A User's Guide to

WINSTEPS

MINISTEP

Rasch-Model
Computer Programs

John M. Linacre

www.winsteps.com
# Table of Contents

2. Control Variable Index ...............................................................................................................................1
3. Control Variable Index by function ................................................................................................................3
4. Control variables from Specification menu .......................................................................................................6
5. Output Table Index .........................................................................................................................................8
6. Output Graphs and Plots Index ........................................................................................................................11
7. Output Files Index ..........................................................................................................................................12
8. Rasch analysis and WINSTEPS .......................................................................................................................12
9. References ..................................................................................................................................................14
10. About the Users’ guide ...................................................................................................................................14
11. Getting further help .....................................................................................................................................15
12. What is supplied ........................................................................................................................................15
13. Installation instructions for WINSTEPS .........................................................................................................15
14. Starting WINSTEPS in Windows ...................................................................................................................16
15. Using Winsteps under Windows ....................................................................................................................18
16. Stopping WINSTEPS ...................................................................................................................................19
17. Uninstalling WINSTEPS ................................................................................................................................20
18. Menu bar ....................................................................................................................................................20
19. Batch menu ................................................................................................................................................20
20. Data Setup menu .........................................................................................................................................21
21. Diagnosis menu ...........................................................................................................................................21
22. Edit menu ....................................................................................................................................................21
23. File menu .....................................................................................................................................................22
24. Graphs menu ...............................................................................................................................................23
25. Help menu ..................................................................................................................................................24
26. Output Files menu .......................................................................................................................................25
27. Output Tables menu ....................................................................................................................................26
28. Plots menu ..................................................................................................................................................27
29. SAS/SPSS menu .........................................................................................................................................28
30. Specification menu ......................................................................................................................................28
31. Control and data file setup window ............................................................................................................29
32. Reading in a pre-existent control file ............................................................................................................29
33. Data display ...............................................................................................................................................30
34. Item labels ...................................................................................................................................................31
35. Category labels ..........................................................................................................................................31
36. Do-it-yourself control and data file construction ..........................................................................................32
37. Control file and template.txt .......................................................................................................................35
38. Data file.....................................................................................................................................................36
39. Data from Excel and other spreadsheets .........................................................................................................37
40. Data from SAS files ...................................................................................................................................37
41. Data from SPSS files ..................................................................................................................................39
42. Data from STATA files ................................................................................................................................41
43. Data file with other delimiters .......................................................................................................................43
44. Example 0: Rating scale data: The Liking for Science data ..........................................................................43
45. Example 1: Dichotomous data: Simple control file with data included .......................................................45
46. Example 2: Control and anchor files ............................................................................................................47
47. Example 3: Item recoding and item deletion ................................................................................................48
48. Example 4: Selective item recoding ...............................................................................................................48
49. Example 5: Scoring key for items, also CAT responses ................................................................................49
50. Example 6: Keys in data record FORMAT ..................................................................................................50
51. Example 7: A partial credit analysis .............................................................................................................51
52. Example 8: Items with various rating scale models ......................................................................................52
53. Example 9: Grouping and modeling items ...................................................................................................52
54. Example 10: Combining tests with common items .....................................................................................53
55. Example 11: Item responses two characters wide .......................................................................................56
56. Example 12: Comparing high and low samples with rating scales ............................................................56
57. Example 13: Paired comparisons as the basis for measurement .................................................................59
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>117.</td>
<td>ISORT column within item name for alphabetical sort in Table 15 ........................................</td>
</tr>
<tr>
<td>118.</td>
<td>ISUBTOTAL columns within item label for subtotals in Table 27 ...........................................</td>
</tr>
<tr>
<td>119.</td>
<td>ITEM title for item labels ........................................................................................................</td>
</tr>
<tr>
<td>120.</td>
<td>ITEM1 column number of first response .....................................................................................</td>
</tr>
<tr>
<td>121.</td>
<td>ITLEN maximum length of item label ..........................................................................................</td>
</tr>
<tr>
<td>122.</td>
<td>IVALUEx recoding of data .............................................................................................................</td>
</tr>
<tr>
<td>123.</td>
<td>IWEIGHT item (variable) weighting ...............................................................................................</td>
</tr>
<tr>
<td>124.</td>
<td>KEYn scoring key ..........................................................................................................................</td>
</tr>
<tr>
<td>125.</td>
<td>KEYFROM location of KEYn ............................................................................................................</td>
</tr>
<tr>
<td>126.</td>
<td>KEYSCR reassign scoring keys ......................................................................................................</td>
</tr>
<tr>
<td>127.</td>
<td>LCONV logit change at convergence ..............................................................................................</td>
</tr>
<tr>
<td>128.</td>
<td>LINLEN length of printed lines in Tables 7, 10-16, 22 ..................................................................</td>
</tr>
<tr>
<td>129.</td>
<td>LOCAL locally restandardize fit statistics .....................................................................................</td>
</tr>
<tr>
<td>130.</td>
<td>LOGFILE accumulates control files ...............................................................................................</td>
</tr>
<tr>
<td>131.</td>
<td>LOWADJ correction for bottom rating scale categories .................................................................</td>
</tr>
<tr>
<td>132.</td>
<td>MAKEKEY construct MCQ key ........................................................................................................</td>
</tr>
<tr>
<td>133.</td>
<td>MATRIX correlation output format ................................................................................................</td>
</tr>
<tr>
<td>134.</td>
<td>MAXPAGE the maximum number of lines per page .......................................................................</td>
</tr>
<tr>
<td>135.</td>
<td>MFORMS reformat input data ........................................................................................................</td>
</tr>
<tr>
<td>136.</td>
<td>MHSLICE Mantel-Haenszel slice width ...........................................................................................</td>
</tr>
<tr>
<td>137.</td>
<td>MISSCORE scoring of missing data codes .....................................................................................</td>
</tr>
<tr>
<td>138.</td>
<td>MJMILE maximum number of JMILE iterations ............................................................................</td>
</tr>
<tr>
<td>139.</td>
<td>MNSQ show mean-square or standardized fit statistics ................................................................</td>
</tr>
<tr>
<td>140.</td>
<td>MODELS assigns model types to items ..........................................................................................</td>
</tr>
<tr>
<td>141.</td>
<td>MODFROM location of MODELS ..................................................................................................</td>
</tr>
<tr>
<td>142.</td>
<td>MPROX maximum number of PROX iterations ................................................................................</td>
</tr>
<tr>
<td>143.</td>
<td>MRANGE half-range of measures on plots .....................................................................................</td>
</tr>
<tr>
<td>144.</td>
<td>NAME1 first column of person label ............................................................................................</td>
</tr>
<tr>
<td>145.</td>
<td>NAMLEN length of person label ..................................................................................................</td>
</tr>
<tr>
<td>146.</td>
<td>NAMLMNP name length on map for Tables 1, 12, 16 ...................................................................</td>
</tr>
<tr>
<td>147.</td>
<td>NEWSCORE recoding values ..........................................................................................................</td>
</tr>
<tr>
<td>148.</td>
<td>NI number of items .......................................................................................................................</td>
</tr>
<tr>
<td>149.</td>
<td>NORMAL normal distribution for standardizing fit ......................................................................</td>
</tr>
<tr>
<td>150.</td>
<td>OSORT option/distractor sort ......................................................................................................</td>
</tr>
<tr>
<td>151.</td>
<td>OUTFIT sort misfits on infit or outfit ..........................................................................................</td>
</tr>
<tr>
<td>152.</td>
<td>PFILE person anchor file .............................................................................................................</td>
</tr>
<tr>
<td>153.</td>
<td>PAIRED correction for paired comparison data .........................................................................</td>
</tr>
<tr>
<td>154.</td>
<td>PANCHQU anchor persons interactively ......................................................................................</td>
</tr>
<tr>
<td>155.</td>
<td>PCORFIL person residual correlation file ..................................................................................</td>
</tr>
<tr>
<td>156.</td>
<td>PDELETE person one-line item deletion .....................................................................................</td>
</tr>
<tr>
<td>157.</td>
<td>PDELOQU delete persons interactively .......................................................................................</td>
</tr>
<tr>
<td>158.</td>
<td>PDFILE person deletion file .........................................................................................................</td>
</tr>
<tr>
<td>159.</td>
<td>PDROPEXTREME drop persons with extreme scores ....................................................................</td>
</tr>
<tr>
<td>160.</td>
<td>PERSON title for person labels ....................................................................................................</td>
</tr>
<tr>
<td>161.</td>
<td>PFILE person output file .............................................................................................................</td>
</tr>
<tr>
<td>162.</td>
<td>PMAP person label on person map: Tables 1, 16 ......................................................................</td>
</tr>
<tr>
<td>163.</td>
<td>PRCOMP residual type for principal components analyses in Tables 23, 24 ..........................</td>
</tr>
<tr>
<td>164.</td>
<td>PSELECT person selection criterion ...........................................................................................</td>
</tr>
<tr>
<td>165.</td>
<td>PSORT column within person label for alphabetical sort in Table 19 .......................................</td>
</tr>
<tr>
<td>166.</td>
<td>PSUBTOTAL columns within person label for subtotals in Table 28 .......................................</td>
</tr>
<tr>
<td>167.</td>
<td>PTBIS compute point-biserial correlation coefficients ..............................................................</td>
</tr>
<tr>
<td>168.</td>
<td>PVALUE proportion correct or average rating ...........................................................................</td>
</tr>
<tr>
<td>169.</td>
<td>PWEIGHT person (case) weighting .............................................................................................</td>
</tr>
<tr>
<td>170.</td>
<td>QUOTED quote-marks around labels ..........................................................................................</td>
</tr>
<tr>
<td>171.</td>
<td>RCONV score residual at convergence ........................................................................................</td>
</tr>
<tr>
<td>172.</td>
<td>REALSE inflate S.E. for misfit ....................................................................................................</td>
</tr>
<tr>
<td>173.</td>
<td>RESCORE response recoding .......................................................................................................</td>
</tr>
<tr>
<td>174.</td>
<td>RESFROM location of RESCORE ...............................................................................................</td>
</tr>
<tr>
<td>175.</td>
<td>RFILE scored response file .........................................................................................................</td>
</tr>
<tr>
<td>Page</td>
<td>Title</td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>278</td>
<td>Disjoint strings of responses</td>
</tr>
<tr>
<td>278</td>
<td>Disordered rating categories</td>
</tr>
<tr>
<td>279</td>
<td>Displacement measures</td>
</tr>
<tr>
<td>281</td>
<td>Edit Initial Settings (Winsteps.ini) file</td>
</tr>
<tr>
<td>283</td>
<td>Equating and linking tests</td>
</tr>
<tr>
<td>287</td>
<td>Estimation bias correction - warnings</td>
</tr>
<tr>
<td>289</td>
<td>Estimation methods: JMLE, PROX, XMLE</td>
</tr>
<tr>
<td>293</td>
<td>Exact Match: OBS% and EXP%</td>
</tr>
<tr>
<td>294</td>
<td>Exporting Tables to EXCEL</td>
</tr>
<tr>
<td>294</td>
<td>Extra Specifications prompt</td>
</tr>
<tr>
<td>295</td>
<td>Extreme scores: what happens</td>
</tr>
<tr>
<td>295</td>
<td>Global fit statistics</td>
</tr>
<tr>
<td>296</td>
<td>GradeMap interface</td>
</tr>
<tr>
<td>297</td>
<td>Guttman patterns</td>
</tr>
<tr>
<td>298</td>
<td>Half-rounding</td>
</tr>
<tr>
<td>299</td>
<td>How big an analysis can I do?</td>
</tr>
<tr>
<td>299</td>
<td>How long will an analysis take?</td>
</tr>
<tr>
<td>299</td>
<td>Information - item and test</td>
</tr>
<tr>
<td>300</td>
<td>Item difficulty: definition</td>
</tr>
<tr>
<td>300</td>
<td>Item discrimination or slope estimation</td>
</tr>
<tr>
<td>302</td>
<td>Iterations - PROX &amp; JMLE</td>
</tr>
<tr>
<td>304</td>
<td>Local Dependence</td>
</tr>
<tr>
<td>304</td>
<td>Logit and probit</td>
</tr>
<tr>
<td>306</td>
<td>Misfit diagnosis: infit outfit mean-square standardized</td>
</tr>
<tr>
<td>310</td>
<td>Missing data</td>
</tr>
<tr>
<td>310</td>
<td>Mixed and Saltus models</td>
</tr>
<tr>
<td>311</td>
<td>Multiple t-tests</td>
</tr>
<tr>
<td>312</td>
<td>Null or unobserved categories</td>
</tr>
<tr>
<td>312</td>
<td>One observation per respondent</td>
</tr>
<tr>
<td>314</td>
<td>Order of elements in Control file</td>
</tr>
<tr>
<td>314</td>
<td>Partial Credit model</td>
</tr>
<tr>
<td>315</td>
<td>Plausible values</td>
</tr>
<tr>
<td>315</td>
<td>Plotting with EXCEL</td>
</tr>
<tr>
<td>316</td>
<td>Poisson counts</td>
</tr>
<tr>
<td>317</td>
<td>Point-measure correlation</td>
</tr>
<tr>
<td>317</td>
<td>Polytomous mean-square fit statistics</td>
</tr>
<tr>
<td>318</td>
<td>Quality-control misfit selection criteria</td>
</tr>
<tr>
<td>319</td>
<td>Rank order data</td>
</tr>
<tr>
<td>319</td>
<td>Rectangular copying</td>
</tr>
<tr>
<td>320</td>
<td>Reliability and separation of measures</td>
</tr>
<tr>
<td>321</td>
<td>Right-click functions</td>
</tr>
<tr>
<td>322</td>
<td>Rules for assigning values to control variables (key-words)</td>
</tr>
<tr>
<td>324</td>
<td>Shortcut Keys</td>
</tr>
<tr>
<td>324</td>
<td>Specifying how data are to be recoded</td>
</tr>
<tr>
<td>324</td>
<td>Standard errors: model and real</td>
</tr>
<tr>
<td>325</td>
<td>Starting Winsteps from the DOS prompt</td>
</tr>
<tr>
<td>325</td>
<td>Subtest scoring</td>
</tr>
<tr>
<td>326</td>
<td>Transposing the data matrix</td>
</tr>
<tr>
<td>326</td>
<td>Unobserved and dropped categories</td>
</tr>
<tr>
<td>327</td>
<td>User-friendly rescaling</td>
</tr>
<tr>
<td>330</td>
<td>Using a word processor or text editor</td>
</tr>
<tr>
<td>330</td>
<td>Weighting items and persons</td>
</tr>
<tr>
<td>331</td>
<td>Winsteps: history and steps</td>
</tr>
<tr>
<td>332</td>
<td>Wordpad or text editor</td>
</tr>
<tr>
<td>333</td>
<td>Data vanishes</td>
</tr>
<tr>
<td>333</td>
<td>Display too big</td>
</tr>
<tr>
<td>333</td>
<td>File misread problems</td>
</tr>
<tr>
<td>335</td>
<td>If Winsteps does not work</td>
</tr>
<tr>
<td>335</td>
<td>Initialization fails</td>
</tr>
<tr>
<td>351</td>
<td></td>
</tr>
</tbody>
</table>
353. Not enough disk space ..............................................................................................................................335
354. Not enough memory ..................................................................................................................................336
355. Plotting problems .......................................................................................................................................336
356. Tab-delimited file problems .......................................................................................................................336
357. Winsteps problems and error codes ..........................................................................................................336
358. Winsteps SPSS error codes ......................................................................................................................337
Control Variable Index ............................................................................................................................................338
2. **Control Variable Index**

