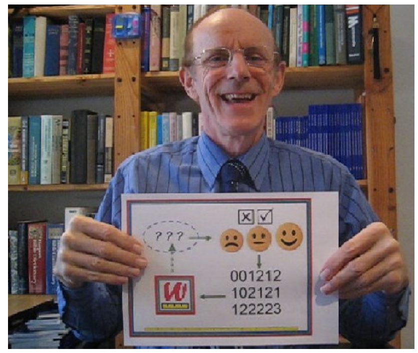
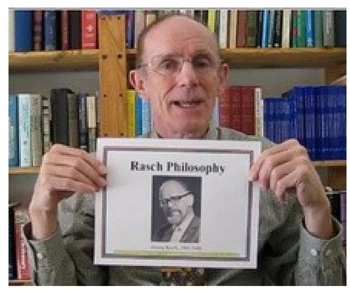
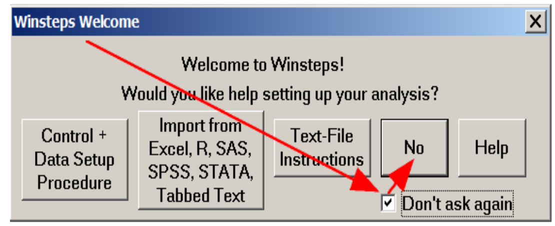
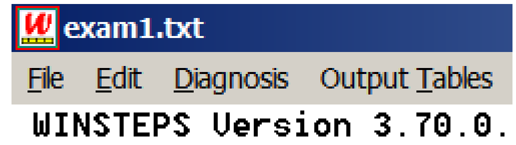
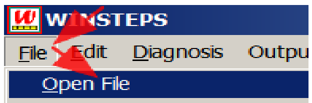
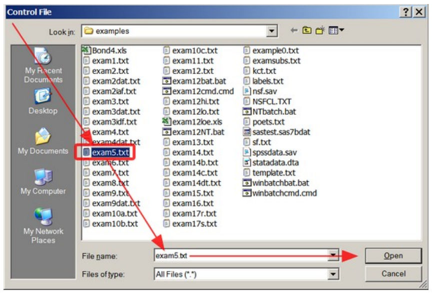
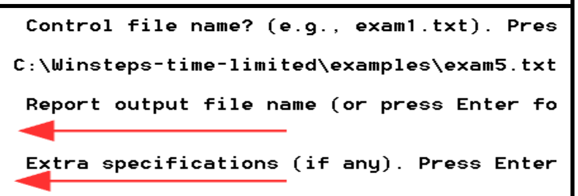
#	Practical Rasch Measurement - Further Topics : www.winsteps.com Mike Linacre, instructor - July 2011	
1.	**Tutorial 1. Analysis of dichotomous (true-false, multiple-choice) data** ***Welcome!*** - and a reminder: *this Course follows on from the "Core Topics" Introductory Course.* • Winsteps software installation and operation • Analysis with missing data • Correlations and local fit-statistics • Item-person maps • Estimating measures from dichotomous data **Each section includes important aspects of theory and practice, so please go through the material carefully.**	
2.	**A. *Introductory Videos***	
3.	***If this is new to you,*** Please view a few minutes of a movie introduction to Rasch and Winsteps by "yours truly" ... ***it is 50MB so may take a few minutes to download ...*** http://www.winsteps.com/a/intro.wmv This is a big file. Add it to your Media Player playlist. Or you may need to use a free download manager such as FDM Lite ***This movie was originally requested by the energetic Rasch folks in Hong Kong - a great idea!*** ***The movie contains slides, separately viewable at*** http://www.winsteps.com/a/slides.htm	
4.	And this http://www.winsteps.com/video/philosophy.htm which digs deeper. ***Strange words?*** Rasch measurement has its own technical words. Please look at http://www.winsteps.com/winman/glossary.htm for explanations. This is also in Winsteps Help	
5.	You can launch *Winsteps* at any time by double-clicking the short-cut on your desktop.	

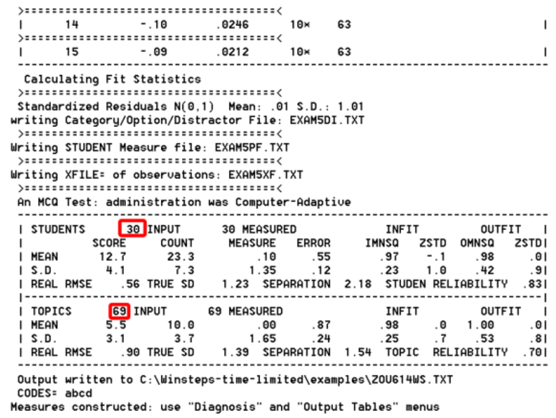
<table>
<tr><th>6.</th><th colspan="2">B. Winsteps Analysis Window</th></tr>
<tr><td></td><td>The main Winsteps window displays.
At the top of the screen is the Winsteps menu bar.

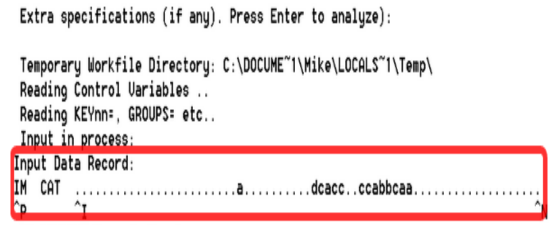
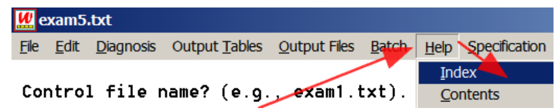
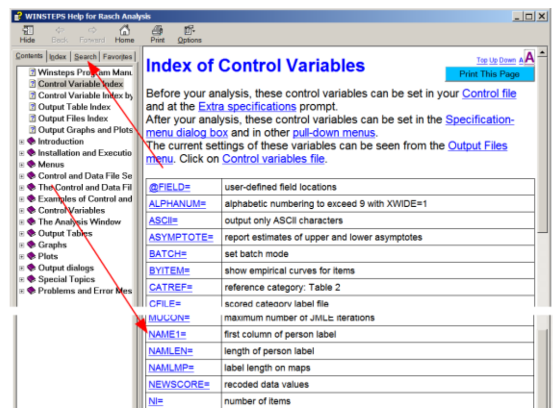
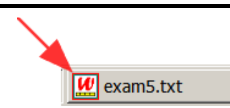
The "Winsteps Welcome" box displays. This shows some start-up options. We will not need these.
Check: "Don't ask again"
Click "No"</td><td>
</td></tr>
<tr><td>7.</td><td>The Winsteps analysis window displays.</td><td>
</td></tr>
<tr><td>8.</td><td>Click on “File” menu
Click on “Open File”</td><td>
</td></tr>
<tr><td>9.</td><td>In the “Control File” box,
Open “Exam5.txt” (you can double-click on it)

Your Winsteps-analysis window shows
C:\Winstep\examples\exam5.txt</td><td>
</td></tr>
<tr><td>10.</td><td>“Report output file name”
Press your Enter key
“Extra specifications”
Press your Enter key</td><td>
</td></tr>
<tr><td>11.</td><td>The estimation of Rasch measures for the Exam5.txt MCQ data is performed.

The summary statistics report that 30 students (persons) and 69 topics (items) have been analyzed.

In your analysis of your own data, if these numbers are not what you expected, then you need to look at the control file immediately.

So far, we are OK.</td><td>
</td></tr>
</table>

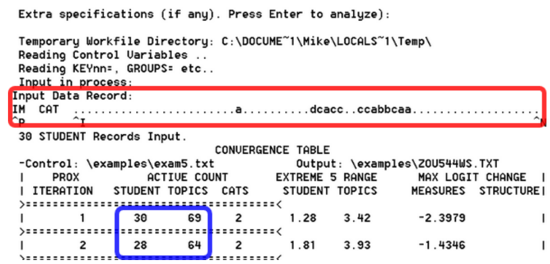
12.	Scroll back up the analysis window to the top. Winsteps data files are **rectangular:** each row is one person, examinee, subject, patient, ... each column is one item, probe, prompt, task, question, .. Do you see "Input Data Record"? This shows the first record (person, subject, case) in your data file, and how Winsteps is interpreting it. ^P means "first column of the person label" This is specified with NAME1= ^I means "the start of the responses to the items" This is specified with ITEM1= ^N means "last column of the responses" This is calculated from ITEM1=, NI= and XWIDE=	
13.	Oops! What's this about NAME1=? What's that? This Course assumes familiarity with basic Winsteps control variables, but we all have gaps in our knowledge, so click on "Help" on the Winsteps menu bar Click on "Index"	
14.	Winsteps Help displays. In the Index of Control Variables, scroll down to see NAME1=. Click on this if you need to know more about it. The Help feature includes the equivalent of about 500 printed pages of material. You can find information quickly using the "Search" feature. Explore in Help as much as you like (whenever you like!) Help is also on line at http://www.winsteps.com/winman/	
15.	Click on the Winsteps icon on the Windows task bar to reactivate Winsteps.	

16. Toward the top of the Winsteps Analysis window, we have looked at the first Input Data record. Make sure that the ^ pointers match the data record.

Then there is the "Convergence Table".

The first line reports 30 students and 69 topics and 2 "CATS" = categories. The 2 categories are "right" and "wrong" exactly what we expect for dichotomous data. "Dichotomous data" are any data scored in two categories including "true-false", "present-absent", multiple-choice.

But what about the second line? 28 students and 64 topics. These numbers tell us how many students and topics have non-extreme (not zero, not perfect) scores, and so can be used to estimate the relative difficulty of the items and the relative ability of the persons. But much more about estimation very soon ...



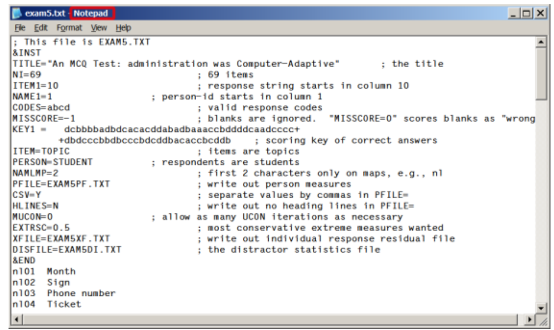
17.

C. Editing a Control File

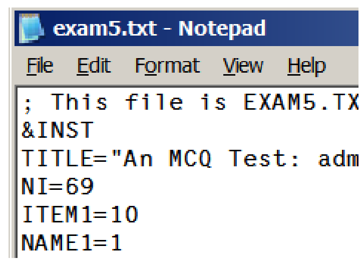
18. In this Course, we will do a lot of tweaking of Winsteps control variables by editing the control file. So let's do that now:
On the Winsteps menu bar, click on "Edit"
Click on "Edit Control File"

19. The Control file is a text file and is displayed in NotePad.

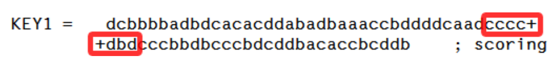
If your NotePad display does not look like mine, with the columns aligned, the please change the NotePad settings. See Appendix 3. Changing the text appearance in NotePad. If you prefer to use a different text-editor (such as WordPad or TextPad), then please see "Editor path" in Winsteps Help or http://www.winsteps.com/winman/editinitialsettings.htm



20. Control file basics:
; starts a comment. Anything after ; is ignored by Winsteps
&INST is now ignored. It is here for backward-compatibility
TITLE= is the same as Title= or title = or titel=. Winsteps allows upper and lower case, and abbreviations, alternate spellings, and spaces. But the first two or three letters must be the same.
"An MCQ Test" - if a variable value includes a space, please put it in "double-quotes" or 'single-quotes'.



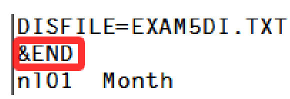
21. Continuation lines: long values can be split by putting + at the end of one section and + at the beginning of the next. The two sections are concatenated (joined) together omitting the +.



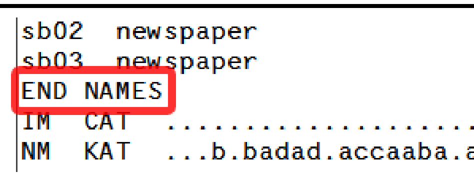
becomes

.....ccccdbd.....

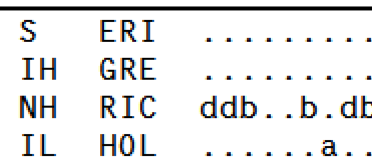
22. &END follows the control variables. The order of the variables does not matter. In Exam5.txt, DISFILE= happens to be the last variable setting.
"nl01 Month" is the label (identification) of the first item.

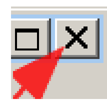


23. Scroll down ...
END NAMES or END LABELS follows the last item label.
Then come the data records, defined by NAME1=, etc.
The data can be in a separate file, specified by DATA=



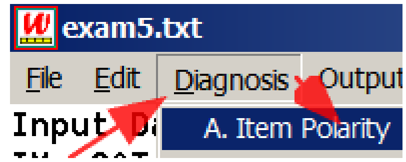
24. Scroll to the bottom of Exam5.txt
The last line is the last data record. Winsteps discovers this.
We don't tell Winsteps how many persons there are, except when we are using the Data Setup input option.



