts)
Data =
1, 2, 3, 0 ; no successes
1, 2, 3, 1 ; one success
1, 2, 3, 2 ; two successes
```

5. Poisson-counts model (with item discrimination), https://www.rasch.org/rmt/rmt132a.htm
Example: spelling mistakes in a long dictation test

```
Models = ?,?,?,P ; where the maximum count is conceptually infinite and
                 ; maximum observable successes = 255
Data =
1, 2, 3, 0 ; no successes
1, 2, 3, 1 ; one success
1, 2, 3, 2 ; two successes
```

6. Glas "Steps" success model, see https://www.rasch.org/rmt/rmt52j.htm
Example: completing a series of tasks

```
Models = ?,?,?,?, D
Labels =
3 = Items facet, A
6 = Item 6 element, 0   ; anchored at 0 to give item fit statistics,
                        ; for interactions, etc.
*
4 =  Decision point facet
61 = first step element for item 6
62 = second step for item 6
*

Data =
```

```
; first step of task failure
1, 2, 6, 61, 0 ;  failure on  item 6 step 1
1, 2, 6, 62, .  ;  item 6 step 2 not attempted

; first step of task success, second step failure
1, 2, 7, 71, 1 ; success on item 7 step 1
1, 2, 7, 72, 0 ; failure on  item 7 step 2

; first step of task success, second step success
1, 2, 8, 81, 1 ;  success on item 8 step 1
1, 2, 8, 82, 1 ; success on item 8 step 2
```

7. Rasch failure model, see https://www.rasch.org/rmt/rmt52j.htm
Example: performing a task within time-limits

```
Models = ?,?,?,?, D
Labels =
3 = Item facet, A
6 = Item 6 element, 0  ; anchored at 0 to give item fit statistics, for interactions, etc.
*
4 = Decision point facet
62 = second step for item 6: task performed immediately
61 = first step for item 6 element: task performed later
*
Data =
; task performed immediately = 2 points
1, 2, 6, 62, 1 ; success on item 6 step 2
1, 2, 6, 61, .  ; item 6 step 2 not attempted

; task performed later = 1 point
1, 2, 7, 72, 0 ; failure on item 7 step 2
1, 2, 7, 71, 1 ; success on item 7 step 1

; task never done = 0 points
1, 2, 8, 81, 0 ; failure on item 8 step 1
1, 2, 8, 82, 0 ; failure on item 8 step 1
```

8. Rasch fractional dichotomous model, see https://www.rasch.org/rmt/rmt233d.htm
Example: percentages: performance assessed as 70% success on an item

```
Models = ?,?,?,D
       Data =
; the data points are weighted using R, data replication.
R0.70, 1, 2, 3, 1 ; for 70% empirically-observed rate of success
                  ; = 0.7 of a dichotomous success
R0.30, 1, 2, 3, 0 ; for 30% empirically-observed rate of failure
                  ; = 0.3 of a dichotomous failure
```

## 16.61 Rater misbehavior

The fit statistics in Facets help us to detect many types of misbehavior. For instance, central tendency usually makes raters too predictable, so that their infit and outfit mean-square statistics are noticeably less than 1.0. Also, if you model each rater to have a unique rating scale, by using # instead of ? for the rater facet in the Models= specification, then you will see in Table 8 that the rater has an unusually high number of ratings in the central categories.

Is there too much rater bias identified? Are there are some persons with idiosyncratic profiles who should be eliminated before the rater bias analysis is taken too seriously?

You also need to identify how big the bias has to be before it makes a substantive difference. Perhaps "Obs-Exp Average Difference" needs to be at least 0.5 score-points.

Then you have to decide what type of rater agreement you want.

Do you want the raters to agree exactly with each other on the ratings awarded? The "rater agreement %".

Do you want the raters to agree about which performances are better and which are worse? Correlations

Do you want the raters to have the same leniency/severity? "1 - Separation Reliability" or "Fixed Chi-squared"

Do you want the raters to behave like independent experts? Rasch fit statistics

Numerous types of rater misbehavior are identified in the literature. Here are some approaches to identifying them. *Please notify us if you discover useful ways to identify misbehavior.*

**A suggested procedure:**
(a) Model all raters to share a common understanding of the rating scale:
Models = ?,?,?,R9 ; the model for your facets and rating scale
Interrater= 2 ; 2 or whatever is the number of your rater facet
In the rater facet report (Table 7):
How much difference in rater severity/leniency is reported? Are there outliers?
Are rater fit statistics homogeneous?
Does inter-rater agreement indicate "scoring machines" or "independent experts"?
In the rating scale report (Table 8):
Is overall usage of the categories as expected?

(b) Model each rater to have a personal understanding of the rating scale:
Models = ?,#,?,R9 ; # marks the rater facet
Interrater = 2 ; 2 or whatever is the number of your rater facet
In the rating scale report (Table 8):
For each rater: is overall usage of the categories as expected?
Are their specific problems, e.g., high or low frequency categories. unobserved categories, average category measures disordered?

(c) Look for rater-item interactions, and rater-demographic interactions:
Models =
?,?B,?,?B,R9 ; Facet 4 is a demographic facet (e.g., gender, sex): rater-gender interaction (bias)
?,?B,?B,?,R9 ; Facet 3 is the items: rater-item interaction (bias)
*
In the bias/interaction report (Table 14):
Are any raters showing large and statistically significant interactions?

**Known rater misbehaviors:**
1. Leniency/Severity/Generosity.
This is usually parameterized directly in the "Rater" facet, and measures are automatically adjusted for it.

2. Extremism/Central Tendency.
Tending to award ratings in the extreme, or in the central, categories.
This can be identified by modeling each rater to have a separate rating scale (or partial credit). Those with very low central probabilities exhibit extremism. Those with very high central probabilities exhibit central tendency.

3. Halo/"Carry Over" Effects.
One attribute biases ratings with respect to other attributes. This requires that we know the order in which ratings are assigned for each person. If we know this, then we can measure all the persons and raters using only the rating of the first item rated. Then anchor everything, including anchoring the other items at the difficulty of the first item. In this anchored analysis of all the data, the raters with the lowest mean-squares are the ones most likely to have a halo effect.

4. Response Sets.
The ratings are not related to the ability of the subjects.
Anchor all persons at the same ability, usually 0. Raters who best fit this situation are most likely to be exhibiting response sets.

5. Playing it safe.
The rater attempts to give a rating near the other raters, rather than independently.
Specify the Inter-rater= facet and monitor the "agreement" percentage. The rater also tends to overfit.

6. Instability.
Rater leniency changes from situation to situation.
Include the "situation" as a dummy facet (e.g., rating session), and investigate rater-situation interactions using "B" in the Models= statements.

## 16.62 Ratio scale of logit measures

*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, e.g., 999=Reference Person. Include in the data ratings for this dummy person against real items, raters, tasks, etc.

Similarly a dummy rater at the raters's reference point, 999=Reference rater. Include in the data ratings for this dummy rater against real persons, tasks, etc.

Downweight the dummy ratings in the data using "R0.0001".
e.g.,
```
Facets=4
Data=
R0.0001, 999, 19,1,1,1   ; reference person (give the reference batter at least 20 ratings against different raters, etc.)
R0.0001, 1, 99,1,1,1   ; reference rater (give the reference pitcher at least 20 ratings against different persons, etc.)
```

You can now use the person or rater at the reference point as the basis for your "twice as much" statements.

You can take this a step further for convenience. Anchor the Dummy person and Dummy rater both at 0 logits.
Then make sure that an unanchored facet is non-centered
```
Noncenter=3 ; facet 3, items
```
or include a dummy facet that can be non-centered:
```
Facets=5
Noncenter=5 ; dummy facet for non-centering
Labels=
1,...
....
*
5, Dummy facet
1=Dummy element
*
Dvalues=
5,1   ; Facet 5 has element number 1 for every observation
*
```

## 16.63 Recoding non-numeric and other data values

Each observation is expected to be a numerical rating or a count of events, but non-numeric codes can be used, if they are recoded in a Rating (or partial credit) scale= specification. Numeric values can also be recoded to other values.

Blank observations in the data file are ignored. Leading and trailing blanks are removed from recorded observations before they are compared with recoded values. The data value of "+" is invalid.

Example 1:

Person 23 made 8 spelling mistakes in Essay 5. No recoding is required.
data=
23,5,8 ; 8 is the count of errors

Example 2:

The correct answer to item 3 is "B", but "A", "C", "D", and "*" are incorrect

Model=?,3,Item3 ; Item 3 has a unique model and rating scale (or partial credit)
Rating (or partial credit) scale=Item3,D ; define a dichotomous item
0,Wrong,,,A+C+D+* ; recode all "wrong" codes to 0
; ",,," indicates logit value and anchor status are omitted
1,Right,,,B ; recode correct, "B" to 1
*
data=
5,3,A ; Person 5 on Item 3 responded "A"
6,3,D ; Person 6 on Item 3 responded "D"

Example 3: A rating scale originally coded 1 through 6 is to be recoded 0,1,2

Model=?,?,Scale ; "Scale" is to be redefined
Rating (or partial credit) scale=Scale,R2 ;define scale as 0,1,2
0,Low,,,1+2 ; recode 1 and 2 to 0. ",,," omits logit value and anchor status
1,Medium,,,3+4+5 ; recode 3,4,5 to 1
2,High,,,6 ; recode 6 to 2
*
data=
13,5-8,4,5,2,6 ; Person 13's responses to items 5 through 8 were 4,5,2,6,
; recoded 1,1,0,2

## 16.64 Rectangular copying

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, 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
or use the free text editor TextPad

## 16.65 Reliability - separation - strata

**(Separation) Reliability and Strata**
These are reporting "reliably different". These are the opposite of inter-rater reliability statistics that are intended to report "reliably the same."

The reported "Separation" Reliability is the Rasch equivalent of the KR-20 or Cronbach Alpha "test reliability" statistic, i.e., the ratio of "True variance" to "Observed variance" for the elements of the facet. This shows how reproducible is the ordering of the measures. This may or may not indicate how "good" the test is in other respects. High (near 1.0) person and item reliabilities are preferred. This "separation" reliability is somewhat the opposite of an interrater reliability, so low (near 0.0) judge and rater separation reliabilities are preferred.

Since the "true" variance of a sample can never be known, but only approximated, the "true" reliability can also only be approximated. All reported reliabilities, such as KR-20, Cronbach Alpha, and the Separation Reliability etc. are only approximations. These approximations are all attempts to compute:

**"Separation" Reliability = True Variance / Observed Variance**

Facets computes upper and lower boundary values for the region in which the true reliability lies. When SE=Model, the upper boundary, the "Model" reliability, is computed on the basis that all unexpectedness in the data is Rasch-predicted randomness.

When SE=Real, The lower boundary, the "Real" reliability is computed on the basis that all unexpectedness in the data contradicts the Rasch model. The unknowable True reliability generally lies somewhere between these two. As contradictory sources of noise are remove from the data, the reported Model and Real reliabilities become closer, and the True Reliability approaches the Model Reliability.

The "model" reliability is based on the model standard errors, which are computed on the basis that all superfluous unexpectedness in the data is the randomness predicted by the Rasch model.
The "real" reliability is based on the hypothesis that superfluous randomness in the data contradicts the Rasch model:

Real S.E. = Model S.E. * sqrt(Max(INFIT MnSq, 1))

Conventionally, only a Person Reliability is reported and called the "test reliability". *Facets* reports separation reliabilities for all facets. Separation reliability is estimated based on the premise that the elements are locally independent. Specifically that raters are acting as "independent experts", not as "scoring machines". But when the raters act as "scoring machines", then Facets overestimates reliability. It would be the same as running MCQ bubble sheets twice through an optical scanner, so doubling the amount of "items" per person, and then claiming

that we had increased test reliability! To assist in identifying this situation, Facets reports to what extent the raters are acting as "independent experts", as aspect of inter-rater reliability, see Table 7 Agreement Statistics.

**Separation = True S.D. / Average measurement error**
This estimates the number of statistically distinguishable levels of performance in a normally distributed sample with the same "true S.D." as the empirical sample, when the tails of the normal distribution are modeled as due to measurement error. www.rasch.org/rmt/rmt94n.htm

**Strata = (4*Separation + 1)/3**
This estimates the number of statistically distinguishable levels of performance in a normally distributed sample with the same "true S.D." as the empirical sample, when the tails of the normal distribution are modeled as extreme "true" levels of performance. www.rasch.org/rmt/rmt163f.htm

So, is sample separation is 2, then strata are (4*2+1)/3 = 3.
Separation = 2: The test is able to statistically distinguish between high and low performers.
Strata = 3: The test is able to statistically distinguish between very high, middle and very low performers.

**Strata vs. Separation:** this depends on the nature of the measure distribution.

Statistically:
If it is hypothesized to be normal, then separation.
If it is hypothesized to be heavy-tailed, then strata.

Substantively:
If very high and very low scores are probably due to accidental circumstances, then separation.
If very high and very low scores are probably due to very high and very low abilities, then strata.

If in doubt, assume that outliers are accidental, and use separation.

*Example:* I have 3 criteria in my analysis. *Facets* reports 32 Strata.
Explanation: "Strata" is a conceptual number, based on a hypothetical normal distribution of the criteria, with the same mean and S.D. as the observed criteria. Each of the infinity of criteria in the hypothetical distribution has the same precision (S.E.) as the average S.E. of the observed criteria. The result is that there are 32 statistically different levels of difficulty in the hypothetical distribution. The large number is because the S.E. of an observed criterion is small due to the large number of observations of each criterion.

## 16.66 Scoring table construction

*Facets and Winsteps or Ministep (easier):*

1. Output Files menu: Output a Winsteps file from Facets with anchor values. Specify the "item" facet as the Winsteps item facet and any other facet as the person facet.
2. In the Winsteps file, remove PAFILE= and leave IAFILE= and SAFILE=.
3. Run the Winsteps analysis (Ministep is good for 25 or fewer items).
4. Ignore everything but produce Table 20.