**Commonly-used Control Variables**

- **CFILE=** scored category label file
- **CLFILE=** codes label file
- **CODES=** valid & missing data codes
- **CSV=** comma-separated values in output files
- **DATA=** name of data file
- **EDFILE=** edit data file
- **GROUPS=** assigns items to rating scale or partial credit groupings (same as ISGROUPS=)
- **IAFILE=** item anchor file
- **IDFILE=** item deletion file
- **FILE=** item output file
- **REFER=** identifying items for recoding
- **SELECT=** item selection criterion
- **ISGROUPS=** assigns items to rating scale or partial credit groupings (same as GROUPS=)
- **ISORT=** sort column in item label
- **ISUBTOTAL=** subtotal items by specified columns
- **ITEM=** title for item labels
- **ITEM1=** column number of first response
- **IVALUEx=** recoding for items
- **KEYn=** scoring key
- **MRANGE=** half-range of measures on plots
- **NAME1=** first column of person label
- **NAMLEN=** length of person label
- **NEWSCORE=** recoding values
- **NI=** number of items
- **PAFILE=** person anchor file
- **PDFILE=** person deletion file
- **PERSON=** title for person labels
- **PFILE=** person output file
- **PSELECT=** person selection criterion
- **PSORT=** sort column in person label
- **PSUBTOTAL=** subtotal persons by specified columns
- **RESCORE=** response recoding
- **SAFILE=** structure anchor file
- **SFILe=** structure output file
- **STKEEP=** keep non-observed categories in structure
- **TITLE=** title for output listing
- **UANCHOR=** anchor values in user-scaled units
- **UASCALE=** anchor user-scale value of 1 logit
- **UDECIMALS=** number of decimal places reported
- **UMEAN=** reported user-set mean of item measures
- **UPMEAN=** reported user-set mean of person measures
- **USCALE=** reported user-scaled value for 1 logit
- **XFILE=** analyzed response file
- **XWIDE=** columns per response

**Special-purpose Control Variables**

- **@FIELD=** user-defined field locations
- **ALPHANUM=** alphabetic numbering to exceed 9 with XWIDE=1
- **ASCII=** output only ASCII characters
- **ASYMPTOTE=** report estimates of upper and lower asymptotes
- **BATCH=** set batch mode
- **BYITEM=** show empirical curves for items
- **CATREF=** reference category: Table 2
CHART=  graphical plots of measures
CONVERGE= select convergence criteria
CURVES= probability curve selection: Tables 21, 2
CUTHI= cut off responses with high probability
CUTOLO= cut off responses with low probability
DELMITER= delimiter of input response data fields
DIF= person label columns for DIF
DISCRIM= display item discrimination estimates
DISFILE= category/distractor/response option count file
DISTRT= output category/distractor/response option counts
DPF= item label columns for DPF
EQFILE= code equivalences
EXTRSCORE= extreme score adjustment
FITHIGH= lower bar in charts
FITI= item misfit criterion
FITLOW= lower bar in charts
FITP= person misfit criterion
FORMAT= reformat data
FORMFD= form-feed character
FRANGE= half-range of fit statistics on plots
G0ZONE= percent of 0's within 0-zone among which all 1's are turned to 0's
G1ZONE= percent of 1's within 1-zone among which all 0's are turned to 1's
GRFILE= graphing file for probability curves
GRPFROM= location of ISGROUPS=
GUFILE= Guttmanized file of responses
HEADER= display or suppress Sub-Table Headings after the first
HIADJ= correction for top categories
HLINES= heading lines in output files
IANCHQU= anchor items interactively
ICORFILE= inter-item residual correlations
IDELETE= one-line item deletion list
IDELQU= delete items interactively
IDROPEXTR= remove items with extreme scores
ILFILE= item label file
IMAP= item label for item maps
INUMB= label items by sequence number
IPMATRIX= response matrix (Output Files menu only)
ISFILE= item-structure output file
ITLEN= maximum length of item label
IWEIGHT= item (variable) weighting
KEYFORM= KeyForm skeleton for Excel plot
KEYFROM= location of KEYn=
KEYSCR= reassign scoring keys
LCONV= logit change at convergence
LINLEN= length of printed lines
LOCAL= locally restandardize fit statistics
LOGFILE= accumulates control files
LOWADJ= correction for bottom categories
MAKEKEY= construct an MCQ scoring key
MAXPAGE= maximum number of lines per page
MFORMS= reformattng input records & multiple forms
MHSLICE= size of Mantel or Mantel-Haenszel slice
MISSCORE= scored value of missing data: not your missing-data code
MJMLE= maximum number of JMLE iterations
MNSQ= show mean-square or t standardized fit
MODELS= assigns model types to items
MODFROM= location of MODELS=
MPROX= maximum number of PROX iterations
MUCON= maximum number of JMLE iterations
NAMLMP= label length on maps
NORMAL= normal distribution for standardizing fit
OSORT= category/option/distractor sorted by score, count or measure
OUTFIT= sort misfits on outfit or infit
PAIRED= correction for paired comparison data
PANCHQU= anchor persons interactively
PCORFIL= inter-person residual correlations
PDELETE= one-line person deletion list
PDELQU= delete persons interactively
PDROPEXTR= remove persons with extreme scores
PMAP= person label for person maps
PRCOMP= residual type for principal components/contrast analysis
PTBISERIAL= report point-biserial correlations
PVALUE= report item p-values
PWEIGHT= person (case) weighting
QUOTED= quote-marks around labels
RCONV= score residual at convergence
REALSE= inflate S.E. of measures for misfit
RESFRM= location of RESCORE=
RFILE= scored response output file
SAITEM= multiple ISGROUP= format in SFILE= & SAFILE=
SANCHQU= anchor structure interactively
SCOREFILE= score-to-measure file
SDELQU= delete structure interactively
SDFILE= structure deletion file
SEPARATOR= delimiter of input response data fields
SIFILE= simulated data file
SPFILE= supplemental control file
STBIAS= correct for estimation bias
STEPT3= structure summary in Table 3 or 21
T1I#= items per # in Table 1
T1P#= persons per # in Table 1
TABLES= output table selection
TARGET= information-weighted estimation
TFILE= input file of table numbers to be output
TOTALSCORE= show total observed score and count
UCOUNT= most unexpected responses in Tables 6 and 10.
W300= produce IFILE= and PFILE= in 3.00 format
XMLE= consistent, almost unbiased, estimation (experimental)

3. Control Variable Index by function

Data file layout:
DATA= name of data file
DELIMITER= delimiter of input response data fields
FORMAT= reformat data
ILFILE= item label file
ITEM1= column number of first response
ITLEN= maximum length of item label
INUMB= label items by sequence number
MFORMS= reformatting input records & multiple forms
NAME1= first column of person label
NAMLEN= length of person label
NI= number of items
SEPARATOR= delimiter of input response data fields
XWIDE= columns per response
@FIELD= user-defined field locations

3
Data selection and recoding:
ALPHANUM= alphabetic numbering to exceed 9 with XWIDE=1
CODES= valid & missing data codes
CUTHI= cut off responses with high probability
CUTLO= cut off responses with low probability
EDFILE= edit data file
IREFER= identifying items for recoding
IVALUEx= recoding for items
IWEIGHT= item (variable) weighting
KEYn= scoring key
KEYFROM= location of KEYn=
KEYSCR= reassign scoring keys
MAKEKEY= construct an MCQ scoring key
MISSCORE= scored value of missing data: not your missing-data code
NEWSCORE= recoding values
PWEIGHT= person (case) weighting
RESCORE= response recoding
RESFRM= location of RESCORE=

Items: deleting, anchoring and selecting:
IAFILE= item anchor file
IANCHOU= anchor items interactively
IDELETE= one-line item deletion list
IDELQU= delete items interactively
IDFILE= item deletion file
IDROPEXTR= remove items with extreme scores
ISELECT= item selection criterion

Person: deleting, anchoring and selecting:
PAFILE= person anchor file
PANCHOU= anchor persons interactively
PDELETE= one-line person deletion list
PDELQU= delete persons interactively
PDFILE= person deletion file
PDROPEXTR= remove persons with extreme scores
PSELECT= person selection criterion

Rating scales, partial credit items and polytomour response structures:
GROUPS= assigns items to rating scale or partial credit groupings (same as ISGROUPS=)
GRPFROM= location of ISGROUPS=
ISGROUPS= assigns items to rating scale or partial credit groupings (same as GROUPS=)
MODELS= assigns model types to items
MODFROM= location of MODELS=
STKEEP= keep non-observed categories in structure

Category structure: anchoring, labeling, deleting:
CFILE= scored category label file
CLFILE= codes label file
SAFILE= structure anchor file
SAITEM= multiple ISGROUP= format in SFILE= & SAFILE=
SANCHQU= anchor structure interactively
SDELQU= delete structure interactively
SDFILE= structure deletion file

Measure origin, anchoring and user-scaling:
UANCHOR= anchor values in user-scaled units
UASCALE= anchor user-scaled value for 1 logit
UDECIMALS= number of decimal places reported
UIMEAN= reported user-set mean of item measures
UMEAN = reported user-set mean of item measures
UPMEAN = reported user-set mean of person measures
USCALE = reported user-scaled value for 1 logit

Output table selection and format:
ASCII = output only ASCII characters
FORMFD = form-feed character
HEADER = display or suppress Sub-Table Headings after the first
ITEM = title for item labels
LINLEN = length of printed lines
MAXPAGE = maximum number of lines per page
PERSON = title for person labels
TABLES = output table selection
TFILE = input file of table numbers to be output
TITLE = title for output listing
TOTALSCORE = show total observed score and count

Output tables, files and graphs: specific controls
ASYMPTOTE = report estimates of upper and lower asymptotes
BYITEM = show empirical curves for items
CATREF = reference category: Table 2
CHART = graphical plots of measures
CURVES = probability curve selection: Tables 21, 2
DIF = person label columns for DIF
DISCRIM = display item discrimination estimates
DISTRT = output category/distractor/option counts
DPF = item label columns for DPF
EQFILE = code equivalences
FITHIGH = lower bar in charts
FITI = item misfit criterion
FITLOW = lower bar in charts
FITP = person misfit criterion
FRANGE = half-range of fit statistics on plots
HIADJ = correction for top categories
IMAP = item label for item maps
ISORT = sort column in item label
ISUBTOTAL = subtotal items by specified columns
LOWADJ = correction for bottom categories
MNSQ = show mean-square or t standardized fit
MRANGE = half-range of measures on plots
NAMLMP = label length on maps
OSORT = category/distractor/option sorted by score, count or measure
OUTFIT = sort misfits on outfit or infit
PMAP = person label for person maps
PSORT = sort column in person label
PSUBTOTAL = subtotal persons by specified columns
PVALUE = report item p-values
STEPT3 = structure summary in Table 3 or 21
T1I# = items per # in Table 1
T1P# = persons per # in Table 1
UCOUNT = most unexpected responses in Tables 6 and 10.

Output files:
DISFILE = category/distractor/response option count file
GRFILE = graphing file for probability curves
GUFILE = Guttmanized file of responses
ICORFILE = inter-item residual correlations
IFILE = item output file
IPMATRIX = response matrix (Output Files menu only)
ISFILE= item-structure output file
KEYFORM= KeyForm skeleton for Excel plot
LOCAL= locally restandardize fit statistics
LOGFILE= accumulates control files
PCORFIL= inter-person residual correlations
PFILE= person output file
RFILE= scored response output file
SFILE= structure output file
SCOREFILE= score-to-measure file
SIFILE= simulated data file
XFILE= analyzed response file

Output file format control:
CSV= comma-separated values in output files
G0ZONE= percent of 0’s within 0-zone among which all 1’s are turned to 0’s
G1ZONE= percent of 1’s within 1-zone among which all 0’s are turned to 1’s
HLINES= heading lines in output files
QUOTED= quote-marks around labels
W300= produce IFILE= and PFILE= in 3.00 format

Estimation, operation and convergence control:
CONVERGE= select convergence criteria
EXTRSCORE= extreme score adjustment
LCONV= logit change at convergence
LOCAL= locally restandardize fit statistics
MJMLE= maximum number of JMLE iterations
MPROX= maximum number of PROX iterations
MUCON= maximum number of JMLE iterations
NORMAL= normal distribution for standardizing fit
PAIRED= correction for paired comparison data
PRCOMP= residual type for principal components/contrast analysis
PTBISERIAL= compute point-biserial correlations
RCONV= score residual at convergence
REALSE= inflate S.E. of measures for misfit
STBIAS= correct for estimation bias
TARGET= information-weighted estimation
WHEXACT= Wilson-Hilferty exact normalization
XMLE= consistent, almost unbiased, estimation (experimental)

Program operation:
BATCH= set batch mode
MAKEKEY= construct an MCQ scoring key
SPFILE= supplemental control file
&END= end of control variable list
&INST= start of control variable list (ignored)
END LABELS= end of item label list

4. Control variables from Specification menu

These control variables can be changed using the Specification pull-down menu after measures have been estimated.

