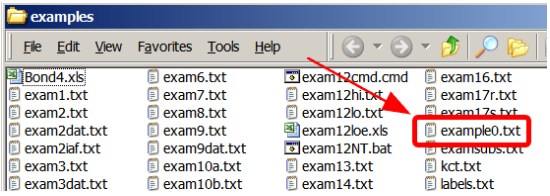
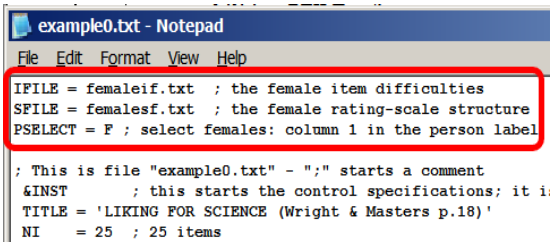
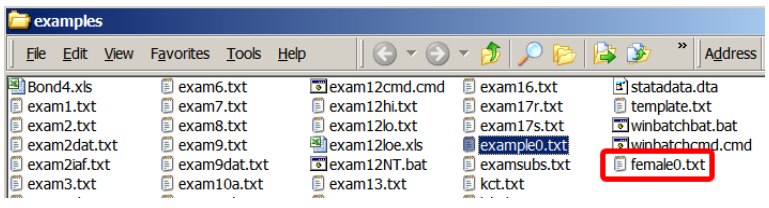
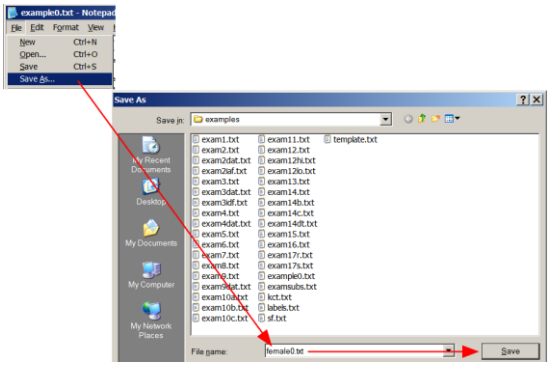
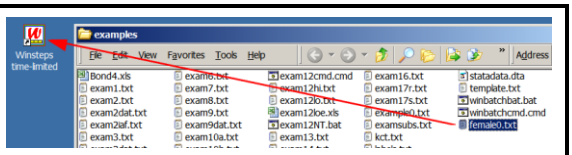
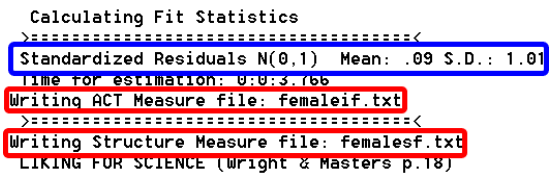
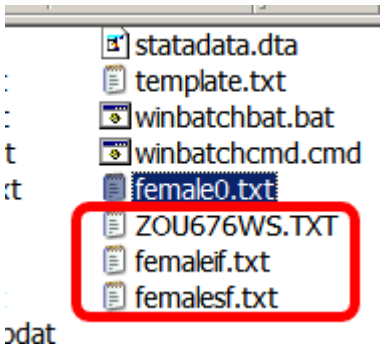
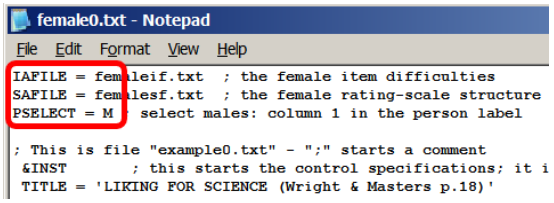
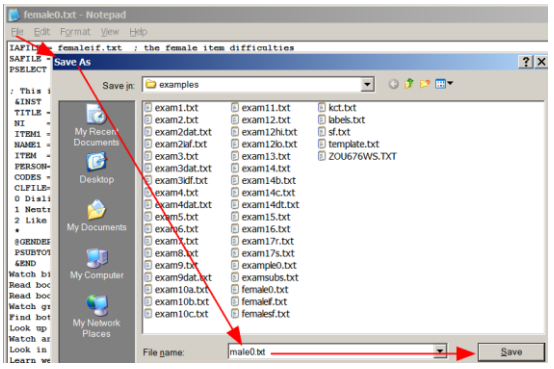
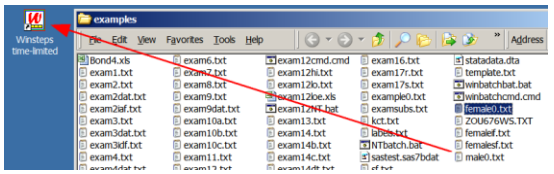
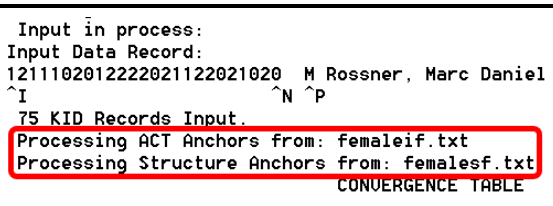
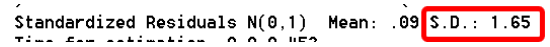
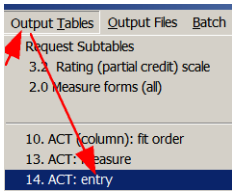
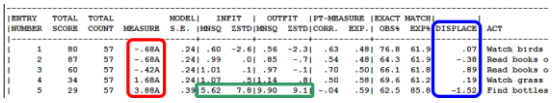
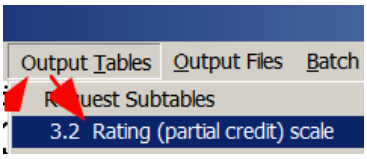
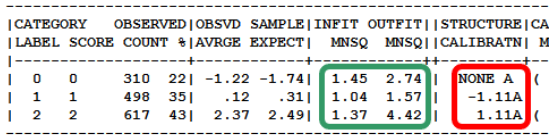
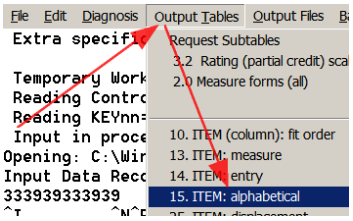
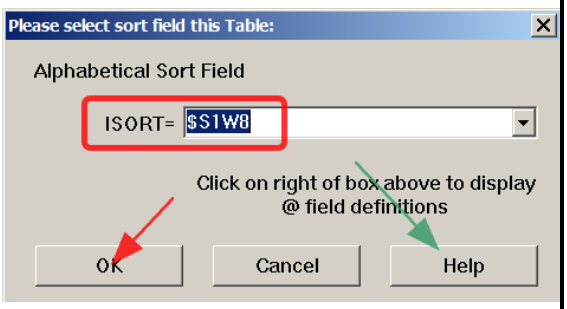
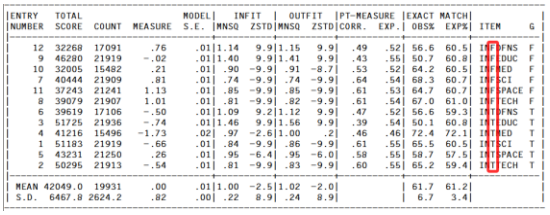
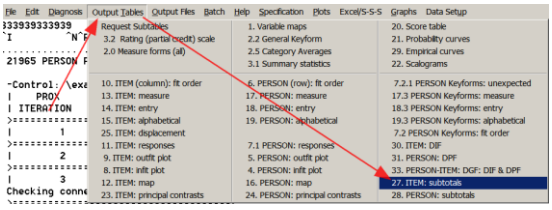
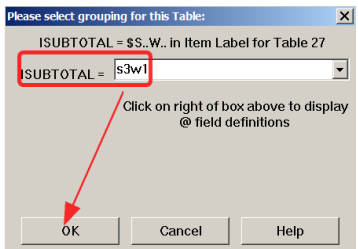
#	Practical Rasch Measurement - Further Topics : www.winsteps.com Mike Linacre, instructor - July 2011	
1.	**Tutorial 3. Investigating test functioning** • Subtotals • Differential Item Functioning • Dimensionality If you don't know the meaning of a word, then please look at the "Glossary" Lesson.	
2.	**A. Anchoring (fixing) abilities, difficulties and rating-scale structures**	
3.	There are many situations in which we want to apply the person abilities, item difficulties, or rating-scale structures from one dataset to the analysis of another dataset. For instance: we have analyzed and reported 1,000 students (or survey respondents, or patients, or ...). Then we find 10 more were overlooked. We want to measure these, without changing the other measures. We do this by anchoring (fixing) then analysis of the 10 using the measures for the 1000	Analysis of 1,000 students + Analysis of 10 students = Analysis of 10 students in the frame of reference of 1,000 students
4.	The process is simple: from the first (free) analysis, output item difficulties: IFILE=if.txt (you choose the file name) person abilities: PFILE=pf.txt rating-scale structures: SFILE= sf.txt	In the anchored (fixed) analysis, then anchored item difficulties: IAFILE=if.txt anchored person abilities: PAFILE=pf.txt anchored rating-scale structures: SAFILE=sf.txt
5.	Let's do this with Example0.txt We will measure the Females. Estimate the item difficulties and rating-scale structure for the females. Then measure the Males in the Female frame-of-reference. Double-click on Example0.txt in the Examples folder	
6.	Add these control-variable lines to Example0.txt : IFILE = femaleif.txt ; the female item difficulties SFILE = femalesf.txt ; the female rating-scale structure PSELECT = F ; select females: column 1 in the person label These lines can go anywhere before &END. I have put them at the top of the file.	
7.	Save example0.txt as female0.txt Display the Examples folder. It now contains "female0.txt' 	

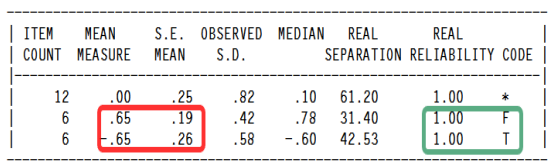
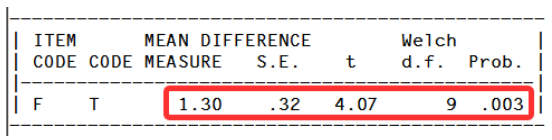
8.	Drag female0.txt from the Examples folder **onto** the Winsteps shortcut-icon Winsteps launches	
9.	Run the analysis. Red boxes: The analysis screen shows that the IFILE= and SFILE= have been output. Blue box: notice the "Standardized Residuals" - they are approximately N(0,1). We will refer to this number soon.	
10.	Display the Examples folder, notice that the there are three more files. femaleif.txt and femalesf.txt are the two measure files from our female analysis. ZOU...WS.TXT is a Winsteps work file. It will disappear when the Winsteps analysis closes. If you see any stray ZOU... files on your computer, please delete them. They are not needed.	
11.	We want to anchor the male analysis using the female measures. Double-click on female0.txt. Change it so that: IAFILE = femaleif.txt ; the female item difficulties SAFILE = femalesf.txt ; the female rating-scale structure PSELECT = **M** ; select males: column 1 in the person label	
12.	Save "female0.txt" as "male0".txt	
13.	Drag "male0.txt" onto the Winsteps shortcut-icon, and then Do the Winsteps analysis	
14.	Look at the Winsteps analysis window for "male0.txt" Winsteps reports that the anchor files have been processed	

15.	Look further down at the "Standardized Residuals". These are distributed N(0, $1.65^2 = 2.7$), with more than twice the expected variance. Anchoring the measures has considerably increased the misfit in the data!	
16.	Always verify that the IAFILE= anchor values have been applied correctly. For the male0.txt analysis, display "Output Tables", "Table 14" - Items in entry order	
17.	Red box: The measures are anchored "A". Green box: The misfit is huge for item 5. Blue box: "Displacement" "Anchor value" + "Displacement" = "Measure for these data" "Displacement" is one way of reporting "Differential Item Functioning" = difference in item difficulty for males and females. But there is an easier way soon :-)	 Item 5 is anchored at the female measure (3.88). The male measure is (3.88 + -1.52 = 2.36)
18.	Also verify that the SAFILE= rating-scale structures have been applied correctly. Display "Output Tables", "Table 3.2" - Items in entry order	
19.	Red box: The "structure calibrations" (Rasch-Andrich thresholds) are anchored "A". Green box: All the categories misfit.	
20.	Close all the Winsteps analyses You can delete female0.txt, male0.txt, etc.	

21.	**B. Item Subtotals**	
22.	Launch *Winsteps*	

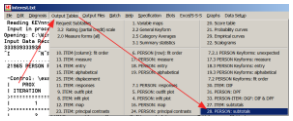
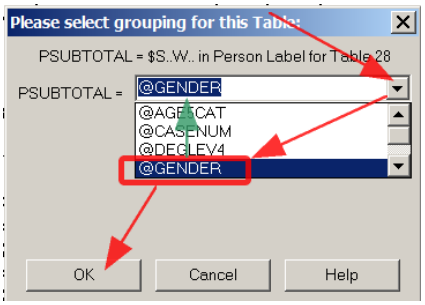
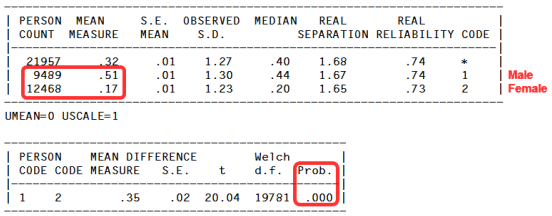
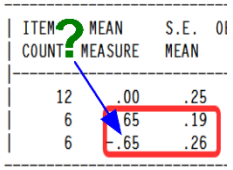
23.	Analyze
	c:\Winsteps-time\further**interest.txt**
	Beneath the summary table at the end of the estimation process, you should see this brief summary of the crucial control variables.
	If yours differs, there is a new copy of interest.txt in
	c:\Winsteps\further**further-data.zip**
	or www.winsteps.com\a\further-data.zip
		`CODES= 123456789`
		`IVALUET= 321******`
		`IVALUEF= 321******`
		`IREFER= TTTTTTFFFFFF`
		`GROUPS= TTTTTTFFFFFF`