```
              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 |
```

*Facets only (much more difficult):* Constructing a score-to-measure table for a many-faceted situation can be challenging. Here is a procedure:

1. Do an analysis of your data. Let's assume it has 3 facets: 1. Examinees. 2. Raters. 3. Items. We want a score-to-measure table for the items with a standard rater.

2. If "disconnected" subsets are reported, then resolve that through group-anchoring or direct anchoring. We need to have the set of definitive measures.

3. Write out an anchorfile= with everything anchored.

4. Delete the current data= reference. We will construct a new data file

5. Replace the raters with a "standard" rater in the rater facet, with measure 0 (or whatever leniency you want your standard rater to have)

```
2, Rater, A ; assuming facet 2 is the rater facet
10, standard rater, 0 ; this is to be the standard rater for the score table
*
```

6. Identify the items that are to be part of the standard test. They must have anchor values. For convenience, let's assume there are 5 items. Renumber them with element numbers 1 to 5. The previous element numbers won't be needed here. Let's say they are on a rating scale 0-4. So the possible scores range from 0 to 5*4 = 20.

```
3, Item, A ; assume facet 3 is items
1, (whatever its name is), (whatever its anchor measure value is)
....
5, (whatever its name is), (whatever its anchor measure value is)
*
```

7. delete all current examinee element labels. Replace them with artificial examinee labels. One for every possible score on the test, e.g.,

```
1, Examinee ; these aren't anchored
1000 = Examinee who scored 0
1001-1020 ; I am assuming that 20 is the maximum score on a 5 item test.
*
```

8. set up a new data file, giving the corresponding scores. There are 5 items in my example, numbered 1-5, and the possible scores are 0 to 20 assigned to examinees 1000 to 1020. Rater 10 is the standard rater. So this is the data file:

```
data=
1000, 10, 1-5, 0, 0, 0, 0, 0 ; raw score of 0
1001, 10, 1-5, 1 ,0, 0, 0, 0 ; it does not matter which item is item 1. It is only the raw score that matters.
1002, 10, 1-5, 2 ,0, 0, 0, 0
1003, 10, 1-5, 3 ,0, 0, 0, 0
1004, 10, 1-5, 4 ,0, 0, 0, 0
1005, 10, 1-5, 4, 1, 0, 0, 0
1006, 10, 1-5, 4, 2, 0, 0, 0
1007, 10, 1-5, 4, 3, 0, 0, 0
1008, 10, 1-5, 4, 4, 0, 0, 0
1009, 10, 1-5, 4, 4, 1, 0, 0
1010, 10, 1-5, 4, 4, 2, 0, 0
1011, 10, 1-5, 4, 4, 3, 0, 0
1012, 10, 1-5, 4, 4, 4, 0, 0
```

1013, 10, 1-5, 4, 4, 4, 1, 0
1014, 10, 1-5, 4, 4, 4, 2, 0
1015, 10, 1-5, 4, 4, 4, 3, 0
1016, 10, 1-5, 4, 4, 4, 4, 0
1017, 10, 1-5, 4, 4, 4, 4, 1
1018, 10, 1-5, 4, 4, 4, 4, 2
1019, 10, 1-5, 4, 4, 4, 4, 3
1020, 10, 1-5, 4, 4, 4, 4, 4 ; raw score of 20

9. Perform this analysis. Items, raters and the rating scale are anchored. Examinees are not. Examinees 1000 to 1020 will have scores 0 to 20. And so their measures correspond to 0 to 20. Ignore all other statistics.

10. The examinee scores and measures provide your score-to-measure table. Use the Output Files menu, Scorefile= to write the scores and measures to Excel for easy import wherever they are needed.

11. If you want the measure range to match the original raw score range, then
USCALE= (max possible score - minimum possible score) / (measure for max score - measure for min score)
UIMEAN = ( minimum possible score) ) - ( measure for min score * USCALE )
and, if you want output to look neat
UDECIM = 0

## 16.67 Selecting elements

*Question:* Is there a simple way to select subsets of cases from a data file for Facets analysis? Specifically, if I had a dataset in which I embedded a group ID in one of the facet labels or in the examinee ID, could I select only the cases where the group ID is a single number?
For example, suppose I created a data set like this:
1ID1,…
1ID2,…
2ID3,…
2ID4,…
Is there a simple way to tell Facets only to use the cases where the first column of the first variable is 1?

*Answer:* DVALUES= can help you.
Define a dummy facet of "Group ID"
Facets = 4 ; examinee, item, rater, Group ID
Models =
?, ?, ?, 1, R4 ; selects only group 1 for analysis
*
Labels =
1, Examinee
1= 1ID1,…
2= 1ID2,…
3 =2ID3,…
4 = 2ID4,…
….
*
2, item
…
*
3, rater
….
*
4, Group ID, A or D ; dummy facet used only for selection (can also be used for interactions, etc.)
1 = Group 1, 0

2 = Group  2, 0
3 =  Group 3, 0
4 =  Group 4, 0
*
Dvalues =
4, 1, 1    ; the element identifier for facet 4 is in facet 1,  the first character of the element label
*
Data =
examinee, item, rater, observation  ; do not specify an element number for facet 4 here
.....

## 16.68 Silent installation of Facets from the DOS prompt

1. Download FacetsPasswordInstall(version number).exe
or MinifacPasswordInstall.exe

Example: `FacetsPasswordInstall371.exe`

2. Obtain your installation password.

3. Installing in folder c:\Facets

In the folder which contains FacetsPasswordInstall(version number).exe
At a DOS prompt or in a batch file, type in:
`FacetsPasswordInstall(version number).exe /serial:InstallationPassword`

Example: `FacetsPasswordInstall371.exe /serial:abcd1234`

Press enter at the DOS prompt, or save and launch the batch file.
Facets or Minifac should install

or 3. Installing in a different folder:

**Silent Install of Facets**

1. Download FacetsPasswordInstallxxx.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:\Facets
%AppShortcutFolderName%=Facets
```
4. In your work folder, create text file: install.bat
5. The contents of install.bat are:
`FacetsPasswordInstallxxx.exe /S:setupvars.ini /serial:123456789` ← your installation password goes here
6. To install Facets silently, double-click on install.bat in the Explorer view of your work folder.
7. View c:\Facets to verify that Facets.exe has installed correctly
8. Back up these files in your work folder, then delete them: FacetsPasswordInstallxxx.exe, setupvars.ini, install.bat