They do not alter the estimates from the main analysis. They only change how it is reported.

@FIELD= user-defined field locations
ASCII= output only ASCII characters
ASYMPTOTE= report estimates of upper and lower asymptotes
BYITEM= show empirical curves for items
CATREF= reference category: Table 2
CFILE= scored category label file (file name only, blank deletes category labels)
CHART= graphical plots of measures
CLFILE= codes label file (file name only, blank deletes code labels)
CSV= comma-separated values in output files
CURVES= probability curve selection: Tables 21, 2
DISCRIM= display item discrimination estimates
DISTRT= output category/distractor/option counts
FITHIGH= lower bar in charts
FITI= item misfit criterion
FITLOW= lower bar in charts
FITP= person misfit criterion
FORMFD= form-feed character
FRANGE= half-range of fit statistics on plots
G0ZONE= percent of 0’s within 0-zone among which all 1’s are turned to 0’s
G1ZONE= percent of 1’s within 1-zone among which all 0’s are turned to 1’s
HEADER= display or suppress Sub-Table Headings after the first
HIADJ= correction for top categories
HLINES= heading lines in output files
IDDELETE= one-line item deletion list
IDFILE= item deletion file (file name only: blank resets temporary item deletions)
ILFILE= item label file (file name only: blank not allowed)
IMAP= item label for item maps
ISELECT= item selection criterion
ISORT= sort column in item label
ITEM= title for item labels
LINLEN= length of printed lines
LOWADJ= correction for bottom categories
MAXPAGE= maximum number of lines per page
MHSLICE= size of Mantel or Mantel-Haenszel slice
MNSQ= show mean-square or t standardized fit
MRANGE= half-range of measures on plots
NAMLMP= label length on maps
OSORT= category/option/distractor sorted by score, count or measure
OUTFIT= sort misfits on outfit or infit
PDELETE= one-line person deletion list (blank resets temporary person deletions)
PDFILE= person deletion file (file name only: blank resets temporary person deletions)
PERSON= title for person labels
PMAP= person label for person maps
PSELECT= person selection criterion
PSORT= sort column in person label
PVALUE= report item p-values
PWEIGHT= person (case) weighting
QUOTED= quote-marks around output labels
STEPT3= structure summary in Table 3 or 21
T1I#= items per # in Table 1
T1P#= persons per # in Table 1
TITLE= title for output listing
TOTALSCORE= show total observed score and count
UDECIMALS= number of decimal places reported
UIMEAN= reported user-set mean of item measures
UMEAN= reported user-set mean of item measures
UPMEAN= reported user-set mean of person measures
USCALE= reported user-scaled value for 1 logit
W300= produce IFILE= and PFILE= in 3.00 format

Control variables that can be set with other pull-down menus
DIF= person label columns for DIF (from Output Tables menu)
DISFILE= category/distractor/response option count file (from Output Files menu)
DPF= item label columns for DPF (from Output Tables menu)
GRFILE = graphing file for probability curves (from Output Files menu)
GUFILE = Guttmannized file of responses (from Output Files menu)
ICORFILE = inter-item residual correlations (from Output Files menu)
IFILE = item output file (from Output Files menu)
IPMATRIX = response matrix (from Output Files menu)
ISFILE = item-structure output file (from Output Files menu)
ISUBTOTAL = subtotal items by specified columns (from Output Tables menu)
KEYFORM = KeyForm skeleton for Excel plot (from Plots menu)
PCORFIL = inter-person residual correlations (from Output Files menu)
PFILE = person output file (from Output Files menu)
PSUBTOTAL = subtotal persons by specified columns (from Output Tables menu)
RFILE = scored response output file (from Output Files menu)
SCOREFILE = score-to-measure file (from Output Files menu)
SFILE = structure output file (from Output Files menu)
SIFILE = simulated data file (from Output Files menu)
XFILE = analyzed response file (from Output Files menu)

5. Output Table Index

Table  Description
1  Maps of person and item measures. Show Rasch measures.
   1.0 One page map with names.
   1.1 Map of distributions - persons and items
   1.2 Item labels with person distribution (squeezed onto one page)
   1.3 Person labels with item distribution (squeezed onto one page)
   1.4 Rating scale or partial credit map of distributions: persons with items at high, mean, low
   1.10 One page map with person names by measure, item names by easiness.
   1.12 Item labels, by easiness, with person distribution (squeezed onto one page)

2  Measures and responses plots. Response categories for each item, listed in measure order, plotted
against person measures, shown as modal categories, expected values and cumulative probabilities.  
Table 2 for multiple-choice items.
   By observed categories
   2.6 Observed average measures of persons (empirical averages)

      By scored categories (illustrated by an observed category code for each score)
      2.1 Modal categories (most probable)
      2.2 Mean categories (average or expected: Rasch-half-point thresholds)
      2.3 Median categories (cumulative probabilities: Rasch-Thurstone thresholds)
      2.4 Structure calibrations (Rasch model parameters: rating scale, partial credit, "restricted", "unrestricted": Rasch-Andrich thresholds)
      2.5 Observed average measures of persons (empirical averages)
      2.7 Expected average measures of persons

   By category scores (if category scores differ from category codes in the data):
   2.11 Modal categories (most probable)
   2.12 Mean categories (average or expected: Rasch-half-point thresholds)
   2.13 Median categories (cumulative probabilities: Rasch-Thurstone thresholds)
   2.14 Structure calibrations (Rasch model parameters: rating scale, partial credit, "restricted", "unrestricted")
   2.15 Observed average measures of persons (empirical averages)
   2.16 Observed average measures of persons (empirical averages)
   2.17 Expected average measures of persons

3  Summary statistics. Person, item, and category measures and fit statistics.
   3.1 Summaries of person and items: means, S.D.s, separation, reliability.
   3.2 Summary of rating categories and probability curves.
4 **Person infit plot.** Person infits plotted against person measures.
   4.1 Person infit vs. person measure plot.

5 **Person outfit plot.** Person outfits plotted against person measures.
   5.1 Person outfit vs. person measure plot.
   5.2 Person infit vs. person outfit plot.

6 **Person statistics - fit order.** Misfitting person list. Scalogram of unexpected responses.
   6.1 Table of person measures in descending order of misfit. (Specify \( \text{FITP}=0 \) to list all persons)
   6.2 Chart of person measures, infit mean-squares and outfit mean-squares. \( \text{Chart} = \text{Yes} \)
   6.3 (Not produced for persons)
   6.4 Scalogram of most misfitting person response strings.
   6.5 Scalogram of most unexpected responses.
   6.6 Most unexpected responses list.

7 **Misfitting Persons.** Lists response details for persons with \( t \) standardized fit greater than \( \text{FITP} \).
   7.1 Response strings for most misfitting persons.
   7.2 KeyForms of responses of misfitting persons.

8 **Item infit plot.** Item infits plotted against item calibrations.
   8.1 Item infit vs. item measure plot.

9 **Item outfit plot.** Item outfits plotted against item calibrations.
   9.1 Item outfit vs. item measure plot.
   9.2 Item infit vs. item outfit plot.

10 **Item statistics - fit order.** Misfitting item list with option counts. Scalogram of unexpected responses.
    10.1 Table of item measures in descending order of misfit. (Specify \( \text{FITI}=0 \) to list all persons)
    10.2 Chart of item measures, infit mean-squares and outfit mean-squares. \( \text{Chart} = \text{Yes} \)
    10.3 Item response-structure categories/options/distractors: counts and average measures. \( \text{Distractors}=\text{Yes} \)
    10.4 Scalogram of most misfitting item response strings.
    10.5 Scalogram of most unexpected responses.
    10.6 Most unexpected responses list.

11 **Misfitting Items.** Response details for items with \( t \) standardized fit greater than \( \text{FITI} \).
    11.1 Response strings for most misfitting items.

12 **Item distribution map.** Horizontal histogram of item distribution with abbreviated item names.
    12.1 Item labels with person distribution (same as 1.12)
    12.2 Item labels with expected score zones
    12.3 Item labels with 50% cumulative probabilities
    12.4 Item labels, by easiness, with person distribution (same as 1.12)

13 **Item statistics - measure order** list and graph with category/option/distractor counts.
    13.1 Table of items in descending measure order.
    13.2 Chart of item measures, infit mean-squares and outfit mean-squares. \( \text{Chart}=\text{Yes} \)
    13.3 Item response-structure categories/options/distractors: counts and average measures. \( \text{Distractors}=\text{Yes} \)

14 **Item statistics - entry order** list and graph with category/option/distractor counts.
    14.1 Table of items in entry number (sequence) order.
    14.2 Chart of item measures, infit mean-squares and outfit mean-squares. \( \text{Chart}=\text{Yes} \)
    14.3 Item response-structure categories/options/distractors: counts and average measures. \( \text{Distractors}=\text{Yes} \)

15 **Item statistics - alphabetical order** list and graph with category/option/distractor counts.
    15.1 Table of item measures in alphabetical order by label. (Specify \( \text{ISORT}= \) for sort column)
    15.2 Chart of item measures, infit mean-squares and outfit mean-squares. \( \text{Chart}=\text{Yes} \)
15.3 Item response-structure categories/options/distractors: counts and average measures.  
(Distractors=Yes)

16 Person distribution map. Horizontal histogram of person distribution, with abbreviated person-ids.  
16.3 Person labels with item distribution (same as 1.3)

17 Person statistics - measure order list and chart.  
17.1 Table of persons in descending measure order.  
17.2 Chart of person measures, infit mean-squares and outfit mean-squares. (Chart=Yes)  
17.3-... KeyForms of responses of persons.

18 Person statistics - entry order list and chart.  
18.1 Table of persons in entry number (sequence) order.  
18.2 Chart of person measures, infit mean-squares and outfit mean-squares. (Chart=Yes)  
18.3-... KeyForms of responses of persons.

19 Person statistics - alphabetical order list and chart.  
19.1 Table of person measures in alphabetical order by label. (Specify PSORT= for sort column)  
19.2 Chart of person measures, infit mean-squares and outfit mean-squares. (Chart=Yes)  
19.3-... KeyForms of responses of persons.

20 Measures for all scores on a test of all calibrated items, with percentiles.  
20.1 Table of person measures for every score on complete test. (Specify ISELECT= for subtests).  
20.2 Table of measures for every score, with sample percentiles and norm-referenced measures.  
20.3 Table of item difficulty measures (calibrations) for every score (p-value) by complete sample.

21 Category probability curves. Category probabilities plotted against the difference between person and item measures, then the expected score and cumulative probability and expected score ogives.  
21.1 Category probability curves (modes, structure calibrations).  
21.2 Expected score ogive (means, model Item Characteristic Curve).  
21.3 Cumulative category probability curves (medians, shows 50% cumulative probabilities).

22 Sorted observations. Data sorted by person and item measures into scalogram patterns.  
22.1 Guttman scalogram of sorted scored responses.  
22.2 Guttman scalogram showing out-of-place responses.  
22.3 Guttman scalogram showing original responses.

23 Item principal components/contrasts. Identifies structure in response residuals (BIGSTEPS Table: 10.3)  
23.0 Scree plot of variance components.  
23.2 Plot of loadings on first contrast in residuals vs. item measures.  
23.3 Items in contrast loading order.  
23.4 Persons exhibiting contrast.  
23.5 Items in measure order.  
23.6 Items in entry order.  
23.7 etc. Subsequent contrasts.  
23.99 Tables of items with highly correlated residuals. (Reported as last subtable in Table 23)

24 Person principal components/contrasts. Identifies structure in residuals (not in BIGSTEPS)  
24.0 Scree plot of variance components.  
24.2 Plot of loadings on first contrast in residuals vs. person measures.  
24.3 Persons in contrast loading order.  
24.4 Items exhibiting contrast.  
24.5 Persons in measure order.  
24.6 Persons in entry order.  
24.7 etc. Subsequent contrasts.  
24.99 Tables of persons with highly correlated residuals. (Reported as last subtable in Table 24)

25 Item statistics - displacement order list and graph with category(option/distractor counts.  
25.1 Table of items in descending displacement order.
25.2 Chart of item measures, infit mean-squares and outfit mean-squares. (Chart=Yes)
25.3 Item response-structure categories/options/distractors: counts and average measures. (Distractors=Yes)

26 **Item statistics - correlation order** list and graph with category/optioan/distractor counts.
   26.1 Table of items in ascending correlation order (Point-biserial, if PTBIS=Yes, else Point-measure).
   26.2 Chart of item measures, infit mean-squares and outfit mean-squares.
   26.3 Item response-structure categories/options/distractors: counts and average measures.