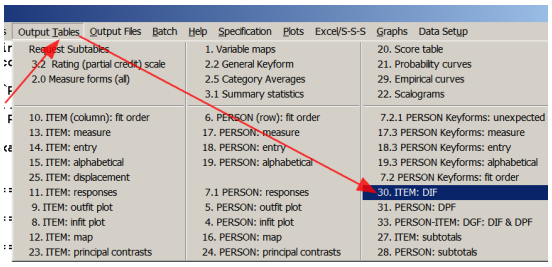
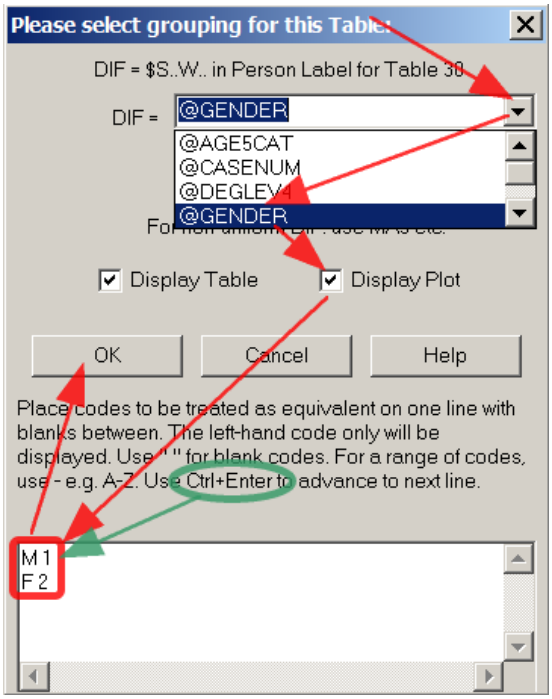
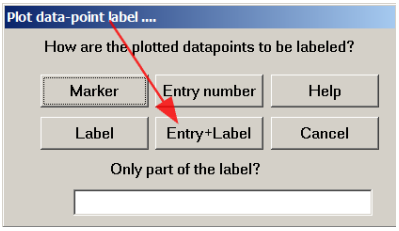
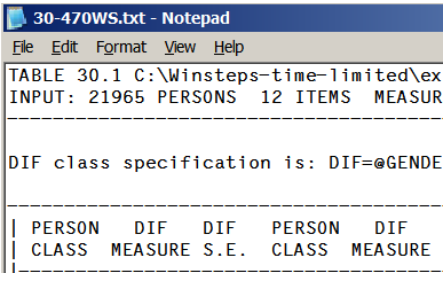
24.	We are going to investigate the Interest and Information items in the NSF survey. Output Tables: Click on Table15, the items in alphabetical order. The numbering of the tables was established around 20 years ago when we all the output was on computer paper.	
25.	A dialog box displays, asking for the sorting information. $S1W8 means $S1 = starting in column 1 of the item label. W8 = 8 characters wide This is the same as: $S1E8 = starting in column 1, ending in column 8 Click on Help to see other options. Click on **OK**	
26.	In Table 15, the items are listed in alphabetical order. There are two types: InFormation: with F in the 3rd character of the item label InTerest: with T in the 3rd character.	
27.	Let's produce a subtotal of the item of each type: Output Tables: Click on Table 27, item subtotals.	
28.	In the "grouping for this Table" dialog box, Type in: s3w1 which means "subtotal based on column 3 of the item label, one character wide". This is the same as $S3W1. See the "selection rules" in Winsteps Help: http://www.winsteps.com/winman/columnselection.htm Click on OK.	

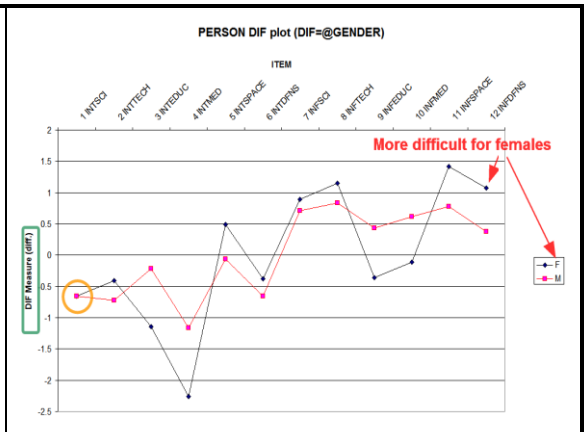
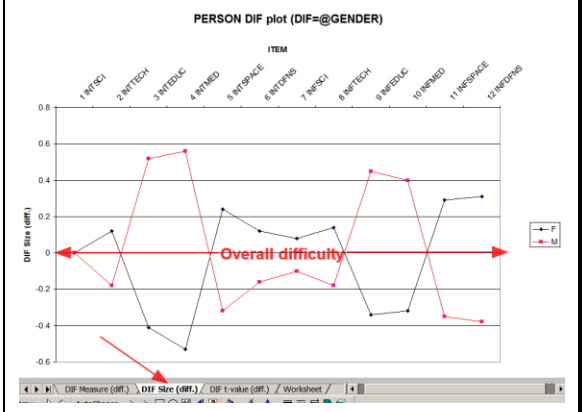
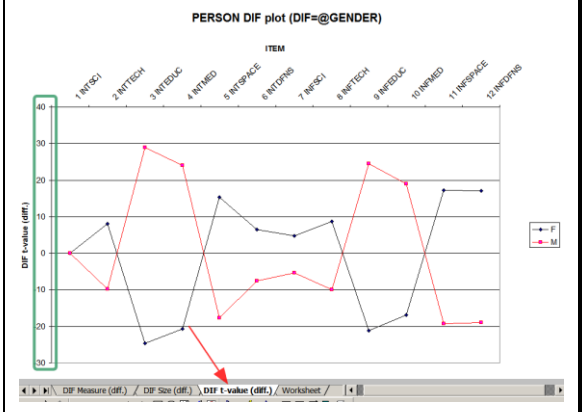
29.	Table 27 displays. *Green box:* The item reliabilities (reproducibilities) for the two item types are close to 1.00 because of the large sample size. *Red box:* On average, the inFormation items are 1.3 logits more "difficult" than the inTerest items. Is this difference significant, based on the standard errors of the means?	
30.	Beneath the subtotals is displayed Welch's two-sided *t*-test of the **statistical difference between the average difficulties of the two sets of items**. Welch's t-test is an improved version of Student's t-test. p=.003, which is p<.01, so the result is highly statistically significant.	

ITEM COUNT	MEAN MEASURE	S.E. MEAN	OBSERVED S.D.	MEDIAN	REAL SEPARATION	REAL RELIABILITY	CODE
12	.00	.25	.82	.10	61.20	1.00	*
6	.65	.19	.42	.78	31.40	1.00	F
6	-.65	.26	.58	-.60	42.53	1.00	T

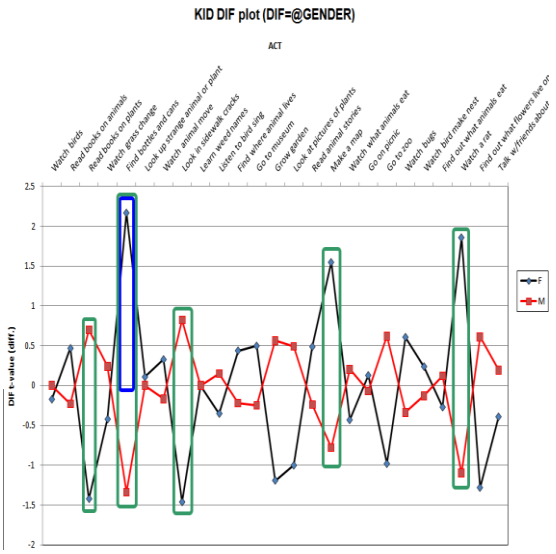
ITEM CODE	ITEM CODE	MEAN DIFFERENCE MEASURE	S.E.	t	Welch d.f.	Prob.
F	T	1.30	.32	4.07	9	.003

31.	**C. Person Subtotals**	
32.	Let's do the same for a person classification. Click on Table 28. Person subtotals.	
33.	Again we see a grouping selection box. We could look at the person labels to identify the columns of the demographic codes, but this has been done for us during the conversion from SPSS. Click on the drop-down menu indicator:  Click on @GENDER, this variable contains the column information for gender (sex) in the person label. Click on OK	
34.	Table 28 displays. The coding for "GENDER" is in nsfcl.txt, so I typed it in here. The males are considerably more interested and informed than the females. The difference is statistically highly significant (very unlikely to have happened by chance). *Notice that males and females have different group sizes, and different average abilities.*	
35.	If you are communicating measures and sub-totals to a non-specialist audience, negative logit values with decimal places can be difficult for them to understand. Linear rescaling of the logits into user-friendly numbers with UIMEAN= and USCALE= is helpful. See **Appendix 1.** *User-Friendly* Rescaling of Rasch Measures.	

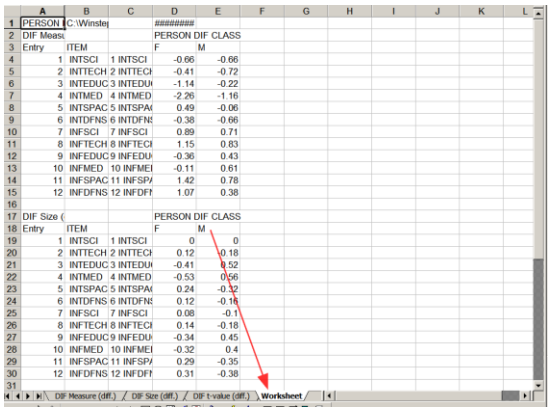
36.	**D. Uniform Differential Item Functioning**	
37.	Is the difference between males and females the same across all the items or do the genders differ across items? This is "Differential Item Functioning" (DIF), also called "Item bias". It is "Uniform" because it is imagined to impact all ability levels equally. "Uniform" means that the item has the same amount of extra difficulty for everyone in the classification group. For instance, "Item 3 is one logit more difficult for all the girls" Click on Table 30. Item DIF.	
38.	In the "grouping" dialog box: Select @GENDER again. We want to see both Table 30 and the Excel DIF plot: Check (tick) "Display Table and "Display Plot" In the SPSS file, 1=Male and 2=Female. It will be easier for us if the genders display as "M" and "F", so we enter this information in the lower section of the box. M 1 means "show data code 1 as M" to advance to the next line in this section, press **Ctrl+Enter**.	
39.	Click on "Entry+Label" We want the Excel plots to be identified with the item entry numbers and the item labels.	
40.	Table 30 displays first. It is packed with numbers, so it is easier to think about the Excel plots first. These may take a some seconds to display. *If you don't have Excel, please do look at the plots I will show you here.*	

41.	The first Excel DIF plot is "DIF Measure (Diff*iculty*)". This shows the absolute logit difficulty of each item for each person classification-group. We can see that items 11 and 12 are noticeably more difficult for the females than for the males. Item 4 is noticeably easier for the females than the males. These difficulties are in the same frame-of-reference as the difficulties reported in the item Tables (14, etc.), so zero logits is the same zero as in those Tables. Notice that Item 1 is reported to have the same difficulty for both groups.	
42.	On item 4, is the item ***easier*** for females, or do the females ***have greater "ability"*** when they respond to that item? The statistics can't answer that question. Lawyers would contend that it is the item difficulty that changes, and the ability of the classification-group stays constant.	
43.	Click on Chart "DIF Size (diff.*iculty*)". This shows the size in logits of the item DIF for each group relative to the overall difficulty of each item. This plot is useful when several classification groups are being analyzed.	
44.	Click on Chart "DIF t-value (diff.)". This shows the statistical significance (t-value) associated with each size shown in the DIF Size chart. Here, the sample sizes are large, so these t-values can be interpreted as unit-normal deviates. Except for item 1, all the DIF Sizes are all highly significant (green box) because of the large sample sizes. Highly significant is p<.01 (double-sided) t outside of ±2.58. This chart tells us where to look. For precise significance tests, we need to look at Table 30.	

45. #44 shows the statistical significance of the DIF size. How likely it is to occur by chance. In that Figure, almost everything is significant, so here is a different Figure from *example0.txt.*
In this Figure, the blue box shows an item which is significantly more difficult for Females (t-value >2) than its average Female+Male difficulty.
The green boxes show items which are significantly different in difficulty for Females against Males (t-value difference >2).
When there are only two classes (as in this example) we usually compare one group with the other (green boxes).
When there are many classes we compare each group with the biggest group or the average (which is usually close to the biggest group).
This Figure is not exact, but it tells us what to investigate in Table 30. Table 30 shows the exact values and the probabilities of the t-tests, which depend on the sample sizes. Table 30 can be large, and in large tables of numbers it is easy to miss something that is important, or to perceive something as important which is really much like everything else.



46. Click on "Worksheet".
This shows the numbers and descriptions on the plots.
You have complete control to edit the data and reformat the plots.



47. Winsteps Table 30 in NotePad is on your Windows taskbar

48.

Table 30.1

PERSON CLASS	DIF MEASURE	DIF S.E.	PERSON CLASS	DIF MEASURE	DIF S.E.	DIF CONTRAST	JOINT S.E.	t	Welch d.f.	Welch Prob.	Mantel Prob.	Hanzl Size	ITEM Number	ITEM Name
F	-.66	.02	M	-.66	.02	.00	.02	.00	INF	1.000	.3041	-.03	1	INTSCI
F	-.41	.02	M	-.72	.02	.31	.02	12.66	INF	.0000	.0000	.37	2	INTTECH
F	-1.14	.02	M	-.22	.02	-.93	.02	-37.9	INF	.0000	.0000	-.64	3	INTEDUC
F	-2.25	.03	M	-1.16	.02	-1.09	.03	-31.5	INF	.0000	.0000	-1.24	4	INTMED

49. This Table is useful when comparing a "focal" group to a "reference" group.
In Table 30.1, each pair of classification-groups is shown twice on each item. F-M, then M-F. The numbers mean the same thing on both lines.
PERSON CLASS is the person classification group, F or M.
Remember that in person sub-totals we discovered that F and M have different group sizes, and different mean measures. DIF computations adjust for these differences.

First is a maximum-likelihood estimate of DIF size:
DIF MEASURE is the absolute difficulty of the item for the group, which we saw in the first Excel plot. Each measures has a standard error (precision).
DIF CONTRAST is the difference between the two DIF measures (left-hand DIF measure - right-hand DIF measure). Its S.E. is the JOINT S.E. of the two DIF MEASURE S.E.s.
t is Welch's t-statistic testing the hypothesis that the DIF size is zero, apart from measurement error.
d.f. are the degrees of freedom of the t-statistic. "INF" means "effectively infinite", so that the t-statistic can be investigated as though it is as a unit-normal deviate. If p<.05, then we reject the hypothesis that the t-statistic is part of the t-distribution corresponding to the null hypothesis.
Prob. is the resulting probability (statistical significance).

Second is the Mantel-Haenszel (MH) DIF statistic (for dichotomies) or Mantel DIF statistic (for polytomies).
Prob. is the probability of the test that the DIF is zero.
Size is an MH estimate of the "contrast" size of the DIF.

50. "DIF Contrast" and Mantel-Haenszel "Size" are two different estimates of the size of the DIF. They partition the data somewhat differently. DIF Contrast is more robust against missing data. Mantel-Haenszel has far greater acceptance among Paper reviewers.