<table>
<tr>
<td>25.</td>
<td>A reminder about the most common control variables:
TITLE= is a brief heading which will be shown at the top of every output report
NI= is the number of items
ITEM1= is the starting column of the items in the data
NAME1= is the starting column of the persons in the data
CODES= are all valid response codes in the data file</td>
<td>TITLE="An MCQ Test: administration was Computer-Adap
NI=69 ; 69 items
ITEM1=10 ; response string st
NAME1=1 ; person-id starts in column
CODES=abcd ; valid response cod
MISSCORE=-1 ; blanks are ignored
KEY1 = dcbbbbadbdcacacddabadbaaaccbddddcaadcccc+
+dbdcccbbdbcccbdcddbacaccbcddb ; scoring
ITEM=TOPIC ; items are topics
PERSON=STUDENT ; respondents are students</td>
</tr>
<tr>
<td>26.</td>
<td colspan="2">MISSCORE= value (Missing-scored-as = value) is the action for responses not in CODES=.
MISSCORE=-1 (the default) means “ignore this response, treat as Not Administered”.
MISSCORE=0 means “score as 0, treat as wrong”.
KEY1= is the scoring key for the items. The correct answers are in the same order as the item columns in the rectangular data.
ITEM= allows you to choose a descriptive word for your items
PERSON = allows you to choose a descriptive word for your persons</td>
</tr>
<tr>
<td>27.</td>
<td>You can make any changes you want to this text file.
Then Save it. After analysis complete, any changes to Exam5.txt will alter the next analysis, not this one.
Close the Editor window.</td>
<td></td>
</tr>
<tr>
<td></td>
<td colspan="2"></td>
</tr>
</table>

28. **D. Diagnosing the Data: Diagnosis menu, A. Item Polarity & Table 26**

29. In the Winsteps Analysis window for Exam5.txt, let's investigate these data:
On the Winsteps menu bar,
Click on Diagnosis
Click on A. Item Polarity



30. The "Item Polarity" table, Table 26.1, displays.
This Table is full of exciting information.
First, look at the PT-MEASURE (point-measure correlation) columns. We expect to see positive correlations (positive polarity). So that higher-valued responses to the items (correct answers) correlate positively with the person measures.
When the data are a complete rectangle (with no missing responses), then the point-measure correlation is similar to the point-biserial correlation. But many observations are missing, so that these data are not complete. The point-biserial correlation malfunctions.

```
TABLE 26.1 An MCQ Test: administration was Comput ZOU484WS.TXT Aug  9 20:38 2008
INPUT: 30 STUDENTS  69 TOPICS  MEASURED: 30 STUDENTS  69 TOPICS  2 CATS   3.65.1
--------------------------------------------------------------------------------

STUDENT: REAL SEP.: 2.18  REL.: .83 ... TOPIC: REAL SEP.: 1.54  REL.: .70

     TOPIC STATISTICS:  CORRELATION ORDER

-------------------------------------------------------------------------------------------
|ENTRY   RAW                    MODEL|   INFIT  |  OUTFIT  |PT-MEASURE |EXACT MATCH|
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| TOPIC
|------------------------------------+----------+----------+-----------+-----------+--------
|    37     11     13   -1.77    .82|1.41    .9|3.57   2.1| -.26   .32| 84.6  84.6| im05  library
|    51      3      4    -.31   1.19|1.26    .6|1.30    .6| -.20   .24| 75.0  75.2| al03  airport
|    12      7      8   -3.16   1.08|1.16    .5|1.54    .8| -.15   .17| 87.5  87.4| nm07  sign on
|    13      8     13   -1.04    .61|1.36   1.5|1.48   1.6| -.11   .35| 46.2  67.3| nh01  superma
|    52      2      7    2.03    .92|1.55   1.1|1.41    .9| -.07   .42| 57.1  77.9| al04  postal
|    28     12     16   -1.49    .61|1.36   1.2|1.40    .9| -.06   .31| 68.8  75.6| il05  on stre
|    19      5      9   -1.13    .74|1.40   1.5|1.64   1.4|  .02   .41| 55.6  66.1| nh07  Taipei
```

31. Red box: CORR. is the observed correlation. EXP. is the expected correlation.
When the data fit the Rasch model, these values will be the same.
When CORR. is greater than EXP., the item is over-discriminating between high and low performers.
When CORR. is less than EXP., the item is under-discriminating between high and low performers.
When EXP. is near to zero, then the item is very easy or very hard. It is off-target to the person distribution

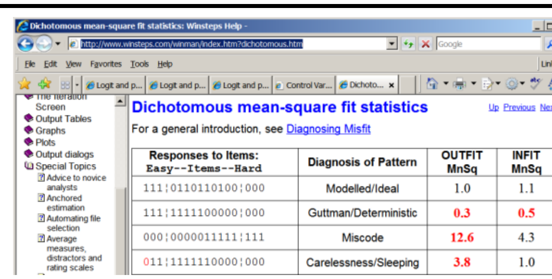
32. For item "im05 library", the point-measure correlation (polarity) is negative, CORR. = -.26. It is **contradicting** our expectation that "the higher the ability of the person, the more probable the person will succeed on this item". This item may be destroying measurement, not constructing it. If the persons had answered this item how the Rasch models expects, the point-measure correlation would have been EXP. = 0.32

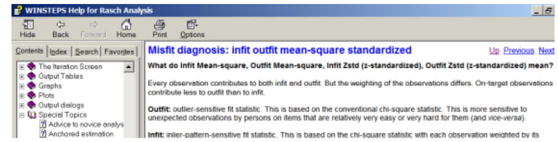
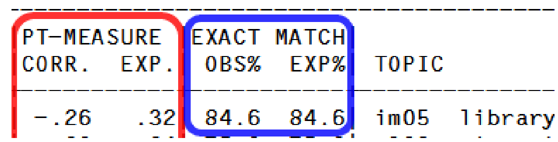
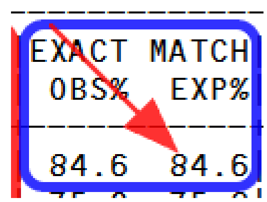
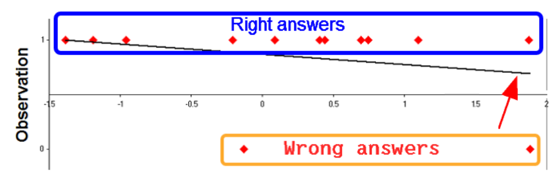
```
-------------------------------------------
|PT-MEASURE |EXACT MATCH|
|CORR.  EXP.| OBS%  EXP%| TOPIC
|-----------+-----------+-------------
| -.26   .32| 84.6  84.6| im05   library
```

33. To the left, the OUTFIT MNSQ, the Outfit Mean-Square statistic, is huge, and ZSTD, its standardized value, indicate that this is not likely to be due to chance, so there is a serious flaw in the data.

```
-----------------------
|   INFIT  |  OUTFIT  |
|MNSQ  ZSTD|MNSQ  ZSTD|
+----------+----------+
|1.41    .9|3.57   2.1|
|1.26    .6|1.30    .6|
```

34. If interpreting outfit and infit mean-square statistics is new to you, then please see Winsteps Help: "Dichotomous mean-square fit statistics"
http://www.winsteps.com/winman/dichotomous.htm



35.	If ZSTD standardized statistics are new to you please see Winsteps Help, “Misfit diagnosis: infit outfit mean-square standardized”, which is also at http://www.winsteps.com/winman/diagnosingmisfit.htm	
36.	In Table 26.1, look a little to the right. We see something very surprising. “EXACT MATCH”. 84.6% (OBS%) of the responses in the data to “im05” match their predicted values (“right” or “wrong”) based on the Rasch measures. If the data fit the Rasch model, we would expect (EXP%) 84.6% of the responses to match their predicted values.	
37.	The EXP% is not 100%, because the Rasch predicts both the response values and also that there will be some unpredictability (randomness) associated with the response values. When we combine the predicted values with their predicted randomness, the expected exact matches become EXP% = 84.6%.	
38.	So the correlation is negative, but 84.6% of the responses agree with our expectations. So our diagnosis is that the 15.4% of observations (which disagreed with what we expected) were very surprising: wrong answers to very easy items, or right answers to very difficult answers. The graph here shows the negative correlation due to the two unexpected wrong answers. (2/13 = 15.4%)	 *I drew this with Excel using the Winsteps XFILE= output*

39. **E. Table 26.3 Category/Option/Distractor Subtable**

40. Scroll down to Table 26.3. This gives us more information about item "im05." Option "d", scored 0, was chosen by 2 people. Option "c", scored 1, was chosen by 11 people. 17 people have "MISSING" because they were not given the item.

```
-------------------------------------------------------------------------------------
|ENTRY   DATA  SCORE |     DATA    | AVERAGE  S.E.  OUTF PTMEA|
|NUMBER  CODE  VALUE |  COUNT    % | MEASURE  MEAN  MNSQ CORR.| TOPIC
|--------------------+-------------+--------------------------+----------------------
|    37  d         0 |      2   15 |     .88  1.01  4.0   .26 |im05  library
|        c         1 |     11   85 |     .15*  .30  1.3  -.26 |
|        MISSING *** |     17   57*|     .56   .55         .08 |
```

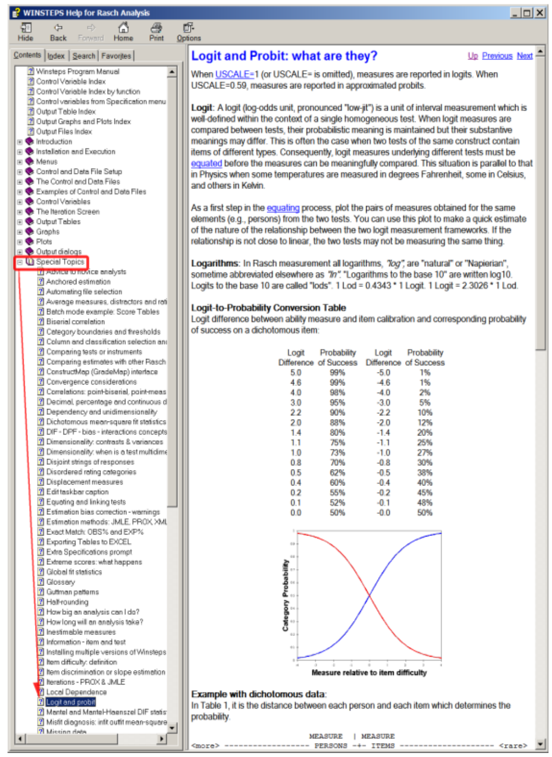
41. The AVERAGE ability MEASURE of the people choosing the two incorrect answers is .88 logits (log-odds units, pronounced "low-jit"). This is higher than average ability of those choosing the correct answer, .15. The "*" warns us that this contradicts our premise that "higher ability = higher score".

```
SCORE |     DATA     | AVERAGE
VALUE | COUNT     %  | MEASURE
------+--------------+---------
    0 |      2   15  |     .88
    1 |     11   85  |     .15*
```

42. ***Logits:*** if these are not your friends,
Winsteps menu bar
Click on Help
Click on Contents
Click on Special Topics
Click on Logit and Probit.
or http://www.winsteps.com/winman/whatisalogit.htm

Notice the useful table. This is fundamental to interpreting the distances between Rasch measures for dichotomies:

Logit Difference	Probability of Success	Logit Difference	Probability of Success
5.0	99%	-5.0	1%
4.6	99%	-4.6	1%
4.0	98%	-4.0	2%
3.0	95%	-3.0	5%
2.2	90%	-2.2	10%
2.0	88%	-2.0	12%
1.4	80%	-1.4	20%
1.1	75%	-1.1	25%
1.0	73%	-1.0	27%
0.8	70%	-0.8	30%
0.5	62%	-0.5	38%
0.4	60%	-0.4	40%
0.2	55%	-0.2	45%
0.1	52%	-0.1	48%
0.0	50%	-0.0	50%



43. Scroll down to item 55, "al07" almost at the bottom of the subtable. This item is well-behaved. The average ability of "1", the correct answer. (4.31) is higher than for the incorrect answers.

```
-------------------------------------------------------------------------
|ENTRY   DATA  SCORE |     DATA    | AVERAGE  S.E.  OUTF PTMEA|
|NUMBER  CODE  VALUE |  COUNT    % | MEASURE  MEAN  MNSQ CORR.| TOPIC
|--------------------+-------------+--------------------------+------
.....
|                    |             |                          |
|    55  a         0 |      3   30 |     .41   .18    .4  -.50 |al07
|        b         0 |      2   20 |     .49   .54    .5  -.36 |
|        c         0 |      2   20 |    1.41   .66   1.3  -.11 |
|        d         1 |      3   30 |    4.31   .63    .2   .91 |
|        MISSING *** |     20   67#|    -.26   .30        -.54 |
```

44. In this Table, a point-measure (PTMEA) correlation is shown for each Data Code (a,b,c,d). For its correlation, the data code is scored "1" and the other data codes are scored "0". The PTMEA of the correct answer is positive, and the PTMEA of the lowest-ability incorrect answer is the most negative.

```
-------------------------------------------------------------------------
|ENTRY   DATA  SCORE |     DATA    | AVERAGE  S.E.  OUTF PTMEA|
|NUMBER  CODE  VALUE |  COUNT    % | MEASURE  MEAN  MNSQ CORR.| TOPIC
|--------------------+-------------+--------------------------+------
.....
|                    |             |                          |
|    55  a         0 |      3   30 |     .41   .18    .4  -.50 |al07
|        b         0 |      2   20 |     .49   .54    .5  -.36 |
|        c         0 |      2   20 |    1.41   .66   1.3  -.11 |
|        d         1 |      3   30 |    4.31   .63    .2   .91 |
|        MISSING *** |     20   67#|    -.26   .30        -.54 |
```

Why are most correlations negative?