**Silent Install of Minifac**

1. Download Minifac.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:\Facets
%AppShortcutFolderName%=Facets
```

4. In your work folder, create text file: install.bat
5. The contents of install.bat are:

```
MinifacInstall.exe /S:setupvars.ini
```

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

## 16.69 Specification format for Facets

Each specification is of the form:

**Specification = Value**

Specifications have particular meanings to Facets,

e.g., Title= specifies the heading line for the report tables.

Value is what you assign to the specification, e.g., Knox Cube Test

Thus, to title your Facets report Knox Cube Test, specify

Title = Knox Cube Test

**Syntax Rules:**

Refer to these if you can't get your specifications to work:

a. Each specification goes on a separate line in the specification file.

b. Only the first one or two letters of each specification line matter. Title= or Titel= or T= all mean the same thing. Just enough letters are needed to identify the specification uniquely.

c. Upper or lower case letters are treated the same in defining specifications. Title= or title= mean the same thing.

d. Whatever follows a semi-colon ";" is ignored as a comment.

e. Blank lines are ignored, also ASCII 160 (hex A0) codes.

f. Leading and trailing blanks around any value are ignored. "3,Mary,2.3,1" is the same as "  3 , Mary, 2.3,  1"

g. Equal signs, "=", and commas, "," are separators which can be used interchangeably.

h. Each line must terminate with carriage return+line feed, (the standard option for printed or DOS text files). If using a word processor to prepare your specifications or data, be sure to output the files in DOS or ASCII format.

i. Facets= instruction should precede any other instructions that contain facet numbers.

| Code | Action |
|---|---|
| = , tabs | separators |
| blanks | ignored, but can act as separators in data lines |
| ; | comments - ignored |
| ;; | comments - ignored but listed in output file. |

Values in control specifications can be separated by commas "," or equal signs "=" or tabs (ASCII 9 codes), so that these control instructions have the same meaning:

positive = 1,2,3,4 ; blanks around values are ignored.
positive = 1=2=3=4
positive , 1 , 2 , 3 , 4
positive = 1 2 3 4 ; these numbers are separated by Tab codes

## 16.70 Speeding up slow estimation

**On this run:**

The "Bigger" and "Smaller" options on the Estimation pull-down menu may be helpful.
Press Ctrl+F to force an iteration to cease, and fit computation to begin.

**On later runs:**

1) Specify the final values from previous runs as the starting values for new runs. This can be done by generating an anchor file from each run, and then use this as the specification file for the next run, after removing all ",A" anchoring flags.
See Labels= and Rating (or partial credit) scale= for further information on starting values.

2) Specify liberal convergence criteria, particularly if this is an exploratory analysis to see whether things make sense, e.g., "Convergence=1,.1"

3) Do not specify output that you don't really want, e.g., do not specify "Inter-rater=" (rater agreement) unless you need it.

4) Specify "Subset detection=No" once subset

## 16.71 Subdimension person measurement

Facets supports two methods for building a profile of person measures across subdimensions or subtests in one analysis.

1. Person-by-subdimension/subtest interactions. (Easy to implement.)
a. each item is identified with its subdimension
b. a standard person facet is specified
c. the subdimensions are specified as a facet
d. the data contain element numbers for item, person, subdimension
e. a standard analysis is performed
f. a person-by-subdimension bias/interaction analysis is performed.
g. the person profile across subdimensions is:
person measure + person-by-subdimension interaction measure

2. Person subdimension measures. (Complex to implement)
a. each item is identified with its subdimension
b. a standard person facet is specified
c. a facet is specified with an element for every person-by-subdimension pairing.
d. the person-by-subdimension elements are group-anchored at zero for each person
e. the data contain element numbers for item, person, person-by-subdimension
f. a standard analysis is performed.
g. the person profile across subdimensions is:
person measure + person-by-subdimension measure

Method 2 approximates the approach in S. Brandt (2010) "Estimating Tests Including Subtests", *Journal of Applied Measurement,* 11, 4, 352-367.

## 16.72 Subscale measures

In your Facets specification file, group each item with its subscale number.
In Labels=, include a dummy "subscale" facet, with the subscales as elements.
In Dvalues=, the subscale element number is the item element group number:

Example:
Labels=
1, Students
.....
*
2,items ; Items are facet 2
1-16 = addition,,1 ; addition items in group 1, are numbered 1 through 16.
17-32 = subtraction,,2 ; subtraction items, in group 2, are numbered 17 through 32.
33-45 = multiplication,,3 ; multiplication items in group 3, are numbered 33 through 45.

*

3, subscales, D ; dummy facet for subscale reporting - all elements anchored at 0
1, addition
2, subtraction
3, multiplication
*

Dvalues =  ; data file has student number and 45 item responses.
1, 1-45
3, 2, $GROUP
*

Then,
1. Run the Facets analysis
2. "Output Tables" menu - Table 13.
Select Table 13 and temporary output file.

3. When Table 13 appears, it will look like this:

Table 13.4.1 Bias/Interaction Report (arranged by mN).
Bias/Interaction: 1. Children, 3. subscales (higher score = higher bias measure)

```
+-------------------------------------------------------------------------------------------------
       ----------------------------------+
|Observd  Expctd  Observd  Obs-Exp| +Bias  Model                    |Infit Outfit|
       Children             subscales         |
|  Score   Score    Count  Average|  Size   S.E.     t   d.f. Prob. | MnSq  MnSq | Sq     Num
         Childr measr+ N subscale measr- |
|--------------------------------+---------------------------------+------------
       +-----------------------------------------------|
|   13       7.33    13         .44|   3.53  1.80   1.96    12 .0731 |   .0    .0 | 216572
       16691 16691     .04 3 addition    .00 |
|   13       7.33    13         .44|   3.53  1.80   1.96    12 .0731 |   .0    .0 | 254516
       54657 54657     .04 3 addition    .00 |
|   13       7.88    13         .39|   3.00  1.53   1.96    12 .0740 |   .0    .0 | 204335
       4451 4451      .23 3 addition    .00 |
```

Important columns are Observd score, +Bias Size, Children num,  measr+, subscales N

4. At the top of the output file: delete all rows down to "Observd"
5. At the bottom of the output file: delete all rows up to ---- row above Mean
6. Select all and copy
7. paste into a new Excel worksheet: it will all go into column A
8. in Excel: Select column A
   Data - Text to columns - Fixed width - Next
   (The columns we want are identified) Next
  Finish
  (the table is now in Excel columns
9.  Delete all columns except: Observed Score, +Bias Size, Children Num, measr+, N
    There are 5 columns left
10. Delete Rows 1 and 3, so only heading are Score, Size, Num, overall measr (edit this), N
11. make new column F with top row: subscale measure
12.  cell F2: =B2+D2  (this is the subscale measure)
13. copy cell F2 down the other rows of column F
14. select all the rows and sort sort by N then Num
 15. the overall measures are now in column D.
       the values for subscale 1 in column F

16. copy and paste columns E and F for subscale 2 by value into columns G and H (also copy column A if raw scores are needed)
17. copy and paste columns E and F for subscale 3 by value into columns I and J (also copy column A if raw scores are needed)
18. delete all rows for subscale 2 and 3
19. the Excel spreadsheet now has the overall measure and 3 subscale measures for each person

## 16.73 Subset connectedness

**Resolving Subsetting problems**

*Suggestion:* eliminate one facet at a time from the analysis. Replace ? by X in Models= to isolate the problem facet.

1. Be sure that your measurement model is unambiguous:
Unambiguous: John Student + Item 1 -> Data
Facets can estimate a unique ability for John and a difficulty for Item 1.

Ambiguous: John Student + Male gender + Item 1 -> Data
Facets can estimate a difficulty for Item 1, but does not know how to split "ability" between "John Student" and "Male gender". Possible actions:

a.) If male gender is only for computing bias/interactions or for obtaining summary fit statistics for the males, then make Gender a Dummy (Demographic) Facet:

```
Labels=
....
*
2, Gender, D
1 = Female gender
2= Male gender
*
```

b) If male gender is only for obtaining a sub-total of the male abilities, then
i) specify a "male" group number for all the John Student and all the other male students.
or ii) output the student measures to Excel, and sub-total there.
For the S.E. of the mean: compute the S.E. of the mean of the measures for each subtotal as though they are point estimates (no measurement error). Square this value to give the variance of the mean. Then take the S.E. of each measure in a subtotal, square it to give the error variance of each measure, then average the error variances for the measures in a subtotal. Add this average variance to the variance of the mean, then square-root to give the S.E. of the mean of sub-total measures with measurement error.

c) If a Gender effect must be estimated, then group-anchor the student elements by gender at zero.

2. The data are too thin, too sparse or other problems with the data collection.
There may be subsets because the combination of elements needed to make a fully connected design did not happen. We can make a reasonable guess about what the missing data would be. Then, impute hypothetical data.
Example: in an observational study, some combinations of conditions did not occur. The results was disconnected subsets of data.
Remedy: impute observations for all combinations of conditions, but give them very small weights:
Data=
R0.01 (elements) (hypothetical observation)

---

**Judging Plan (Experimental Designs)**

A continuing practical problem in rating performances is eliminating ambiguity introduced by deficient judging plans. These plans are called "non-linked assessment networks" by George Engelhard, Jr. (Constructing rater and task banks for performance assessments. *Journal of Outcome Measurement,* 1997, 1(1), 19-33). They have a lack of identifiability to produce a unique set of estimates.

In general, a sufficient condition for uniqueness is that the response strings for every pair of persons (A,B,) (and pair of items and pair of ..) contains at least one response where A is higher than B and at least one response where B is higher than A responding to the same pair of items (and tasks and ...), either directly or indirectly through the response strings of their pairings with other persons (items, ...). Also for Andrich polytomies, a similar direct or indirect overlap exists across the higher-lower categories of each rating scale in the response string of a person, (item, ...).

| | Judge A | | | Judge B | | |
|---|---|---|---|---|---|---|
| Items: | P | Q | R | P | Q | R |
| Person | Task X | | | | | |
| 101 | 2 | 2 | 1 | 3 | 1 | 2 |
| 102 | 3 | 2 | 2 | | | |
| 103 | 2 | 3 | 2 | 2 | 3 | 2 |
| 104 | 4 | 3 | 3 | | | |
| | Task Y | | | | | |
| 201 | | | | 2 | 3 | 3 |
| 202 | 2 | 3 | 3 | 2 | 4 | 3 |
| 203 | | | | 3 | 2 | 3 |
| 204 | 4 | 2 | 3 | 2 | 3 | 4 |

Consider the data shown in the Table. At first glance, all seems well. The three items, P, Q, R, can be in one frame of reference, because they share the same judge-person-task combinations. The two judges, A, B, can be in the same frame of reference, because they rate every second person together. Now comes the problem. The persons seem to share the same frame of reference because so many of them are rated on the same tasks. But there are two tasks. Why are the four 100-group people rated lower on Task X than the four 200-group people on Task Y? Are the 100-group people less able than the 200-group? Is Task X is harder than Task Y? These data cannot say which!

Resolving this ambiguity requires perception and decision. There is a worked example at Subset Reconnection. The first step is to notice the problem. This requires Subset=Yes, the default. If you detect it during data collection, a slight change to the judging plan can remedy the situation. For instance, some people could be asked to perform both tasks. Nevertheless, continue to be on the look out for this ambiguity during analysis.

"Complete data" such as when every judge rates every person on every item is almost always connected. Lack of connectedness is usually a result of the accidental or deliberate manner in which the data was collected, e.g., the judging plan.
Two elements are connected if there exist connections through
either i) patterns of non-extreme high ratings
and ii) patterns of non-extreme low ratings
or iii) constraints, such as anchor values.

*Facets* examines the data for connectedness using a much enhanced version of a joining algorithm described in *Weeks D.L. and Williams D.R., 1964, A note on the determination of connectedness in an N-way cross classification. Technometrics, 6/3, 319-324.*

There are exotic forms of connectedness which Facets may falsely report as disconnected. Please alert us if this happens in a practical situation.

See also: *G.H. Fischer (1981), On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. Psychometrika 46, pp. 59–77*

---

**What lack of connectedness (subsetting) implies:**

**Beware!** Lack of connectedness (= subsets) means that Facets output is ambiguous, perhaps even misleading.

(a) "Disconnection" has no effect on standard errors and fit statistics, nor the measures of elements within facets that are completely connected, such as the items of a test where everyone takes every item.

(b) Disconnection has no effect on relative measures within subsets of elements that are connected. For instance, under the usual "nested" judging plans, the item facet is centered on zero, and all items are connected, so the disconnected subsets of judges would have no effect on the item facet at all.

(c) The only effect of "disconnection" is on the relationship between measures in different subsets. Under these circumstances, **Facets reports one of the infinite number of different possible solutions, all equally likely.**

Only measures in the same subset are directly comparable. A separate set of vertical rulers is produced for each disjoint subset. These help you identify causes and remedies. See the example data analysis.

When a lack of connectivity is discovered, *Facets* reports subsets of connected elements:

```
-------------------------------------------------------------------------------
|Obsvd   Obsvd  Obsvd  Fair  |  Calib Model | Infit       Outfit    |            |
|Score   Count Average Avrge |  Logit Error | MnSq Zstd   MnSq Zstd | Nu student |
-------------------------------------------------------------------------------
|   16      10     1.6   1.5 |   0.09  0.64 |  0.8    0    0.8    0 |  1 1       | in subset: 1
|   11      10     1.1   1.0 |  -2.25  0.85 |  0.5    0    0.4   -1 |  2 2       | in subset: 1
|   16      10     1.6   1.3 |  -0.45  0.64 |  0.9    0    0.8    0 | 11 11      | in subset: 2
|    8      10     0.8   0.9 |  -3.67  0.76 |  0.8    0    0.6    0 | 12 12      | in subset: 2
```

Students 1 and 2 are connected in subset 1. Students 11 and 12 are connected in subset 2. The relationship between subsets 1 and 2 is ambiguous. This means that all logit values in subset 1 can be increased or decreased by the same amount, relative to subset 2, without altering the fit of the data to the measurement model. Student 1 is 0.09+2.25=2.34 logits more able than student 2, but student 1's relationship to student 11 is not known, and may not be 0.09+0.45=0.54 logits more able.

---

**Assuring data connectedness**

1. Before data collection:
   a. Map out the judging plan / experimental design. Look for unintended nesting of elements.
   b. Make up dummy data to match the design and use *Facets* to analyze it. You could make every observation a "1" and specify every model as "B2" (two binomial trials) so that Facets thinks that every element is estimable.

2. During data collection:
   Start data analysis simultaneously with data collection. If they had done this at the Salt Lake City Winter Olympics, they would have caught the problems when they were still solvable, and have avoided an embarrassing judging scandal.

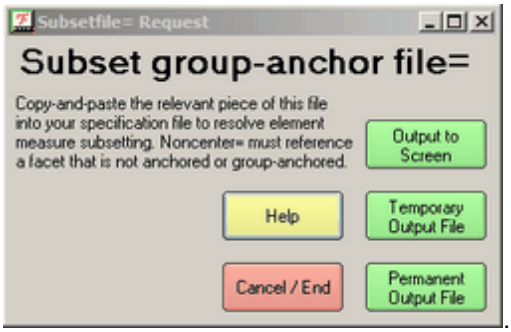
---

**Connecting final data**
Data collection may have already concluded before the first Facets analysis is made. Consequently, when Facets warns you of lack of connectedness, as in this example, there are two choices for resolving the problem. Either the

tasks are "said to be alike" or the people are "said to be alike". It is wise to try both options. The subset group-anchor file will assist here.

If Task X and Task Y were intended to have the same difficulty, then anchor them together at the same measure, usually 0. This resolves the ambiguity, and interprets the overall score difference between the 100-group and the 200-group of persons as a difference in ability levels.

On the other hand, you may have intended that the tasks be different by an amount unknown as yet, but have allocated persons to the tasks more or less at random, intending to obtain two randomly equivalent groups. Then a solution is to treat the two groups of persons as though they estimate the same mean ability. Code each person element with a 0 logit ability and a group number. Then specify group-anchoring to set the mean ability level of the 100-group at the same value as the mean ability level of the 200-group. Now the overall score difference between the 100-group and the 200-group will express a difference in difficulty between Task X and Task Y




This dialog box, called from the Output Files menu, provides pre-linked elements which can be edited in to your specification file to assist with establishing connections:

```
Copy-and-paste the relevant sections of this into your specification file to resolve subset
problems
Non-center= must reference a facet that is not anchored or group-anchored.
1,Senior scientists, G ; group-anchoring at Umean = 50
1,Avogadro,50, 1
2,Brahe,50, 1
3,Cavendish,50, 2
4,Davey,50, 2
*
2,Junior Scientists, G ; group-anchoring at Umean = 50
1,Anne,50, 1
2,Betty,50, 1
3,Chris,50, 1
4,David,50, 2
5,Edward,50, 2
6,Fred,50, 2
7,George,50, 2
*
```

---

**Connecting intermediate data**

Whenever possible, Facets should be run on available data even before data collection has concluded. Then elements identified as disconnected can be targeted for inclusion in the rating process. Thus, if it is discovered that one panel of judges has been rating the boys and another panel the girls, then some judges can be switched between panels, or some boys rated by the "girls" panel and some girls by the "boys" panel. In the example, some of these examinees, or other students like these examinees, could perform both Task X and Task Y. This would establish the relative difficulty of the tasks.

---

**Double subsetting**

Here is an example where elements are reported to be in two subsets. What has happened?