51. **Table 30.2**

PERSON CLASS	OBSERVATIONS COUNT	AVERAGE	BASELINE EXPECT	MEASURE	DIF SCORE	DIF MEASURE	DIF SIZE	DIF S.E.	DIF t	DIF Prob.	ITEM Number	ITEM Name
F	12270	1.31	1.31	-.66	.00	-.66	.00	.02	.00	1.000	1	INTSCI
M	9322	1.40	1.40	-.66	.00	-.66	.00	.02	.00	1.000	1	INTSCI
F	12264	1.22	1.26	-.54	-.04	-.41	.12	.02	8.08	.0000	2	INTTECH
M	9323	1.41	1.36	-.54	.06	-.72	-.18	.02	-9.75	.0000	2	INTTECH

52. Table 30.2 shows the numbers reported in the Excel plots.
Red box: This shows the raw-score computation for the DIF. It shows the observed average rating on the item for each group. Then expected average rating based on the overall "baseline" difficulty of the item (as reported in Table 14).
DIF Score is the difference between the observed average and the expected average of the ratings for each group. It is the DIF Score that is the basis for estimating the DIF Size.
If you wonder "Is this item biased in favor of, or against, this group?", then **if the DIF Score is positive, the item bias is in favor of the group. If it is negative, the item bias is against the group.**

53. The ETS Table of DIF impact is in Winsteps Help: DIF Concepts.
A rule-of-thumb: DIF must be at least 0.5 logits and with probability p<.05 to merit further investigation.
But the impact of DIF on the person measures is
Impact = DIF Size / Test length

ETS DIF Category	DIF Effect Size (Logits)	DIF Statistical Significance
C = moderate to large	|DIF| >=1.5 / 2.35 = 0.64	p(|DIF| ≤1/2.35 = 0.43) < .05
B = slight to moderate	|DIF| >= 1/2.35 = 0.43	p(|DIF| <0) < .05

54.

Table 30.3 is Table 30.2 in a different order ....

PERSON CLASS	OBSERVATIONS COUNT	AVERAGE	BASELINE EXPECT	MEASURE	DIF SCORE	DIF MEASURE	DIF SIZE	DIF S.E.	DIF t	Prob.	ITEM Number	Name
F	12270	1.31	1.31	-.66	.00	-.66	.00	.02	.00	1.000	1	INTSCI
F	12264	1.22	1.26	-.54	-.04	-.41	.12	.02	8.08	.0000	2	INTTECH
F	12281	1.46	1.33	-.74	.13	-1.14	-.41	.02	-24.6	.0000	3	INTEDUC
F	8492	1.75	1.64	-1.73	.11	-2.26	-.53	.03	-20.7	.0000	4	INTMED

55.

Table 30.4

DIF class specification is: DIF=@GENDER

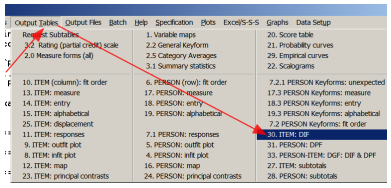
PERSON CLASSES	SUMMARY DIF CHI-SQUARE	D.F.	PROB.	ITEM Number	Name
2	.0000	1	1.0000	1	INTSCI
2	159.735	1	.0000	2	INTTECH

Table 30.4 tests the hypothesis: "This item exhibits no DIF beyond statistical randomness".
We accept the hypothesis for item 1, but reject it for item 2, p<.05

56.

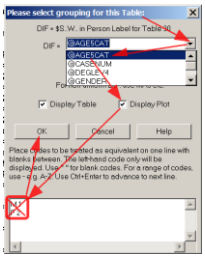
Now let's perform a more elaborate DIF investigation of different age-groups:

Winsteps menu bar
Click on Table 30. Item DIF.



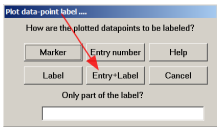
57.

Select @AGE5CAT
Click on "Display Plot"
Delete unwanted codes
Click on OK



58.

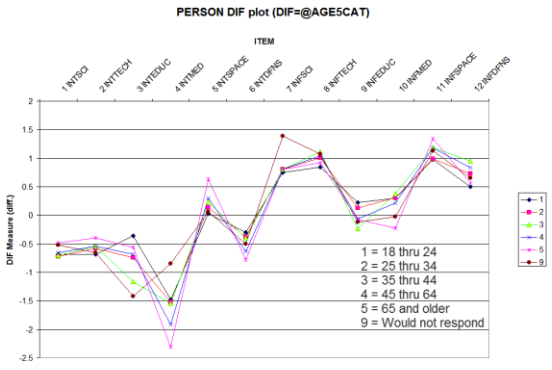
Click on "Entry+Label"

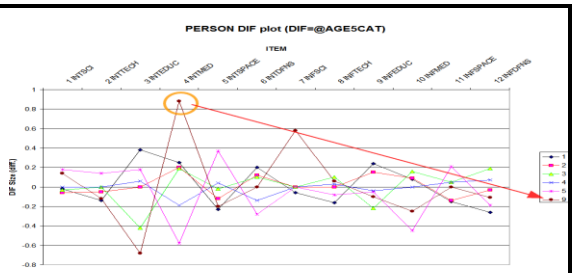
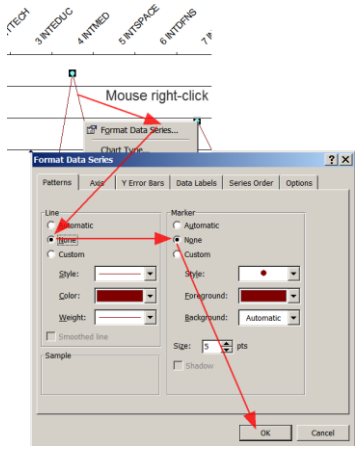
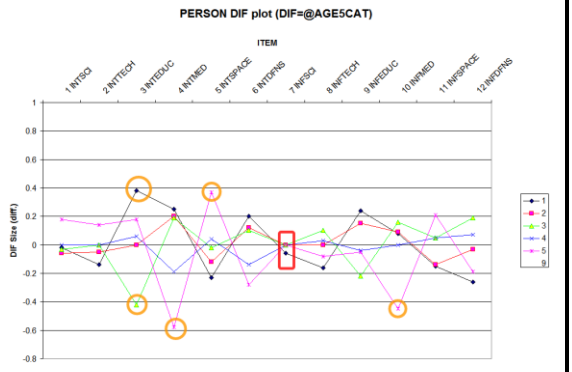
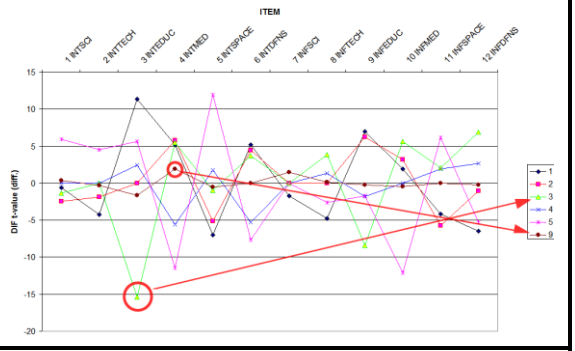
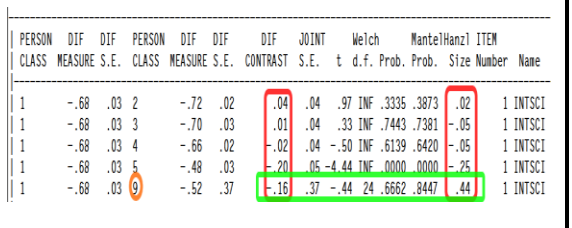


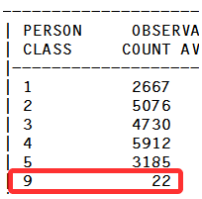
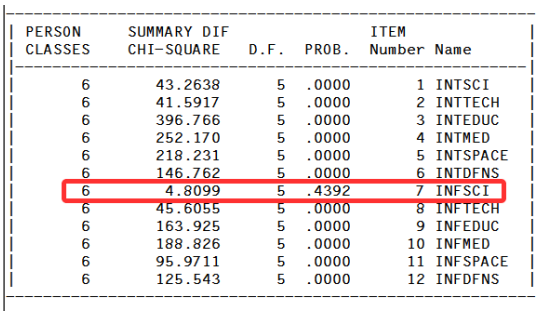
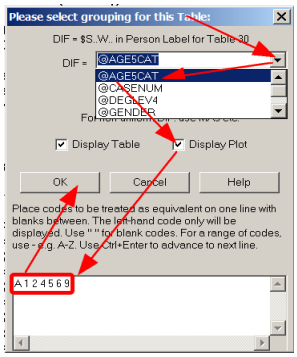
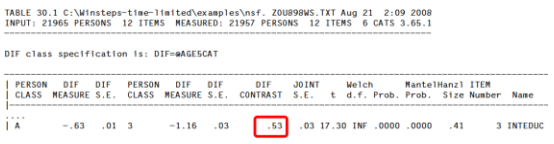
59.

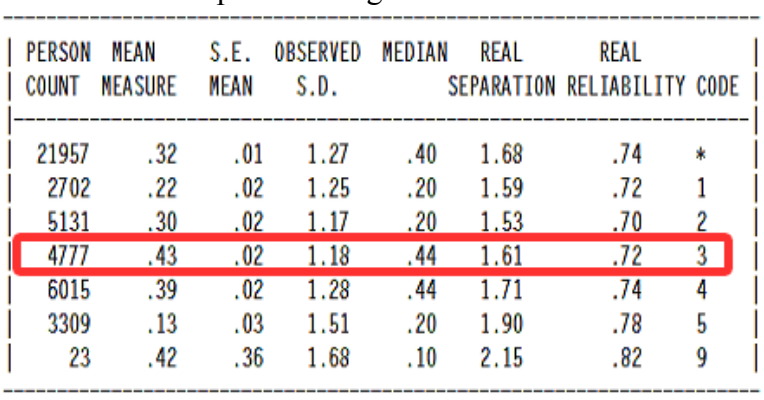
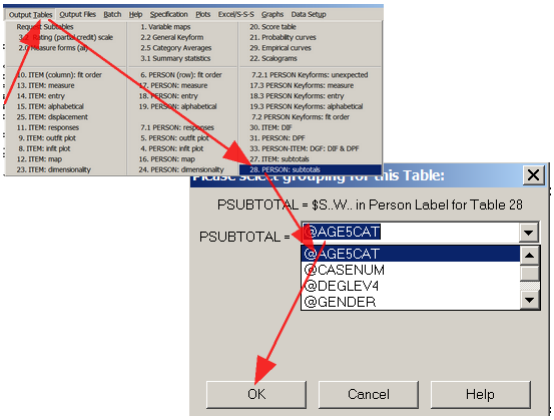
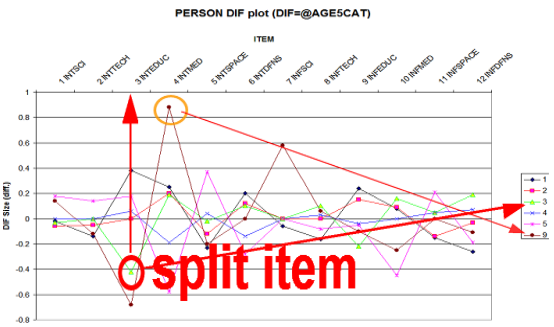
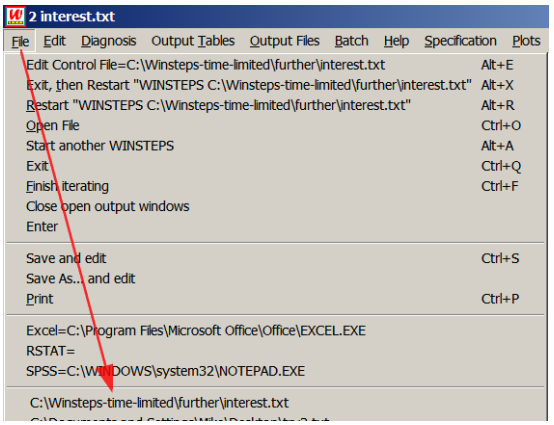
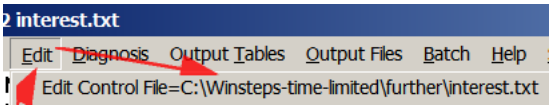
The Excel "DIF Measure" plot displays.

According to nsfcl.txt:
AGE5CAT =
1 = 18 thru 24
2 = 25 thru 34
3 = 35 thru 44
4 = 45 thru 64
5 = 65 and older
9 = Would not respond



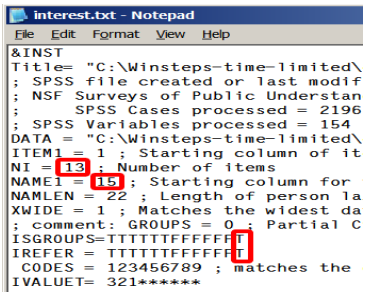
60.	We can see that there appears to be some DIF, but not much. Look at the Excel "DIF Size" plot. The most conspicuous DIF is for group "9", "Would not respond". There is something peculiar about them, but we don't know what.	
61.	Let's remove the line for "9" from the picture: Mouse pointer to the line for "9" Right-click Click on "Format Data Series". If you don't see "Format Data Series" Left-click on the background Move the mouse slightly closer to the line Right-click Patterns Tab. Line: click on "None" Marker: click on "None" Click on OK	
62.	*Orange circles:* We see some DIF of about 0.4 logits relative to the overall item difficulty, but is it random chance or perhaps something real? DIF studies tend to produce non-replicable findings, so we want to be as certain as possible that we are really observing a change in item difficulty. *Red rectangle:* Item 7 appears to exhibit the least DIF.	
63.	In the "DIF t-value" plot, we see that the large values of DIF relative to the overall difficulty for group "9" were mainly due to chance. They are relatively statistically insignificant. The most statistically significant DIF effect is **in favor of** classification-group 3 (35 thru 44 years old) on item 3 "INTEDUC", but, statistically with these large sample sizes (giving our hypothesis test great statistical power), almost everything is statistically significant.	
64.	Table 30.1 gives us the opportunity to compare DIF pairwise between classification groups. At the top of the Table, we are comparing the classification groups with group 1. The DIF Contrasts and MH Sizes align fairly well - except for group 9. Here the two DIF effect estimates are in opposite directions!	