45. For this item the fit to the Rasch model is too good. The OUTF MNSQ, outfit mean-squares, are near to, or below, their expected values, 1.0. According to the Rasch model, the responses to item 55 are somewhat over-predictable. For more about mean-square statistics, please see Appendix 4. Exploring INFIT and OUTFIT Mean-Square Statistics**.**

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	AVERAGE MEASURE	S.E. MEAN	OUTF MNSQ	PTMEA CORR.	TOPIC
.....									
55	a	0	3	30	.41	.18	.4	-.50	al07
	b	0	2	20	.49	.54	.5	-.36	
	c	0	2	20	1.41	.66	1.3	-.11	
	d	1	3	30	4.31	.63	.2	.91	
	MISSING	***	20	67#	-.26	.30		-.54	

Mean-squares less than 1.0 (overfit to the Rasch model) are rarely a problem. They can mean the item is redundant (duplicating other items)

46. Scroll to the item at the top of Table 26.3
The PTMEA (point-measure correlation) between "1" and the abilities is -.26 as we saw in Table 26.1.
We expect the PTMEA for "0" to be **negatively** correlated with ability. Here it is surprisingly positive, 0.26, because higher ability people failed on this item.

SCORE VALUE	DATA COUNT	%	AVERAGE MEASURE	S.E. MEAN	OUTF MNSQ	PTMEA CORR.
0	2	15	.88	1.01	4.0	.26
1	11	85	.15*	.30	1.3	-.26

47. The Outfit mean-square for the two incorrect answers is 4.0. This is a huge underfit of the data to the model. These two responses were exceedingly unpredictable. Is item "im05" a bad item? Or is the problem 2 high performers who may be making careless mistakes? Or what?
Statistics can tell us where to look, but they cannot tell us what to do. We must decide that for ourselves!

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	AVERAGE MEASURE	S.E. MEAN	OUTF MNSQ	PTMEA CORR.	TOPIC
37	d	0	2	15	.88	1.01	4.0	.26	im05 library
	c	1	11	85	.15*	.30	1.3	-.26	
	MISSING	***	17	57*	.56	.55		.08	

Why is the misfit so bad?

48. Why look at **Diagnosis A. Item Polarity** first?
If the scoring key for a multiple-choice test is wrong, or the responses to a rating scale are incorrectly coded, then the point-measure correlations will not be positively correlated with the latent trait. The analysis will not make sense. **We need to verify that the data are correct before going further.**

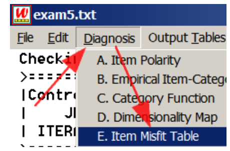
49. **Example:**

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	AVERAGE ABILITY	S.E. MEAN	OUTF MNSQ	PTMEA CORR.	ITEM
12	X	0	3	21	-1.31	1.99	6.7	-.34	1 7+8
	r	1	11	79	.41	.52	1.3	.34	

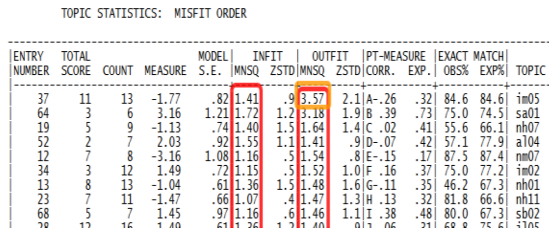
"6.7" is huge misfit. We will definitely need to investigate this misfit .....

50. F. Diagnosis Menu E. Item Misfit Table & Table 10.

51. Re-analyze Exam5.txt (if it is not on your Windows taskbar).
In the Winsteps Analysis window,
On the Winsteps menu bar,
Click on Diagnosis menu
Click on E. Item Misfit Table



52. Table 10.1 displays. It looks the same as Table 26.1, but this Table is ordered by **infit** and **outfit**, not by correlation. The item with the highest mean-square (infit or outfit) is the first listed. This item is the one with the worst fit to the Rasch model. Here it is item "im05" again. It is first because its outfit mean-square of 3.57 is the worst one.



```
TOPIC STATISTICS:  MISFIT ORDER

-----------------------------------------------------------------------------------------
|ENTRY   TOTAL                  MODEL|   INFIT  |  OUTFIT  |PT-MEASURE |EXACT MATCH|
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| TOPIC
|------------------------------------+----------+----------+-----------+-----------+------
|    37     11     13    -1.77    .82|1.41    .9|3.57   2.1|A-.26   .32| 84.6  84.6| im05
|    64      3      6     3.16   1.21|1.72   1.2|3.18   1.9|B .39   .73| 75.0  74.5| sa01
|    19      5      9    -1.13    .74|1.40   1.5|1.64   1.4|C .02   .41| 55.6  66.1| nh07
|    52      2      7     2.03    .92|1.55   1.1|1.41    .9|D-.07   .42| 57.1  77.9| al04
|    12      7      8    -3.16   1.08|1.16    .5|1.54    .8|E-.15   .17| 87.5  87.4| nm07
|    34      3     12     1.49    .72|1.15    .5|1.52   1.0|F .16   .37| 75.0  77.2| im02
|    13      8     13    -1.04    .61|1.36   1.5|1.48   1.6|G-.11   .35| 46.2  67.3| nh01
|    23      7     11    -1.47    .66|1.07    .4|1.47   1.3|H .13   .32| 81.8  66.6| nh11
|    68      5      7     1.45    .97|1.16    .6|1.46   1.1|I .38   .48| 80.0  67.3| sb02
```

53. Interpretation of INFIT and OUTFIT mean-square fit statistics:

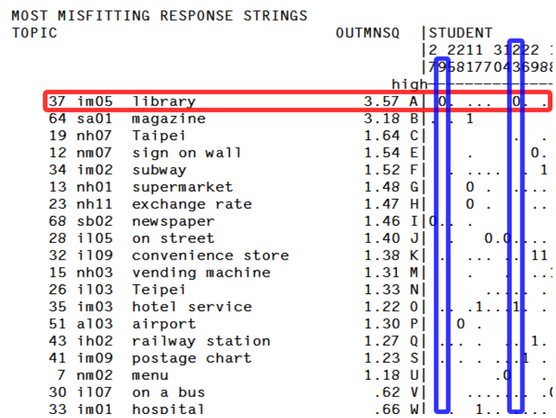
>2.0 Distorts or degrades the measurement system
1.5 - 2.0 Unproductive for construction of measurement, but not degrading
0.5 - 1.5 Productive for measurement
<0.5 Less productive for measurement, but not degrading. May produce misleadingly good reliabilities and separations

http://www.rasch.org/rmt/rmt83b.htm

Reasonable Item Mean-square Ranges for INFIT and OUTFIT

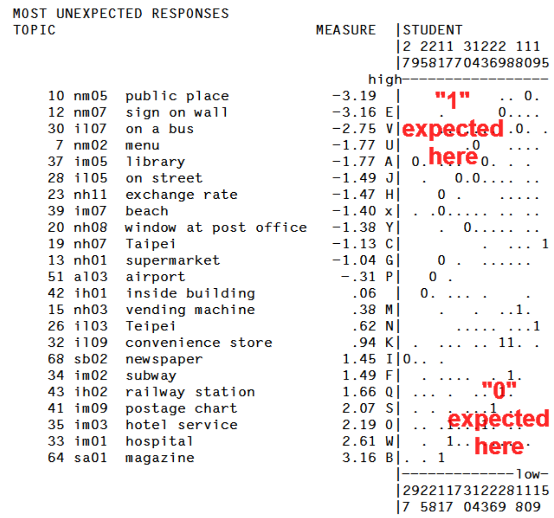
Type of Test	Range
MCQ (High stakes)	0.8 - 1.2
MCQ (Run of the mill)	0.7 - 1.3
Rating scale (survey)	0.6 - 1.4
Clinical observation	0.5 - 1.7
Judged (agreement encouraged)	0.4 - 1.2

54. Scroll down to **Table 10.4**, the "Most Misfitting Response Strings". This subtable shows the responses that caused the large mean-square values in Table 10.1 The values of the surprising responses are shown. "." means "unsurprising response. " " mean "not administered".
Item "im05" is listed first. Its two incorrect answers of "0" are shown. One was by student 9. The other was by student 23 (vertically in the blue boxes).
We can see that there is little evidence that students 9 and 23 were behaving unexpectedly on other items (such as guessing or a response-set), so item *im05* is probably the cause of the misfit.



```
MOST MISFITTING RESPONSE STRINGS
TOPIC                                 OUTMNSQ   |STUDENT
                                                |2 2211 31222
                                                |7958177043698
                                            high-------------
  37 im05  library                       3.57 A| 0. ...  0. .
  64 sa01  magazine                      3.18 B| .. 1
  19 nh07  Taipei                        1.64 C|         .  ..
  12 nm07  sign on wall                  1.54 E|     .      0.
  34 im02  subway                        1.52 F| .. ....  . 1
  13 nh01  supermarket                   1.48 G|     0 .  ....
  23 nh11  exchange rate                 1.47 H|     0 .    ..
  68 sb02  newspaper                     1.46 I|0..  .
  28 il05  on street                     1.40 J| ..    0.0....
  32 il09  convenience store             1.38 K| .   ...  .. 11
  15 nh03  vending machine               1.31 M|     .   .  ..
  26 il03  Teipei                        1.33 N|      ...... .
  35 im03  hotel service                 1.22 O| ..  .1...1.  .
  51 al03  airport                       1.30 P|   0 .
  43 ih02  railway station               1.27 Q| ... .   .. 1.
  41 im09  postage chart                 1.23 S| .  . .  ..1  .
   7 nm02  menu                          1.18 U|        .0    .
  30 il07  on a bus                       .62 V|     .....  .
  33 im01  hospital                       .66 W| ..  1.. .....
```

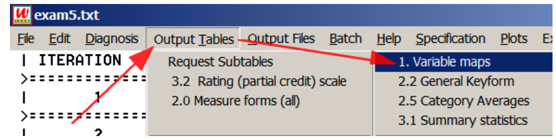
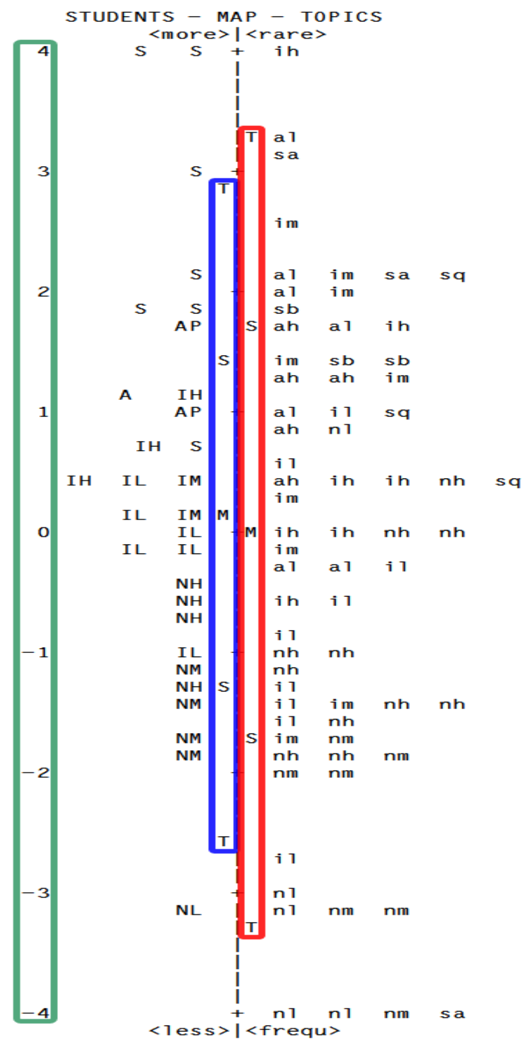
55. Immediately below is Table 10.5.
This shows the most unexpected, the most surprising responses in the data. It is arranged as a Guttman Scalogram.
Left-most column in the data matrix: Student 27 (left-most column) has the highest ability.
Right-most column Student 5 (right-most column) has the lowest ability.
Top row: Item 10 is the easiest item.
Bottom row: Item 64 is the most difficult item.
We expect to see scores of "1" in the top left corner where the most able students meet the easiest items,
and "0" in the bottom right corner where the least able students meet the hardest items.
Since these data are from a computer-adaptive test, items that were too easy for a student, or too difficult were not administered to a student, so the extreme corners contain missing data. Those are shown as blanks.
"." means that the scored response matched Rasch model predictions.
"1" and "0" show unexpected successes and failures.