27 **Item subtotals.**
   27.1 Measure sub-totals, controlled by ISUBTOT=
   27.2 Measure sub-totals histograms, controlled by ISUBTOT=
   27.3 Measure sub-totals summary statistics, controlled by ISUBTOT=

28 **Person subtotals.**
   28.1 Measure sub-totals, controlled by PSUBTOT=
   28.2 Measure sub-totals histograms, controlled by PSUBTOT=
   28.3 Measure sub-totals summary statistics, controlled by PSUBTOT=

29 **Empirical item character curves and response frequency plots.**
   29.1 Empirical and model ICCs (see also Graph Menu)
   29.2 Use of response categories by measure

30 **Differential Item Function across Person classifications**
   30.1 DIF report (paired), controlled by DIF=
   30.2 DIF report (measure list: person class within item)
   30.3 DIF report (measure list: item within person class)

31 **Differential Person Function across Item classifications**
   31.1 DPF report (paired), controlled by DPF=
   31.2 DPF report (measure list: item class within person)
   31.3 DPF report (measure list: person within item class)

32 **Control Variable Listing** of the current settings of all Winsteps control variables - appears on the Output Files pull-down menu.

33 **Differential Item Classification vs. Person Classification interactions/biases**
   33.1 DIF report (paired person classifications on each item classification), controlled by DIF= and DPF=
   33.2 DIF report (measure list of item classification differential difficulties)
   33.3 DIF report (paired item classifications on each person classification)
   33.4 DPF report (measure list of person classification differential abilities)

34 Control Variables and Convergence report. Lists the control variables and shows estimation convergence. (Only appears at end of Output Report File).
   0.0 Title page
   0.1 Analysis identification
   0.2 Convergence table
   0.3 Control file

6. **Output Graphs and Plots Index**

**Graphs** - from the Graphs pull-down menu
   Category Probability Curves
   Expected Score ICC (Item Characteristic Curve)
   Cumulative Probabilities
   Information Function
   Category Information
   Conditional Probability Curves
   Test Characteristic Curve
Test Information

Plots - from the Plots pull-down menu
Compare statistics generates an Excel-based cross-plot of the values in Table 34.
Bubble chart generates a Bond & Fox-style bubble chart.
Keyform plot - Horizontal generates a horizontal keyform layout.
Keyform plot - Vertical generates a vertical keyform layout.
DIF plot - the DIF values in Table 30.
DPF plot - the DPF values in Table 31.
DIF+DPF plot - the DIF+DPF values in Table 33.

7. Output Files Index

These can be accessed from the Output Files pull-down menu

Output Files - in control file or pull-down menu
DISFILE= category/distractor/response option count file
GRFILE= graphing file for probability curves
GUFILE= Guttmanized response file
ICORFILE= inter-item residual correlations
IFILE= item output file (use for anchoring)
ISFILE= item-structure output file (do not use for anchoring)
PFFILE= person output file (useful or anchoring)
PCORFILE= inter-person residual correlations
RFILE= scored response output file
SCOREFILE= score-to-measure file
SFFILE= structure output file (use for anchoring)
SIFILE= simulated data file
XFILE= analyzed response file

Output files - only from pull-down menu
Control variable file control variable listing (same as Table 32)
IPMATRIX= matrix of response-level data
TRANSPOSE= transposed (rows-columns) control file
GradeMap GradeMap model specification and student data files

To control output file formatting in the control file:
CSV= fixed fields, tab-delimited or comma-separated values in output files
HLINES= heading lines written to output files

8. Rasch analysis and WINSTEPS

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

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

The person and item total raw scores are used to estimate linear measures. Under Rasch model conditions, these measures are item-free (item-distribution-free) and person-free (person-distribution-free). So that the measures are statistically equivalent for the items regardless of which persons (from the same population) are analyzed, and for the items regardless of which items (from the same population) are analyzed. Analysis of the data at the response-level indicates to what extent these ideals are realized within any particular data set.
The Rasch models implemented in Winsteps include the Georg Rasch dichotomous, Andrich "rating scale", Masters "partial credit", Bradley-Terry "paired comparison", Glas "success model", Linacre "failure model" and most combinations of these models. Other models such as binomial trials and Poisson can also be analyzed by anchoring (fixing) the response structure to accord with the response model. (If you have a particular need, please let us know as Winsteps is continually being enhanced.)

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

The Rasch Family of Models

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

**The dichotomous model:**
\[ \log\left(\frac{Pni1}{Pni0}\right) = Bn - Di \]

**The polytomous "Rating Scale" model:**
\[ \log\left(\frac{Pnij}{Pni(j-1)}\right) = Bn - Di - Fj \]

**The polytomous "Partial Credit" model:**
\[ \log\left(\frac{Pnij}{Pni(j-1)}\right) = Bn - Di - Fij = Bn - Dij \]

**The polytomous "Grouped response-structure" model:**
\[ \log\left(\frac{Pnij}{Pni(j-1)}\right) = Bn - Dig - Fgj \]

where
- \( Pnij \) is the probability that person \( n \) encountering item \( i \) is observed in category \( j \),
- \( Bn \) is the "ability" measure of person \( n \),
- \( Di \) is the "difficulty" measure of item \( i \), the point where the highest and lowest categories of the item are equally probable.
- \( Fj \) is the "calibration" measure of category \( j \) relative to category \( j-1 \), the point where categories \( j-1 \) and \( j \) are equally probable relative to the measure of the item.

Also models with the form of "Continuation Ratio" models, such as the "Success" model and the "Failure" model.

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

**Work-flow with Winsteps**

Control + Data file or Control file and Data file(s)

User-interaction → **Winsteps** ← Anchor Files

Report Output File + Output Tables + Graphs + Output Files

Word Processor, Spreadsheet, Statistical Package

↓

Actions

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

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

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

9. References

Please cite the current Winsteps computer program as:

• RSA means Wright B.D. & Masters G.N. Rating Scale Analysis, Chicago: MESA Press, 1982, especially p. 100:
  www.rasch.org/books.htm
• BTD means Wright B.D. & Stone M.H. Best Test Design, Chicago: MESA Press, 1979:
  www.rasch.org/books.htm

Other recommended sources:
• Rasch Measurement Transactions: www.rasch.org/rmt/
• Journal of Applied Measurement: www.jampress.org
• "Applying the Rasch Model: Fundamental Measurement in the Human Sciences", by Trevor G. Bond & Christine
  Fox examples
"Introduction to Rasch Measurement", Everett V. Smith, Jr. & Richard M. Smith (Eds.) JAM Press, 2004
  www.jampress.org

10. About the Users' guide

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

Most of this Guide is in proportionately-spaced type.

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

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

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

Please cite the current Winsteps computer program as:

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

An Ode from a user:

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

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

Jim Houston, Nov. 30, 2001

11. Getting further help

*Common installation problems are solved at: [www.winsteps.com/problems.htm](http://www.winsteps.com/problems.htm)*

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

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

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

12. What is supplied

WINSTEPS is supplied in three forms:

1) MinistepInstall.exe   To install MINISTEP, the student/evaluation version of Winsteps.
2) WinstepsInstall.exe  To install WINSTEPS under Windows
or 3) WinstepsPasswordInstall.exe To install WINSTEPS under Windows with password-protected installation.

These create directory, C:\WINSTEPS, and install in it WINSTEPS or MINISTEP.

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

- KCT.TXT is the Knox Cube Test data (BTD p.31 - see Section 1.1) The results in BTD were obtained with more approximate algorithms and do not agree exactly with WINSTEPS results.
- SF.TXT is the Liking For Science data (RSA p.18)
- There are many more EXAMPLE files described later in this manual.

13. Installation instructions for WINSTEPS

Under Windows XP/98/NT/ME/2000/...:
Run WinstepsInstall.exe from the downloaded file or from the CD-ROM.

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

To Run WINSTEPS:

Click on *Start* button (or use desktop Winsteps icon)
Point to *Programs*
Point to *WINSTEPS*  
Click on *WINSTEPS* icon
16

Type SF.TXT
Press Enter key
Press Enter key
Press Enter key

Winsteps will run. Examine its output with the pull-down menus, particularly Output Tables

Additional Notes:

a) WINSTEPS is usually installed in the C:\WINSTEPS directory.
b) When WINSTEPS ends, pull down the "Edit" menu to edit control and output files.
c) All information on the screen is also in the Report output file.
d) Files in the C:\TEMP directory and with the suffix .TMP may be deleted.

Macintosh Computer: A Winsteps users reports: I've successfully installed Winsteps on a PowerMac G3 running Virtual PC and OS 9.2. It seems to work fine. Surprisingly fast in fact, considering the Mac is emulating a Windows environment.

14. Starting WINSTEPS in Windows

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

1) Double-click the Winsteps Icon on the Desktop.

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

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

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

Setup procedure: takes you to the Control and data file setup window
Instructions only: takes you to the do-it-yourself instructions
No: WINSTEPS asks you for the names of your input and report output files. There are example files already in the Winsteps\examples directory.
Don’t ask again: makes "No" the standard option here. You can reset this using Edit Initial Settings

If you need help to set up a new control file, go to control and data file set-up.

To select your control file:

Winsteps asks: Please enter name of WINSTEPS control file:
(a) You can type it in ...
Please enter name of WINSTEPS control file: **KCT.TXT** *(Enter)*

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

<table>
<thead>
<tr>
<th>Open File</th>
<th>Ctrl+O</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start/another WINSTEPS</td>
<td>Alt+A</td>
</tr>
<tr>
<td>Exit</td>
<td>Ctrl+Q</td>
</tr>
<tr>
<td>Finish iterating</td>
<td>Ctrl+F</td>
</tr>
<tr>
<td>Enter</td>
<td></td>
</tr>
<tr>
<td>Save</td>
<td>Ctrl+S</td>
</tr>
<tr>
<td>Save As...</td>
<td></td>
</tr>
<tr>
<td>Print</td>
<td>Ctrl+P</td>
</tr>
<tr>
<td>Excel=C:\Program Files\Microsoft Office\Office\EXCEL.EXE</td>
<td></td>
</tr>
<tr>
<td>C:\WINSTEPS\kct.txt</td>
<td></td>
</tr>
<tr>
<td>C:\WINSTEPS\example0.txt</td>
<td></td>
</tr>
</tbody>
</table>

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

or (d) Press your **Enter key**
This displays a standard Windows open dialog box - try it.

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

**Example Analysis:**

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

Please enter name of report output file: *(Enter)*
*If you only press Enter, a temporary output file will be used.*
*If you type in a file name, that file will be created and used.*
*If you want the file dialog box, use 'Open File' on the File pull-down menu.*

Extra specifications? (e.g., MJMLE=1), or press Enter: *(Enter)*
*Usually you have none, so merely press Enter.*

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

Use the *Edit pull-down menu* to simplify inspection and editing of the control, input and output files. This is done with *WordPad* or your own text editor.
At the end of the run, using the Output Tables pull-down menu you can request further output Tables to be produced interactively and displayed on the screen. If you want to save any of these, use the Save As option when they are displayed. If you omit to save them, they are written as "ws.txt" files, and can be recovered from the Recycle Bin.

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

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

15. Using Winsteps under Windows

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

(a) You respond to prompts on the screen.
   There are two frequent ones:
   **Name of control file:**
   This is the "DOS-text with line breaks" or ASCII file where your control specifications reside.
   You can press Enter to browse for it.

   **Report output file name:**
   Press Enter for a temporary output file, or
   Type in an output file name or use the pull-down file menu

   **Extra specifications (or press Enter):**
   Press Enter!
   This is used for making just-in-time changes to your control file instructions, for this run only.
(b) You use the pull-down menus.
A frequent choice is the first choice on both the File and Edit menus:

**Edit Control File=**
Select this to edit your Winsteps control file.

(c) You use the WordPad text editor.
All editing and display of Winsteps output is done using text files and Wordpad or your own text editor.
This gives you great flexibility to:
- modify control and anchor files
- view, copy and paste output into Word (or other) files

16. Stopping WINSTEPS

The WINSTEPS program ceases execution when

1) The program stops itself:
The estimation procedure has reached an acceptable level of convergence and all pre-specified output has been produced. This happen when:
   a) The estimates are within the convergence criteria (LCONV= and RCONV= as controlled by CONVERGE=)
   b) The maximum number of iterations has been reached (MPROX= and then MJMLE=)

   *To instruct WINSTEPS to run indefinitely (up to 2,000,000,000 iterations), set*
   MJMLE=0
   LCONV=0
   RCONV=0
   CONVERGE=F

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

2) You stop the iterative process:
   a) If you press Ctrl with F (or use File menu) during PROX iterations:
      PROX iteration will cease as soon extreme scores have been identified and point-biserial correlations have been calculated. JMLE iterations then start.
      b) If you press Ctrl with F during JMLE iterations:
         JMLE iteration will cease at the end of this iteration. Fit statistics will then be calculated and output tables written to disk.
      c) If you press Ctrl with F during the output phase:
         Output will cease at the end of the current output operation.

   Acknowledgment of your Ctrl with F instruction is shown by the replacement of = by # in the horizontal bar drawn across you screen which indicates progress through the current phase of analysis.

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

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

Yes: close all open windows. If some windows have been modified, but not saved, you may be asked if you want
to save those.  
**No:** leave all windows as they are, but close the Winsteps analysis window. To make **Yes** or **No** the automatic standard, click "and from now on". This choice may be reset in **Edit Initial Settings**

17. Uninstalling WINSTEPS

Depending on the installation procedure:

(a) Select "Uninstall" from the **Programs** menu  
(b) Go to "Settings", "Control Panel", "Add/Remove Programs" and double-click on "WINSTEPS Uninstall"

(c) Delete WINSTEPS directory, and delete the WINSTEPS entries from "Windows\Start Menu\Programs"

(d) Use a Windows clean-up utility to tidy up loose ends.