```
Table 7.1.1  Examinees Measurement Report
|- +----------------------|
|   | 1 Mary               | in subset: 1 4
|   | 2 George             | in subset: 2 3
|- +----------------------|

Table 7.2.1  Time-Point Measurement Report
|- +----------------------|
|   | 1 Before             | in subset: 1 2
|   | 2 After              | in subset: 3 4
|- +----------------------|

Table 7.3.1  Test-Type Measurement Report
|- +----------------------|
|   | 1 Paper-and-Pencil   | in subset: 1 3
|   | 2 Computer-Adaptive  | in subset: 2 4
|- +----------------------|
```

Each Examinee has been tested at two time-points: Before and After.
Each Examinee has been tested using two test formats, a different one at each time-point: Paper-and-Pencil and Computer-Adaptive.

Subset 1 are all examinees tested "Before" with "Paper-and-Pencil"
Subset 2 are all examinees tested "Before" with "Computer-Adaptive"
Subset 3 are all examinees tested "After" with "Paper-and-Pencil"
Subset 4 are all examinees tested "After" with "Computer-Adaptive"

Mary was a member of the group of examinees that were tested "Before" with "Paper-and-Pencil" then "After" with "Computer-Adaptive".
George was a member of the group of examinees that were tested "Before" with "Computer-Adaptive" then "After" with "Paper-and-Pencil".

Since the two groups of examinees (1,4 and 2,3) were probably assigned at random, we can use group-anchoring to make them randomly equivalent. Let's call the two groups 14 and 23.

Labels=
1, Examinees, G  ; group-anchor the examinees
1 = Mary, 0, 14 ; Mary's group is group-anchored at 0
2 = George, 0, 23 ; George's group is group-anchored at 0
....
*

Since, in this example, we imagine that the examinees were assigned at random to their subset, we can use group anchoring to make the connections.

**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.

## 16.74 Subset connection demonstration

Here is the "Minimal Effort Judging Plan":

| Rater<br>Essay | 1<br>ABC | 2<br>ABC | 3<br>ABC | 4<br>ABC | 5<br>ABC | 6<br>ABC | 7<br>ABC | 8<br>ABC | 9<br>ABC | 10<br>ABC | 11<br>ABC | 12<br>ABC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Person 1 | ... | ... | ... | ... | ... | ... | .6. | ..7 | 5.. | ... | ... | ... |
| 2 | ... | ... | ... | 5.. | .3. | ... | ... | ... | ... | ..5 | ... | ... |
| 3 | 4.. | ... | ... | ... | ... | ... | ... | ... | ... | .5. | ..3 | ... |
| 4 | ... | .2. | ... | ... | ... | ..5 | ... | ... | ... | 4.. | ... | ... |
| 5 | ... | ... | ... | ... | ..7 | ... | ... | ... | ... | .4. | 5.. | ... |
| 6 | 5.. | ..6 | ... | ... | ... | ... | ... | ... | ... | ... | .6. | ... |
| 7 | ... | ... | ... | ... | .5. | ... | ..3 | ... | ... | ... | ... | 4.. |
| 8 | ... | .6. | ... | ... | ... | ... | ... | ..6 | ... | 5.. | ... | ... |
| 9 | ... | ... | ... | ... | .3. | ... | 6.. | ..4 | ... | ... | ... | ... |
| 10 | ... | ... | ... | .7. | ..5 | ... | ... | ... | 7.. | ... | ... | ... |
| 11 | ... | ... | ... | ... | ... | ... | ... | ... | ..6 | ... | 4.. | .3. |
| 12 | 4.. | ... | ... | ..6 | ... | ... | ... | ... | ... | ... | ... | .7. |
| 13 | .4. | ... | ... | ..6 | ... | ... | ... | ... | 4.. | ... | ... | ... |
| 14 | ..6 | .3. | ... | ... | ... | ... | 4.. | ... | ... | ... | ... | ... |
| 15 | .4. | ... | ... | ... | ... | ... | ... | ... | ..4 | 6.. | ... | ... |
| 16 | ... | 6.. | .4. | ... | ... | ... | ... | ... | ... | ... | ..8 | ... |
| 17 | ... | ... | .2. | 6.. | ..6 | ... | ... | ... | ... | ... | ... | ... |
| 18 | ... | ... | ... | ... | ... | 4.. | .4. | ... | ... | ... | ... | ..4 |
| 19 | ... | 7.. | ... | ..6 | ... | .4. | ... | ... | ... | ... | ... | ... |
| 20 | ... | ... | 7.. | .5. | ... | ... | ... | ... | ..6 | ... | ... | ... |
| 21 | ... | ... | ... | ... | ... | 2.. | .6. | ... | ... | ..3 | ... | ... |
| 22 | ... | ... | ..7 | ... | 8.. | ... | ... | ... | .6. | ... | ... | ... |
| 23 | ... | ..6 | 7.. | ... | ... | ... | ... | .8. | ... | ... | ... | ... |
| 24 | ... | ... | ... | ... | ... | 2.. | ..3 | ... | ... | ... | ... | .2. |
| 25 | ... | ... | ... | ... | ... | .4. | ..4 | ... | ... | ... | 4.. | ... |
| 26 | ... | ... | ... | ... | ... | ... | ... | ... | .7. | ..6 | ... | 8.. |
| 27 | ... | ... | .3. | ... | ... | ... | ... | 2.. | ... | ... | ..6 | ... |
| 28 | .4. | ..3 | ... | ... | ... | ... | ... | ... | ... | ... | ... | 2.. |
| 29 | ... | ... | 3.. | ... | ... | ... | ... | .3. | ... | ... | ... | ..3 |
| 30 | ..2 | ... | ... | ... | ... | ... | ... | 5.. | ... | ... | .3. | ... |
| 31 | ... | ... | ... | .2. | ... | ..4 | ... | 2.. | ... | ... | ... | ... |
| 32 | ... | ... | ..5 | ... | 5.. | .5. | ... | ... | ... | ... | ... | ... |

Figure A.8. Minimal effort judging plan from www.rasch.org/rn3.htm

In this plan, each person writes 3 essays (A,B,C). For each person, each essay is rated by a different rater, so that each essay is rated only once, resulting in the minimum rater effort. However, all persons, raters and essays are comparable because all ratings form one connected set of data. Here is the procedure to verify the data connection:

1. The ratings are processed consecutively.
   *Example:* the ratings in the plan are stored by columns so each of the 12x3 = 36 columns of data is scanned in turn.

2. Lists of connected elements are constructed. When the elements of two facets for an observation are the same as the elements of two facets for another observation, the two observations are connected in the same subset and their elements are added to the subset's list.
   *Example:* Persons 3 and 6 are connected because they share Rater 1 and Essay A, as does Person 12. So they are added to subset list 1. Rater 1, Essay B and Persons 13, 15, 28 are added to list 2. Rater 1, Essay C and Persons 14 and 30 are added to list 3.

3. Scanning down the columns of observations in this way, 36 lists of connected elements are constructed. A list may contain the elements for only one observation.

4. Pairs of lists are compared and lists with connected elements are combined. When the same elements of two facets appear in two lists, then the lists are combined.
   *Example:* List 1 has Rater 1, Essay B, Persons 3, 6, 12. List 29 has Rater 10, Essay B, Persons 3, 5. Essay B and Person 3 are common to the two lists, so the two lists, and so the two subsets of ratings, are combined. The combined list is Raters 1, 10, Essay B, Person 3, 5, 6, 12.

5. Paired comparison of lists continues until no more lists can be connected. The final lists contain the details of the disconnected subsets. An element can be in several disconnected subsets.
   *Example:* for this plan, the 36 lists connect and are combined until there is only one list. The ratings in the data are connected.

Here is the *Facets* specification and data file:

```
Title = "Minimal effort judging plan"
Facets=3
Models=?,?,?,R9
Subsets=Yes
Labels=
1,Person
1-32
*
2, Judge
1-12
*
3, Essay
1=A
2=B
3=C
*
Dvalues=3,1-3
Data=
01,01,.,.,.
02,01,.,.,.
03,01,4,.,.
04,01,.,.,.
05,01,.,.,.
06,01,5,.,.
07,01,.,.,.
08,01,.,.,.
09,01,.,.,.
10,01,.,.,.
11,01,.,.,.
12,01,4,.,.
13,01,.,4,.
14,01,.,.,6
```

```
15,01,.,4,.
16,01,.,.,.
17,01,.,.,.
18,01,.,.,.
19,01,.,.,.
20,01,.,.,.
21,01,.,.,.
22,01,.,.,.
23,01,.,.,.
24,01,.,.,.
25,01,.,.,.
26,01,.,.,.
27,01,.,.,.
28,01,.,4,.
29,01,.,.,.
30,01,.,.,2
31,01,.,.,.
32,01,.,.,.

01,02,.,.,.
02,02,.,.,.
03,02,.,.,.
04,02,.,2,.
05,02,.,.,.
06,02,.,.,6
07,02,.,.,.
08,02,.,6,.
09,02,.,.,.
10,02,.,.,.
11,02,.,.,.
12,02,.,.,.
13,02,.,.,.
14,02,.,3,.
15,02,.,.,.
16,02,6,.,.
17,02,.,.,.
18,02,.,.,.
19,02,7,.,.
20,02,.,.,.
21,02,.,.,.
22,02,.,.,.
23,02,.,.,6
24,02,.,.,.
25,02,.,.,.
26,02,.,.,.
27,02,.,.,.
28,02,.,.,3
29,02,.,.,.
30,02,.,.,.
31,02,.,.,.
32,02,.,.,.

01,03,.,.,.
02,03,.,.,.
03,03,.,.,.
04,03,.,.,.
05,03,.,.,.
06,03,.,.,.
07,03,.,.,.
08,03,.,.,.
09,03,.,.,.
10,03,.,.,.
11,03,.,.,.
12,03,.,.,.
13,03,.,.,.
14,03,.,.,.
15,03,.,.,.
16,03,.,4,.
17,03,.,2,.
18,03,.,.,.
19,03,.,.,.
20,03,7,.,.
21,03,.,.,.
22,03,.,.,7
23,03,7,.,.
24,03,.,.,.
25,03,.,.,.
26,03,.,.,.
27,03,.,3,.
28,03,.,.,.
29,03,3,.,.
```

```
30,03,.,.,.
31,03,.,.,.
32,03,.,.,5

01,04,.,.,.
02,04,5,.,.
03,04,.,.,.
04,04,.,.,.
05,04,.,.,.
06,04,.,.,.
07,04,.,.,.
08,04,.,.,.
09,04,.,.,.
10,04,.,7,.
11,04,.,.,.
12,04,.,.,6
13,04,.,.,6
14,04,.,.,.
15,04,.,.,.
16,04,.,.,.
17,04,6,.,.
18,04,.,.,.
19,04,.,.,6
20,04,.,5,.
21,04,.,.,.
22,04,.,.,.
23,04,.,.,.
24,04,.,.,.
25,04,.,.,.
26,04,.,.,.
27,04,.,.,.
28,04,.,.,.
29,04,.,.,.
30,04,.,.,.
31,04,.,2,.
32,04,.,.,.