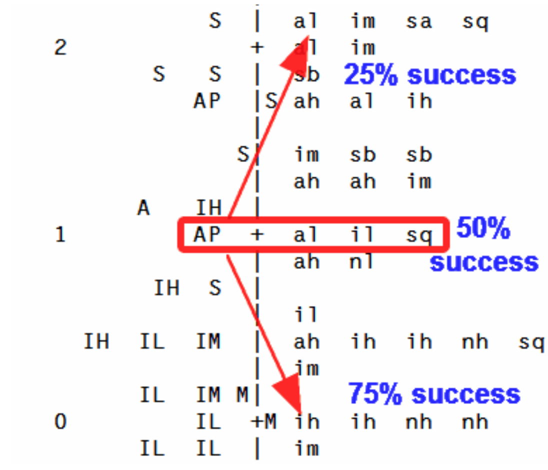
```
MOST UNEXPECTED RESPONSES
TOPIC                                    MEASURE  |STUDENT
                                                  |2 2211 31222 111
                                                  |7958177043698809 5
                                              high-----------------
   10 nm05  public place                   -3.19  |          .. 0.
   12 nm07  sign on wall                   -3.16 E|   .      0....
   30 il07  on a bus                       -2.75 V|          .0. .
    7 nm02  menu                           -1.77 U|       .0   ....
   37 im05  library                        -1.77 A| 0.     0. . .
   28 il05  on street                      -1.49 J|  .    0.0.... ..
   23 nh11  exchange rate                  -1.47 H|    0 .     .....
   39 im07  beach                          -1.40 x| . .0..... .. ..
   20 nh08  window at post office          -1.38 Y|     .  0..... ..
   19 nh07  Taipei                         -1.13 C|          .  ... 1
   13 nh01  supermarket                    -1.04 G|    0 .   ......
   51 al03  airport                         -.31 P|   0 .
   42 ih01  inside building                  .06  |  0. ... .     .
   15 nh03  vending machine                  .38 M|     .   .  ..1.
   26 il03  Teipei                           .62 N|        ..... ...1
   32 il09  convenience store                .94 K| .  ... .. 11. .
   68 sb02  newspaper                       1.45 I|0.. .
   34 im02  subway                          1.49 F|  . ....   . 1.
   43 ih02  railway station                 1.66 Q| ... .   .. .
   41 im09  postage chart                   2.07 S| . .  .   1
   35 im03  hotel service                   2.19 O| ..  .  .  1. ..
   33 im01  hospital                        2.61 W|  .  1..      .
   64 sa01  magazine                        3.16 B|. . 1
                                                  |--------------low-
                                                  |29221173122281115
                                                  |7 5817 04369 809
```


Top left: We can see the places where 0s are observed (but 1's were expected).
Bottom right: the places where 1s are observed (but 0's were expected).
This picture looks usual for this type of data. Nothing is seriously wrong.

56. So what was wrong the item *im05* "library"?
The students answering these questions had visited different parts of the Chinese-speaking world. An expert tells me that the characters for "Library" mean "Bookshop" in some places. This is an ambiguous item. The test would be slightly better without it.

57.	**G. Displaying Rasch Measures**	
58.	The results shown in Table 26 and Table 10 are based on the Rasch measures, our estimates of the ability of the persons and the difficulty of the items. Let's see them: Winsteps Analysis window Click on Output Tables menu Click on Table 1. Variable maps	
59.	Table 1.1 displays in NotePad. It a map of the latent trait, "knowledge of Chinese street signs". On the left, in the green box, is the equal-interval logit scale. One more logit means the same amount extra, exactly as one liter more means the same amount extra. On the left side, between the green and the blue boxes are the students. Only the first letters of their person labels are shown. The letters are demographics. They report the ability level of the student according to a teacher. "S" = Superior. "A" = Advanced. "I" = Intermediate. "N"=Novice. The students are located at their estimated ability measures on the logit scale. In the blue box, "M" = Mean, "S" = 1 S.D. from the mean, "T" = 2 S.D.s from the mean. We can see that there is a strong correlation between the ability levels according to the teachers and the ability levels on the logit scale. This confirms the "predictive validity" of the test. On the right-hand side are the items, located by difficulty. The most difficult item is at the top. The items were labeled with their intended difficulties by the test constructor before the test was administered: "s" = superior. "a" = advanced, "i" = intermediate, "n" = novice. The item identifiers show a strong correlation with the logit scale, confirming the "construct validity" of the test. This item map is very encouraging. It suggests that the instrument is measuring very effectively what it was intended to measure.	

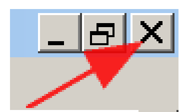
60. When an item is aligned with a person (red box for person AP), then the person is predicted to have a 50% probability of succeeding on the item. So the more difficult items at +3 logits align with the most able student, and the easier items at -3 logits with the least able student. When an item is at the level as a person, then the item is “targeted” on the person.
Do you remember the Table of logits in #42? If an item is 1.1 logits more difficult than the student is able, then the student has a 25% probability of success on the item.
If an item is -1.1 logits less difficult than the student is able, then the student has a 75% probability of success on the item.



61. But where do the logit measures and their standard errors (precision) come from?
We wish we could know the true values for all the measures, but we cannot. We must estimate the measures from the data. Until the arrival of electronic computers, this was difficult, painstaking work. But now it is easy.

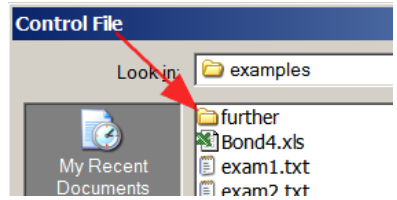
ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.
37	11	13	-1.77	.82
51	3	4	-.31	1.19
12	7	8	-3.16	1.08
13	8	13	-1.04	.61
52	2	7	2.03	.92

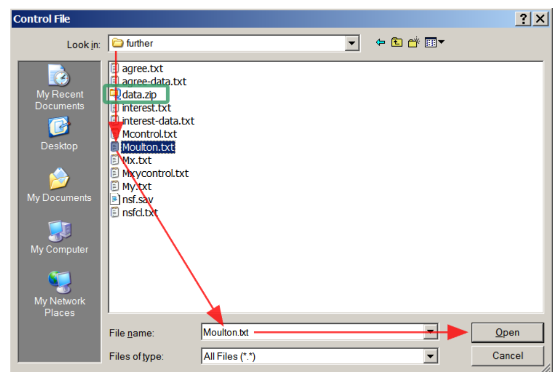
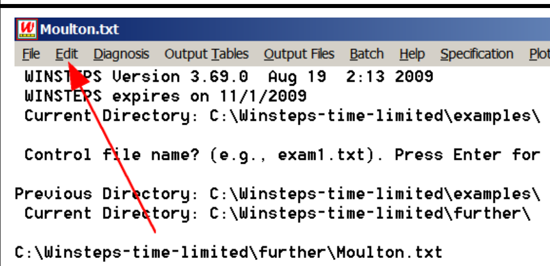
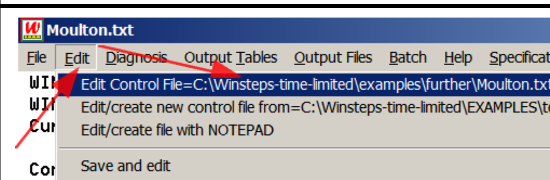
62. Close all windows



<table>
<tr><td>63.</td><td colspan="2">H. Estimating Rasch Measures</td></tr>
<tr><td>64.</td><td>First, we will do a standard analysis of a small dataset.
Launch Winsteps
Winsteps menu bar: Open file
Double-click on folder: “Further”

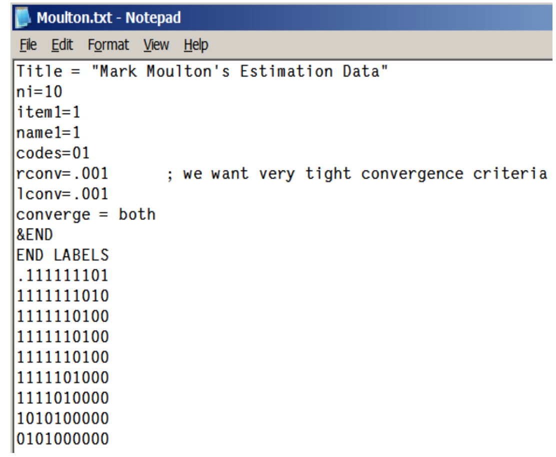
If you do not have folder c:\Winsteps\Further, then please create it. Download www.winsteps.com/a/further-data.zip into your Further folder. This contains a copy of all the data files for this Course. Extract the files from further-data.zip into the Further folder.

Please keep further-data.zip in case you need new copies of any of the files.</td><td>
</td></tr>
<tr><td>65.</td><td>Folder “Further” contains datasets for this Course

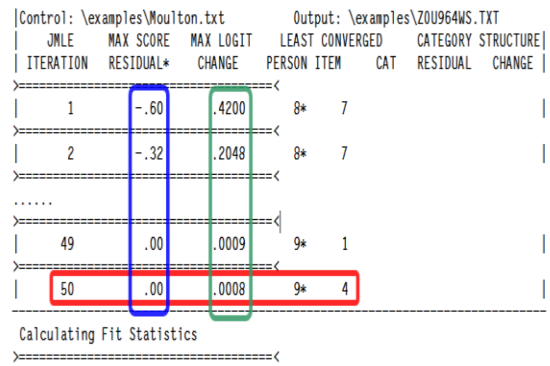
Double-click on the file we want now : Moulton.txt</td><td>
</td></tr>
<tr><td>66.</td><td>Before we analyze this dataset, we want to look at it:
Immediately:
Click on the Winsteps Edit menu</td><td>
</td></tr>
<tr><td>67.</td><td>Click on: Edit control file = ... Moulton.txt</td><td>
</td></tr>
</table>

68. Mark Moulton's small dataset, Moulton.txt, displays in NotePad. The data are dichotomous, 0,1, with NI=10 items and 9 persons. The response by person 1 to item 1 is missing. Overall the data have an almost Guttman pattern.

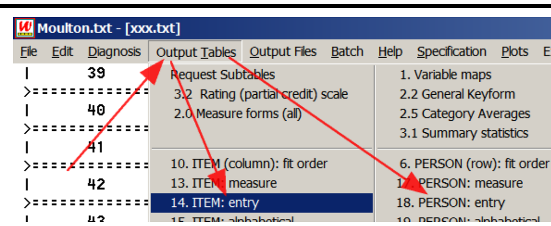
RCONV=, LCONV= and CONVERGE= control the Winsteps Convergence Table.
RCONV=.001, stop estimation when the biggest marginal score residual is less than .001 score-points.
LCONV=.001, stop estimation when the biggest change in any estimated measure is less than .001 logits.
CONVERGE=Both, stop when both RCONV= and LCONV= say "stop!"



69. Click back on the Winsteps analysis window.
Run the Winsteps analysis. (Enter - Enter)
Look at the "Convergence Table" on your screen.
The "Maximum Score Residual" has reduced from .60 score points to .00 score points. This is much less than the smallest difference in raw scores, which is 1 score point.
The "Maximum logit change" has reduced from .4200 logits to .0008. The observable smallest change in the printed computer output is .01 logits.
The estimation is very exact. It has taken 50 iterations.



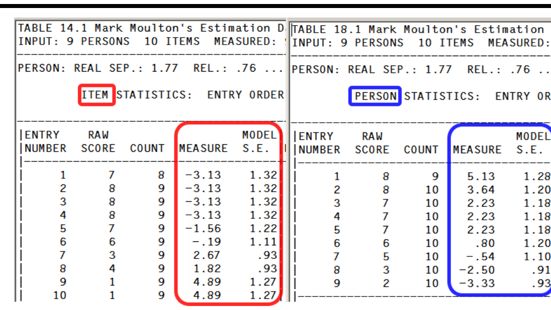
70. We will need the measures shortly:
On the Output Tables menu, display
Table 14. ITEM: entry
Table 18. PERSON: entry



71. The measure tables display in NotePad.

Keep these values for reference.

We are going to see how these numbers are estimated by using Excel to compute the same numbers!



72. **If you want to know exactly how Winsteps works inside, then study this Tutorial carefully.** We will do with Excel the same mathematics that Winsteps does.
If you only want the general idea, then glance through this Tutorial down to #117

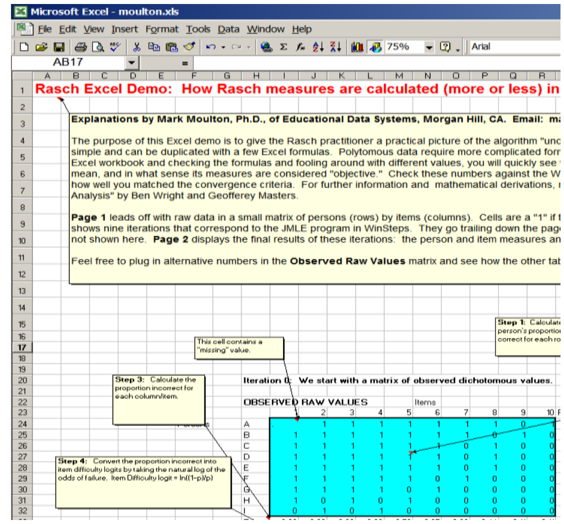
73. Let's experiment with Mark Moulton's Excel spreadsheet:
http://www.winsteps.com/a/Moulton.xls
I have tweaked the version on his website for this Course.

Please "enable macros"
http://www.mdmproofing.com/iym/macros.php

In the top left corner of his worksheet is the data matrix. I copied this one to use in Moulton.txt

There are explanatory comments everywhere! Thank you, Mark Moulton!

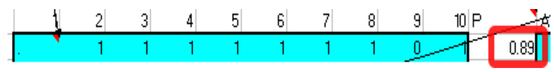
The **"missing" observation** by person 1 to item 1 is treated as "not administered", so the first person will be scored and measured on a 9-item test, not a 10-item test like everyone else.



74. To start the estimation process, we need an initial estimate of each person's ability. The dichotomous Rasch model is:

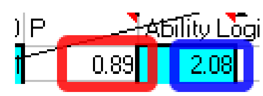
$$\log(P_{ni}/(1-P_{ni})) = B_n - D_i$$

where Pni is the probability that person *n,* of ability B_n, scores "1" on item *i,* of difficulty D_i.
For our initial person estimates, let's imagine that all the items have difficulty, 0, then, for person 1, Pni ≈ 0.89

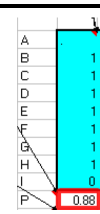


Correct "1" = 8
Incorrect "0" = 1
P("1") = 8/9 = 0.89 (in the red box)
P("0") = 1- P("1") = 1/9 = 0.11

75. Thus an initial Rasch-measure estimate for person 1 is $B_1 \approx \log(P_{1i}/(1-P_{1i})) = \log(0.89/0.11) = 2.08$ logits (blue box).
For more about logarithms, see Appendix 5. What are Logarithms? *and* Appendix 6. Probabilities, Logarithms and the Rasch Model.