(e) Delete files in C:\TEMP and C:\WINDOWS\TEMP (or your Windows temporary file) and files ending .....ws.txt

18. Menu bar

Winsteps has a useful set of pull-down menus:

<table>
<thead>
<tr>
<th>File</th>
<th>overall control of the analysis/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edit</td>
<td>display and editing of input and output files and tables.</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Tables for understanding, evaluating and improving your measurement system.</td>
</tr>
<tr>
<td>Output Tables</td>
<td>produces all output Tables produced by Winsteps.</td>
</tr>
<tr>
<td>Output Files</td>
<td>produces output primarily intended for input into other software.</td>
</tr>
<tr>
<td>Batch</td>
<td>facilitates running Winsteps in batch mode</td>
</tr>
<tr>
<td>Help</td>
<td>displays Help file</td>
</tr>
<tr>
<td>Specification</td>
<td>allows entry of specifications after the analysis, one at a time, in the form of specification=value.</td>
</tr>
<tr>
<td>Plots</td>
<td>uses Excel to display and compare analyses.</td>
</tr>
<tr>
<td>SAS/SPSS</td>
<td>reformat SAS .sas7bdat and SPSS .sav files into Winsteps control and data files.</td>
</tr>
<tr>
<td>Graphs Menu</td>
<td>bit-mapped graphics for test, item and category display.</td>
</tr>
<tr>
<td>Data Setup</td>
<td>provides a immediate means for setting up control and data files</td>
</tr>
</tbody>
</table>

19. Batch menu

This facilitates running Winsteps in batch mode.

<table>
<thead>
<tr>
<th>Batch</th>
<th>Help</th>
<th>Specification</th>
<th>Compare Files</th>
<th>SPSS</th>
<th>Graphs</th>
<th>Data Setup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Running WINSTEPS in Batch mode</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Help for Batch mode</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edit;create batch file from=C:\WINSTEPS\winbatch.bat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edit;create batch file from=C:\WINSTEPS\winbatch.cmd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edit batch file</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run batch file: right-click on file name, then Open on menu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Running Winsteps in Batch mode** Summary instructions for running Winsteps in batch mode.  
**Help** Displays help information for batch mode.  
**Edit** Edit batch file  
**Run** Run batch file: done by right-clicking on batch file name (.bat or .cmd), then clicking on **open on the right-click menu.**
20. **Data Setup menu**

<table>
<thead>
<tr>
<th>Data Setup</th>
<th>Start Control and data file setup</th>
<th>Ctrl+D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exit to Control and data file setup</td>
<td></td>
</tr>
</tbody>
</table>

This is described at Control and data file setup.

21. **Diagnosis menu**

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

A. **Item Polarity** check that all items are aligned in the same direction on the latent variable, same as Table 26. Check that all items have positive correlations. Use IREFER= and IVALUE= to point all items in the same direction, or KEY1= to correct a multiple-choice key error.

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

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

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

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

F. **Construct KeyMap** check that the item hierarchy is as intended (construct validity), same as Table 2. This locates items, response categories and your sample in one picture. Does your item measure hierarchy make sense? What is the typical person in your sample saying?

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

22. **Edit menu**

Display and editing of input and output files and tables.
Edit Control File= Display and edit the current control file. Alters this analysis if no computation has been done, otherwise the next analysis done with this control file.

Edit Report Output File= display and edit the report output file written during the main analysis phase. This contains Table 0 and output specified with TABLES= and TFILE=.

Edit/create new control file from= ....\template.txt template.txt is a generic control file which can be edited and saved under another name to give you a flying start at setting up your own analysis. There is a control and data file setup procedure. It is easier!

Edit/create file with wordpad launches WordPad or your own text editor.

Undo undo most recent change to the Output screen
Cut copy characters from an output screen line to the Windows clipboard and delete them from the screen
Copy copy characters from an output screen line to the Windows clipboard
Paste paste characters from the Windows clipboard to a screen line
Delete delete character from a screen line
for more substantial editing, save the screen using the File pull-down menu.

Edit initial settings change standard files and settings in Winsteps.ini
Edit Table ... display and edit a Table produced from the Diagnosis, Output Tables or other pull-down menu
Edit ... File display and edit a file produced from the Output Files or other pull-down menu

23. File menu

This menu launches and terminates Winsteps analysis.
Restart "WINSTEPS ..."  Restart this analysis, leaving the current one running.
Open File  Select control file for this analysis
Start another Winsteps  Launch a new copy of Winsteps. More than one copy of Winsteps can run at the same time.
Exit  Exit from Winsteps immediately
Finish iterating  Finish the current phase as quickly as possible.
Close open output windows  close any open windows for output tables, files or plots.
Enter  Acts as the Enter key
Save  Save the information displayed on the processing screen to disk.
Save As...  Save the screen output to a named disk file.
Print  Print the screen output
Excel= Location of the EXCEL program (if installed on your computer): can be change in Edit Initial Settings
SPSS= Location of the SPSS program (if installed on your computer): can be change in Edit Initial Settings
C:\WINSTEPS\examples\exam1.txt  Previously used Winsteps control files, select these to analyze them

24. Graphs menu

Winsteps produces bit-mapped images, using the Graphs menu. Winsteps produces character-based graphs in Table 21

<table>
<thead>
<tr>
<th>Graphs</th>
<th>Data Setup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category Probability Curves</td>
<td>Expected Score ICC</td>
</tr>
<tr>
<td>Cumulative Probabilities</td>
<td>Information Function</td>
</tr>
<tr>
<td>Category Information</td>
<td>Conditional Probability Curves</td>
</tr>
<tr>
<td>Test Characteristics Curve</td>
<td>Test Information Function</td>
</tr>
</tbody>
</table>

Initially, select which type of curves you want to see. You can look at the others later without going back to this menu. Graphs are plotted relative to the central difficulty of each item or response structure. Model-based curves (such as probability and information functions) are the same for all items which share the same model definition in ISGROUPS=. Empirical curves differ across items.

**Category Probability Curves:** model-based probability of observing each category of the response structure at each point on the latent variable (relative to the item difficulty)

**Empirical Category Curves:** data-based relative frequencies of categories in each interval along the latent variable

Expected Score ICC shows the model-based Item Characteristic Curve (or Item Response Function IRF) for the item or response structure. This is controlled BYITEM= or the last two entries in this menu.

Empirical ICC shows the data-based empirical curve.

**Empirical randomness** shows the observed randomness (mean-square fit) in each interval on the variable with logarithmic scaling. The model expectation is 1.0

Cumulative Probabilities plot the model-based sum of category probabilities. The category median boundaries are the points at which the probability is .5. Click on a line to obtain the category accumulation.

Item Information Function plots the model-based Fisher statistical information for the item. This is also the model variance of the responses, see RSA p. 100.

**Category Information** plots the model-based item information partitioned according to the probability of observing the category. Click on a line to obtain the category number.
Conditional Probability Curves show the model-based relationship between probabilities of adjacent categories. These follow dichotomous logistic ogives. Click on a line to obtain the category pairing.

Test Characteristic Curve is the model-based test score-to-measure characteristic curve.

Test Information Function plots the model-based test information function, the sum of the item information functions.

Test randomness shows the observed randomness (mean-square fit) in each interval on the variable with logarithmic scaling. The model expectation is 1.0.

Multiple Item ICCs supports the display of several model and empirical ICCs simultaneously.

Only on the Graphs menu:

Display by item shows these curves for individual items, also controlled by BYITEM=. Model-based output is the same for all items with the same ISGROUPS= designation.

Display by scale group for each ISGROUPS= code, a set of curves is shown. An example item number is also shown - all other items in the grouping are included in the one set of grouping plots. Also controlled by BYITEM=.

25. Help menu

Index displays the index of control variables

Contents displays the Table of Contents of the Help file.

About shows version number and compile date. Please mention these when reporting problems.

www.winsteps.com takes you to our website.

Bongo is the Winsteps "Adjutant's Call" - play this when you are summoning the data in preparation for constructing measures!

Scaling calculator is designed to help you linearly rescale your measures in the way most meaningful for your audience:

Under Current measure: enter two measures from your current analysis.
Under Desired measure: enter the values with which you want them to be reported.
Under Decimals: enter the number of decimal places for the measure Tables, Udecimals=.

Press Compute New to calculate the revised values of Uimean= and Uscale=.
The current values of Uimean= and Uscale= are displayed and also the revised New values. The New values can be altered if you wish.
Press **Specify New** to action the **New** values. Or the values can be copied (Ctrl+c) and pasted into your Winsteps control file.

### 26. Output Files menu

It produces output primarily intended for input into other software. These files are referenced in the **Output Files Index**

Most of these files can be written in several formats, so a dialog box is shown:

- **Display the Output File with:**
  - **Text Editor:** this is usually WordPad or your own text editor.
  - **Excel:** the file is automatically input into Excel. Tab-delimited format is recommended.
  - **SPSS:** SPSS is launched, if available: if this malfunctions, check that the path to SPSS is correct with **Edit Initial Settings**.
  - **Don't display:** the file is written to disk, but no further action is taken.

- **File format:**
  - **Text: space-separated: fixed field:** this is usually easiest to look at.
  - **Text: tab-delimited fields:** columns are separated by Tab characters, which EXCEL expects.
  - **Text: comma-separated fields:** columns are separated by commas or their international equivalents.
  - **Labels in "quotation marks"** place non-numeric values within quotation marks
  - **SPSS: .sav format:** this can be input directly into SPSS (compatible with SPSS 6.0 and later). To see its...
contents as plain text, use the SPSS menu. If this format is not available: Write the file in Excel (tab-separated). Then with variable names in row 1 and the data beginning in row 2, save the Excel data as an Excel spreadsheet (not a workbook) with the extension .xls. Use the SPSS Open command to import this into SPSS, and ask it to use the first row as variable names.

Column Headings:
Include column headings convenient when looking at the file, or reading into EXCEL
No column headings convenient when the file is to be processed directly by other software.

File status:
Permanent file: the file will not be deleted when Winsteps terminates. A file name is requested.
Temporary file: the file will be deleted when Winsteps terminates. "Save as" to make the file permanent.

27. Output Tables menu

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

<table>
<thead>
<tr>
<th>Output Tables</th>
<th>Output Files</th>
<th>Batch</th>
<th>Help</th>
<th>Specification</th>
<th>Compare Files</th>
<th>SPSS</th>
<th>Graphs</th>
<th>Setup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Request Subtables</td>
<td>Weight: selection</td>
<td>3.2 Rating (partial credit) scale</td>
<td>2.0 Measure forms (all)</td>
<td>1. Variable maps</td>
<td>2.2 General Keyform</td>
<td>2.5 Category Averages</td>
<td>3.1 Summary statistics</td>
<td>20. Score table</td>
</tr>
</tbody>
</table>

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

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

| SUMMARY OF CATEGORY STRUCTURE. Model="R" |
| | CATEGORY | OBSERVED | OBSVD SAMPLE | INFIT | OUTFIT | STRUCTURE | CATEGORY |
| | | | | | | | |
| | | | | | | |
|---------------------------------|------------|-------------|-------|
| 0 0 667 33 | -1.30 -1.30 | .96 .95 | NONE | (.204) 00 dislike |
| 1 1 757 37 | -.08 -.09 | .90 .78 | -.82 | .00 01 neutral |
| 2 2 609 30 | 1.40 1.41 | 1.09 1.33 | .82 | (.204) 02 like |

AVERAGE MEASURE is mean of measures in category.

Tables 27, 28, 30, 31, 33
These all allow the user to change the relevant control command on execution. ISUBTOTAL= controls the sub-total segments for Table 27 with a selection command so you are asked to confirm or change this value, before the Table is produced.
Request Subtables
Any Table (except Table 0) can be displayed using this command. It also accepts the special fields available with TFILE=

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

28. Plots menu

Here is the Plots menu.
Plotting problems? - these are usually due to the Winsteps-Excel interface. See www.winsteps.com/problems.htm
Compare statistics - enables you to draw scatterplots of Winsteps statistics within or between analyses. It also produces the tabular output of Table 34.
Bubble chart generates a Bond & Fox-style bubble chart.
Keyform Plot - Horizontal generates a horizontal keyform layout.
Keyform Plot - Vertical generates a vertical keyform layout.
Plot 30 - DIF plots the DIF values in Table 30.
Plot 31 - DPF plots the DPF values in Table 31.
Plot 33 - DIF & DPF plots the DIF+DPF values in Table 33.
29. SAS/SPSS menu

SAS: This is described under Data from SAS files
SPSS: This is described under Data from SPSS files

30. Specification menu

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

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

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

Specifications do not support "*".

So, instead of
- CLFILE=*  
  1 strongly disagree  
  2 disagree  
  3 agree  
  4 strongly agree

use the Edit menu, "Create/Edit with Wordpad", then, in Wordpad, type
- 1 strongly disagree  
- 2 disagree  
- 3 agree  
- 4 strongly agree

save as "clfile.txt" and in the Specification box, enter:

  CLFILE = clfile.txt
31. **Control and data file setup window**

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

Select "Setup Procedure" on Winsteps startup, or use the **Setup pull-down menu.**

![Setup Procedure Screen]

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

- **Data Setup**
  - Start Control and data file setup
  - Exit to Control and data file setup

This displays the Control File Set-Up screen:

![Control File Set-Up Screen]

A multiple-choice test key, \( \text{KEY1=} \), can be specified, if desired. Items can be clustered into similar response-structure groupings using \( \text{ISGROUPS=} \), using a one character code for each grouping.

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

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

32. **Reading in a pre-existent control file**

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

From the **Setup** screen, use the **File** pull-down menu:
Select your desired pre-existing control or data file:

This fills in the Setup screen:

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

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

33. Data display

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

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

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

Click on "Refresh Data Display" if the display does not show the current specification settings.
Press shift + left-click on the Person or Item No. row to dynamically select person label or item response columns.

### 34. Item labels

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

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

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

### 35. Category labels

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

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

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

Double-click on the bottom line for another blank line.
36. Do-it-yourself control and data file construction

There is a control and data file setup procedure. It is even easier!

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

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

1. Obtain your data

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

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

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

Your list will look like this:

Eating
Dressing
....
Walking
Stair climbing
END LABELS

You can use WordPad or your own text editor or pull-down the Winsteps "Edit" menu, and select "Create/Edit file with WordPad"

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

You will need some useful identifying codes for each row such as age, gender, demographics, diagnosis, time-point. Winsteps doesn't require these, but it is much more useful when they appear. Give each of these identifiers one or two letter codes, e.g., F=Female, M=Male, and give each identifier a column of pigeon-holes.