01,05,.,.,.
02,05,.,3,.
03,05,.,.,.
04,05,.,.,.
05,05,.,.,7
06,05,.,.,.
07,05,.,5,.
08,05,.,.,.
09,05,.,3,.
10,05,.,.,5
11,05,.,.,.
12,05,.,.,.
13,05,.,.,.
14,05,.,.,.
15,05,.,.,.
16,05,.,.,.
17,05,.,.,6
18,05,.,.,.
19,05,.,.,.
20,05,.,.,.
21,05,.,.,.
22,05,8,.,.
23,05,.,.,.
24,05,.,.,.
25,05,.,.,.
26,05,.,.,.
27,05,.,.,.
28,05,.,.,.
29,05,.,.,.
30,05,.,.,.
31,05,.,.,.
32,05,5,.,.

01,06,.,.,.
02,06,.,.,.
03,06,.,.,.
04,06,.,.,5
05,06,.,.,.
06,06,.,.,.
07,06,.,.,.
08,06,.,.,.
09,06,.,.,.
10,06,.,.,.
11,06,.,.,.
```

12,06,.,.,.
13,06,.,.,.
14,06,.,.,.
15,06,.,.,.
16,06,.,.,.
17,06,.,.,.
18,06,4,.,.
19,06,.,4,.
20,06,.,.,.
21,06,2,.,.
22,06,.,.,.
23,06,.,.,.
24,06,2,.,.
25,06,.,4,.
26,06,.,.,.
27,06,.,.,.
28,06,.,.,.
29,06,.,.,.
30,06,.,.,.
31,06,.,.,4
32,06,.,5,.

01,07,.,6,.
02,07,.,.,.
03,07,.,.,.
04,07,.,.,.
05,07,.,.,.
06,07,.,.,.
07,07,.,.,3
08,07,.,.,.
09,07,6,.,.
10,07,.,.,.
11,07,.,.,.
12,07,.,.,.
13,07,.,.,.
14,07,4,.,.
15,07,.,.,.
16,07,.,.,.
17,07,.,.,.
18,07,.,4,.
19,07,.,.,.
20,07,.,.,.
21,07,.,6,.
22,07,.,.,.
23,07,.,.,.
24,07,.,.,3
25,07,.,.,4
26,07,.,.,.
27,07,.,.,.
28,07,.,.,.
29,07,.,.,.
30,07,.,.,.
31,07,.,.,.
32,07,.,.,.

01,08,.,.,7
02,08,.,.,.
03,08,.,.,.
04,08,.,.,.
05,08,.,.,.
06,08,.,.,.
07,08,.,.,.
08,08,.,.,6
09,08,.,.,4
10,08,.,.,.
11,08,.,.,.
12,08,.,.,.
13,08,.,.,.
14,08,.,.,.
15,08,.,.,.
16,08,.,.,.
17,08,.,.,.
18,08,.,.,.
19,08,.,.,.
20,08,.,.,.
21,08,.,.,.
22,08,.,.,.
23,08,.,8,.
24,08,.,.,.
25,08,.,.,.

```
26,08,.,.,.
27,08,2,.,.
28,08,.,.,.
29,08,.,3,.
30,08,5,.,.
31,08,2,.,.
32,08,.,.,.

01,09,5,.,.
02,09,.,.,.
03,09,.,.,.
04,09,.,.,.
05,09,.,.,.
06,09,.,.,.
07,09,.,.,.
08,09,.,.,.
09,09,.,.,.
10,09,7,.,.
11,09,.,.,6
12,09,.,.,.
13,09,4,.,.
14,09,.,.,.
15,09,.,.,4
16,09,.,.,.
17,09,.,.,.
18,09,.,.,.
19,09,.,.,.
20,09,.,.,6
21,09,.,.,.
22,09,.,6,.
23,09,.,.,.
24,09,.,.,.
25,09,.,.,.
26,09,.,7,.
27,09,.,.,.
28,09,.,.,.
29,09,.,.,.
30,09,.,.,.
31,09,.,.,.
32,09,.,.,.

01,10,.,.,.
02,10,.,.,5
03,10,.,5,.
04,10,4,.,.
05,10,.,4,.
06,10,.,.,.
07,10,.,.,.
08,10,5,.,.
09,10,.,.,.
10,10,.,.,.
11,10,.,.,.
12,10,.,.,.
13,10,.,.,.
14,10,.,.,.
15,10,6,.,.
16,10,.,.,.
17,10,.,.,.
18,10,.,.,.
19,10,.,.,.
20,10,.,.,.
21,10,.,.,3
22,10,.,.,.
23,10,.,.,.
24,10,.,.,.
25,10,.,.,.
26,10,.,.,6
27,10,.,.,.
28,10,.,.,.
29,10,.,.,.
30,10,.,.,.
31,10,.,.,.
32,10,.,.,.

01,11,.,.,.
02,11,.,.,.
03,11,.,.,3
04,11,.,.,.
05,11,5,.,.
06,11,.,6,.
07,11,.,.,.
```

```
08,11,.,.,.
09,11,.,.,.
10,11,.,.,.
11,11,4,.,.
12,11,.,.,.
13,11,.,.,.
14,11,.,.,.
15,11,.,.,.
16,11,.,.,8
17,11,.,.,.
18,11,.,.,.
19,11,.,.,.
20,11,.,.,.
21,11,.,.,.
22,11,.,.,.
23,11,.,.,.
24,11,.,.,.
25,11,4,.,.
26,11,.,.,.
27,11,.,.,6
28,11,.,.,.
29,11,.,.,.
30,11,.,3,.
31,11,.,.,.
32,11,.,.,.

01,12,.,.,.
02,12,.,.,.
03,12,.,.,.
04,12,.,.,.
05,12,.,.,.
06,12,.,.,.
07,12,4,.,.
08,12,.,.,.
09,12,.,.,.
10,12,.,.,.
11,12,.,3,.
12,12,.,7,.
13,12,.,.,.
14,12,.,.,.
15,12,.,.,.
16,12,.,.,.
17,12,.,.,.
18,12,.,.,4
19,12,.,.,.
20,12,.,.,.
21,12,.,.,.
22,12,.,.,.
23,12,.,.,.
24,12,.,2,.
25,12,.,.,.
26,12,8,.,.
27,12,.,.,.
28,12,2,.,.
29,12,.,.,3
30,12,.,.,.
31,12,.,.,.
32,12,.,.,.
```

## 16.75 Teaching Rasch measurement, Rasch analysis

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 ....

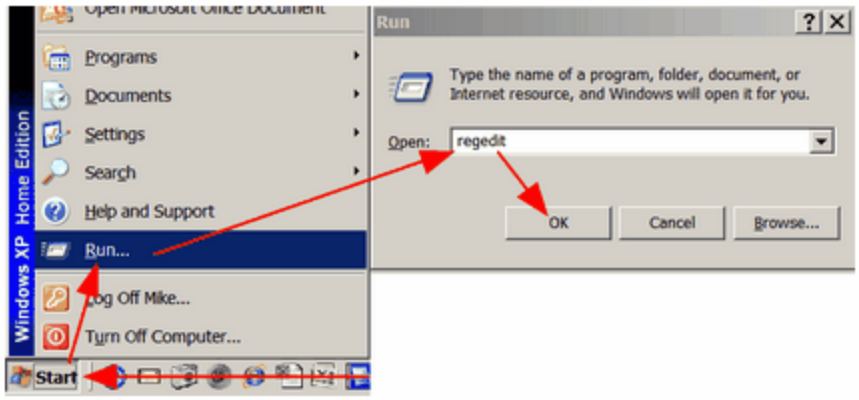
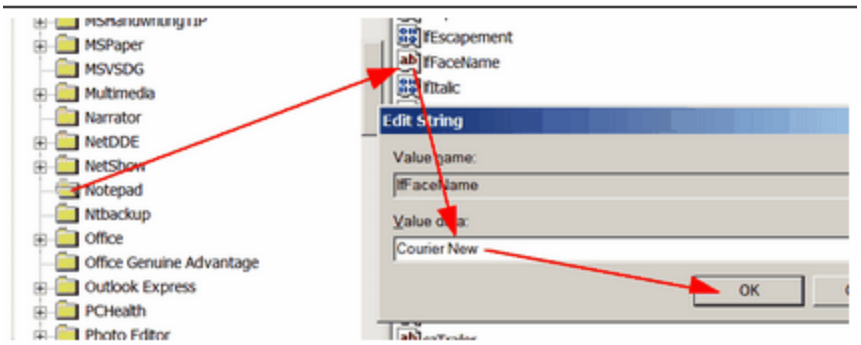
So avoid CTT pitsfalls by starting from Rasch first principles with lots of examples and classroom interactions.

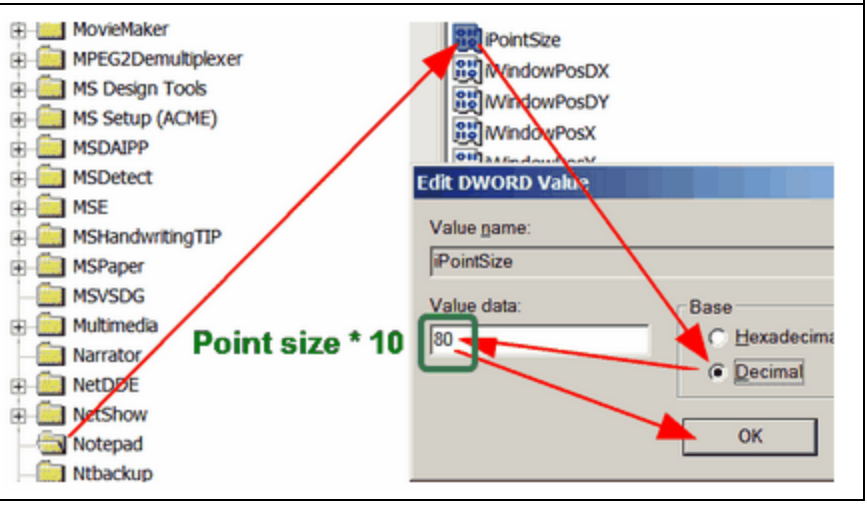
1. What do we want to measure? Can we express it along a line (latent variable) from less (easy, beginner) to more (hard, expert)?
2. Can we convert this progression into dichotomous items? (relevant items = content validity, item hierarchy = construct validity)
3. How do we expect people of different 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

## 16.76 Text Editor

Winsteps uses a text editor to display and edit control files, data files, Tables and output files. The default editor is NotePad. You can change to another editor, such as WordPad or TextPad in Edit Initial Settings.

| **To make permanent (default) changes in NotePad font face and/or size:** | |
|---|---|
| Windows "Start"<br>Click on "Run"<br>Type in "regedit"<br>Click on "OK" |  |
| Registry Editor:<br>Click on the + in front of "HKEY_CURRENT_USER"<br>Click on the + in front of "Software"<br>Click on the + in front of "Microsoft"<br>Click on "Notepad"<br>For the type face:<br>Double-click on "lfFaceName"<br>Type in "Courier New"<br>Click on "OK" |  |

| To make permanent (default) changes in NotePad font face and/or size: | |
|---|---|
| For the font size:<br>Double-click on "iPointSize"<br>Click on "Decimal"<br>Type in 80 (for point-size 8 multiplied by 10)<br>Double-click on "lfFaceName"<br>Click on "OK"<br><br>Close registry<br>Click on top right ☒ |  |

## 16.77 t-statistics

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.

| *Table of the two-sided t distribution:* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **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) | | |

**Welch's refinement of Student's t-test for possibly unequal variances:**

For sample 1 (or estimate 1),
$M_1$ = mean of the sample (or the estimate)
$SS_1$ = sum of squares of observations from the individual sample means
$N_1$ = sample size (or number of observations)
$N_1 - 1$ = degrees of freedom
$SS_1 / (N_1 - 1)$ = sample variance around the mean (or the measure)
$SS_1 / ((N_1 - 1)(N_1))$ = standard error variance = $EV_1 = SE_1^2$

$SE_1$ = Sqrt($EV_1$) = standard error of the mean (or the standard error of the measure estimate)

($M_1$ - $M_2$) = difference between means (or estimates)
sqrt ($SE_1^2$ + $SE_2^2$) = joint standard error of the means (or of the estimates

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 https://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

## 16.78 Unmeasurable

If an element is reported as "Unmeasurable" in Table 7 or Status "-4" in a Score file, then ...

Observations for the element are part of an extreme score. One way is to impute non-extreme elements.

For instance:

Original specifications and data:
Models = ?, ?, R5 ; persons, items, rating-scale is 1-5

Data =
1, 1-2, 3, 4
2, 1-2, 3, 2
3, 1-2, 5, 5
4, 1-2. 1, 2

Person 3 is reported as "Unmeasurable". This is because it is the only person element rated 5, and all its ratings are 5. We know that 5 is better than 4, but the data do not tell us how much.

Let's say that 5 is only a little better than 4. Then we impute observations of 4 to Person 3:

```
Models =
3, ?, RScale, 0.5  ; weight observations of 3 at o.5
?, ?, RScale ; everyone else remains as before
*
Rating-scale = RScale, R5, General  ; Person 3 and everyone else share the same rating scale

Data =
1, 1-2,  3, 4
2, 1-2,  3, 2
3, 1-2,  5, 5  ; original data, now weighted 0.5
3, 1-2,  4, 4  ; imputed data, now weighted 0.5
4, 1-2.  1, 2
```

Person 3 will now be estimated with a measure corresponding to a rating of 4.5

## 16.79 Unobserved or null categories

There are two types of unobserved or null categories: structural zeroes and incidental/sampling zeroes.

**Structural** null categories occur when rating scale categories are numbered 10, 20, 30,... instead of 1,2,3. Facets ordinarily eliminates unobserved categories.

incidental/sampling zeroes occur when occur when rating scale categories are numbered 1, 2, 3, ... but a category such as 2 is not observed this time. Since Facets ordinarily eliminates unobserved categories, the unobserved categories must be "kept". Extreme unobserved categories can only be kept by anchoring.

For intermediate **incidental** or **sampling** 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, somewhat agree, agree". We can imagine that no child in this sample selected the half-smile of somewhat agree.
The category frequencies of categories 0,1,2,3 are 378, 620, 0, 852
The three Rasch-Andrich threshold parameters are -.89, +infinity, -infinity.
The +infinity is because the second parameter is of the order log(620/0). The -infinity 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 the Anchorfile=. 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. To force unobserved intermediate categories into the analysis, use:
Models = ?,?,?, R9K
or
Models = ?,?,?, myscale
Rating scale = myscale, R9, Keep

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. .

Our recommendation is that structural zeroes be rescored out of the data.

**Unobserved extreme categories:**
If these categories **may be** observed next time, then it is better to **include dummy data records** in your data file which include observations of the missing category and reasonable values for all the other item responses that accord with that missing category. Give the data dummy records a very small weight using R weighting. These few data records will have minimal impact on the rest of the analysis.
If there are too many unobserved categories, then it may be better to impute a rating-structure using anchor values in Rating scale=, or model the rating scale as a binomial trial.

---

*Question 1:* The items in my data have different scoring systems. For example,
Items 1 and 2 have categories: 0, 3, 6, 8
Items 3 and 4 have categories: 0, 4, 8, 12

*Answer:* Modeling these depends on how you conceptualize these scales. Unobserved categories are always difficult to model.

A. If 0,3,6,8 really mean 0,1,2,3

and 0,4,8,12 really means 0,1,2,3, then, if items are facet 2
models=*
?, 1-2, ?, R8
?, 3-4, ?, R12
*

B. Or if 0,3,6,8 really mean 0,1,2,3,4,5,6,7,8
and 0,4,8,12 really means 0,1,2,3,4,5,6,7,8,9,10,11,12, then
models=*
?, 1-2, ?, scalea
?, 3-4, ?, scaleb
*
rating scale=scalea,R8,K
rating scale=scalea,R12,K

C. Or if 0,3,6,8 really mean 0,1,2,3,4,5,6,7,8, 9, 10,11,12
and 0,4,8,12 really means 0,1,2,3,4,5,6,7,8,9,10,11,12, both on the same 0-12 scale, then
models=*
?, ?, ?, scalec
*
rating scale=scalec,R12,K

D. Or if 0,3,6,8 and 0,4,8,12 correspond to 0,3,4,6,8,12 and really mean 0,1,2,3,4,5, then
models=*
?,?,?,R12
*

---

*Question 2:* I am using the Partial Credit Model, #, and my raters have used different parts of the rating scale. How can I get the full range of the rating scale reported for every rater?

*Answer:* Include dummy data rated in the highest and lowest category by every rater with very small weight. Also include "K" (for Keep) next to the rating scale specification.

*Example:* the rating scale should go from 1 to 7, but some raters have missing categories:
Models = ?,?,#, R7K ; raters are facet 3

Data =
(your data, then)
R0.00001, 0, 0, 1, 1 ; rater 1 (facet 3) gives a rating of 1
R0.00001, 0, 0, 2, 7 ; rater 1 (facet 3) gives a rating of 7
and so on for all raters.

## 16.80 Weighting the data

There are 3 methods of weighting:
1) Models= model weight: Model = ?,?,..., R, model weight
2) Labels= element weight: element number = element label, anchor value, group number, element weight
3) Data= observation weight: R...

These multiply to give a combined weight to each observation.

---

In *Facets,* all the weights (in Models=, in Labels= or R.... at the observation level) are applied as replications of the observation. So for instance:

Models=?,?,?, 2 ; weight of two
or
Labels=*
1, Facet 1
1, element 1, , , 2 ; weight of two
or
R2, 1,1,1, 3 ; weight of two
are all processed internally to mean the same as
1,1,1, 3
1,1,1, 3

Fractional weights are allowed.
If two or more weights apply to the same observation, then the weights are multiplied.

---

**Reliability Index:** 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 = Ru. The reliability of the weighted analysis, using an initial set of weights, = Rw. We can then scale the weights using the Spearman-Brown Prophecy Formula: S = Ru * (1-Rw) / ((1-Ru)*Rw)). Multiply the initial set of weights by S. Then the weighted and unweighted reliabilities should be the same.

**Weighting using Models=:** Example: Two Cases: A and B. Four aspects: Taste, Touch, Sound, Sight.
Case A Taste weight twice as important as the rest.
Case B Sound weight twice as important as the rest.

Labels =
1, Examinees
1-1000
*
2, Case
1=A
2=B
*
3, Aspect
1=Taste
2=Touch
3=Sound
4=Sight
*
Models=
?, 1, 1, MyScale, 2 ; Case A Taste weighted 2
?, 2, 3, MyScale, 2 ; Case B Sound weighted 2
?, ?, ?, MyScale, 1 ; everything else weighted 1
*
Rating scale = MyScale, R9, General ; this rating scale is the same for all models

If you want to keep the "reliabilities" and standard errors meaningful then adjust the weights:

Original total weights = 2 cases x 4 aspects = 8
New total weights = 2 + 2 + 6 = 10
Weight adjustment to maintain total weight is 8/10.

So adjusted weighting is:
Models=
?, 1, 1, MyScale, 1.6 ; Case A Taste

?, 2, 3, MyScale, 1.6 ; Case B Sound
?, ?, ?, MyScale, 0.8 ; everything else
*

**Weighting using Labels=**: individual elements can be weighted

element number = element label, anchor value, group number, element weight

Labels=
...
*
3, facet name
1 = first element, , , 0.8 ; all observations with this element weighted 0.8
....
*

**Weighting of a data point using R.... weights:** can be specified by R (or another replication character) and the number of replications, for instance:
R3,2,23,6,4 means that the value of 4 was observed in this context 3 times.
Fractional replication permits flexible observation-weighting:
R3.5,2,23,6,4 means that the value of 4 was observed in this context 3.5 times.

*Example:* We want to construct response data according to the known probabilities of being observed:
Person 3 has a 60% probability of succeeding on item 4:
Person 7 has a 25% probability of succeeding on item 11:
Data=
R0.60, 3, 4, 1 ; 60% probability of success
R0.40, 3, 4, 0 ; 40% probability of failure
R0.25, 7, 11, 1 ; 25% probability of success
R0.75, 7, 11, 0 ; 75% probability of failure

**Weighting summaries using element weights and R.... weights:**
*Example:* for 25 people scoring 32, there was 0.57 success on item 4:
Facets=2
Models= ?,?,D
Labels=
1, Raw Scores
....
32=32, , , 25 ; weight by number of people at score
.....
*
2, item
1-9
*
Data=
....
R0.57 , 32,4, 1 ; weight correct answer by its probability
R0.43 , 32,4, 0 ; weight incorrect answer by it probability
....

**Weighting specific observations:** We want to give some incorrect answers a smaller penalty than other incorrect answers. There are two ways to do this:

1) in the data:

3 facets + correct