76. We are going to estimate the items and persons in the same way, so we go through the same process for each person. For each item, we imagine that each person has ability zero. Then for item 1, $P_{n11} \approx 0.875$

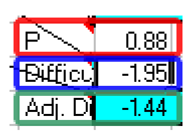


Correct "1" = 7
Incorrect "0" = 1
P("1") = 7/8 = 0.875 (red box)
P("0") = 1-P("1") = 1/8
= 0.125

77. Thus an initial Rasch-measure estimate for item 1 is

$D_1 \approx - \log(P_{n1}/(1-P_{n1}))$ ← the negative is because we are thinking of item difficulty (not item easiness)

$$D_1 \approx \log((1-P_{n1})/P_{n11})$$
$$= \log(0.125/0.875) = -1.95 \text{ logits (blue box).}$$



78. *Everyone notice this!*
Now we must make a decision that happens in every measurement process. Where are we measuring from? What is the local origin? The zero point?
A useful starting-point is the average difficulty of the items. So we will set the average difficulty of the items at zero, so we can measure away from that point.
We don't need to make any adjustment to the persons because the adjustment will happen automatically soon.

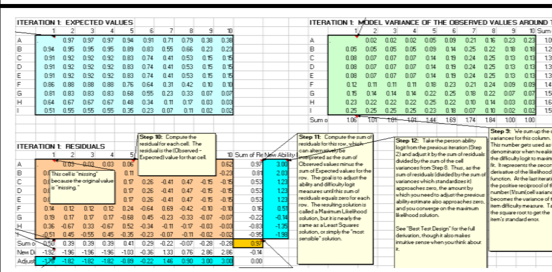
Difficu	-1.95	-2.08	-2.08	-2.08	-1.25	-0.69	0.69	0.22	2.08	2.08	-0.51
Adj. Di	-1.44	-1.57	-1.57	-1.57	-0.75	-0.19	1.20	0.73	2.58	2.58	0.00

-0.51 is the average of the initial item estimates. We subtract this from every estimate (green box), so that the initial item estimates sum to zero.

79. We have now concluded "iteration 0". We have looked at the data and obtained our initial estimates.
Winsteps uses the *PROX* (Normal Approximation) algorithm to do the same thing.
Initial estimates of 0.0 could also be used!

```
                          CONVERGENCE TABLE
-Control: \examples\Moulton.txt         Output: \examples\ZOU964WS.TXT
|    PROX          ACTIVE COUNT        EXTREME 5 RANGE      MAX LOGIT CHANGE  |
| ITERATION    PERSONS ITEMS   CATS     PERSONS ITEMS       MEASURES  STRUCTURE|
>=====================================<
|        1          9    10      2      1.40    2.76          2.0794           |
>=====================================<
|        2          9    10      2      1.99    3.29           .9990           |
>=====================================<
|        3          9    10      2      2.18    3.76           .4156           |
```

80. We begin JMLE (Joint Maximum Likelihood Estimation), which is the method used in Winsteps. There are other methods, each with strengths and weaknesses.
The spreadsheet shows 3 components for Iteration 1: Expected values (light blue rectangle), Residuals (orange rectangle), and Model variances (light green rectangle).



81. Here is the JMLE Newton-Raphson estimation equation for person n,

$$B'_n = B_n + \Sigma (X_{ni} - P_{ni}) / \Sigma (W_{ni})$$

$B'_n = B_n$ + (sum of orange values)/(sum of green values)
B'_n is a better estimate of B_n
Σ means sum-together everything involving person *n*.
X_{ni} are the "observed raw score values" in the **darker blue rectangle** at iteration 0.
P_{ni} are the "expected values" based on the current values of all the B_n and D_i. They are in the **blue rectangle**.
$(X_{ni} - P_{ni})$ are the residuals. They are in the **orange rectangle.**
W_{ni} is the model variance of the observed value X_{ni} around the expected value P_{ni}. They are in the **light green rectangle.**

The Rasch model for computing the probabilities of success:

$$P_{ni} = \frac{e^{(B_n - D_i)}}{1 + e^{(B_n - D_i)}}$$

For the dichotomous model, the model variance is the binomial variance:

$$W_{ni} = P_{ni} * (1 - P_{ni})$$

Notice that $P_{ni} * (1 - P_{ni})$ is biggest when P = 0.5, so that P*(1 - P) = 0.5*0.5 = 0.25 and smallest when P = 0 or 1, so that P*(1 - P) = 0*1 = 0.0

82. For item i, it is almost the same thing:

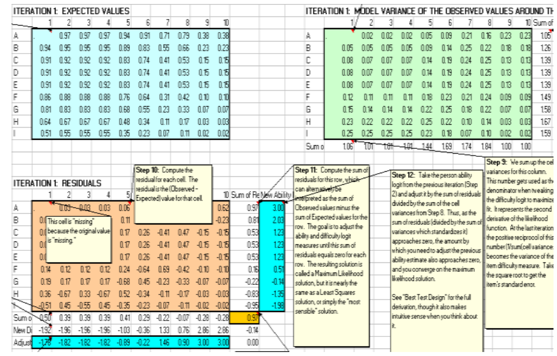
$$D'_i = D_i - \Sigma (X_{ni} - P_{ni}) / \Sigma (W_{ni})$$

Σ means sum-together everything involving item *i*.

$D'_i = D_i$ - (sum of orange values)/(sum of green values)
Do you see the "-" minus sign? This is because we are measuring item "difficulty", not item "easiness".

But we must not forget to maintain the local origin, so there is the additional constraint that: $\Sigma (D'_i) = 0$ for all *i*.

83. Here is the process in Iteration 1. We use the relationships in #81
1. In the light blue rectangle, the initial estimates in iteration 0 are used to compute the probabilities of success for every observation in the data using the Rasch dichotomous model
2. In the light green rectangle, the model variances corresponding to each observation are computed
3. In the orange rectangle, the residuals are computed for each observation.

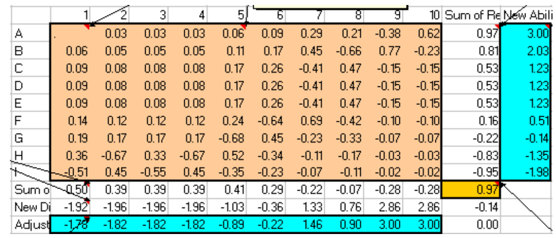


84. At the right-side and bottom-side of the orange rectangle, the residuals are summed to become the marginal residuals. The "New Abilities" (blue column to the right) are the B'_n and the "New Difficulties" (white row below) are the D'_i.

$B'_n = B_n$ + (sum of orange values)/(sum of green values)

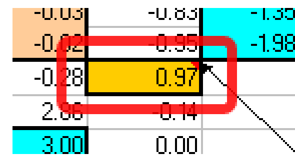
$D'_i = D_i$ - (sum of orange values)/(sum of green values)

Then the "Adjusted Difficulties" (blue row below) sum to zero to maintain the local origin.

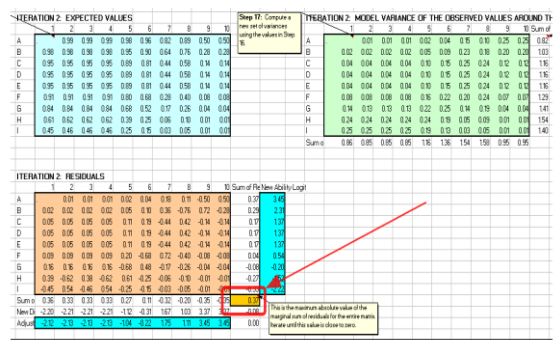


	1	2	3	4	5	6	7	8	9	10	Sum of Re	New Abili
A	.	0.03	0.03	0.03	0.06	0.09	0.29	0.21	-0.38	0.62	0.97	3.00
B	0.06	0.05	0.05	0.05	0.11	0.17	0.45	-0.66	0.77	-0.23	0.81	2.03
C	0.09	0.08	0.08	0.08	0.17	0.26	-0.41	0.47	-0.15	-0.15	0.53	1.23
D	0.09	0.08	0.08	0.08	0.17	0.26	-0.41	0.47	-0.15	-0.15	0.53	1.23
E	0.09	0.08	0.08	0.08	0.17	0.26	-0.41	0.47	-0.15	-0.15	0.53	1.23
F	0.14	0.12	0.12	0.12	0.24	-0.64	0.69	-0.42	-0.10	-0.10	0.16	0.51
G	0.19	0.17	0.17	0.17	-0.68	0.45	-0.23	-0.33	-0.07	-0.07	-0.22	-0.14
H	0.36	-0.67	0.33	-0.67	0.52	-0.34	-0.11	-0.17	-0.03	-0.03	-0.83	-1.35
I	-0.51	0.45	-0.55	0.45	-0.35	-0.23	-0.07	-0.11	-0.02	-0.02	-0.95	-1.98
Sum o	-0.50	0.39	0.39	0.39	0.41	0.29	-0.22	-0.07	-0.28	-0.28	0.97	
New Di	-1.92	-1.96	-1.96	-1.96	-1.03	-0.36	1.33	0.76	2.86	2.86	-0.14	
Adjust	-1.78	-1.82	-1.82	-1.82	-0.89	-0.22	1.46	0.90	3.00	3.00	0.00	

85. Are these "better estimates" good enough?
The yellow cell tells us. This is the biggest marginal-residual. Here it is 0.97, almost a score-point. We would like this to be much smaller, certainly less than 0.5 score-points, the smallest observable difference. So, onward to iteration 2.

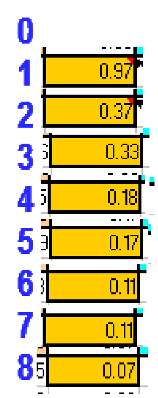


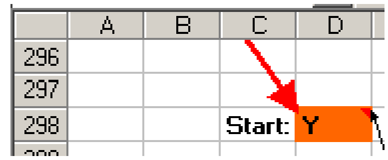
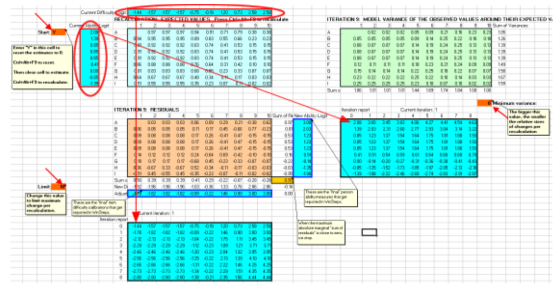
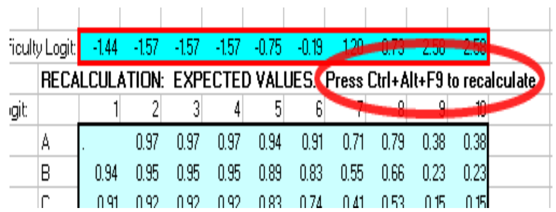
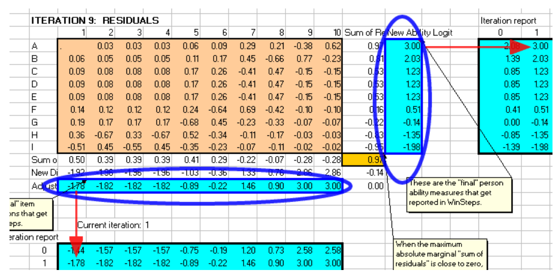
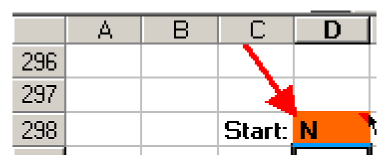
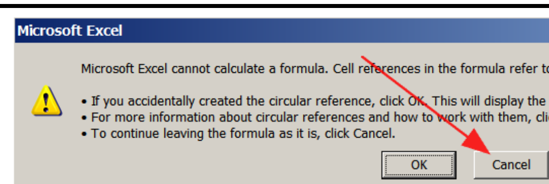
86. Iteration 2 performs the same computations as iteration 1 in #83 using the better estimates of ability and difficulty from iteration 1. After this iteration, the yellow box shows that the biggest marginal residual is 0.37 score-points. It has reduced from 0.97. We are converging toward the maximum likelihood estimates. When the value in the yellow box is very near to zero, then estimation process will have *converged,* and the estimates of ability and difficulty will be as good as these data can provide.



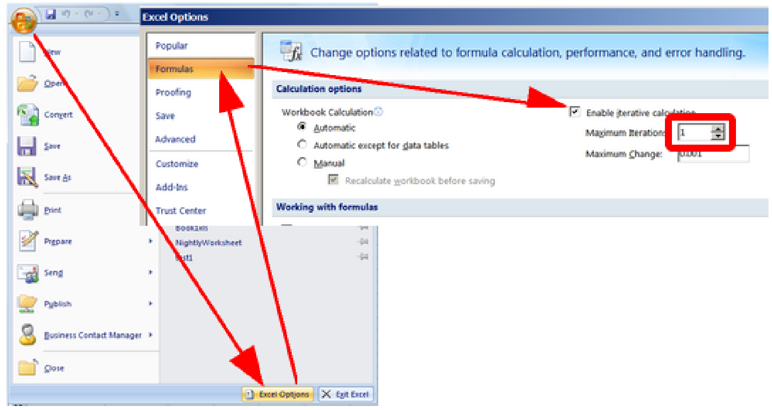
87. Continue down Mark Moulton's spreadsheet. You will see iterations 3, 4, 5, 6, 7, 8.