(c) The Data must be carefully lined up.
It is simpler if each data point, observation, response, rating can be squeezed into one character - numeric or alphabetic.

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

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

or, less conveniently,

M29B0012102110200102
F27W002122121010201020
After the END LABELS line, or in a separate file, on each line enter the *person* identifying codes. Line them up so that each column has a meaning. This is easier if you set the font to Courier.

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

**Excel, SPSS, SAS, ACCESS**

Your data may already be entered in a spreadsheet, statistics program or database. "Copy and Paste", Save As, Export or Print to disk the data from that program into "**DOS-text with line breaks**" or ASCII file.

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

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

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

In **SPSS**, see SPSS pull-down menu.

### 2. Set up your Winsteps Control and Data file

![Winsteps Interface](image)

(a) **Edit Template.txt**

Pull-down the Winsteps "Edit" menu, select "Create new control file from= ...\Template.txt"

The Template.txt will be displayed on your screen by WordPad or your own text editor.

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

Find the three sections:

- top: down to &END are Control Specifications
  
- middle: between &END and END LABELS are the Item Labels
  
- bottom: below END LABELS are the data
copy and paste the person labels and responses into this area.
one person per line (row).

(c) Edit the Control Specifications

Find the line "Title=
Replace Put your page heading here with your own page heading.

Look at a data line, and count across:

In which column does the person identifying label start, the first position of the person name?
This is the Name1= value e.g., if it is column 4, then Name1=-4
How long is the person identifying label, the name length?
This is the Namlen= value e.g., if it is 10 columns, then Namlen=10
In which column is the response to the first item?
This is the Item1= value e.g., if the first response is in column 12, then Item1=12
How many items are there, the number of items?
This is the NI= value e.g., if the number of items is 50, then NI=50
What are the valid item response codes in the data file?
This is the Codes= value e.g., if the codes are 1,2,3,4, then Codes=1234

If your codes are not numeric, then you will need to rescore them.
See Data recoding

This is usually enough for your first Winsteps run.

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

3. Run Winsteps

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

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

Extra specifications (or press Enter):
Press the Enter key

4. Your analysis commences

5. Your analysis concludes.

If there is an error message:
select "Edit Control File=
from the Winsteps Edit menu
correct the control file
save it
select "Exit, then Restart Winsteps" from the Winsteps File menu
If "Measures constructed" -
use the Output Tables pull-down menus to look at the Output Tables
here is the list of output tables.

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

37. Control file and template.txt

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

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

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

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

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

&INST ; optional
TITLE = "Put your page heading here"

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

; DATA = ; data after control specifications

; For rescoring
; 0 1 2 3 4 5 6 7
; 123456789012345678901234567890123456789012345678901234567890
38. Data file

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

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

By the term “response” is meant a data value which can be a category label or value, score on an item or a multiple-choice option code. The responses can be one or two characters wide. Every record must contain responses (or missing data codes) to the same items. The response (or missing data code) for a particular item must be in the same position in the same format in every record. If every person was not administered every item then mark the missing responses blank or make them some otherwise unused code, so that the alignment of item responses from record to record is maintained.
A table of valid responses is entered using the CODES= character string. Any other response found in your data is treated as missing. By using the CODES=, KEYn=, NEWSCORE= and IVALUE= options, virtually any type of response, e.g., "01", "1234", "1 2 3 4", "abcd", "a b c d", can be scored and analyzed. Missing responses are usually ignored, but the MISSCORE= control variable allows such responses to be treated as, say, "wrong".

When writing a file from SPSS, the syntax is:
FORMATS ITEM1 ITEM2 ITEM3 (F1). i.e., FORMATS varlist (format) [varlist..]
The procedure is FORMATS and then the variable list. Enclosed in parentheses is the format type. F signifies numeric while 1 signifies the width. (F2) would signify a numeric with a width of 2 columns for XWIDE=2. See pages 216 and 217 of the SPSS Reference Guide (1990). See also the SPSS pull-down menu.

39. Data from Excel and other spreadsheets

It is easy to copy data from an Excel spreadsheet into a Winsteps data file.

(i) Organize your data.
   Transform all item responses into columns one or two columns wide, e.g., "1" or "23"
   Transform all demographics into columns, one column wide, e.g., "M" and "F" for male and female.

(ii) Organize your Excel spreadsheet.
   Put all item responses (one item per column) into one block to the left of your spreadsheet.
   Put all person identifiers (one item per column) into one block, immediately to the right of the last item column.

(iii) Organize your column widths.
   Make all item column widths the same (usually one or two columns).
   Person identifier widths can match the identifiers, but these are best at one column wide.

(iv) Replace missing data with "**" or "," or "."
   Global replace nothing in a cell with a convenient clear missing data indicator, which is **not a number**.

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

(vi) Select all cells.

(vii) Copy into clipboard (Ctrl+C), or write to a tab-delimited file
      or write to a "Formatted Text (space delimited) (*.prn)" file

(viii) Open WordPad or your own text editor.
      Paste (Ctrl+V) or open the tab-delimited file.

(ix) Removing tabs
    Highlight a tab (area between two columns)
    Copy (Ctrl+C)
    Replace all tabs: (Ctrl+V) tab if necessary
    with nothing.

(x) The file should now look like a standard Winsteps rectangular data file.
    Save as a text file.

40. Data from SAS files

There are two approaches to analyzing SAS data with Winsteps:

1. Use the Winsteps SAS/SPSS menu: SAS option:
This may require you to install Microsoft MDAC and SAS OLE DB Local Provider freeware, see SAS Conversion Problems.

Clicking "Construct Winsteps file from SAS file" displays:

**Select SAS file:** choose the SAS .sas7bdat file that you want to convert to a Winsteps control and data file.

If this fails, you may not have installed the free SAS interface software or your version of Windows may not fully support OLE. See www.winsteps.com/problems.htm