```
2,3,4, 1
3 facets + incorrect
2,3,4, 0
3 facets + half-weight incorrect
R0.5, 2,3,4, 0
```

2) with a Models= specification and a weighting facets

```
Models =
; 3 facets + dummy indicator facet + correct/incorrect
?,?,?,1,D,1   ; full weight
?,?,?,2,D,0.5   ; half weight
*
Labels=
....
*
4, Weighting, A
1 = Full weight, 0
2 = Half weight, 0
*
Data =
3 facets + indicator +correct
2,3,4, 1, 1
3 facets + indicator + incorrect
2,3,4, 1, 0
3 facets + indicator +half-weight incorrect
2,3,4,  2, 0
```

3) with a Labels= specification and a weighting facets

```
Models =
; 3 facets + dummy indicator facet + correct/incorrect
?,?,?,?,D
*
Labels=
....
*
4, Weighting, A
1 = Full weight, 0, , 1          ; element weight
2 = Half weight, 0, , 0.5
*
```

## 16.81 Zeroes: Structural and Incidental: Ordinal or Keep

Unobserved categories can be dropped from rating scales (or partial credit items) 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 scale (or partial credit) logic or are usually observed, even though they happen to have been unused this time. Category transitions for which anchor Rasch-Andrich threshold-values (step calibrations) are supplied are always maintained wherever computationally possible, even when there are no observations of a category in the current data set.

Use "Rating Scale= .... Keep" when there may be intermediate categories in your rating scale (or partial credit) that aren't observed in this data set, i.e., incidental zeroes.

Use "Rating Scale= .... Ordinal" when your category numbering deliberately skips over intermediate categories, i.e., structural zeroes.

"Rating Scale= .... Ordinal" Eliminate unused categories and close up the observed categories.

"Rating Scale= .... Keep" Retain unused non-extreme categories in the ordinal categorization.

When "Rating Scale= .... Keep", missing categories are retained in the rating scale (or partial credit), so maintaining the raw score ordering. But missing categories require arbitrarily extreme Rasch-Andrich thresholds. If these threshold values are to be used for anchoring later runs, compare these thresholds with the thresholds obtained by an unanchored analysis of the new data. This will assist you in determining what adjustments need to be made to the original threshold-values in order to establish a set of anchor threshold-values that maintain the same rating scale (or partial credit) structure.

Example 1: Incidental unobserved categories. Keep the developmentally important rating scale (or partial credit) categories, observed or not. Your small Piaget scale goes from 1 to 6. But some levels may not have been observed in this data set.
```
Models = ?,?,?, Piagetscale
Rating Scale = Piagetscale, R6, Keep
```

Example 2: Structural unobserved categories. Responses have been coded as "10", "20", "30", "40", but they really mean 1,2,3,4
```
Models = ?, ?, ?, Tensscale
Rating Scale = Tensscale, R40, Ordinal
; if "Rating Scale= .... Keep", then data are analyzed as though categories 11, 12, 13, 14, etc. could exist, which would distort the measures.
; for reporting purposes, multiply Facets reported raw scores 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 "Rating Scale= .... Keep". 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
(b) Set "Rating Scale= .... Keep", so that the observed 1,3,4 and unobserved 2 are treated as 1,2,3,4
(c) For reporting purposes, multiply the reported Facets scores by 2 using Excel or similar software.
Models = ?,?,?,Evenscale
Rating Scale = Evenscale, R8, Keep
1 = original 2, , , 2
2 = original 4, , , 4
3 = original 6, , , 6
4 = original 8, , , 8
*
```

---

**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 "Rating Scale= .... Ordinal". 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 "Rating Scale= .... Keep". 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 "Rating Scale= .... Keep", unobserved intermediate categories are imputed using a mathematical device noticed by Mark Wilson. This device can be extended to runs of unobserved categories.

# Index of Specifications and Files


#File: 3mile.txt ....56
#File: Baseball.txt ....57
#File: Chinese.txt ....83
#File: Creativity.sav ....65, 95
#File: Creativity.txt ....65
#File: Creativity.xls ....65, 87
#File: Dives.txt ....71
#File: Dives.xls ....71, 87
#File: Divesb.txt ....72
#File: Divesint.txt ....72
#File: Essayday.txt ....80
#File: Essays.txt ....79
#File: Facets95.bat ....109
#File: Facetsnt.bat ....109
#File: Facetsxp.cmd ....109
#File: Gel.txt ....81
#File: Guilford.sav ....65, 95
#File: Guilford.txt ....65
#File: Guilford.xls ....65, 87
#File: Guilfordsb.txt ....72
#File: Kct.txt ....58
#File: Kcta.txt ....63
#File: Kctinter.txt ....75
#File: Lfs.txt ....62
#File: Meas2anc.txt ....69
#File: Measanc.txt ....69
#File: Measure.txt ....69
#File: Olympics.txt ....79
#File: Pair.txt ....77
#File: SASdata.sas7bdat ....92
#File: Simul.bat ....109
#File: Sportcas.txt ....60
#File: STATAdata.dta ....96
#File: Subsets.txt ....67
#File: Woodcut.txt ....64

%include file (included specification and data files) ....102

Anchor output file = "" ....103
Arrangement by Number/Alpha/Measure/Fit/PBS (Asc./Desc.) = N ....106
ASCII output table display format = Yes ....107

Barcharts output = Yes ....108
Batch file processing = Yes ....109
Beep sound = No ....111
Bias (direction: difficulty/ability/omit) = Ability ....111
Boxshow (show boxes around tables) = Yes ....111

Center facet = ....112

Convergence (criteria to end iteration) = .5, .01 ..... 113
CSV output format = " " ..... 114

Data file name = facets.dat ..... 114
Delements element type = ..... 118
Dvalues for data facets = ..... 121

Entry order of facets in data = 1,2,3,4,5,... ..... 132

Facets in data = 2 ..... 133
Fair scores based on = Mean ..... 133

General statistics = No ..... 154
Glabel group labels = ..... 134
Graph plotting file name = " "," " ..... 136

Hardware = 80 ..... 137
Heading lines = Yes ..... 137

Inter-rater Agreement Coefficients = 0 ..... 138
Iterations (stop after) = 0 (unlimited) ..... 138

Juxtapose (column headings after every this many lines) = 0 ..... 138

Keep as null element zero = 0 ..... 149

Labels of facets and elements = ..... 138
Left-hand placement of row labels = No ..... 140

Missing data codes = ..... 141
Model to be used in the analysis = ?,?,D,1 ..... 142

Negatively-oriented facet = ..... 148
Newton-Raphson = 0 ..... 148
Noncenter facet = 1 ..... 149
Null element = 0 ..... 149

Output results/anchors/residuals = ..... 151

Positively-oriented facet = 1 ..... 152
Pt-biserial = No ..... 154

QM quotation marks around labels = ..... 157
Query to monitor matching of data to models = No ..... 158

Rating scale (or Response model) = ..... 159
Replication character = R ..... 167
Residuals Output File = "" ..... 168
Reversed scoring = Yes ..... 170

Score/Measure output files name and CSV status = " "," " ..... 171
Show unobserved elements = No ..... 150, 174
Simulated data file = "" ..... 174
Standard (error is Asymptotic/Real) = Asymptotic ..... 174
Subset detection = Yes ..... 176

T3onscreen show only one line on screen iteration report = Y ..... 178
T4maximum numer of residuals in Table 4 = 100 ..... 178
T7LGS Table 7 List-Group details-Group Summary = LGS ..... 179
T8NBC Table 8 Numbers-Barcharts-Curves = NBC ..... 179

Tables output from main analysis = Yes....179
Title for each results table =....179
Total score = No....180
Udecimals (user decimals) = 2....180, 182
Umean (user mean and scale and decimals) = 0,1,2....180, 182
Unexpected (standardized residuals reported at least) = 3....181
Uscale (user scale) = 1....180, 182
Usort sort order of unexpected responses = u....183
UTF8enc encoding = ?....184
UTF8sub substitute character = ....184
Vertical rulers in Table 6 = 1A, 2A, 3A,....184
Whexact Wilson-Hilferty standardization = Yes....187
Write iteration details in output file = No....188
www.rasch.org/rmt/rmt221j.htm....201
Xtreme (extreme score adjustment) = 0.3, 0.5....188
Yardstick (Table 6 columns lines low high) = 80,,,....189
Zscore minimum for reported bias terms (bias/t-statistic) = 0,0....190