65.	Table 30.2 - group 9: it has only 22 observations! The small number of persons has strongly influenced the two DIF detection methods. DIF researchers suggest that group sizes of at least 200 are needed for DIF studies.	
66.	Table 30.4 Only for Item 7 (the item exhibiting the least DIF) do we accept the hypothesis that "this item exhibits no DIF beyond that due to random error". *There is more about DIF in Winsteps Help.*	
67.	Let's contrast item 3 with all the other items. Winsteps menu bar Click on Table 30 Select AGE5CAT In the code box: recode all age codes to "A" except "3". A 1 2 4 5 6 9	
68.	Table 30.1 shows that item 3. INTEDUC is 0.53 logits easier for age-category 3 than for the other age-categories. This is so large that we will treat item 3 as two items.	
69.	Please close all windows	

70.	**E. Splitting a DIF item**
71.	If an item exhibits conspicuous DIF, there are several actions we could take: 1. Adjust all the ability measures of the focal group up (DIF against) or down (DIF in favor) by (DIF contrast / test length). 2. Omit the item from the analysis (IDELETE=). 3. Split the item into two items: one item for the reference group and one item for the focal group.
72.	Table 28 person-subtotals for all AGE5CAT: Notice that Group 3 is the highest: .43  
73.	**Let's make item 3 in interest.txt a separate item for age-category 3.** This item appears to be biased in favor of age-category 3, so it would be biased against most other age-categories. We want to "level the playing field". Look at the Table in #72. Guess what will happen to age-category 3 when item 3 is split. How will it move relative to the other age-categories? *Will age-category 3 be relatively higher or lower in Table 28 (#72)?*	
74.	Launch Winsteps Look at the File menu Can you see interest.txt ? Click on it for the control file Wait! Do not click on "Report file"	
75.	Edit Control File: Interest.txt	

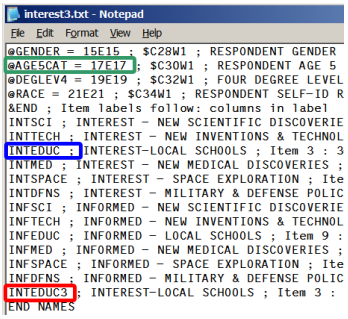
PERSON COUNT	MEASURE	S.E. MEAN	OBSERVED S.D.	MEDIAN	REAL SEPARATION	REAL RELIABILITY	CODE
21957	.32	.01	1.27	.40	1.68	.74	*
2702	.22	.02	1.25	.20	1.59	.72	1
5131	.30	.02	1.17	.20	1.53	.70	2
4777	.43	.02	1.18	.44	1.61	.72	3
6015	.39	.02	1.28	.44	1.71	.74	4
3309	.13	.03	1.51	.20	1.90	.78	5
23	.42	.36	1.68	.10	2.15	.82	9

76. In our control file, interest.txt,
we want one more copy of item 3. Change
NI = 13 ; one more item
NAME1 = 15 ; name moved one column right
ISGROUPS= (add)T ; groups for extra item 3
IREFER=(add)T ; rescoring for extra item 3

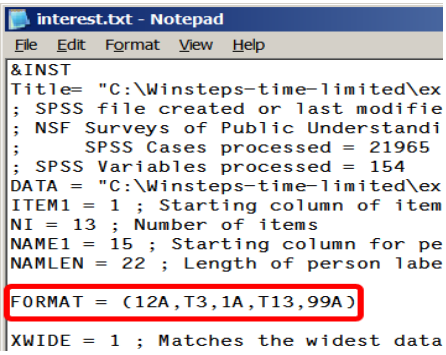


77. Scroll down to the list of item labels
Blue box: Copy the label for item 3, INTEDUC
Red box: Paste it after the last item label, before END NAMES
Change the item label. Mine is INTEDUC3.

Green box: notice that the age-category is in column 17 of the person label. *We will need this later ....*

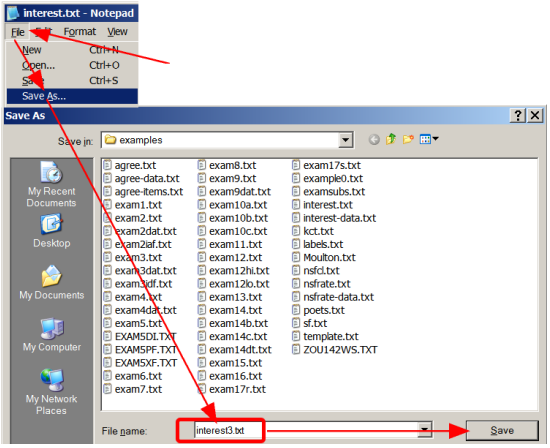


78. Now for the tricky part ....
Scroll back to the top of interest.txt
Type in:
FORMAT = (12A,T3,1A,T13,99A)
This instruction says: Reformat the input data record:
12A - read in the first 12 characters (the item responses)
T3 - go to column 3 of the input record
1A - read in 1 character (item 3 again)
T13 - go to column 13 (next after the 12 responses)
99A - read in 99 characters (the person label etc.)



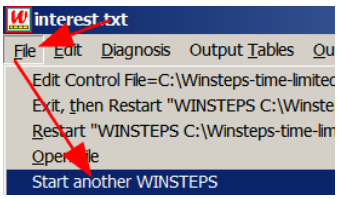
79. "Save as" **interest3.txt**

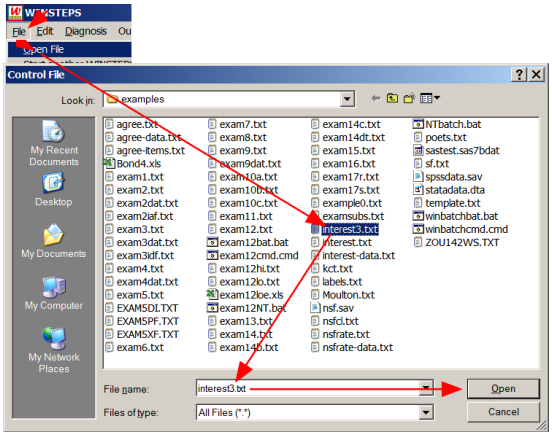
This is the control file for our new data analysis.



80. Let's check if this has worked properly so far ....

Winsteps menu bar
Click on "Start another Winsteps"



81.	New Winsteps Click on File menu Click on Open File Double-click on **interest3.txt**	
82.	Run the analysis The Analysis windows show the Input Data Record before FORMAT= this is the first line of our data file: interest-data.txt Input Data Record after FORMAT= *red box:* the response for item 13 is the same as *blue box:* the response for item 3	Input in process: Opening: C:\Winsteps-time-limited\ex Input Data Record before FORMAT=: 333939333939 1 1979 1 5 1 9 Input Data Record after FORMAT=: 3339393339393 1 1979 1 5 1 9 ^I N^P ..... 21965 PERSON Records Input.
83.	Click on "Output Tables" Click on Table 13. Items in measure order Item 3 INTEDUC and item 13 INTEDUC3 should be exactly the same. *If they are not exactly the same, please confirm each step of this process*	2 50295 21913 -.46 .01 .84 -9.9 .86 -9.9 .57 .54 63.2 58.8 INTTECH T 1 51183 21919 -.58 .01 .86 -9.9 .89 -9.9 .57 .54 62.9 59.0 INTSCI T 3 51725 21936 -.66 .01 1.17 9.9 1.16 9.9 .51 .54 54.7 59.4 INTEDUC T 13 51725 21936 -.66 .01 1.17 9.9 1.16 9.9 .51 .54 54.7 59.4 INTEDUC3 T 4 41216 15496 -1.62 .02 .95 -3.9 .98 -.9 .45 .45 72.5 72.0 INTMED T

84. *Congratulations! You have succeeded on the difficult part!*
In interest3.txt:
For item 3, we want to code all the responses for age-level 3 as missing.
For item 13, we want to code the responses for all age-levels except age-level 3 as missing.

Winsteps menu bar
Edit Control File: interest3.txt
Type (or copy-and-paste) in the extra lines:
EDFILE=*
"???????????????3" 3 M
"????????????????{~3}" 13 M
*

EDFILE= means: edit data file
"???????????????3" 3 M
"...." means: person selection
???????????????? means: anything in the first 16 columns of the person label (one column for each ?)
....?3 means: 3 in column 17 of the person label (the age-category)
then
3 means: item 3
M (or any non-numeric) means: treat as missing data
"????????????????{~3}" 13 M
{~3} means: anything not 3 in column17 of the person label
13 means: item 13

interest3.txt - Notepad

File Edit Format View Help

```
&INST
Title= "C:\Winsteps-time-limited\examples\nsf.sav"
; SPSS file created or last modified: 8/13/2008 7:5
; NSF Surveys of Public Understanding of Science an
;     SPSS Cases processed = 21965
; SPSS Variables processed = 154
DATA = "C:\Winsteps-time-limited\examples\interest-
ITEM1 = 1 ; Starting column of item responses
NI = 13 ; Number of items
NAME1 = 15 ; Starting column for person label in da
NAMLEN = 22 ; Length of person label

FORMAT = (12A,T3,1A,T13,99A)

EDFILE=*
"???????????????3" 3 M
"????????????????{~3}" 13 M
*

XWIDE = 1 ; Matches the widest data value observed
```

85. Save interest3.txt
Exit and restart Winsteps
Produce Table 13
Does yours look like this?
Item 3. INTEDUC. Measure -.53
Item 13. INTEDUC3. Measure -1.12
Difference: -.53 - -1.12 = 0.59 logits.
INTEDUC3 is 0.59 logits than the edited INTEDUC.
This approximates the DIF contrast of 0.53 logits we found earlier at #68! *Excellent!*
Age-category 3 are now being measured more fairly!

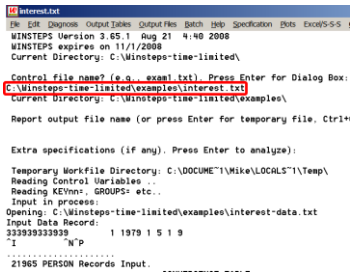
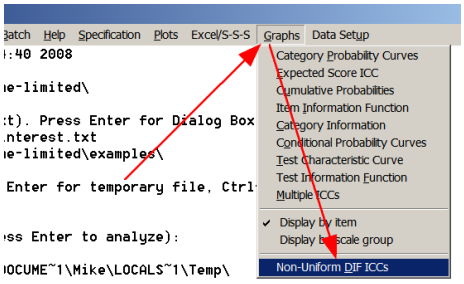
Bottom half of the Table ...

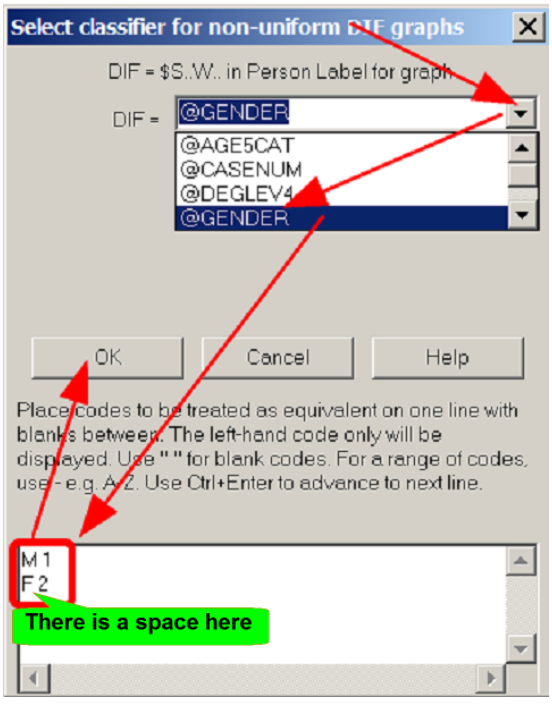
ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM	G
....														
2	50295	21913	-.45	.01	.81	-9.9	.83	-9.9	.60	.55	65.3	59.4	INTTECH	T
3	39703	17163	-.53	.01	1.45	9.9	1.54	9.9	.40	.55	48.9	59.5	INTEDUC	T
1	51183	21919	-.58	.01	.84	-9.9	.86	-9.9	.61	.55	65.2	60.4	INTSCI	T
13	12022	4773	-1.12	.03	1.47	9.9	1.64	9.9	.33	.50	56.6	65.1	INTEDUC3	T
4	41216	15496	-1.65	.02	.97	-2.6	1.00	.1	.46	.46	72.5	72.1	INTMED	T

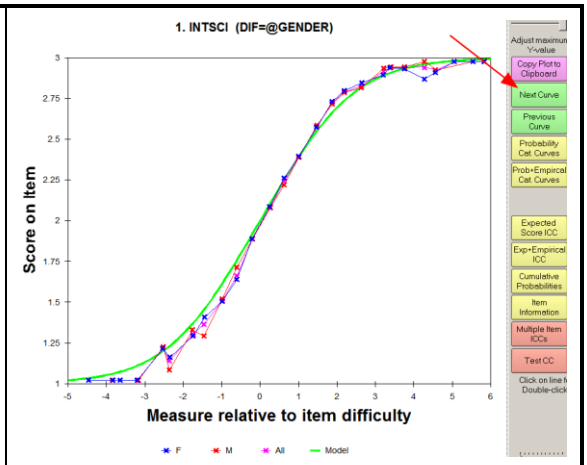
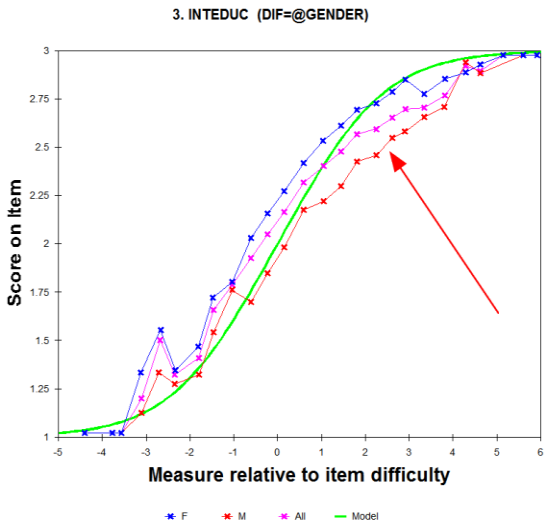
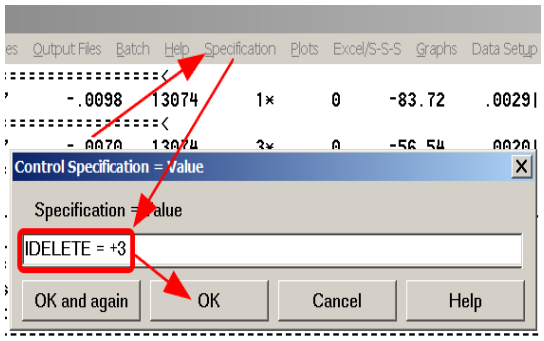
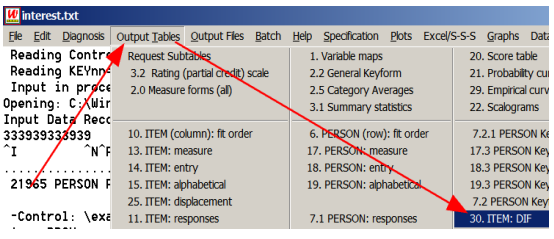
86.	Now look at Table 28. Person sub-totals for the new ability measures on all the items. Compare this to the original subtotals (#72).	Now, all numbers are higher, but Group 3 is relatively lower. It is the third highest. *Are you surprised?*
87.	Please close all windows	