This displays in the text box the details of the SAS file:

```
;     SAS dataset name: sastest
;     Number of Cases: 200
;     Number of SAS Variables: 6
;
; Move SAS variables under "! Person Label Variables"
; and "! Item Response Variables"
; SAS variables can be placed in both sections.
;   Numeric item variables are truncated to integers.
;   Constant fields may be specified with " "
; "XWIDE= is set according to the biggest item response value.
"Create..." when completed

! SAS File (do not delete this line): C:\WINSTEPS\examples\sastest.sas7bdat

! Person Label Variables. (Do not delete this line)

! Item Response Variables. (Do not delete this line)

!Other SAS Variables (ignored)
;Variable   Format
SN      ; double-precision floating-point value
item 1  ; double-precision floating-point value
item 2  ; double-precision floating-point value
ita     ; double-precision floating-point value
itb     ; double-precision floating-point value
itc     ; double-precision floating-point value

Cut-and-paste the SAS variables you want as Winsteps person and item variables:

! SAS File (do not delete this line): C:\WINSTEPS\examples\sastest.sas7bdat

! Person Label Variables. (Do not delete this line)
SN      ; double-precision floating-point value
Save and Exit: save the text box and exit from the SAS conversion. The contents of the text box will automatically redisplay if SAS conversion is requested from the same Winsteps run.

Permanent Output File: A Winsteps control and data file is created with a file name you select.

Temporary Output File: A Winsteps control and data file is created with a temporary name which will automatically be deleted when the conversion screen is closed.

Display Output File: display the converted Winsteps control and data file, temporary or permanent. This can be edited and saved or "saved as".

&INST
Title= "C:\WINSTEPS\examples\sastest.sas7bdat"
ITEM1= 1 ; Starting column of item responses
NI= 5 ; Number of items
; SAS Cases processed = 200
; datum: 0 count: 450
; datum: 1 count: 550
XWIDE = 1 ; this matches the biggest data value observed
CODES= 01 ; matches the data
NAME1 = 7 ; Starting column for person label in data record
; Person Label variables: columns in label: columns in line
@SN = 1E3 ; SN 1-3 7-9
@sascase = 5E7 ; case 5-7 11-13
NAMLEN = 7 ; Length of person label
&END ; Item labels follow: columns in label
item 1 ; Item 1 : 1-1
item 2 ; Item 2 : 2-2
ita ; Item 3 : 3-3
itb ; Item 4 : 4-4
itc ; Item 5 : 5-5
END NAMES
01000 70 1
11101 121 2
01101 86 3
.....

Launch Winsteps: launch a new copy of Winsteps using the permanent or temporary control and data file.

2. SAS provides an environment within which Winsteps can run. See Kazuaki Uekawa's instructions, www.estat.us/id2.html.

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

41. Data from SPSS files

Winsteps control and data files can easily be constructed from SPSS .sav files (compatible with SPSS 6.0 and
later). This can be done using the SPSS pull-down menu. Winsteps uses an interface routine provided by SPSS which should work for all recent SPSS versions.

There are two approaches:
(a) Make the Winsteps control and data files directly. This option also allows inspection of SPSS variable definitions.
(b) Convert the SPSS file into an EXCEL file for manipulation when SPSS is not available. This option can be used for examining the contents of any SPSS .sav file.

1. Select SPSS file and variables. This displays the format of the SPSS variables. This is a utility function, and can be used inspecing any SPSS .sav file.

Choose the variables you want in the person labels and the item response strings. Copy and paste the wanted variables under:

! Person Label Variables. (Do not delete this line)
and/or
! Item Response Variables. (Do not delete this line)

Constant fields can be added using " " or '

then click Save. Not "Save as"

Example of display:

; Move SPSS variables under "!Person Label Variables"
; and "!Item Response Variables"
; SPSS variables can be placed in both sections.
; Numeric variables are truncated to integers.
; XWIDE= is set according to the biggest item response value.
; "Save" when completed

! SPSS File (do not delete this line): C:\WINDOWS\Desktop\older driver.sav
;
Number of Cases: 144
;
Number of SPSS Variables: 55

! Person Label Variables. (Do not delete this line)
INITIALS ; A8  Pasted in
SUMMARY ; F8.2
"A" ; a constant field containing "A" used to identify these records when this data file is analyzed with other data files using DATA= file1 + file2 + ...

! Item Response Variables. (Do not delete this line)
"." ; a constant of . used to indicate missing data for the CLOCK0 item in other data files.
CLOCK1 ; F1.0 closed circle Pasted in
CLOCK2 ; F1.0 numbers in correct positions
!Other SPSS Variables (ignored)
;Variable Format Label
IDNUMBER ; F6.0
INITIALS ; A8
SUMMARY ; F8.2
IFNOT ; F8.2 if not complete
CLOCK1 ; F1.0 closed circle
CLOCK2 ; F1.0 numbers in correct positions

2. Edit SPSS variable selection
This permits you to change your variable selection.

3. Construct Winsteps file from SPSS file
This uses your variable selection to construct a Winsteps control and data file.
If "No SPSS variables selected", then back to step 1, and be sure to "Save", not "Save as"

Your SPSS file has been converted into a temporary .txt file. "Save as" your own permanent .txt file.

Example of Winsteps control and data file
; Save this file as your Winsteps control and data file
Title="C:\WINDOWS\Desktop\older driver.sav"
ITEM1=1
NI=2
XWIDE = 3
CODES = ; Please supply your values here
; SPSS Cases processed = 144
NAME1 = 8
; Person Label variables
; INITIALS 8-10
; SUMMARY 11-13
NAMLEN = 6
&END
CLOCK3 ; Item 1
CLOCK4 ; Item 2
END NAMES
0 1 kah6 1
1 1 b-f2 2
1 1 mbh1 3
....

Note: that SPSS variables are truncated to integers in the range 0-254 for Winsteps items. They are converted with decimal places for person variables.

4. Convert SPSS file to EXCEL file
The selected SPSS file is converted into a Tab-delimited file and EXCEL is launched. EXCEL automatically reads the SPSS variables into columns. "Save as" to keep this worksheet.

5. Convert SPSS file to EXCEL-compatible Tab-separated file
The selected SPSS file is converted into a Tab-delimited file. Copy and paste this file, or "Save as". This is a utility-function for converting SPSS files for any purpose.

42. Data from STATA files
Fred Wolfe has provided a Stata module which produces WINSTEPS control and data files. It is at ideas.uqam.ca/ideas/data/Softwares/bocbocodeS423302.html and has an accompanying help file.

Here is what this file looked like on 12-25-2001:

*! raschcvt version 1.2.1 fw 8/1/00 Prepares data file and control file for Winsteps
*! 5/26/01 update to version 7. turns off log output, 1.2.2 (12/20/01) corrects
"! decimal point formatting error and removes " = 4 0".
program define raschcvt
version 7.0
syntax varlist, outfile(string) id(varlist) max(integer) [min(integer 0)] [Xwide(integer 2)]
tokenize `varlist'
set more off
preserve
confirm variable `id'
di in gr "Building Rasch files"
local itemno = 0
local counter = 0
while "$i" != "$id" {
local itemno = `itemno' + 1
mac shift
}
capture log close
qui log using `outfile'.con, replace
di "; The id variable is " "id"
di "; There are `itemno' items"
di "; Items and lengths"
tokenize `varlist'
while "; `1'" != ";id"
  {  
capture assert `1' == int("`1'")
    if _rc != 0 {
      di "non-integer value found"
      exit _rc
    }
    local counter = `counter' + 1
    qui su `1'
    local f = length("r(max)"")
    di "; `1' " "f' = " g'
    if `counter' <= `itemno' {
      local f = `xwide'
      format `1' %0`f'.0f
      qui replace `1' = 99 if `1' == . & `f' == 2
      qui replace `1' = 999 if `1' == . & `f' == 3
    }
    macro shift
  }
order `varlist'
qui outfile `varlist' using `outfile'.dat, nolabel wide replace
di "; outfile' has been written to disc"
di in gr "; Start control file below"
di "TITLE=
  "DATA=`outfile'.dat"
di "ITEM1=1"
di "NT= " "itemno"
  
di "NAME1= `itemno' +1"
  
di "DELIMITER = SPACE"
  
di ";XWIDE= `xwide'
  
di "CODES=" _c
  
if `xwide' == 1{
  for num `min' / `max', noheader: di X _c
  }
if `xwide' == 2{
  for num `min' / `max', noheader: if X == 0(di "00"_c) \ if X > 0 & X < 10(di "0"X _c) \ if X > 9 & X < 99 (di "0"X _c) 
  }
if `xwide' == 3{
  for num `min' / `max', noheader: if X == 0(di "000"_c) \ if X > 0 & X < 10(di "00"X _c) \ if X > 9 & X < 99 (di "0"X _c) 
  }
di
  
di "MAXPAG=60"
  
di "PRCOMP=5"
  
di "MJMLE=0"
  
di "; CODES=" _c
  
di ";LCONV= .001
  
di ";RCONV= .00
  
di ";ISGROUPS=
  
di ";HLINES=Y
  
di ";PSELECT= ?????1*
  
di ";TABLES=1111011001111100000111111
  
di ";ISFILE=`outfile'.isf"
  
di ";IFILE=`outfile'.IFL"
  
di ";XFILE=`outfile'.XFL"
  
}
di "&END"
di  
tokenize `varlist'
while "; `1'" != ";id"
  {
    local lbl:variable label `1'
di `"lbl"'
macro shift
}
di "END LABELS"
di
tokenize `varlist'
while `"1"' != `"id"' {
   di `"1"'
   macro shift
}
di "END NAMES"
qui log close
restore
end

43. **Data file with other delimiters**

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

44. **Example 0: Rating scale data: The Liking for Science data**

*Rather than attempting to construct a control file from scratch, it is usually easier to find one of these examples that is similar to your problem, and modify it.*

Control and data file, Example0.txt, for the Liking for Science data (see RSA) contains the responses of 75 children to 25 rating-scale items. The responses are 0-dislike, 1-neutral, 2-like. To analyze these data, start WINSTEPS, then:

Control file name?:
"Files" pull-down menu
"Control file name"
"Examples" folder
Example0.txt
Open
Report output file: (Enter) press Enter for a temporary file.
Extra specifications:(Enter) no extra specifications at present.

; This is file "example0.txt" - ";
TITLE= 'LIKING FOR SCIENCE (Wright & Masters p.18)'
NI= 25 ; 25 items
ITEM1= 1 ; responses start in column 1 of the data
NAME1= 30 ; person-label starts in column 30 of the data
ITEM= ACT ; items are called "activities"
PERSON= KID ; persons are called "kids"
CODES= 012 ; valid response codes (ratings) are 0, 1, 2
CLFILE= * ; label the response categories
0 Dislike ; names of the response categories
1 Neutral
2 Like
* ; "*" means the end of a list
&END ; this ends the control specifications
; These are brief descriptions of the 25 items
WATCH BIRDS
READ BOOKS ON ANIMALS
READ BOOKS ON PLANTS
WATCH GRASS CHANGE
FIND BOTTLES AND CANS
LOOK UP STRANGE ANIMAL OR PLANT
WATCH ANIMAL MOVE
LOOK IN SIDEWALK CRACKS
LEARN WEED NAMES
LISTEN TO BIRD SING
FIND WHERE ANIMAL LIVES
GO TO MUSEUM
GROW GARDEN
LOOK AT PICTURES OF PLANTS
READ ANIMAL STORIES
MAKE A MAP
WATCH WHAT ANIMALS EAT
GO ON PICNIC
GO TO ZOO
WATCH BUGS
WATCH BIRD MAKE NEST
FIND OUT WHAT ANIMALS EAT
WATCH A RAT
FIND OUT WHAT FLOWERS LIVE ON
TALK W/FRIENDS ABOUT PLANTS

END NAMES ;this follows the item names: - the data follow:
1211102012222021122021020 ROSSNER, MARC DANIEL
2222222222222222222222222 ROSSNER, LAWRENCE F.
2211011012222122121221111 ROSSNER, TOBY G.
1010010121221111122021111 ROSSNER, MICHAEL T.
101010100111012211111110 ROSSNER, REBECCA A.
1011211112101221101210 ROSSNER, TR CAT
2220022202222222222222222 WRIGHT, BENJAMIN
221002102222020122022022 LAMBERT, MD., ROSS W.
0110100112122001121120111 SCHULZ, MATTHEW
2100010122221122011021 SCHIEH, DANIEL SEB
222001110222220222022022 Hsieh, Paul Fred
0100220101210011021000101 LIEBERMAN, DANIEL
121100010212112112111000 LIEBERMAN, BENJAMIN
2110002012200210002200012 HWA, NANCY MARIE
1111111111111111211111111 DYSON, STEPHIE NINA
2221121222222222222222222 BUFF, MARGE BABY
2222012222222222222222222 SCHATTNER, GAIL
2222012222222222222222222 ERNST, RICHARD MAX
2221011002222211220202022 FONTANILLA, HAMES
110001002122122122020202 ANGUIANO, ROB
1111011112122111122011111 EISEN, NORM L.
12111111121221122121121212 HOogan, KATHLEEN
2211111212222212222222222 VROOM, JEFF
12210121022221022222222222 TOZER, AMY ELIZABETH
2221122202222222222222222 SEILER, KAREN
1111111111111111111111111 NEIMAN, RAYMOND
1110011011221111022120221 Denny, Don
221110200122121022012011 ALLEN, PETER
10001010011000022110200 LANDMAN, ALAN
1000000002122000220121000 NORDGREN, JAN SWEDE
2211011112221011122121111 SABILE, JACK
121010011222110121022021 ROSSNER, JACK
22120121212222222222222222 ROSSNER, BESS
22212212222222222222221222 PASTER, RUTH
1110221011111211221120111 RINZLER, JAMES
1110201001122111122011111 AMIRAULT, ZIPPI
121001001221211122021021 AIRHEAD, JOHN
221102112211222112222121 MOOSE, BULLWINKLE
221121221222222212222222 SQUtREL, ROCKY J.
221101122222222222222222 BADENOV, BORIS
2222222222222222222222222 FATALE, NATASHA
212101111222112212011021 LEADER, FEARLESS
222011022222212222202202 MAN, SPIDER
121002102122112110221012 CIANCI, BUDDY
2221222222222222222222222 MCLoughlin, BILLY
2200020002222202221220210 MULLER, JEFF
1110221101221111022101000 VAN DAM, ANDY
2212022222222222122122022 CHAZELLE, BERNIE
110011010111010111111110 BAUDET, GERARD
2212122222222222222222122 DOEPNER, TOM
When analysis completes, use the "Output Tables" pull-down menu to examine the measures and quality-control fit statistics. Start with Table 1, which gives you a quick summary. Here are the Science Activity items in measure order, Table 13. The hierarchy of item labels, in the right-hand column, defines the construct that has been measured. "Go on picnic" is easiest to like. "Find bottles" is hardest to like.

### ACT STATISTICS: MEASURE ORDER

<table>
<thead>
<tr>
<th>ENTRY</th>
<th>RAW</th>
<th>MODEL</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>PTEMA</th>
<th>ACT</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUMBER</td>
<td>SCORE</td>
<td>COUNT</td>
<td>MEASURE</td>
<td>S.E.</td>
<td>MNSQ</td>
<td>ZSTD</td>
</tr>
<tr>
<td>5</td>
<td>35</td>
<td>74</td>
<td>2.42</td>
<td>.22</td>
<td>.30</td>
<td>6.414.11</td>
</tr>
<tr>
<td>23</td>
<td>40</td>
<td>74</td>
<td>2.18</td>
<td>.21</td>
<td>.41</td>
<td>6.314.11</td>
</tr>
<tr>
<td>20</td>
<td>48</td>
<td>74</td>
<td>1.83</td>
<td>.20</td>
<td>1.33</td>
<td>2.011.82</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>74</td>
<td>1.75</td>
<td>.20</td>
<td>.89</td>
<td>.71</td>
</tr>
<tr>
<td>8</td>
<td>52</td>
<td>74</td>
<td>1.67</td>
<td>.20</td>
<td>1.10</td>
<td>.71</td>
</tr>
<tr>
<td>7</td>
<td>67</td>
<td>74</td>
<td>1.10</td>
<td>.19</td>
<td>.97</td>
<td>.11</td>
</tr>
<tr>
<td>9</td>
<td>78</td>
<td>74</td>
<td>.71</td>
<td>.19</td>
<td>1.18</td>
<td>1.311.17</td>
</tr>
<tr>
<td>16</td>
<td>81</td>
<td>74</td>
<td>.60</td>
<td>.19</td>
<td>.97</td>
<td>.21</td>
</tr>
<tr>
<td>25</td>
<td>83</td>
<td>74</td>
<td>.53</td>
<td>.19</td>
<td>.80</td>
<td>1.5</td>
</tr>
<tr>
<td>3</td>
<td>86</td>
<td>74</td>
<td>.42</td>
<td>.19</td>
<td>.57</td>
<td>3.5</td>
</tr>
<tr>
<td>14</td>
<td>86</td>
<td>74</td>
<td>.42</td>
<td>.19</td>
<td>.82</td>
<td>1.3</td>
</tr>
<tr>
<td>6</td>
<td>89</td>
<td>74</td>
<td>.31</td>
<td>.19</td>
<td>.81</td>
<td>1.4</td>
</tr>
<tr>
<td>17</td>
<td>93</td>
<td>74</td>
<td>.16</td>
<td>.19</td>
<td>.65</td>
<td>2.7</td>
</tr>
<tr>
<td>22</td>
<td>95</td>
<td>74</td>
<td>.08</td>
<td>.19</td>
<td>.83</td>
<td>1.2</td>
</tr>
<tr>
<td>24</td>
<td>105</td>
<td>74</td>
<td>.31</td>
<td>.20</td>
<td>.90</td>
<td>1.6</td>
</tr>
<tr>
<td>1</td>
<td>107</td>
<td>74</td>
<td>.40</td>
<td>.21</td>
<td>.55</td>
<td>3.5</td>
</tr>
<tr>
<td>15</td>
<td>109</td>
<td>74</td>
<td>.48</td>
<td>.21</td>
<td>.78</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>114</td>