At iteration 8, the biggest marginal residual (the largest difference between the observed raw score and the expected raw score for any person or item) is 0.07 score points. This is less than 1/10th of a score point, far smaller than can be observed in a data set.
So we can say that the estimates have converged, and the "better estimates" from iteration 8 are exact enough measures for practical purposes.

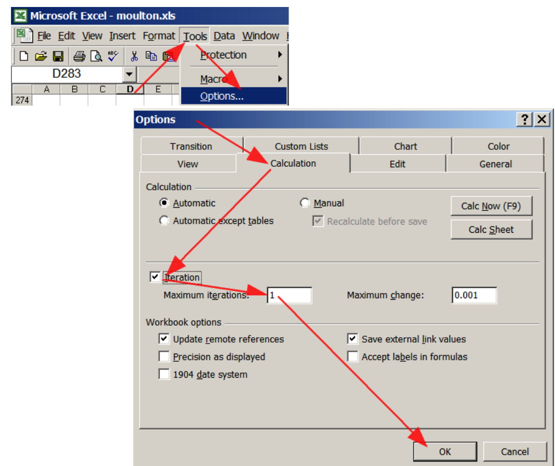


88.	**I.** ***Now for some "estimation" fun!***	
89.	Look at orange cell D298, if it does not have "Y" in it: Click on orange cell D298 Type "Y" Press Enter	
90.	The last Excel computation on this worksheet is not the same as the others. It is a mini-Winsteps! It will compute one iteration after another. Look at the values in the box outlined in red. These are the initial estimates of the measures from the top of the worksheet. You can also see them in the two blue boxes where there are the values from the 8 iterations above.	
91.	Press: Ctrl+Alt+F9 at the same time (or perhaps F9 or shift+F9). This forces the spreadsheet to recalculate, so that all the numbers are at their initial values. The keyboard combination on the Mac is: **Command** ⌘ + = www.wallst-training.com/WST_Excel_Shortcuts_Mac.pdf	
92.	The estimates for iteration 1 are now in the boxes outlined in blue. You can compare them with our previous values in the bigger blue boxes.	
93.	Now for iteration 2: Click on orange cell D298 with "Y" in it Type "N" Press Enter	
94.	Oops! If you see the "Circular Reference" warning, click "Cancel"	

95. *Excel 2007:*
Click the Office Button
Click Excel Options
Click Formulas Tab
Check "Enable iterative calc..."
Maximum iterations: 1
Click on "OK"

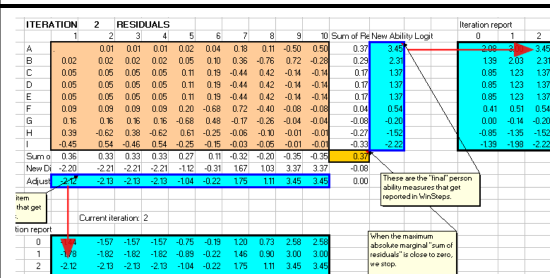


96. *Excel other versions:*
Excel menu bar:
Click on "Tools"
Click on "Options"
Click on "Calculation" tab
Check "Iterations"
"Maximum iterations", type "1"
Click on "OK"



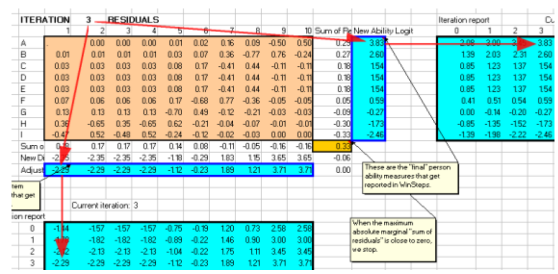
97. The estimates in the red-outlined boxes should now be for iteration 1, and in the blue-outlined boxes for iteration 2.

Excel has performed a Newton-Raphson iteration.



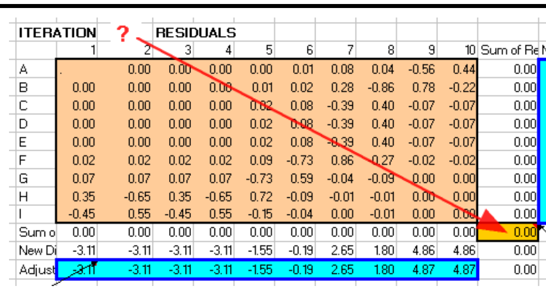
98. Now for iteration 3:
Press Ctrl+Alt+F9

Do you see the iteration 3 values in the blue-outlined boxes.



99. *Now you are on your own!*
Keep on pressing Ctrl+Alt+F9
The iterations continue. This is what Winsteps does!

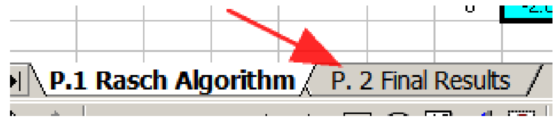
How many iterations does it take you to make the biggest marginal residual display as .00?



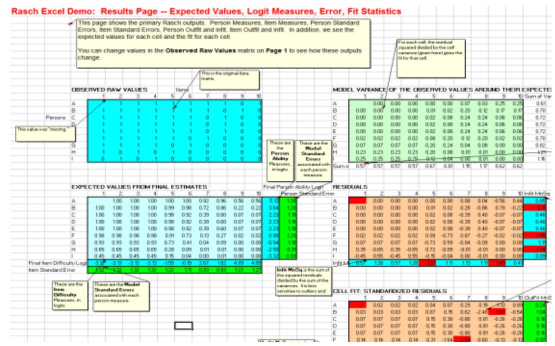
100.	Let's compare results: Excel (on the left-side) and Winsteps (from above). Here are the person measures. Almost the same - but not exactly. What has happened? The convergence criteria! We stopped at .00 in the yellow box, but Winsteps kept going. Keep pressing Ctrl+Alt+F9 until the Excel person measures (on the right-hand side) match Winsteps. *How many iterations does it take?*	5.11 5.13 5.13 3.62 3.64 3.64 2.21 2.23 2.23 2.21 2.23 2.23 2.21 2.23 2.23 0.80 .80 0.80 -0.54 -.54 -0.54 -2.48 -2.50 -2.50 -3.31 -3.33 -3.33
101.	Did you notice that Winsteps takes 50 iterations in its Convergence Table, but Excel a lot less? Is Excel better than Winsteps? *Yes and No!*	>==================================== | 50 .00 .0008 ------------------------------------ Calculating Fit Statistics ------------------------------------
102.	This dataset has a neat structure, so Newton-Raphson iteration works very smoothly. But most datasets are much messier. So Winsteps has to be prepared for anything. Consequently Winsteps uses a more robust estimation method called "logistic curve-fitting", but the final estimates are the same JMLE estimates. We have JMLE estimates when "observed raw scores = expected raw scores" for every person, item, etc.	.111111101 1111111010 1111110100 1111110100 1111110100 1111101000 1111010000 1010100000 0101000000

103. **J. Computing Fit Statistics with Excel**

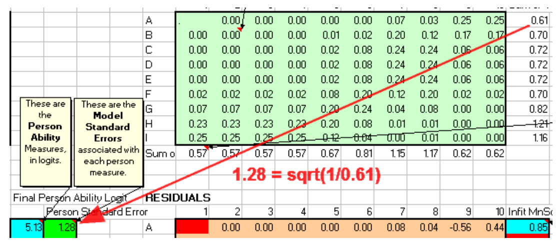
104. We now have a very exact set of measures of person ability and item difficulty. Let's see the fit statistics.
In the spreadsheet, click on the tab: "P.2 Final Results"



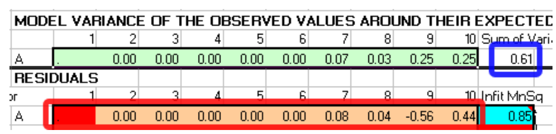
105. Most of the "Final Results" should look familiar.
Dark blue box: original data.
Light blue box: expected values, with the final estimates on the right-hand-side and bottom of the box.
Light green box: model variances
Orange box: the residuals.
There is a new orange box: the "standardized residuals".
The results also include: "standard errors", "infit mean-square" and "outfit mean-square".



106. Dark green boxes: **Standard Errors** (S.E.s) of the measures (their precision) are the inverse square-roots of their accumulated model variances. So we can see that:
1. the more responses a person makes, the more model variances there are to sum, and the smaller the S.E.s.
2. The better the targeting of the items on the person, the bigger the variance terms, and the smaller the S.E.s.

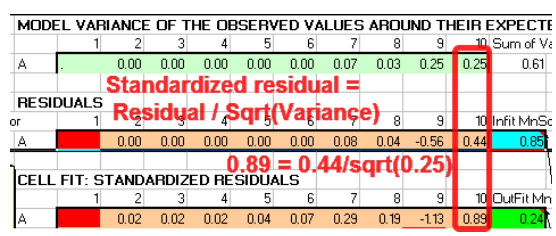


107. Blue boxes around big orange box: **Infit Mean-Square** is the ratio of the observed residual variance for a person (or item) and the model variance for the same person (or item). Its expected value is 1.0 (when the observed variance = model variance).

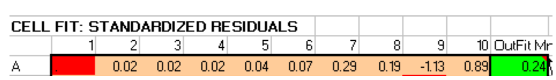


108. The residual variance is the sum of the squared residuals in the residual box
$= 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0.08^2 + 0.04^2 + (-0.56)^2 + 0.44^2 = 0.51$
The model variance (blue box) = 0.61.
The Infit mean-square = 0.51/0.61 ≈ 0.85 (the difference is rounding when displaying the numbers)

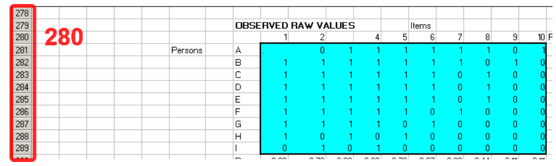
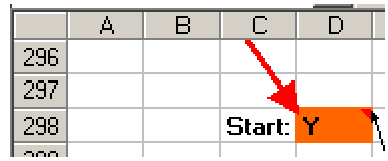
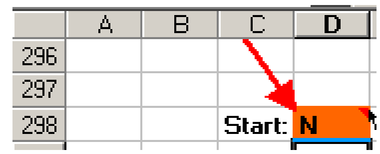
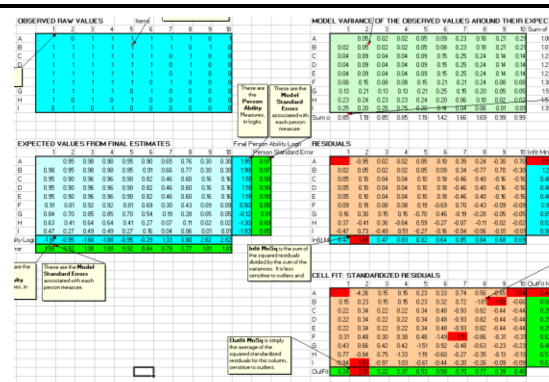
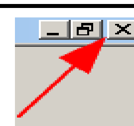
109. **Standardized Residuals** tell us how surprising each observation is. They are the Residuals / Sqrt (Model Variance). *Which is the most unexpected observation?*
The Standardized Residuals are modeled to approximate a unit-normal distribution (Appendix 7. Unit Normal Deviates), so their values can be interpreted as unit-normal-deviates. The expected value of a standardized residual is 0.0, and the expected value of a squared standardized residual is 1.0.



110. **Outfit Mean-Square** is the average of the squared standardized-residuals, so its expected value is their expected value, 1.0.



111. Outfit mean-square
$= (0.02^2 + 0.02^2 + 0.02^2 + 0.04^2 + 0.07^2 + 0.29^2 + 0.19^2 + (-1.13)^2 + 0.89^2) /9 = 2.19/9 = 0.24$

112.	***K.*** ***More fun!***	
113.	Now, back to the first worksheet. Change some 0s and 1s in the **data matrix at row 280**. If you get lost, you can restore the original Moulton.xls from [*www.winsteps.com/a/further-data.zip*](http://www.winsteps.com/a/further-data.zip)	
114.	Reset the bottom computation by entering "Y" in cell D298 Press Ctrl+Alt+F9 You have started your own estimation	
115.	Enter "N" in cell D298 Press Ctrl+Alt+F9 for your own iterations, until the estimates have converged. *Watch the yellow cell!*	
116.	*Can you guess what the fit statistics for your data will be?* Look at P.2 to find out if you are right!	
117.		

1.	Appendix 1. Logit differences and dichotomous probabilities			
2.	Logit Difference	Probability of Success	Logit Difference	Probability of Success
	5.0	99%	-5.0	1%
	4.6	99%	-4.6	1%
	4.0	98%	-4.0	2%
	3.0	95%	-3.0	5%
	2.2	90%	-2.2	10%
	2.0	88%	-2.0	12%
	1.4	80%	-1.4	20%
	1.1	75%	-1.1	25%
	1.0	73%	-1.0	27%
	0.8	70%	-0.8	30%
	0.5	62%	-0.5	38%
	0.4	60%	-0.4	40%
	0.2	55%	-0.2	45%
	0.1	52%	-0.1	48%
	0.0	50%	-0.0	50%

1.	Appendix 2. Displaying file name suffixes
2.	**Windows 7 and Vista:** 1. click the Start button , 2. click Control Panel, 3. click Appearance and Personalization, 4. click Folder Options. 5. click the View tab, 6. under Advanced settings 7. click the Hide extensions for known file types check box to clear it 8. click OK.
3.	**Windows XP:** 1. open any folder, such as "My Computer" or "My Documents." 2. Tools menu 3. Folder Options 4. View tab 5. Advanced settings 6. Hide extensions for known file types - uncheck this box 7. OK

1.