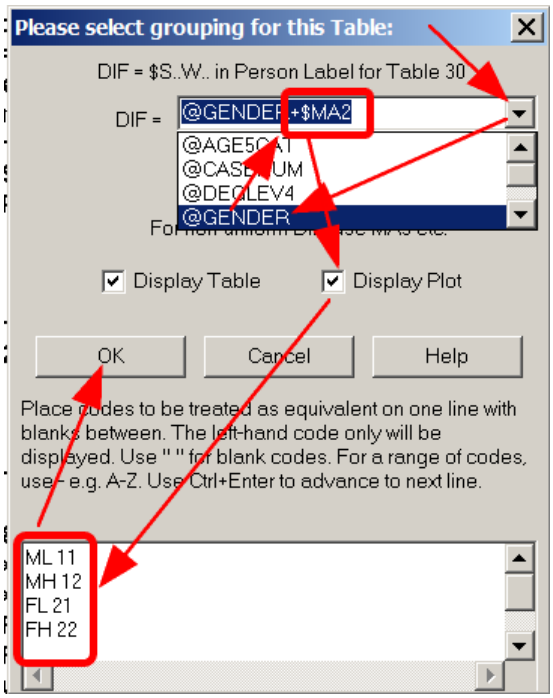
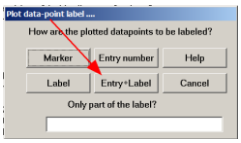
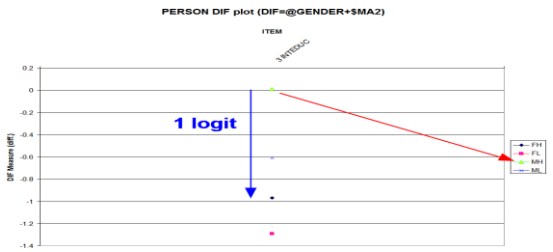
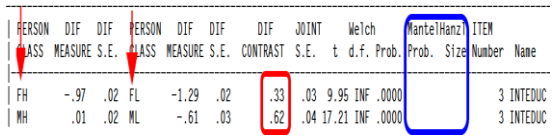
PERSON COUNT	MEAN MEASURE	S.E. MEAN	OBSERVED S.D.	MEDIAN	REAL SEPARATION	REAL RELIABILITY	CODE
21957	.32	.01	1.27	.40	1.68	.74	*
2702	.22	.02	1.25	.20	1.59	.72	1
5131	.30	.02	1.17	.20	1.53	.70	2
4777	.43	.02	1.18	.44	1.61	.72	3
6015	.39	.02	1.28	.44	1.71	.74	4
3309	.13	.03	1.51	.20	1.90	.78	5
23	.42	.36	1.68	.10	2.15	.82	9

PERSON COUNT	MEAN MEASURE	S.E. MEAN	OBSERVED S.D.	MEDIAN	REAL SEPARATION	REAL RELIABILITY	CODE
21957	.40	.01	1.27	.49	1.68	.74	*
2702	.31	.02	1.25	.30	1.60	.72	1
5131	.39	.02	1.17	.30	1.53	.70	2
4777	.47	.02	1.19	.49	1.62	.72	3
6015	.49	.02	1.28	.54	1.71	.74	4
3309	.23	.03	1.50	.30	1.90	.78	5
23	.52	.36	1.68	.19	2.15	.82	9

88.	**F. Non-Uniform Differential Item Functioning**	
89.	DIF may have different sizes for different ability levels in the classification-group. This is called "Non-uniform DIF" (NUDIF). Launch Winsteps Analyze interest.txt	
90.	The first step in investigating Non-Uniform DIF is to see if there is any evidence of it. Winsteps menu bar Click on "Graphs" Click on "Non-Uniform DIF ICCs"	

91.	Click on the pull-down menu
	
	Click on @GENDER - we will look at NUDIF for gender
	
	Type into the code box, the gender codes in the format: code **(space)** code Ctrl+Enter
	
	`M 1 Ctrl+Enter`
	`F 2`
	
	Click on OK
		

92.	Item 1. INTSCI displays. The green line is the Rasch-model prediction. The other lines are the average observed performance on this item by Females, Males, and All = Females+Males. Where these lines depart from the model curve, there is NUDIF. There is a little noise at the top and the bottom, but nothing to alarm us. Click on "Next Curve" and look at the curves	
93.	Item 3. INTEDUC is interesting (yes, that item again!) The Females (blue line) are performing relatively better than the Males (red line) at almost every ability level. The Females (blue line) are showing uniform DIF on the item. Their line is uniformly above the green line for much of the ability range. The Males (red line) are showing non-uniform DIF. High performing Males (red arrow) are performing worse than expected . Low performing males are tracking the green model line. So we want to investigate whether this Male high-low performance difference is statistically significant.	
94.	Winsteps output can become overwhelming, so let's focus only on item 3. Winsteps menu bar Click on "Specification" In the specification box, delete all items except item 3: IDELETE=+3 Click on OK The Analysis window reports: **IDELETE = +3** **CURRENTLY REPORTABLE ITEMS = 1**	
95.	Winsteps menu bar Click on "Output Tables" Click on Table 30. Item DIF.	

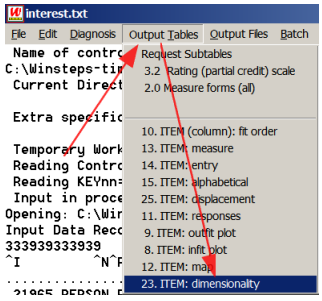
96.	In the "select grouping" dialog box, there is a lot to do: Click on the pull-down menu arrow Click on @GENDER In the DIF= box, after @GENDER type: +$MA2 This specifies that DIF will be reported for gender combined with M (measures) A (ascending) 2 (levels). The lower half of the measures will be labeled "1", and the upper half labeled "2". $MA2 can also be used for obtain subtotals of high and low performers. Click on "Display Plot' To make the output easier to understand, type into the codes box: ML 11 (Male = 1, Lower ability = 1) MH 12 (Male = 1, Higher ability = 2) FL 21 (Female = 2, Lower ability = 1) FH 22 (Female = 2, Higher ability = 2) Click OK	
97.	Click on "Entry + Label" to identify points on the Excel plot.	
98.	The Excel "DIF Measure" plot shows that Item 3. INTEDUC is about 1 logit more difficult for the MH (Males - Higher Ability) then for the other classification-groups.	
99.	Table 30.1 shows the relative difficulty of item 3 for (red arrows) FH (Females-Higher ability) and FL, and also MH and ML (after adjusting for the abilities of the groups). *Red box:* The difference in difficulty for the high and low females is .33 logits, but for the high and low males is .66 logits. There is some NUDIF for the females, but twice as much for the males. *Blue box:* Mantel DIF statistics are not computable under these conditions.	
100.	Please close all windows.	

101. **G. Investigating Dimensionality**

102. Look again at interest.txt - it contains 12 items, 6 "interest" and 6 "information". Do they lie along the same dimension, or different dimensions?

Winsteps: analyze interest.txt

Winsteps menu bar: "Output Tables"
Click on "Table 23: dimensionality"



103. **H. Decomposing the Variance in the Data**

104. *Baseline observation with no measure variance:*
Think of an observation, $\mathbf{X_{ni}}$, by person n on item i. Imagine that all the person abilities were the same ability **B** (so that there is no ability variance), and all the item difficulties were the same difficulty, **D** (so that there is no difficulty variance). Then, instead of X_{ni}, **B** and **D** would produce a baseline observation $\mathbf{C_{ni}}$. C_{ni} usually has the same value for every n and i. We choose **B** and **D** such that $\Sigma(C_{ni}) = \Sigma(X_{ni})$ across all the data, so that the total score of the data is the has not changed.

Table 23.0 Variance Table displays in a NotePad window. It is crammed with diagnostic information about this dataset.

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

		-- Empirical --		Modeled
Total raw variance in observations	=	19.4	100.0%	100.0%
Raw variance explained by measures	=	7.4	38.1%	37.9%
Raw variance explained by persons	=	4.6	23.5%	23.4%
Raw Variance explained by items	=	2.8	14.5%	14.4%
Raw unexplained variance (total)	=	12.0	61.9% 100.0%	62.1%

105. Green box: The total raw (empirical) variance in the data (and in Table 23.0) is
TEV = $\Sigma(Xni-Cni)^2$ = 100% in the Empirical (green) column

106. *Expected value, Residual and residual variance:*
According to the Rasch model, the expected value of X_{ni} is $\mathbf{E_{ni}}$ based on person ability B_n and item difficulty D_i. The residual $\mathbf{R_{ni}} = X_{ni} - E_{ni}$ is the part of the observation not explained by the Rasch measures. The raw unexplained (empirical) variance is
REV = $\Sigma(Rni)^2$ = 61.9% in the Empirical (green) column

Thus the raw (empirical) variance explained by the measures is
EEV = TEV - REV = 38.1% in the Empirical (green) column

Based on the variance of the person measures $\{B_n\}$ and the item measures $\{D_i\}$, EEV (38.1%) is split into the empirical variance explained by the persons = *23.5% in the Empirical (green column)*
and the empirical variance explained by the items = *14.5% in the Empirical (green column)*

107. ***The Rasch measures explain only 38.1% of the variance in the data! Can this be true? Surely they should explain 80% or 90% of the variance in the data!***
We are continually surprised to discover that randomness dominates the world around us. As the ancient philosopher wrote: "Time and chance happens to them all." (Eccl. 9:11).
A recent example is the *Netflix Prize Challenge*. $1,000,000 was the prize to the analyst who could best explain the variance in a set of ratings. After 3 years of effort using the most sophisticated statistical models and huge computer power, the winning team could explain only 42% of the variance in the data!

108. ***But what if the data fit the Rasch model perfectly?***
The raw (modeled) variance explained by the measures would be EMV = *Σ(Eni-Cni)2 = 37.9% in the Modeled (orange) column*

The Rasch-model-predicted random (so unexplained) variance of the observation, X_{ni}, around its expectation, E_{ni}, is $\mathbf{W_{ni}}$. So the raw unexplained (modeled) variance in the data would be RMV = $\Sigma(W_{ni})$ *=100% - 37.9% = 62.1% in the Modeled (orange) column*

Thus the total raw (modeled) variance in the data would be TMV = EMV + RMV = *100% in the Modeled (orange) column.*

109. If the *Empirical* and *Modeled* columns look very different, then the variance in the data has departed considerably from Rasch expectations. *The reasons for the difference need to be discovered.*
One reason could be that some parameter values are anchored (fixed).

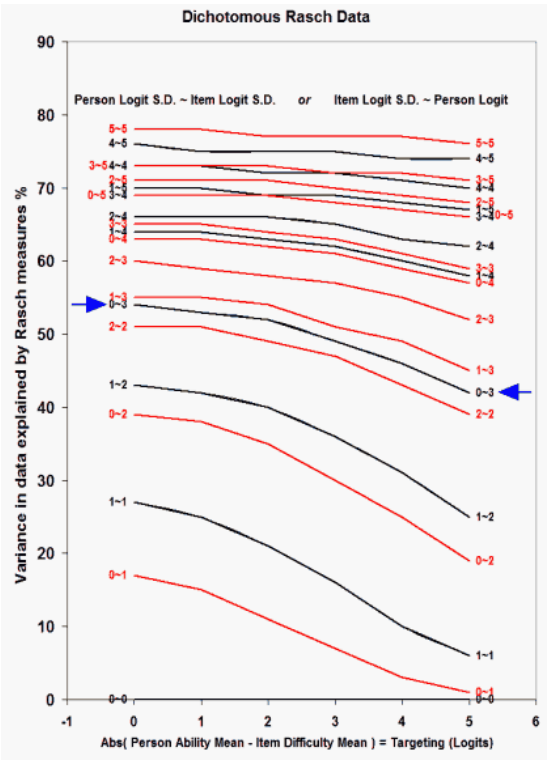
110. *Question:* What percentage of the variance in the data must be explained by the Rasch measures for the data to be unidimensional in the Rasch sense?
Answer: It depends on the dataset. The percentage is *(red box)* "raw variance explained by measures" (Modeled).

"Explained variance" (Modeled):

exam1.txt	71.1% (Knox Cube Test)
example0.txt	50.8% (Liking for Science)
interest.txt	37.5% (NSF data)
agree.txt	30.0% (NSF data)
exam5.txt	29.5% (CAT test)

111. This Figure shows the expected "variance explained" for different person S.D., item S.D. and targeting (person mean - item mean).

Notice that for narrow person and item distributions, the "variance explained" is predicted to be less than 50%.

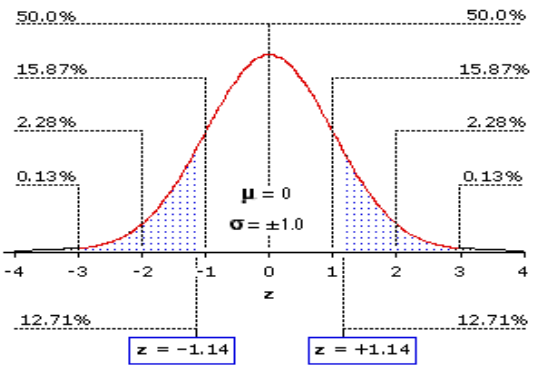


112. **I. Decomposing the Unexplained Variance**

113. If the data fit the Rasch model, then the data will be "locally independent". All the shared variance in the data will be explained by the Rasch measures, and all the remaining variance will be uncorrelated randomness. Then the standardized residuals = residual / (model standard deviation of the observation around its expectation),

$$S_{ni} = R_{ni} / \sqrt{(W_{ni})}$$

This will approximate a unit-normal distribution, and the correlations of the standardized residuals across items will approximate 0. "Standardized" = "Normalized"



Unit-normal distribution

114. This suggests an investigation. Do the standardized residuals have their Rasch-predicted form? Here we follow the exhortation of Karl Popper, the philosopher of science,

> "Once put forward, none of our 'anticipations' are dogmatically upheld. Our method of research is not to defend them, in order to prove how right we were. On the contrary, we try to overthrow them. Using all the weapons of our logical, mathematical, and technical armory, **we try to prove that our anticipations were false** - in order to put forward, in their stead, new unjustified and unjustifiable anticipations." *(The Logic of Scientific Discovery, 1992, p. 278-279)*

115. *So how can we **falsify** Rasch's anticipations about the correlations of the standardized residuals?*
Principal Components Analysis (PCA) of a matrix containing those inter-item correlations!
Unlike Common Factor Analysis (CFA), PCA asserts that all the variance underlying the correlations is shared between the items. None of it is unique to an item. PCA does this by setting all the item auto-correlations (the main diagonal of the correlation matrix) to 1.0.

Correlation matrix of normal deviates for 12 items

1.00	-.09	-.08	-.07	-.10	-.08	-.08	-.08	-.10	-.09	-.09	-.10
-.09	1.00	-.10	-.08	-.09	-.09	-.09	-.09	-.08	-.09	-.09	-.09
-.08	-.10	1.00	-.08	-.10	-.09	-.10	-.08	-.09	-.09	-.08	-.09
-.07	-.08	-.08	1.00	-.07	-.07	-.07	-.07	-.07	-.06	-.08	-.07
-.10	-.09	-.10	-.07	1.00	-.10	-.10	-.10	-.09	-.10	-.09	-.11
-.08	-.09	-.09	-.07	-.10	1.00	-.10	-.09	-.10	-.10	-.09	-.09
-.08	-.09	-.10	-.07	-.10	-.10	1.00	-.10	-.09	-.10	-.09	-.08
-.08	-.09	-.08	-.07	-.10	-.09	-.10	1.00	-.10	-.10	-.11	-.09
-.10	-.08	-.09	-.07	-.09	-.10	-.09	-.10	1.00	-.10	-.09	-.09
-.09	-.09	-.09	-.06	-.10	-.10	-.10	-.10	-.10	1.00	-.09	-.10
-.09	-.09	-.08	-.08	-.09	-.09	-.09	-.11	-.09	-.09	1.00	-.09
-.10	-.09	-.09	-.07	-.11	-.09	-.08	-.09	-.09	-.10	-.09	1.00

116. Consequently, PCA will do its best to construct latent components (secondary dimensions) with which the item residuals correlate and which explain as much as possible of the item variances. But if the inter-item residual-correlations do accord with Rasch-model anticipations, then PCA will find no meaningful components. Its results will be the same as a PCA analysis based on random data.

PCA of the correlation matrix of normal deviates

Raw unexplained variance (total) =	12.0
Unexplned variance in 1st contrast =	1.1
Unexplned variance in 2nd contrast =	1.1
Unexplned variance in 3rd contrast =	1.1
Unexplned variance in 4th contrast =	1.1

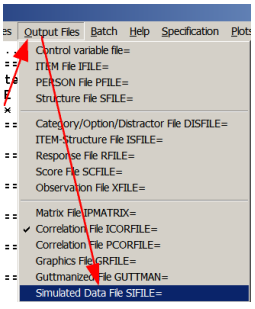
Each item contributes roughly 1 unit of variance

117. If you want to experiment with the decomposition of random residuals for data which fit the Rasch model then simulate a dataset,

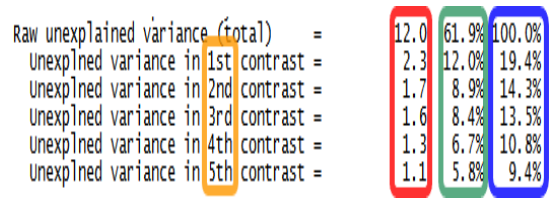
SIFILE = simulated-data.txt

and DATA= to submit it for analysis:

```
Extra specifications (if
data=simulated-data.txt
```



118. Let's return to interest.txt, *Table 23. Item: dimensionality.* 61.9% of the variance in the data was not explained by the Rasch measures. Do the standardized residuals resemble random normal deviates, uncorrelated across the items, as the Rasch model predicts?



119. *Orange box:* Principal Components Analysis decomposes the correlation matrix to find the latent (imagined) component (or factor) with which the items have the highest correlations (factor loadings). After finding that first component, the effect of that component is removed from the correlation matrix, and the PCA looks at the revised correlation matrix to find the second component. Winsteps continues this process to report the first 5 components (when they are estimable).

"Contrast" - to remind us that the components come from the residuals, not from the original data, they are called "Contrasts", also because they are interpreted by contrasting the items with positive and negative correlations (loadings) on the Contrast. The Rasch dimension is extracted from the data *before* the PCA of residuals is performed.

Red box: There are 12 items. The main diagonal of the correlation matrix is set to 1.00, so that each item is modeled to contribute one unit of variance. So there are 12 units of variance in the correlation matrix. Its total "eigenvalue" is 12. Each component explains some of this variance. In this example, the first component has an eigenvalue of 2.3. It explains 2.3 items' worth of variance in the residuals.

Blue box: The percentage of "unexplained variance" for each component is shown.

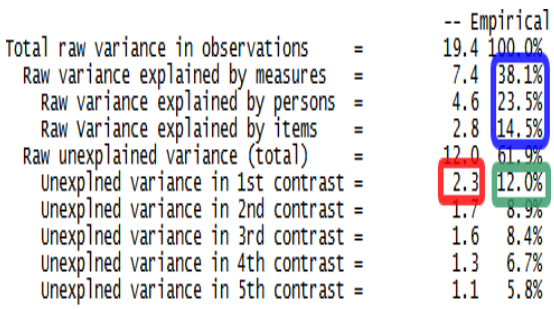
Green box: The percentage of the total variance in the data for each component is shown.

120. *Are these data unidimensional or multidimensional?*

Red box: A secondary dimension in the data must explain at least 2 items' worth of variance:

i) Unless a component has the **strength of at least 2 items** it may be merely due to an idiosyncratic item. We conceptualize a dimension in the data to require more than one item. *The 1st Contrast has the strength of 2.3 items.*

ii) Simulation studies indicate that eigenvalues approach 2.0 by chance: http://www.rasch.org/rmt/rmt191h.htm



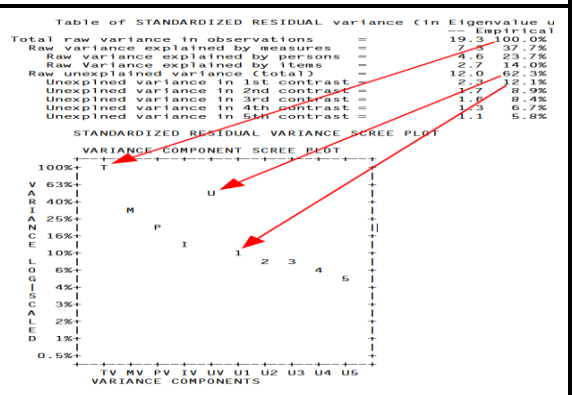
121. *Blue box:* This shows the variance explained by the Rasch measures. It is 38.1%.

Green box: The variance explained by the 1st Contrast is 12.0%. The first sub-dimension is 1/3rd of the Rasch dimension, and almost the same size as the variance explained by the item difficulties (14.5%). We definitely need to investigate further.

122. In Table 23.0, the scree plot shows the variance table in graphical form. The y-axis is *logarithmically-scaled* so that differences between small contrasts are visible.

This plot helps us to picture the relative sizes of the variance components. It is much easier to think about pictures than about tables of numbers.

Whoever reads all those massive tables of numbers printed in the academic Journals?

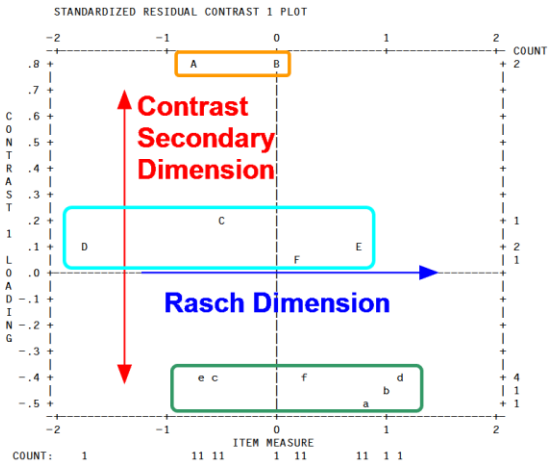


123. *Speculation:* The 12 items are 6 interest items (red box) and 6 information items (blue box), so my guess is that the first contrast in the data is between some of the *red* items and some of the *blue* items.
What is your guess?
In Science, we learn more if we guess at (anticipate) the result of an experiment before we conduct it. Then we know if the experiment confirms or challenges what we suspect. Either way, *we learn* ....

INTSCI
INTTECH
INTEDUC
INTMED
INTSPACE
INTDFNS
INFSCI
INFTECH
INFEDUC
INFMED
INFSPACE
INFDFNS

124. Scroll down to Table 23.2.
The plot tells us the dimensionality story!
The x-axis is the Rasch dimension (in logits), the latent trait, with easy items on the left and hard items on the right.
The y-axis is the Contrast "dimension". The numbers are factor loadings (correlations with the latent component).
At the top are two items, labeled A and B with large positive loadings. Winsteps labels items from the top of the 1st Contrast plot downwards: A, B, C, ...
At the bottom are a cluster of items, a,b,...,e,f with not so large negative loadings. Winsteps labels items from the bottom of the 1st Contrast plot upwards: a, b, c, ...
The plot means the same thing if the signs of the loadings are reversed. By convention the biggest loadings are treated as positive.

For a unidimensional scale, the randomness in each item is unique to that item. So we would expect a series of plots with a big loading on one item and very small loadings on all the other items.
Left-right is decided by the item difficulties. So that does not matter statistically, but is often helpful in "seeing" what the meaning of the contrast is.



Something about items A,B contrasts with something about items a,b,c,d,e,f

125. *Now to investigate my speculation about the 1st Contrast..*
In Table 23.3, scroll down to the item list.
I was wrong!
The first Contrast is "Education" against "Science, Technology, Space". "Defense, Medicine" are in a neutral position.
We will investigate this further ....

CON- TRAST	LOADING	INFIT MEASURE	INFIT MNSQ	OUTFIT MNSQ	ENTRY NUMBER	ITEM	G R
1	.79	-.74	1.46	1.56	A 3	INTEDUC	T
1	.78	-.02	1.40	1.41	B 9	INFEDUC	F
1	.21	-.50	1.09	1.12	C 6	INTDFNS	T
1	.12	-1.73	.97	1.00	D 4	INTMED	T
1	.11	.76	1.14	1.15	E 12	INFDFNS	F
1	.04	.21	.90	.91	F 10	INFMED	F
1	-.48	.81	.74	.74	a 7	INFSCI	F
1	-.47	1.01	.81	.82	b 8	INFTECH	F
1	-.40	-.54	.81	.83	c 2	INTTECH	T
1	-.39	1.13	.85	.85	d 11	INFSPACE	F
1	-.38	-.66	.84	.86	e 1	INTSCI	T
1	-.38	.26	.95	.95	f 5	INTSPACE	T

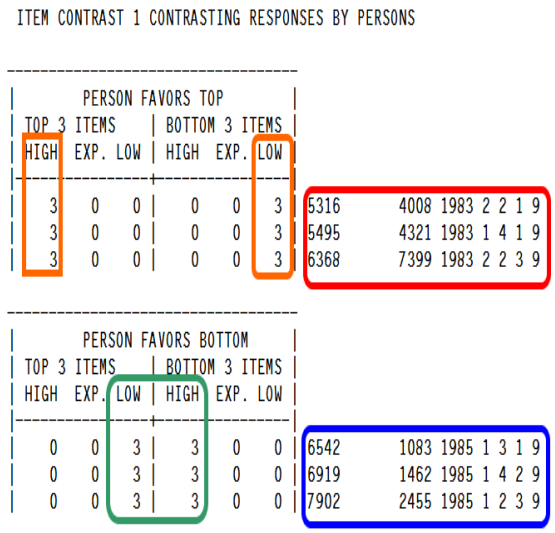
126. *Is there a Contrast between "Information" and "Interest"?*

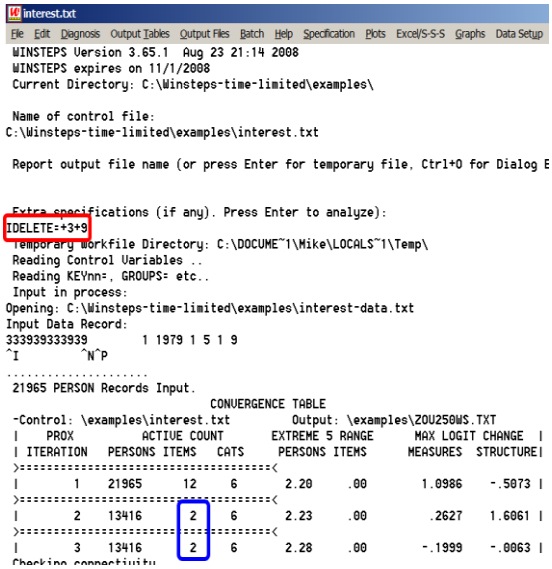
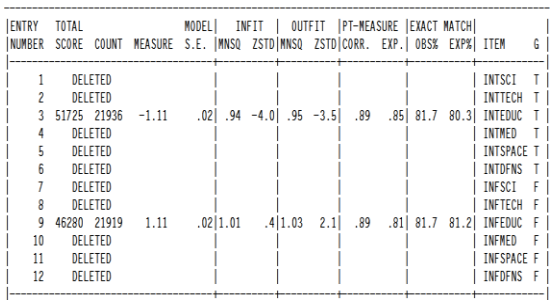
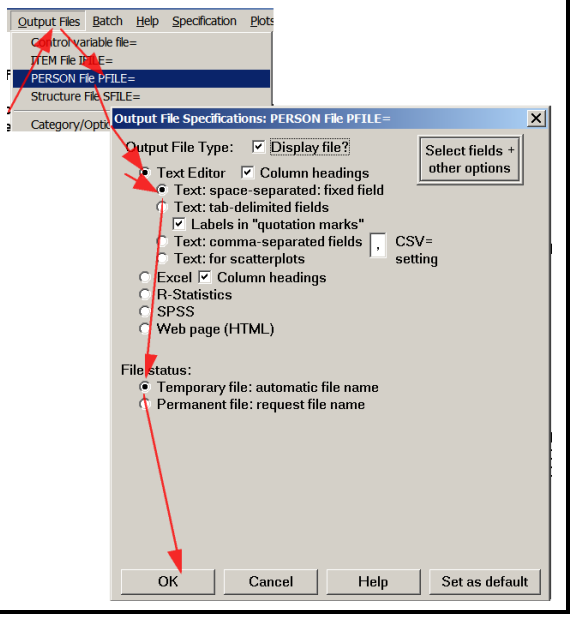
Scroll down to Table 23.13 showing the 3rd Contrast.

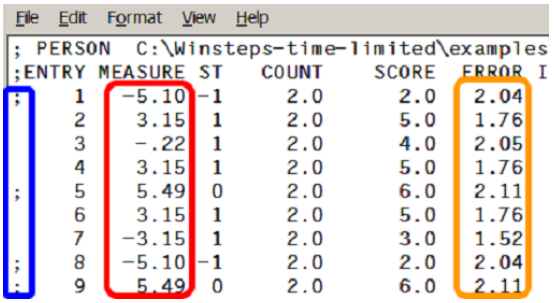
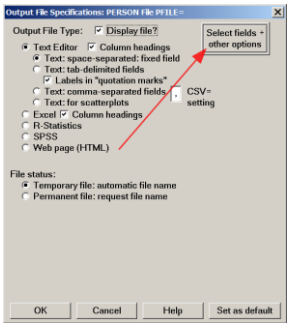
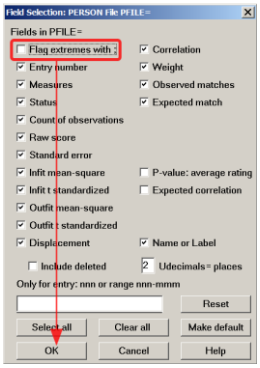
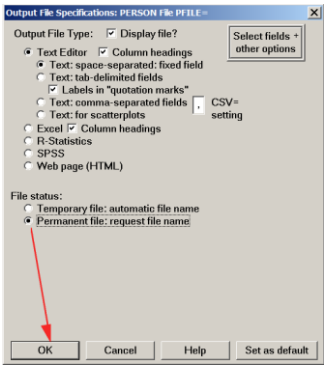
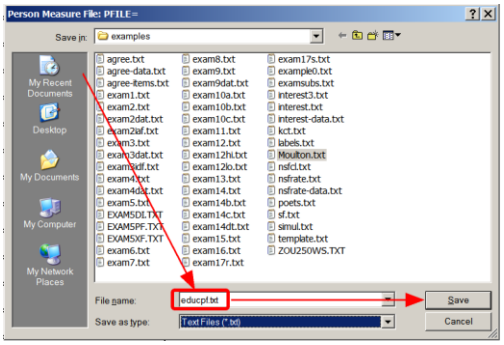
There it is!