Appendix 3. Changing the text appearance in NotePad

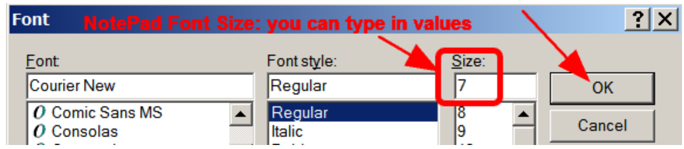
2.

Oops! A Table may display too big or ragged or wrapped. We need to display this text in a fixed-space font, such as Courier New, and also a smaller font so everything fits in the window.

```
-----------------------------------------------------------
|ENTRY   RAW              MODEL|  INFIT  | O
|
|NUMBER  SCORE  COUNT  MEASURE  S.E.
OBS%  EXP%| ACT                |
|------------------------------+----------+----------+-
```

3.

On the NotePad menu bar,
Alt+O or click on Format pull-down menu
Alt+F to change the Font
Font: Courier New
Font style: Regular
Font size: 7 (You can type in values not listed)
Click on OK



4.

The Table now displays neatly

Alter the Font size if the Table is too big or too small.

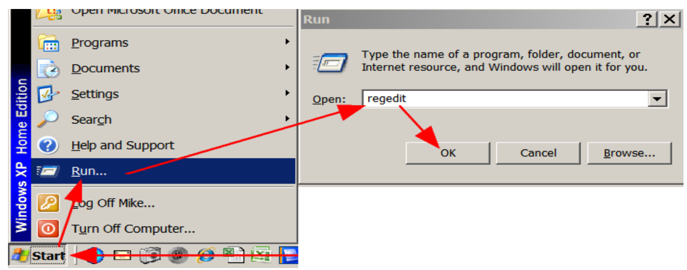
```
------------------------------------------------------------------------------------------------
|ENTRY    RAW                        MODEL|   INFIT  |  OUTFIT  |PTMEA|EXACT MATCH|                                  |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.| OBS%  EXP%| ACT                              |
|------------------------------------+----------+----------+-----+-----------+----------------------------------|
|    23     40     74     2.18     .21|2.41    6.3|4.11    9.0|  .00| 40.5  65.0| Watch a rat                      |
|     5     35     74     2.42     .22|2.30    5.6|3.62    7.3|  .05| 52.7  68.1| Find bottles and cans            |
|    18    143     74    -3.15     .47|1.50    1.2|1.23     .5|  .14| 94.6  93.4| Go on picnic                     |
|    19    139     74    -2.48     .36|1.08     .4|1.10     .4|  .30| 90.5  88.0| Go to zoo                        |
|    20     48     74     1.83     .20|1.33    2.0|1.82    3.7|  .42| 54.1  61.1| Watch bugs                       |
|    12    135     74    -2.04     .31| .70   -1.2| .51   -1.0|  .45| 86.5  83.7| Go to museum                     |
|    13    125     74    -1.29     .25|1.22    1.1| .94     .0|  .47| 78.4  74.3| Grow garden                      |
|    10    128     74    -1.49     .26| .78   -1.1| .57   -1.1|  .50| 78.4  77.1| Listen to bird sing              |
|     8     52     74     1.67     .20|1.10     .7|1.21    1.2|  .51| 47.3  59.0| Look in sidewalk cracks          |
|    16     81     74      .60     .19| .97    -.2| .95    -.3|  .51| 60.8  57.3| Make a map                       |
```

5.

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

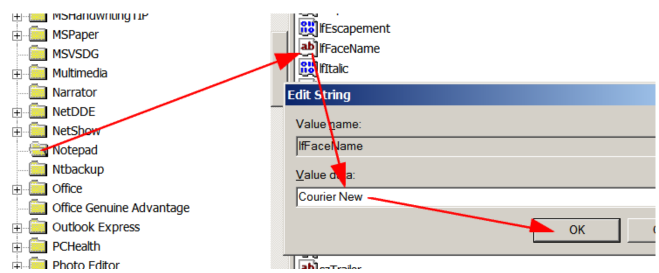
6.

Windows "Start"
Click on "Run"
Type in "regedit"
Click on "OK"



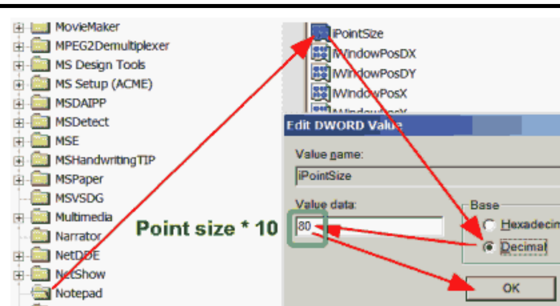
7.

Registry Editor:
Click on the + in front of "HKEY_CURRENT_USER"
Click on the + in front of "Software"
Click on the + in front of "Microsoft"
Click on "Notepad"
For the type face:
    Double-click on "lfFaceName"
    Type in "Courier New" (or "Letter Gothic Line")
    Click on "OK"



8.

For the font size:
Double-click on "iPointSize"
Click on "Decimal"
Type in 80 (for point-size 8 multiplied by 10)
Click on "OK"
Close registry
Click on top right ☒

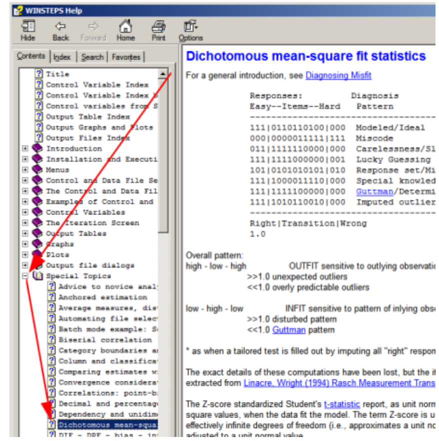


1.

Appendix 4. Exploring INFIT and OUTFIT Mean-Square Statistics

2.

Let's look at some patterns of misfit we would want to identify and diagnose.
To see them:
On the Winsteps Menu Bar
Click on Help
Click on Contents
In the Contents panel,
Click on Special Topics
Click on Dichotomous Mean-Square Fit Statistics
or http://www.winsteps.com/winman/dichotomous.htm



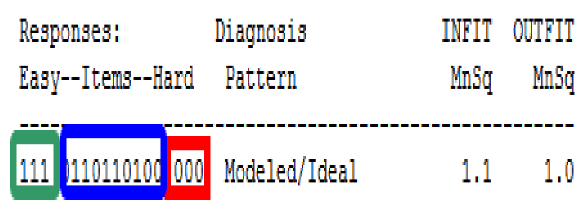
3.

Here they are:
In this Table, we imagine that the items have been arranged from easy to hard and have been administered in ascending order of difficulty as a multiple-choice (MCQ) test with a time limit. A type of test familiar to school children in the USA. The items are scored "1" and "0"

```
Responses:          Diagnosis              INFIT  OUTFIT
Easy--Items--Hard   Pattern                 MnSq    MnSq
----------------------------------------------------------
111|0110110100|000  Modeled/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                 Expectation: 1.0
```

4.

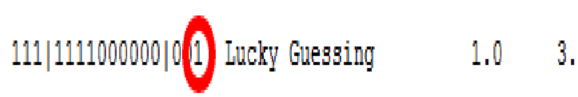
How do we expect a child of medium ability to respond? We expect the child to get the easy items almost always correct (green box) and the hard items almost always incorrect (red box). In between, is a "transition" zone where the item difficulties are targeted on the child's ability. Here we expect the child to succeed on some items and fail on others (blue box). If an item's difficulty exactly corresponds to the child's ability, then the child's probability of success is 0.5, and we expect success or failure (1 or 0) equally.
This is the response pattern predicted by the Rasch model. We can see that this response pattern produces INFIT and OUTFIT mean-square (MnSq) statistics near 1.0.

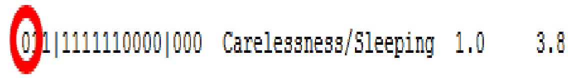
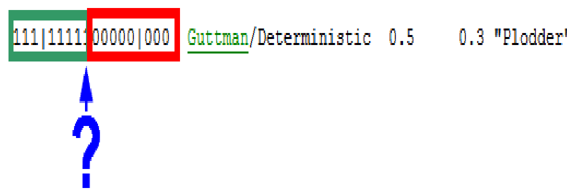
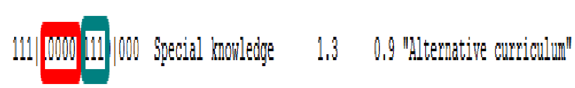


INFIT MnSq = "information-weighted" or "inlier-sensitive" mean-square fit statistic.
OUTFIT MnSq = "outlier-sensitive" mean-square fit statistic.
A mean-square statistic is a chi-square statistic divided by its degrees of freedom.
The expected value of a mean-square is 1.0

5.

What about guessing - a common problem on MCQ items? The only guessing that is of great concern is when the guess is lucky - a correct answer to a hard item (red circle). This is an unexpectedly correct response - an outlier.
The OUTFIT statistic is sensitive to **outliers**. Its value is now 3.8, much bigger than its baseline value of 1.0.
INFIT statistics are relatively insensitive to outliers. Its value is the baseline 1.0.



6.	And careless mistakes? These are incorrect answers to easy items (red circle). Again this is an unexpected response - an outlier. So the OUTFIT statistic is again high, 3.8, but the INFIT statistic is relatively unchanged at its baseline value of 1.0. Values of fit statistics **greater than 1.0 are termed "underfit" - the responses are too unpredictable** from the Rasch model's perspective.	
7.	Let's think about a different behavior: the plodder. He works slowly and carefully through each item, double-checking his answers. He succeeds on every item (green box). But then time runs out. He is automatically scored incorrect (red box) on all the remaining harder items. If we know the cut-point (blue arrow) we can predict all the child's responses exactly. Psychometrician Louis Guttman proclaimed that this is the ideal response pattern. The child's responses seem to tell us that his ability is exactly at the blue arrow. *But, of course, it is not!* He hasn't even reached the start of his "transition zone" predicted by the Rasch model. What we do see is a response pattern that is too predictable. There is no area of uncertainty in it. Accordingly both the INFIT mean-square of 0.5 and the OUTFIT of 0.3 are less than 1.0. This is termed "overfit". The responses are too predictable from the Rasch-model perspective. In fact, if the entire data set has a Guttman pattern, then we can exactly order all the persons and items on the latent trait, but we have no information on which to base linear measurement of the distance between different persons and or different items.	
8.	Let's imagine this situation: most schools teach addition ➔ subtraction ➔ multiplication ➔ division, but my school teaches addition ➔ multiplication ➔ subtraction ➔ division. So when I take a standard arithmetic test, I succeed on the addition items. Fail on the subtraction items (red box). Succeed on the multiplication items (green box) and fail on the division items. Compare this response string to the others. We are not surprised by a failure or two on the subtraction items, or by a success or two on the multiplication items. It is the overall pattern that is surprising. This is what INFIT identifies. So the INFIT mean-square is 1.3, greater than 1.0, indicating underfit, "too much unpredictability". But the OUTFIT mean-square is 0.9, less than 1.0, indicating overfit, my performance on the easy "addition" items and hard "division" items is slightly too predictable.	

9. So what values of the mean-square statistics cause us real concern? Here is my summary table from Winsteps Help “Special Topic” “Misfit Diagnosis ...”

Here’s a story:
When the mean-square value is around 1.0, we are hearing music! The measurement is **accurate**
When the mean-square value is less than 1.0, the music is becoming quieter, becoming muted. When the mean-square is less than 0.5, the item is providing only have the music volume (technically “statistical information”) that it should. But mutedness does not cause any real problems. Muted items aren’t efficient. The measurement is less accurate.

When the mean-squares go above 1.0, the music level stays constant, but now there is other noise: rumbles, clunks, pings, etc. When the mean-square gets above 2.0, then the noise is louder than the music and starting to drown it out. The measures (though still forced to be linear) are becoming distorted relative to the response strings. So **it is mean-square values greater than 2.0 that are of greatest concern. The measurement is inaccurate.**

*But be alert, the **explosion** caused by only one lucky guess can send a mean-square statistic above 2.0. Eliminate the lucky guess from the data set, and harmony will reign!*

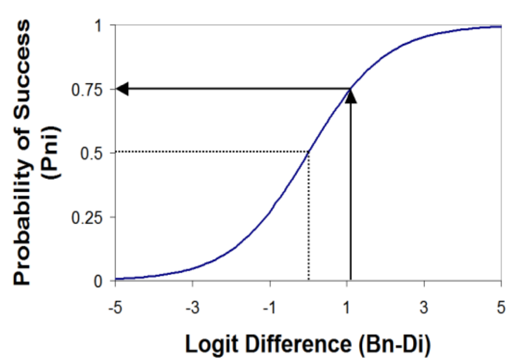
Interpretation of parameter-level mean-square fit statistics:	
>2.0	**Distorts or degrades the measurement system.**
1.5 - 2.0	Unproductive for construction of measurement, but not degrading.
0.5 - 1.5	Productive for measurement.
<0.5	Less productive for measurement, but not degrading. May produce misleadingly good reliabilities and separations.