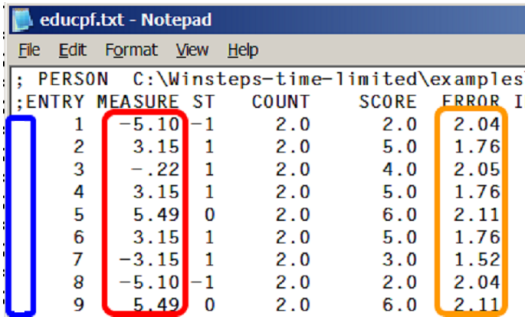
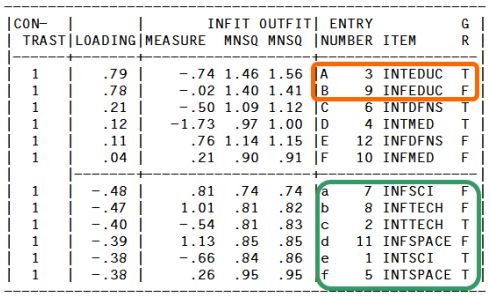
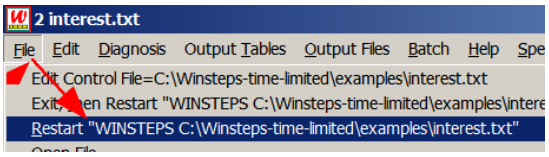
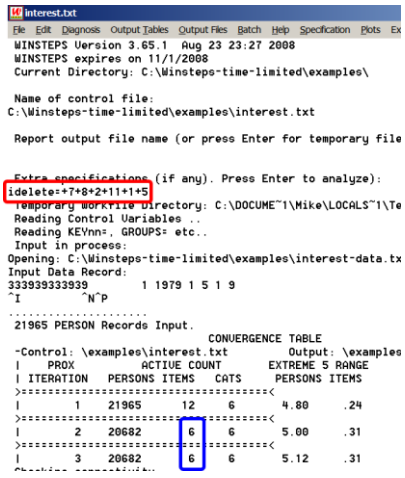
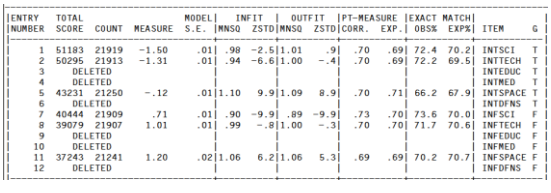
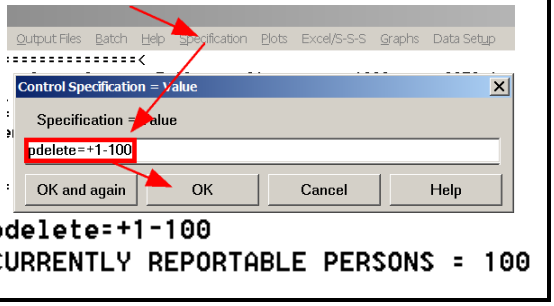
But the size of the Contrast is only 1.6, much less than 2 items. So the Contrast is there, but not enough to prompt us to action.

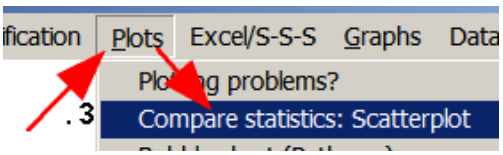
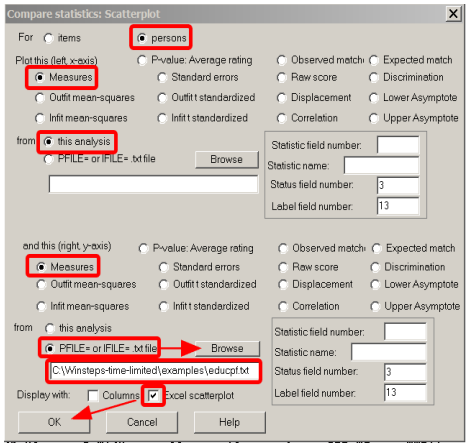
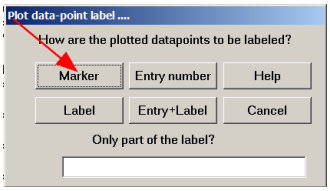
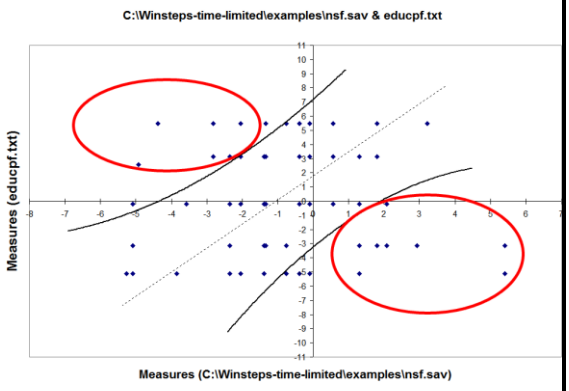
CON-TRAST	LOADING	INFIT MEASURE	INFIT MNSQ	OUTFIT MNSQ	ENTRY NUMBER	ITEM	GR
3	.56	.21	.90	.91	F 10	INFMED	F
3	.49	1.01	.81	.82	b 8	INFTECH	F
3	.46	.81	.74	.74	a 7	INFSCI	F
3	.23	.76	1.14	1.15	E 12	INFOFNS	F
3	.16	1.13	.85	.85	d 11	INFSPACE	F
3	.14	-.02	1.40	1.41	B 9	INFEDUC	F
3	-.51	.26	.95	.95	f 5	INTSPACE	T
3	-.45	-.66	.84	.86	e 1	INTSCI	T
3	-.43	-.54	.81	.83	c 2	INTTECH	T
3	-.26	-.50	1.09	1.12	C 6	INTOFNS	T
3	-.15	-1.73	.97	1.00	D 4	INTMED	T
3	-.13	-.74	1.46	1.56	A 3	INTEDUC	T

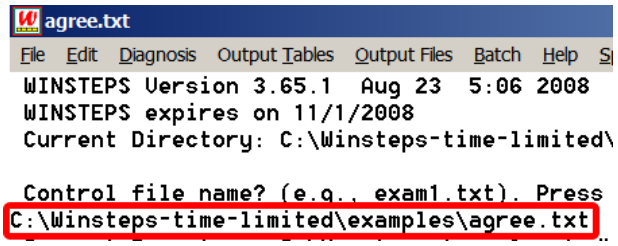
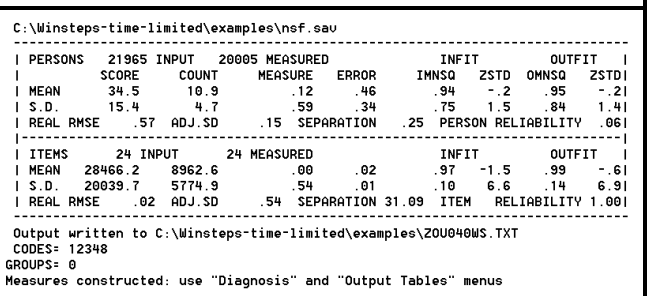
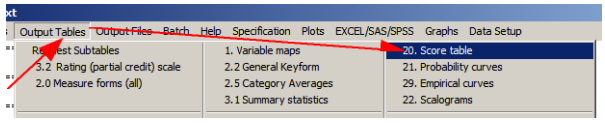
127.	J. Who is Affected by the Sub-dimension?	
128.	Scroll down to Table 23.4. *Red box:* persons who scored *high* on the items at the top of Table 23.2 and *low* on the items at the bottom of Table 23.2. Blue box : persons who scored *low* on the items at the top of Table 23.2 and *high* on the items at the bottom of Table 23.2. This tell us on which persons this Contrast dimension has the greatest impact. If we know something about the persons, this will tells us if this dimension is important to our understanding of person performance. High: scored higher than expected on an item according to the person's overall measure and the item's overall difficulty. Exp: scored as expected Low: scored lower than expected "3" this was observed on all 3 of the items at the top (or bottom) of the plot on Table 23.2	 This shows how the dimensionality impacts the persons. We can investigate this directly in Table 24 "Person dimensionality"
129.	Please close all windows	

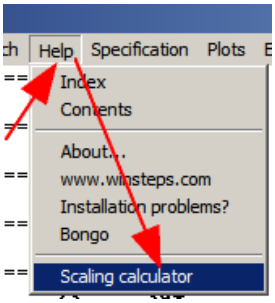
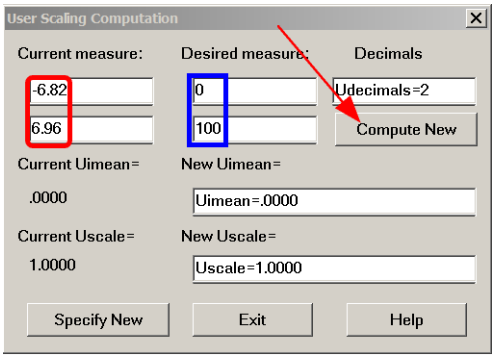
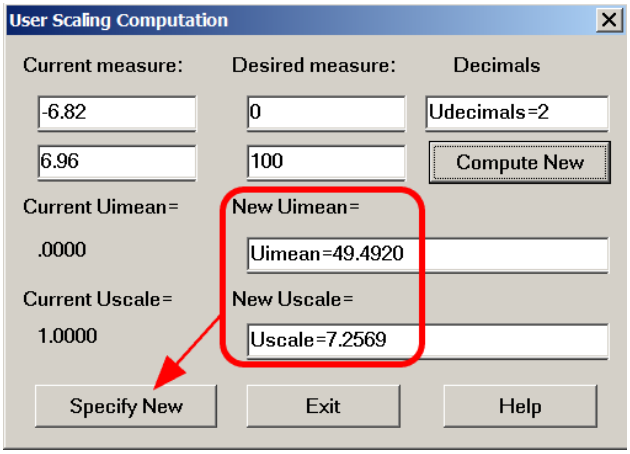
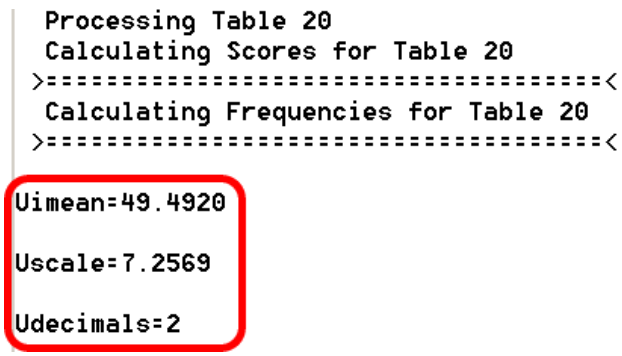
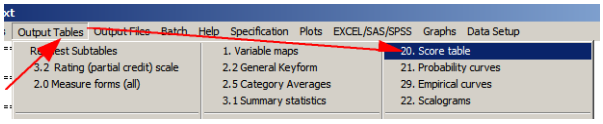
130.	K. Cross-plot of Person Measures	
131.	We have 3 subsets of items in Table 23.2 Let's measure the persons on the opposite subsets. Then cross-plot their measures. First, the two items at the top: Launch Winsteps Open file: interest.txt Report .... *Red box:* Extra specifications: IDELETE=+3+9 *(no spaces)* Press Enter. *Blue box:* only two items analyzed!	
132.	Output Table: Table 14. Items in entry order. Yes, they are the two education items.	
133.	We need to save the person measures on this two-item survey so that we can plot them later. First, let's make sure we will get what we want: Winsteps menu bar: Click on Output Files Click on PERSON file PFILE= Output File dialog box: Click on Text Editor Click on Text: space-separated: fixed fields Click on Permanent file Click on OK	

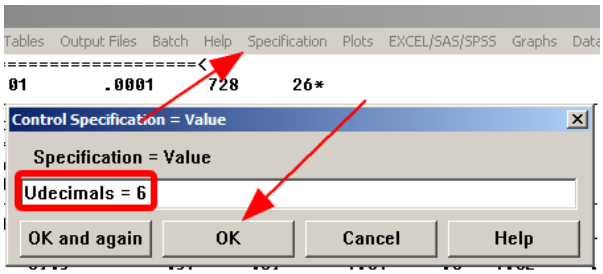
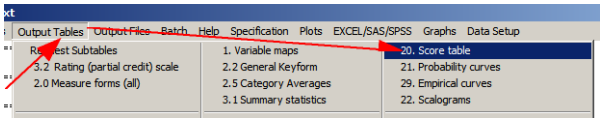
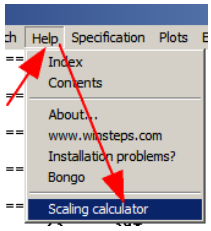
134.	The person-measure file displays in a NotePad window. *Red box:* We will use the measures *Orange box:* We will use the standard errors (the precision of the measures) *Blue box:* we want to include the extreme scores, indicated by ";" in the plots, so we need to produce this file again with the ";"	
135.	Winsteps menu bar: Click on Output Files Click on PERSON file PFILE= Output File dialog box: Click on Select Fields	
136.	Field Selection dialog box: Uncheck the "Flag extremes with ;" box Click on OK	
137.	Output File dialog box: Click on Permanent file Click on OK	
138.	File name dialog box: We need to save the person measures in a data file: Type in: **educpf.txt** Click on Save We have saved the person measures for the two "educational" items in educpf.txt	

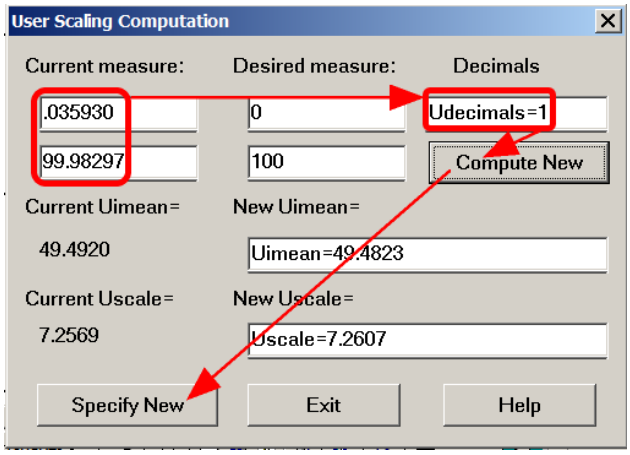
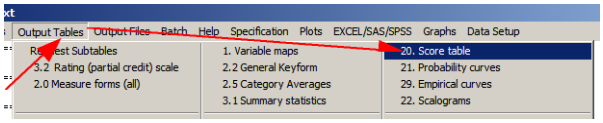
139.	and the format of educpf.txt is what we want.	
140.	Now for the bottom half of Table 23.2. Here is a reminder	
141.	Winsteps menu bar *Restart Winsteps C:\ ....\interest.txt*	
142.	Report .... Press enter Extra specifications: idelete=+7+8+2+11+1+5 Press Enter Blue box: 6 items are analyzed	
143.	*Let's check that this is correct:* Winsteps Menu bar: Output Tables Table 14: items in entry order Yes, they are the Science, Space and Tech items	
144.	There are 21,965 persons we have measured on the two subtests. We could cross-plot them all, but that is much too many. Let's cut down to 100 persons: Winsteps menu bar Click on: Specification menu Type in: PDELETE=+1-100 Click on OK The Winsteps analysis window: PERSONS = 100	

145.	Winsteps Menu bar: Click on Plots menu Click on "Compare statistics: Scatterplot"	
146.	Compare Statistics dialog box: Click on "persons" "Measures" "this analysis" "Measures" "PFILE=" Browse for "educpf.txt" Click on "Excel scatterplot" Click on OK	
147.	Plot data-point label: Click on "Marker" 100 labels or numbers will produce an illegible plot	
148.	Excel scatterplot processing begins	Transferring Scatterplot of 100 data lines to Excel .... Activating Excel interface ... Scatterplot process launched
149.	The Excel plot displays. The measures on the two dimensions are showing considerable disagreement, supporting the finding that these data are not unidimensional.	
150.	That's the end of the Lesson. Congratulations!	