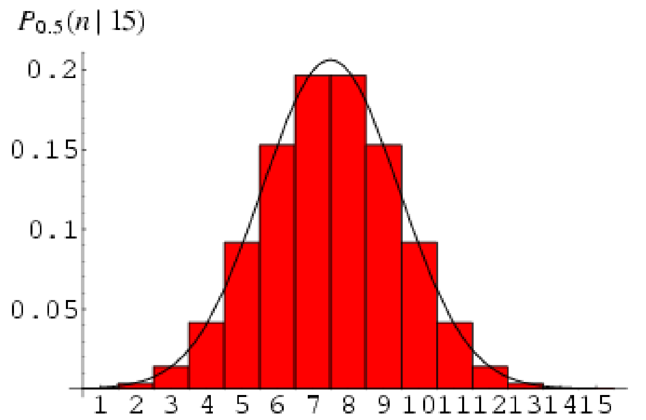
Here are other suggested values from http://www.rasch.org/rmt/rmt83b.htm

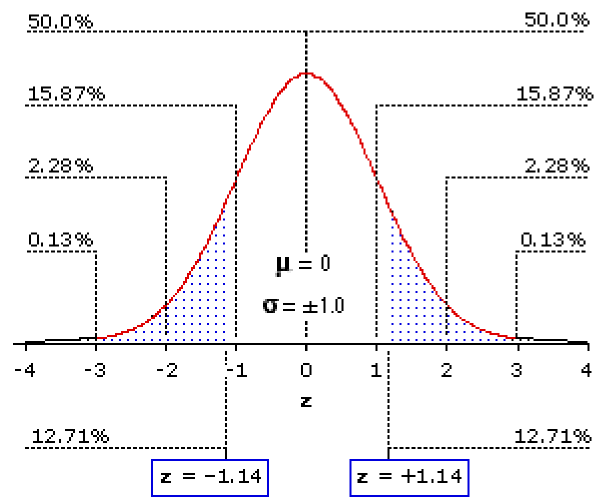
Reasonable Item Mean-square Ranges for INFIT and OUTFIT	
Type of Test	Range
MCQ (High stakes)	0.8 - 1.2
MCQ (Run of the mill)	0.7 - 1.3
Rating scale (survey)	0.6 - 1.4
Clinical observation	0.5 - 1.7
Judged (agreement encouraged)	0.4 - 1.2

1.	**Appendix 5. What are Logarithms?**	
2.	Let's start with squares, 2 x 2 = 4 and cubes, 2 x 2 x 2 = 8 We can see that it takes two 2's to make 4 so let's write them 2^2 and it takes three 2's to make 8 so let's write them 2^3. Then what about $2^{2+3} = 2^5$?	$2^2 = 2 \times 2 = 4$ $2^3 = 2 \times 2 \times 2 = 8$ $2^{2+3} = 2^5 = 2 \times 2 \times 2 \times 2 \times 2 = 32 = 4 \times 8 = 2^2 \times 2^3$
3.	The superscripts ^{2} and ^{3} are called "powers", and we've discovered that multiplying numbers is the same as adding powers. We can do this with non-integers:	$1.4142 \times 1.4142 = 2 = 2^1 = 2^{1/2} * 2^{1/2}$ $1.4142 = 2^{1/2}$ = square-root(2)
4.	And powers work for negative numbers	$2^1 = 2 = 4 / 2 = 2^2 / 2^1 = 2^2 * 2^{-1} = 2^{2-1}$
5.	So, we have the general rule of powers of 2:	$2^{x+y} = 2^x * 2^y$
6.	This works exactly the same way if we change the "base" value from 2 to 10.	$X = 10^x$ $Y = 10^y$ $X * Y = 10^{x+y}$ $10^{x+y} = 10^x * 10^y$
7.	Now x and y are called "exponents" and "10" the base. So this is an "exponential" form.	$X * Y = 10^x * 10^y = 10^{x+y}$
8.	We can rewrite this "exponential" form into "logarithmic" form. "Log" is short for "Logarithm".	$\log_{10}(X) = \log_{10}(10^x) = x$ $\log_{10}(X) = \log_{10}(10^y) = y$ $\log_{10}(X*Y) = \log_{10}(10^{x+y}) = x+y$
9.	This is very useful. We can transform a multiplication X*Y into and addition log(X) + log(Y). This saves a huge amount of effort when the multiplication is done by hand, and was why logarithms were invented around 1617.	$\log_{10}(10^x) + \log_{10}(10^y) = x+y = \log_{10}(10^{x+y})$ $\log_{10}(X) + \log_{10}(Y) = x+y = \log_{10}(X*Y)$
10.	In Rasch work, we use a special base, called "e". This has the value 2.718.... And the logarithms are then called "natural" or "Napierian" logarithms.	$2.3026 * \log_{10}(X) = \mathbf{\log_e(X)}$ $\log_{10}(X) = 0.4343 * \mathbf{\log_e(X)}$
11.	Once we alert the reader what base we are using, "e" from here on, we can omit it.	log(X) + log(Y) = log(X*Y)
12.	Here are some important facts about logarithms: the logarithm of 1 is zero	log(1) + log(1) = log(1*1) = log(1) so log(1) = 0 this is the same as saying $x^0 = 1$
13.	the logarithm of plus infinity is infinity	$\log(\infty) = \infty$
14.	the logarithm of zero is minus infinity	$\log(0) = -\infty$
15.	the logarithm to the base "e" of "e" is 1	$\log_e(e) = 1$
16.	the logarithms of negative numbers don't exist	log (-1) = Oops!

1.	Appendix 6. Probabilities, Logarithms and the Rasch Model	
2.	To help with our understanding of Rasch models, let's be explicit that success is a score of "1", and failure is a score of "0" on an item. Then the **Rasch dichotomous model** specifies the probability, P_{ni1}, that person *n* of ability B_n scores 1 on item *i* of difficulty D_i and P_{ni0} is the probability of scoring 0. "Ability" is what we are looking for in the people. In your situation, you can use other words "Proclivity", "Motivation", "Health status", "Disability". "Difficulty" is what we identify in the items. In your situation, you can use other words, "Challenge", "Impediment"	$\log_e(P_{ni1} / P_{ni0}) = B_n - D_i$ $P_{ni1} + P_{ni0} = 1$
3.	*An essential concept:* "qualitatively-ordered data": "Success" means "more of what we are looking for" "Failure" means "less of what we are looking for" The difference between "Success" and "Failure" is *qualitative.* The ordering of these different qualities is indicated by scoring them "1" and "0". "1" means "indicating more of the latent trait". "0" means "indicating less of the latent trait".	Standard Rasch item-scoring: Success = 1 = more of what we seek Failure = 0 = less of what we seek
4.	$\mathbf{B_n}$ and $\mathbf{D_i}$ are distances in **logits** along the **latent trait** relative to the **local origin** A "**latent trait**" is something which we can have more or less of, but which we cannot measure directly. It is a variable such as "mathematics ability" or "patient quality of life". We conceptualize it to be a straight line marked out in equal-interval units. This line is infinitely long. We can always imagine something (or someone) with more of the attribute than anything (or anyone) we have encountered so far, and also something (or someone) with less of the attribute. We conceptualize each observation in the data to indicate "less" or "more" of this latent trait. "**Logits**" are "log-odds units". Look at the Rasch equations. On the left is "$\log_e$", this means the "natural logarithm". If you don't know about logarithms, please look at Appendix 3 to this document. The logarithm is of P_{ni1} divided by P_{ni0} which is the ratio of two probabilities. A ratio of probabilities is called the "odds". So, on the left-side of the equation we have log(odds). These provide the units for the right-side of the equation, so B_n and D_i are measured in "log-odds units", "logits" - pronounced *low-jits.* In Table 1.0 above, the logit values have been linearly rescaled (multiplied by a constant and then had another constant added) to produce numbers that are easier for most people to think with. This is like going from Celsius to Fahrenheit on a temperature scale. The meaning hasn't changed only the numbering. The "**local origin**" is the place we are measuring from. In physical measurement: for length, it is one end of a tape measure. For mountains, 0 is at sea level. For temperature, it is 0° on a thermometer. In Rasch measurement, 0 is usually in the center of the range of item difficulties, but it can be wherever we prefer it to be - as long as we make it clear to ourselves and everyone else where that 0 point is.	

5.	What if the probability of success is the same as the probability of failure? Then both probabilities are 0.5. The odds of success are 0.5/0.5 = 1, and the logarithm of the odds is log(1) = zero. So the ability and the difficulty are the same. Exactly what we expect! When I encounter an item of exactly the same difficulty as my ability, I can't predict whether I'm going to succeed of fail. My prediction would be like tossing a coin ... *Heads or Tails?*	$P_{ni1} + P_{ni0} = 1$ $P_{ni1} = P_{ni0} = 0.5$ $\log_e(P_{ni1} / P_{ni0}) = \log_e(0.5 / 0.5) =$ $\log_e(1) = 0 = B_n - D_i$ $B_n = D_i$
6.	If I'm sure to succeed, then my probability of success is 1 and my probability of failure is zero. My logit ability will be plus-infinity relative to any item of finite difficulty.	$P_{ni1} + P_{ni0} = 1$ $P_{ni1} = 1, P_{ni0} = 0$ $\log_e(P_{ni1} / P_{ni0}) = \log_e(1 / 0) =$ $\log_e(\infty) = \infty = B_n - D_i$ $B_n = \infty$
7.	If I'm sure to fail, then my probability of success is 0 and my probability of failure is one. My logit ability will be minus-infinity relative to any item of finite difficulty.	$P_{ni1} + P_{ni0} = 1$ $P_{ni1} = 0, P_{ni0} = 1$ $\log_e(P_{ni1} / P_{ni0}) = \log_e(0 / 1) =$ $\log_e(0) = -\infty = B_n - D_i$ $B_n = -\infty$
8.	Suppose that I generally succeed 3 times out of 4 in hitting a target with an arrow. Then my probability of success is P = 3/4 = 0.75. And my probability of failure is 1-P = 1/4 = 0.25. Then my odds of success is P / (1-P) = 0.75 / 0.25 = 3. And my ability is $\log_e(3)$ = 1.1 logits more than the target is difficult.	$P_{ni1} = 0.75$ $P_{ni0} = 0.25$ $P_{ni1} / P_{ni0} = 3$ $\log_e(P_{ni1} / P_{ni0}) = \log_e(3) = 1.1$ $1.1 = B_n - D_i = 1.1$
9.	Extending this idea, we can draw a picture of the relationship between ability (relative to an item) and probability of success. the relationship is called a "logistic ogive". You can see on the plot that a zero logit difference (dotted line) corresponds to a 0.5 probability of success. An ability advantage of 1.1 logits (arrowed) is equivalent to a probability of success of 0.75. This curve is *"monotonic ascending"* - the probability of success always increases with increasing measure difference, monotonously.	

<table>
<tr><td>1.</td><td colspan="2">Appendix 7. Unit Normal Deviates</td></tr>
<tr><td>2.</td><td>The “normal” distribution is fundamental to statistics. It describes what happens when events happen “normally”, purely by chance. The Figure shows the probability of different numbers of “heads” when a coin is tossed 15 items in the red bars: http://mathworld.wolfram.com/NormalDistribution.html We can see that the overall pattern follows a bell-shaped curve the continuous black line. This pattern gets closer to a smooth line, the more coins we toss. The black continuous line for an infinite number of tosses is the “normal distribution”.</td><td>
</td></tr>
<tr><td>3.</td><td>We are interested in a special case of the normal distribution. We want the one when its mean is zero, and its standard deviation is 1.0. This is called the “unit normal distribution”, abbreviated N(0,1). Statisticians use the Greek letter mu, μ, for the mean or average, and the Greek letter sigma, σ, for the standard deviation or spread, so the general normal distribution is $N(\mu, \sigma^2)$.

Look at the plot, the x-axis is labeled “z”. “z” means that these values are “z-scores” also called “unit normal deviates”. They are possible values of the unit normal distribution. The y-axis indicates the probability of observing the z values. Looking at the red curve, values of z near 0 have high probability. Values of z outside ±3 have very low probability.</td><td>

http://faculty.vassar.edu/lowry/ch6pt1.html</td></tr>
<tr><td>4.</td><td colspan="2">The area under the red curve indicates the cumulative probability of observing z values. 68% of the area under the red curve is within ±1, i.e., within 1 S.D. of the mean of the unit normal distribution. So we expect about 2/3 of the values we observe by chance to be statistically close to the mean.</td></tr>
<tr><td>5.</td><td colspan="2">We are usually concerned about values far away from the mean on either side (a 2-sided test). This Figure says that 2.28% of the area under the curve is to the right of +2, and 2.28% is to the left of -2. So, when we sample from random behavior modeled this way, we expect to encounter values outside of ±2.0 only 2.28%+2.28% = 4.56% of the time. This is less than the 5% (in other words, $p<.05$) which is conventionally regarded as indicating statistical significance, i.e., to be contradicting the idea that everything is random.</td></tr>
<tr><td>6.</td><td>The precise value of $p < .05$ is</td><td>$z > |\pm1.96|$ for $p<.05$</td></tr>
<tr><td>7.</td><td>and for $p < .01$ is</td><td>$z > |\pm2.58|$ for $p<.01$</td></tr>
<tr><td>8.</td><td colspan="2">But, remember, just because a value is statistically significant doesn’t mean that it is wrong. We do expect to see those values occasionally. The question to ask ourselves is “Why now?”</td></tr>
</table>

9.	What if we don't have a unit-normal distribution? We can often approximate it by taking our set of numbers, our data, subtracting from them their mean (arithmetic average) and dividing them by their standard deviation)	(the data - their mean) / (their standard deviation) $\rightarrow$ N(0,1)
10.	Residuals from our data, $\{R_{ni}\}$, have a mean of zero, and a modeled standard deviation of $V_{ni}^{0.5}$ so the standardized residuals $\{Z_{ni}\}$ should approximate N(0,1)	$\{R_{ni} / V_{ni}^{0.5}\} = \{Z_{ni}\} \rightarrow N(0,1)$