1.	**Appendix 1**. ***User-Friendly*** **Rescaling of Rasch Measures**	
2.	Rasch measures are computed in logits - log-odds units. Usually the average item difficulty is set at 0, so that measures are generally in the range -5 to +5 logits. Measures are often reported with two decimal places. But a measure such as -3.27 is difficult to communicate to a non-specialist audience, so it can be helpful to rescale the measures to more familiar numbers. Then item maps and person summary statistics can be much more meaningful to your audience.	
3.	Launch Winsteps	
4.	Select your Control file: **Your choice!!** *I'm using agree.txt* "Report output ..." Press: Enter key "Extra specifications ..." Press: Enter key	
5.	The Rasch estimation is performed ...	
6.	Table 20 gives the range of person measures: Winsteps Menu Bar Click on: Output Tables Click on: 20. Score table	
7.	Table 20 displays. It shows the measure range for all possible scores on the complete test. In this example, a score of 24 on the test has a measure or -6.82. A score of 120 on the test has a score of 6.96. **We want to convert this range into 0 to 100. So that the reported measures can be understood as "Percent progress along the range of the test".** The same process applies to any other linear rescaling of the measures. The numbers we will need are shown in the green box, but we will compute them ourselves ...	TABLE OF MEASURES ON COMPLETE TEST | SCORE MEASURE S.E. | SCORE MEASURE S.E. | SCORE MEASURE S.E. | | 24 -6.82E 1.84 | 57 -.77 .23 | 90 .87 .24 | | 25 -5.57 1.03 | 58 -.71 .23 | 91 .93 .25 | | 26 -4.83 .74 | 59 -.66 .23 | 92 .99 .25 | | 27 -4.37 .62 | 60 -.61 .23 | 93 1.06 .25 | | 28 -4.02 .55 | 61 -.55 .23 | 94 1.12 .26 | | 29 -3.74 .51 | 62 -.50 .22 | 95 1.19 .26 | | 30 -3.50 .47 | 63 -.45 .22 | 96 1.26 .27 | | 31 -3.29 .45 | 64 -.41 .22 | 97 1.34 .28 | | 32 -3.09 .43 | 65 -.36 .22 | 98 1.42 .29 | | 33 -2.92 .41 | 66 -.31 .22 | 99 1.50 .29 | | 34 -2.76 .39 | 67 -.26 .22 | 100 1.59 .30 | | 35 -2.61 .38 | 68 -.21 .22 | 101 1.68 .31 | | 36 -2.47 .37 | 69 -.17 .22 | 102 1.78 .32 | | 37 -2.34 .35 | 70 -.12 .22 | 103 1.89 .33 | | 38 -2.22 .34 | 71 -.07 .22 | 104 2.01 .34 | | 39 -2.10 .33 | 72 -.02 .22 | 105 2.13 .35 | | 40 -2.00 .32 | 73 .02 .22 | 106 2.26 .37 | | 41 -1.90 .31 | 74 .07 .22 | 107 2.40 .38 | | 42 -1.80 .30 | 75 .11 .22 | 108 2.54 .39 | | 43 -1.71 .30 | 76 .16 .22 | 109 2.69 .40 | | 44 -1.62 .29 | 77 .21 .22 | 110 2.86 .41 | | 45 -1.54 .28 | 78 .26 .22 | 111 3.03 .42 | | 46 -1.46 .28 | 79 .30 .22 | 112 3.21 .44 | | 47 -1.39 .27 | 80 .35 .22 | 113 3.41 .45 | | 48 -1.32 .27 | 81 .40 .22 | 114 3.63 .48 | | 49 -1.25 .26 | 82 .45 .22 | 115 3.87 .51 | | 50 -1.18 .26 | 83 .50 .22 | 116 4.16 .56 | | 51 -1.12 .25 | 84 .55 .23 | 117 4.50 .63 | | 52 -1.05 .25 | 85 .60 .23 | 118 4.97 .75 | | 53 -.99 .24 | 86 .65 .23 | 119 5.71 1.03 | | 54 -.93 .24 | 87 .70 .23 | 120 6.96E 1.84 | | 55 -.88 .24 | 88 .76 .23 | | 56 -.82 .24 | 89 .81 .24 | CURRENT VALUES, UMEAN=.000 USCALE=1.000 TO SET MEASURE RANGE AS 0-100, UMEAN=49.482 USCALE=7.261

8.	Winsteps menu bar Click on: Help menu Click on: Scaling calculator	
9.	In the Scaling window, enter in the input boxes. the Current measures (in my example) are -6.82 and 6.96. The Desired measures are 0 and 100 Notice that the Current Uimean= is .0000 and the Current Uscale= is 1.0000 Click on: Compute New	
10.	After click on "Compute New", new values of Uimean= 49.4920 and Uscale=7.2568 (in this example) are shown Click on: Specify New	
11.	The Winsteps Analysis window shows that the new values have been actioned.	
12.	Table 20 gives the revised range of person measures: Winsteps Menu Bar Click on: Output Tables Click on: 20. Score table	

13.	Table 20 (in my example) now shows that range from .04 to 99.98 - the peculiar values are because we were using the rounded values from the original Table 20 for the rescaling. We might want a more exact range of .0 to 100.0. To do this, we need to display more decimal places before we do the rescaling.	TABLE OF MEASURES ON COMPLETE TEST SCORE MEASURE S.E. | SCORE MEASURE S.E. | SCORE MEASURE S.E. 24 .04E 13.35 | 57 43.94 1.69 | 90 55.82 1.75 25 9.04 7.45 | 58 44.33 1.68 | 91 56.25 1.78 26 14.47 5.41 | 59 44.71 1.66 | 92 56.69 1.81 27 17.81 4.53 | 60 45.09 1.65 | 93 57.15 1.84 28 20.31 4.02 | 61 45.46 1.64 | 94 57.63 1.88 29 22.35 3.69 | 62 45.83 1.63 | 95 58.12 1.92 30 24.10 3.44 | 63 46.19 1.62 | 96 58.65 1.97 31 25.64 3.26 | 64 46.55 1.61 | 97 59.19 2.02 32 27.04 3.11 | 65 46.91 1.60 | 98 59.77 2.07 33 28.31 2.98 | 66 47.26 1.59 | 99 60.38 2.13 34 29.48 2.86 | 67 47.60 1.59 | 100 61.03 2.20 35 30.57 2.76 | 68 47.95 1.58 | 101 61.71 2.27 36 31.58 2.66 | 69 48.29 1.58 | 102 62.45 2.34 37 32.52 2.57 | 70 48.63 1.57 | 103 63.23 2.42 38 33.40 2.48 | 71 48.97 1.57 | 104 64.06 2.50 39 34.22 2.41 | 72 49.31 1.57 | 105 64.94 2.57 40 35.00 2.34 | 73 49.65 1.57 | 106 65.88 2.65 41 35.73 2.27 | 74 49.99 1.57 | 107 66.88 2.73 42 36.42 2.21 | 75 50.33 1.57 | 108 67.93 2.80 43 37.08 2.16 | 76 50.66 1.57 | 109 69.04 2.88 44 37.70 2.10 | 77 51.00 1.57 | 110 70.22 2.96 45 38.30 2.06 | 78 51.35 1.58 | 111 71.46 3.05 46 38.87 2.01 | 79 51.69 1.58 | 112 72.79 3.16 47 39.41 1.97 | 80 52.04 1.59 | 113 74.23 3.30 48 39.93 1.93 | 81 52.39 1.60 | 114 75.81 3.48 49 40.44 1.89 | 82 52.74 1.61 | 115 77.59 3.71 50 40.92 1.86 | 83 53.10 1.62 | 116 79.65 4.04 51 41.39 1.83 | 84 53.47 1.63 | 117 82.16 4.54 52 41.84 1.80 | 85 53.84 1.65 | 118 85.52 5.42 53 42.29 1.78 | 86 54.22 1.66 | 119 90.96 7.46 54 42.71 1.75 | 87 54.60 1.68 | 120 99.98E 13.36 55 43.13 1.73 | 88 55.00 1.70 | 56 43.54 1.71 | 89 55.40 1.73 |
14.	Winsteps menu bar Click on: Specification Enter: Udecimals = 6 This will display 6 decimal places Click on: OK	
15.	Table 20 gives the revised decimal places of person measures: Winsteps Menu Bar Click on: Output Tables Click on: 20. Score table	
16.	The range is shown as .035930 to 99.98297 We want 0 to 100.	TABLE OF MEASURES ON COMPLETE TEST SCORE MEASURE S.E. | SCORE MEASURE S.E. | SCORE MEASURE S.E. 24 .035930 3.35346 | 57 43.93926 .693586 | 90 55.81854 .750385 25 9.042609 .448342 | 58 44.33059 .677446 | 91 56.24707 .778044 52 41.84460 .801210 | 85 53.83887 .648063 | 118 85.52468 .416787 53 42.28554 .775549 | 86 54.21690 .664376 | 119 90.96492 .456898 54 42.71407 .752125 | 87 54.60291 .682601 | 120 99.98297 3.35758 55 43.13197 .730762 | 88 54.99734 .702886 | 56 43.53968 .711301 | 89 55.40239 .725427 |
17.	Winsteps menu bar Click on: Help menu Click on: Scaling calculator	

18.	And repeat the previous procedure. Enter the Current Measures: *easiest is to copy-and-paste from Table 20* Change the number of decimals back to 1. Udecimals = 1 *(so we don't overwhelm our audience with our decimal places)* Click on: Compute New *Uimean= and Uscale= values change* Click on Specify New	
19.	Table 20 gives the revised decimal places of person measures: Winsteps Menu Bar Click on: Output Tables Click on: 20. Score table	
20.	The range of rescaled measures is now .0 to 100.0 as we desired. This is much more audience-friendly than the original range of measures. All the properties of logits have been maintained except that the inference from these measures to probabilities is: logits = log-odds (probabilities) user-scaled measures / USCALE= value = log-odds(probabilities)	TABLE OF MEASURES ON COMPLETE TEST | SCORE MEASURE S.E. | SCORE MEASURE S.E. | SCORE MEASURE S.E. | | 24 .0E 13.4 | 57 43.9 1.7 | 90 55.8 1.8 | | 25 9.0 7.5 | 58 44.3 1.7 | 91 56.2 1.8 | | 26 14.4 5.4 | 59 44.7 1.7 | 92 56.7 1.8 | | 27 17.8 4.5 | 60 45.1 1.7 | 93 57.1 1.8 | | 28 20.3 4.0 | 61 45.5 1.6 | 94 57.6 1.9 | | 29 22.3 3.7 | 62 45.8 1.6 | 95 58.1 1.9 | | 30 24.1 3.4 | 63 46.2 1.6 | 96 58.6 2.0 | | 31 25.6 3.3 | 64 46.5 1.6 | 97 59.2 2.0 | | 32 27.0 3.1 | 65 46.9 1.6 | 98 59.8 2.1 | | 33 28.3 3.0 | 66 47.2 1.6 | 99 60.4 2.1 | | 34 29.5 2.9 | 67 47.6 1.6 | 100 61.0 2.2 | | 35 30.5 2.8 | 68 47.9 1.6 | 101 61.7 2.3 | | 36 31.6 2.7 | 69 48.3 1.6 | 102 62.4 2.3 | | 37 32.5 2.6 | 70 48.6 1.6 | 103 63.2 2.4 | | 38 33.4 2.5 | 71 49.0 1.6 | 104 64.1 2.5 | | 39 34.2 2.4 | 72 49.3 1.6 | 105 64.9 2.6 | | 40 35.0 2.3 | 73 49.6 1.6 | 106 65.9 2.7 | | 41 35.7 2.3 | 74 50.0 1.6 | 107 66.9 2.7 | | 42 36.4 2.2 | 75 50.3 1.6 | 108 67.9 2.8 | | 43 37.1 2.2 | 76 50.7 1.6 | 109 69.0 2.9 | | 44 37.7 2.1 | 77 51.0 1.6 | 110 70.2 3.0 | | 45 38.3 2.1 | 78 51.3 1.6 | 111 71.5 3.1 | | 46 38.8 2.0 | 79 51.7 1.6 | 112 72.8 3.2 | | 47 39.4 2.0 | 80 52.0 1.6 | 113 74.2 3.3 | | 48 39.9 1.9 | 81 52.4 1.6 | 114 75.8 3.5 | | 49 40.4 1.9 | 82 52.7 1.6 | 115 77.6 3.7 | | 50 40.9 1.9 | 83 53.1 1.6 | 116 79.7 4.0 | | 51 41.4 1.8 | 84 53.5 1.6 | 117 82.2 4.5 | | 52 41.8 1.8 | 85 53.8 1.6 | 118 85.5 5.4 | | 53 42.3 1.8 | 86 54.2 1.7 | 119 91.0 7.5 | | 54 42.7 1.8 | 87 54.6 1.7 | 120 100.0E 13.4 | | 55 43.1 1.7 | 88 55.0 1.7 | | 56 43.5 1.7 | 89 55.4 1.7 |