<td>74</td>
<td>.71</td>
<td>.22</td>
<td>.93</td>
<td>1.4</td>
</tr>
<tr>
<td>21</td>
<td>117</td>
<td>74</td>
<td>.85</td>
<td>.22</td>
<td>.84</td>
<td>1.9</td>
</tr>
<tr>
<td>11</td>
<td>119</td>
<td>74</td>
<td>.96</td>
<td>.23</td>
<td>.63</td>
<td>2.4</td>
</tr>
<tr>
<td>13</td>
<td>125</td>
<td>74</td>
<td>1.29</td>
<td>.25</td>
<td>1.22</td>
<td>1.1</td>
</tr>
<tr>
<td>10</td>
<td>128</td>
<td>74</td>
<td>1.49</td>
<td>.26</td>
<td>1.78</td>
<td>1.1</td>
</tr>
<tr>
<td>12</td>
<td>135</td>
<td>74</td>
<td>2.04</td>
<td>.31</td>
<td>.70</td>
<td>1.2</td>
</tr>
<tr>
<td>19</td>
<td>139</td>
<td>74</td>
<td>2.48</td>
<td>.36</td>
<td>1.08</td>
<td>1.1</td>
</tr>
<tr>
<td>18</td>
<td>143</td>
<td>74</td>
<td>3.15</td>
<td>.47</td>
<td>1.50</td>
<td>1.2</td>
</tr>
</tbody>
</table>

**Example 1: Dichotomous data: Simple control file with data included**

A control file, EXAM1.TXT, for an analysis of the Knox Cube Test (see BTD) a test containing 18 items, each item
is already scored dichotomously as 0, 1. The person-id data begins in column 1 and the item string begins in column 11. No items will be deleted, recoded, or anchored. The number of data lines is counted to determine how many children took the test. Use the pull-down "Diagnosis" and "Output Table" menus to see the output tables. For an explanation of the output obtained, see later in the manual. Run this example with:

Control file: EXAM1.TXT
Report output file: (Enter)
Extra specifications: (Enter)

; This file is EXAM1.TXT - (";" starts a comment)
TITLE='KNOX CUBE TEST' ; Report title
NAME1=1                ; First column of person label in data file
ITEM1=11               ; First column of responses in data file
NI=18                  ; Number of items
CODES=01               ; Valid response codes in the data file
CLFILE=*               ; Labels the observations
0 Wrong                ; 0 in data is "wrong"
1 Right                ; 1 in data is "right"
*                       ; "*" is the end of a list
PERSON=KID             ; Person title: KID means "child"
ITEM=TAP               ; Item title: TAP means "tapping pattern"
&END                    ; Item labels for 18 items follow
1-4                     ; tapping pattern of first item: cube 1 then cube 4 are tapped.
2-3
1-2-4
1-3-4
2-1-4
3-4-1
1-4-3-2
1-4-2-3
1-3-2-4
2-4-3-1
1-3-1-2-4
1-3-2-4-3
1-4-3-2-4
1-4-2-3-4-1
1-3-2-4-1-3
1-4-2-3-1-4
1-4-3-1-2-4
4-1-3-4-2-1-4          ; last tapping pattern: 7 actions to remember!
END NAMES               ; END NAMES or END LABELS must come at end of list
Richard M 111111111000000000    ; Here are the 35 person response strings
Tracie  F 111111111110000000
Walter  M 111111111001000000
Blaise  M 111100101000000000
Ron     M 111111111100000000
William M 111111111100000000
Susan   F 111111111110000000
Linda   F 111111111000000000
Kim     F 111111111100000000
Carol   F 111111111100000000
Pete    M 111011111000000000
Brenda  F 111110101100000000
Mike    M 111110011111100000
Zula    F 111111111110000000
Frank   M 111111111111000000
Dorothy F 111111111100000000
Rod     M 111101111110000000
Britton F 111111111101000000
Janet   F 111111111100000000
David   M 111111111101000000
Thomas  M 111111111110000000
Betty   F 111111111111000000
Bert    M 111111111100110000
Rick    M 111111111110100110    ; best performance
46. Example 2: Control and anchor files

A control file, EXAM2.TXT, for the analysis of a test containing 18 items, each item already scored dichotomously as 0,1. The person-id data begins in column 1 and the item-response string begins in column 11. The standard tables will be appear in the printout. There is user scaling. Items 2, 4, 6 and 8 are anchored at 400, 450, 550 and 600 units respectively, supplied in file EXAM2IAF.TXT. Your data is in file EXAM2DAT.TXT:

; This file is EXAM2.TXT
TITLE='KNOX CUBE TEST - ANCHORED' ; the title for output
NI=18 ; the number of items
ITEM1=11 ; position of first response in data record
NAME1=1 ; first column of person-id in data record
PERSON=KID
ITEM=TAP
DATA=EXAM2DAT.TXT ; name of data file
IAFILE=EXAM2IAF.TXT ; this is item anchor (input) file: it is the IFILE= of an earlier analysis
CONVERGE=L ; use only Logits for convergence criterion
LCONV=.005 ; converged when biggest change is too small to show on any report.
; What follows is equivalent to the IAFILE= above
; IAFILE=* ; item anchor file list
;2 400 ; item 2 anchored at 400 units
;4 450 ; item 4 anchored at 450 units
;6 550 ; item 6 anchored at 550 units
;8 600 ; item 8 anchored at 600 units
; *
UIMEAN=500 ; user scaling - item mean
USCALE=100 ; user scaling - 1 logit = 100 user units
UDECIM=0 ; print measures without decimals
&END
1-4 ; item labels, starting with the first item
4-1-3-4-2-1-4
END NAMES ; End of this file

The anchoring information is contained in file EXAM2IAF.TXT and contains the following lines, starting in column 1:

2 400 ; item 2 anchored at 400 units:
; if logits are user-rescaled, then anchor values are also expected to be user-rescaled.
; for logit anchor values, specify UANCHOR=no
4 450 ; item 4 anchored at 450 units
6 550 ; item 6 anchored at 550 units
8 600 ; item 8 anchored at 600 units

Item calibration files, IFILE=, from prior runs can be used as item anchor files, IAFILE=, of later runs.

Your data is in the separate file, EXAM2DAT.TXT, with person-id starting in column 1, and item responses starting in column 11.
47. Example 3: Item recoding and item deletion

The test has 25 items, specified in EXAM3.TXT. The item response string starts in column 12. Person-id’s start in column 1 (the standard value). Original item codes are "0", "1", "2" and "X". All items are to be recoded and the original-to-new-code assignments will be 0 0, 1 2, 2 1 and X 3. Items 5, 8, and 20 through 25 are to be deleted from the analysis, and are specified in the control. The misfit criterion for person or item behavior is 3.0. Tables 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 13, 15, 17, 19, 20 and 21 are to appear in your report output file EXAM3OUT.TXT. Sequence numbers are used as item names. Data are in file EXAM3DAT.TXT.

48. Example 4: Selective item recoding

The test has 18 items, specified in file EXAM4.TXT. The response string starts in column 1. Person-id’s start in column 41. Original codes are 0,1 in data file EXAM4DAT.TXT. Items 2, 3, and 4 are to be recoded as 1,0. All tables are to appear in report file EXAM4OUT.TXT, in a standardized form.
49. Example 5: Scoring key for items, also CAT responses

A multiple choice adaptive test, in file EXAM5.TXT with responses "a", "b", "c", "d" and a scoring key for 69 items. Your data are in the control file. This was administered as a CAT test, then the response file formatted into a "flat" file with one row per person and one column per item.

; This file is EXAM5.TXT
TITLE="An MCQ Test" ; the title
NI=69 ; 69 items
ITEM1=10 ; response string starts in column 10
NAME1=1 ; person-id starts in column 1
CODES=abcd ; valid response codes
MISSCORE=-1 ; standard scoring of missing data, means that blanks are ignored
KEY1 = dcbbbadbcaccdadaabacbdaccedcddcddbcdbdabcaccaacbcb ; scoring key of correct answers
ITEM=TOPIC ; items are topics
PERSON=STDNT ; respondents are students
NAMLMP=2 ; first 2 characters on maps, e.g., nl
PFILE=EXAM5PF.TXT ; write out person measures
CSV=Y ; separate values by commas in PFILE=
HLINES=N ; write out no heading lines in PFILE=
; Many spreadsheets and statistics programs expect a file of numbers separated by commas.
; Use IFILE= or PFILE= with CSV=Y and HLINES=N.
MJMLE=0 ; allow as many JMLE iterations as necessary
EXTRSC=0.5 ; most conservative (central) extreme measures wanted
XFILE=EXAM5XF.TXT ; write out individual response residual file
&END
n101 Month
n102 Sign
s02 newspaper
s03 newspaper
END NAMES
IM  CAT a dcacc ccabbcac
NM  KAT b badad acacaab aa c dd ab c
NH  RIC ddb b dbcbcaadba ba acd bad db c d cc
IL  HOL a a da d d cbdddc bcd dc ca
50.

Example 6: Keys in data record FORMAT

Do not use FORMAT= unless there is no other way. It is tricky to set up correctly. MFORMS= is much easier.
A test of 165 multiple-choice items with multiple data lines per data record. The scoring key is formatted in the
same way as the data lines:
; This file is EXAM6.TXT
TITLE='Demonstration of KEY1 record'
title
FORMAT=(1X,10A,T23,50A,/,T23,50A,/,T23,50A,/,T23,15A)

;
;
;
;
;
;
;
;
;
;
;
;
;
;
;
;

The first character is ignored,
then 10 characters in first record are person id,
then starting in column 23,
50 columns in first 3 records,
and 15 responses in fourth record.
Using MFORMS= to
MFORMS=*
Data= filename
L=4
P1-10=2
I1-50=23
I51-100=2:23
I101-150=3:23
I151-165=4:23
*

reformat the same data record
;
;
;
;
;
;
;
;
;

put the input data in a separate file
4 data input lines per output record
person label characters 1-10 start in column 2 of line 1 of input data
responses to items 1-50 start in column 23 of line 1 of input data
responses to items 51-100 start in column 23 of line 2 of input data
responses to items 101-150 start in column 23 of line 3 of input data
responses to items 151-165 start in column 23 of line 4 of input data
end of MFORMS=
Note this does not reformat the Keyfrm=, so use KEY1=

; In the reformatted record
NAME1=1
ITEM1=11
NI=165
CODES="ABCD "
MISSCORE=0
KEYFRM=1
RFILE=exam6rf.txt
PTBIS=YES
TFILE=*
1.0
3
6
10
20
*

;
;
;
;
;
;
;
;
;
;
;
;
;

Person-id starts in column 1
Item responses start in column 11 of reformatted record
There are 165 items
The raw responses are ABCD and BLANK.
Put character strings in " " if blanks are to be included.
Blanks and invalid codes scored wrong=0
There is a KEY1 record after &END which is formatted
exactly like a data record specifying the correct responses.
this shows effect of reformatting and scoring
Raw score point-biserial
List of desired tables
Subtable 0 of Table 1
Table 3

; end of list

&END
; KEY1= formatted like your data follows:
Key 1 Record
CDABCDBDABCADCBDBCADBABDDCDABCBABDCACBADACBADBAACD
after &END
CCBDACABDADCBDCABBCACDBAABCDADCDCADBCABCDCADABACDA
in FORMAT= format
BADCDBADCBADCDBACBADBCAADBCBBDCBACDBACBADCDADBACDB
before item names
ABDACDCDBADBCAB
A1
First item name
A2
.
A164
A165
END NAMES
090111000102
10001 BDABADCDACCDCCADBCBDBCDDACADDCACCCBCCADBDABADCAADD
ABDDDCABDADCBDACDBCACADABCDCCDCBDBCCABBCDCADDCDCDA
BDCCDBABCDCDDDCBADCACBDCBDBACBCBCADBABAADCDCBABAAC
DCBCCACABCDDCBC
090111000202
10002 BDCDCDCDADCBCCBDBDCABCBDACDABCAABCAACBBBACAADDAACA
ACBCACBBDADCBDCBBBCDCCDACCBCADCACCAACDBCCDADDBACDA
BADCDCBDBDCDCCBACCCBBAABDBCDBCCBAADBABBADBDDABDCAA

50


or, using continuation lines for the key:

CODES="ABCD " ; The raw responses are ABCD and BLANK
;KEYFRM=  ; omit this, not needed
KEY1 = CDABCBDBACDCCDDBBACDBBCBABCACBADDACBCADDBACADCD
+CCBDBACDCCBDADADDACBCBDDBBDACCBDDBDBBDACBACDBACDBACDBABDBC
KEY2=ABCDABCCBADDABACABBBACCCCDABAB  ; Partially correct
KEY3=DABABAAACC*CCABAC************* ; Some partially correct
; if no matching response, use a character not in CODES=, e.g., *
; the keys are matched in sequence, "B" for item 15 matches Key1=, and
Key3= is ignored
KEYSCR=211  ; KEY1 fully correct (2 points),
; KEY2, KEY3 partially correct (1 point each)
ISGROUPS=0  ; Each item is its own grouping, i.e., the Partial Credit model
MODELS=R  ; Each item has its own Andrich rating scale
STKEEP=Y  ; Keep all intermediate categories in analysis, even if never observed
CURVES=111 ; Print all 3 item curves in Tables 2 and 21
CATREF=2  ; Use category 2 for ordering in Table 2
&END

51. Example 7: A partial credit analysis

A 30 item MCQ Arithmetic Test is to be analyzed in which credit is given for partial solutions to the questions. Each item is conceptualized to have its own response structure, as in the Masters' Partial Credit model. Estimation for the Partial Credit model is described in RSA, p. 87.

In this example, item 1 has 3 scoring levels. "C" is correct, worth 2 points. "A" and "D" are partially correct, worth 1 point. "B" is incorrect, worth 0 points. CODES= identifies all possible valid responses. In this example, KEY1= identifies responses worth 2 points. KEY2= and KEY3= identify responses worth 1 point. The values of KEY1=, KEY2= and KEY3= are set by KEYSCR=. So for item 1, KEY1=C,..., KEY2=A..., and KEY3=D. Response B is not in a KEY= and so is scored 0. Here, invalid responses are treated as not-administered. If invalid responses are to be treated as "wrong", specify MISSCORE=0.

; This file is EXAM7.TXT
TITLE="A Partial Credit Analysis" page heading
NAME1=1 ; Person-id starts in column 1
ITEM1=23 ; Item responses start in column 23
NI=30 ; There are 30 items
CODES=ABCD ; Scores entered as A through D
KEY1=CDABCBDBACDCCDDBBACDBBCBABCACBADDACBCADDBACADCD
+CCBDBACDCCBDADADDACBCBDDBBDACCBDDBDBBDACBACDBACDBACDBABDBC
KEY2=ABCDABCCBADDABACABBBACCCCDABAB  ; Partially correct
KEY3=DABABAAACC*CCABAC************* ; Some partially correct
; if no matching response, use a character not in CODES=, e.g., *
; the keys are matched in sequence, "B" for item 15 matches Key1=, and
Key3= is ignored
KEYSCR=211  ; KEY1 fully correct (2 points),
; KEY2, KEY3 partially correct (1 point each)
ISGROUPS=0  ; Each item is its own grouping, i.e., the Partial Credit model
MODELS=R  ; Each item has its own Andrich rating scale
STKEEP=Y  ; Keep all intermediate categories in analysis, even if never observed
CURVES=111 ; Print all 3 item curves in Tables 2 and 21
CATREF=2  ; Use category 2 for ordering in Table 2
&END

51
52. Example 8: Items with various rating scale models

A 4 item test in which each item has a 9-level scoring rubric. We suspect that there are not really 9 levels of competence. After several WINSTEPS analyses, we choose to recode the rating scale or partial credit items by collapsing categories in order to increase measurement effectiveness (separation) and increase parameter stability. An objection to collapsing is that it violates the Rasch model. This is only true if the uncollapsed data strictly accord with the model. In fact, the collapsed data may fit the Rasch model better than the uncollapsed. We have to compare the collapsed and uncollapsed analyses to decide.

; This file is EXAM8.TXT
TITLE="Success and Failure Items"
NAME1=6
ITEM1=1
NI=4 ; 4 items
ISGROUPS=1234 ; one item per grouping, same as ISGROUPS=0
IREFER=ABCD ; the 4 items are to be recoded differently. Item 1 is type "A", etc.
CODES=123456789 ; the codes in the data file
IVALUEA=333456666 ; the recoding for A-type items in IREFER=, i.e., Item 1
IVALUET=333444444
IVALUED=444456777
&END

Maze
Passengers
Blocks
Egg race

END NAMES

Table 14.3 shows the recoding:

<table>
<thead>
<tr>
<th>ENTRY NUMBER DATA CODE VALUE</th>
<th>DATA SCORE</th>
<th>COUNT %</th>
<th>USED COUNT %</th>
<th>AVERAGE</th>
<th>OUTFIT</th>
<th>ITEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1 3</td>
<td>1 3</td>
<td>-2.39</td>
<td>Maze</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>7 22</td>
<td>6 19</td>
<td>-1.57</td>
<td>.7</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>10 31</td>
<td>10 32</td>
<td>-.54</td>
<td>.5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>10 31</td>
<td>10 32</td>
<td>1.23</td>
<td>.6</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>3 9</td>
<td>3 10</td>
<td>2.42</td>
<td>.7</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>1 3</td>
<td>1 3</td>
<td>2.78</td>
<td>.5</td>
<td></td>
</tr>
</tbody>
</table>

53. Example 9: Grouping and modeling items

A 20 item test. Items 1, 2 are dichotomous items, coded "Y", "N". Item 3 is a "Partial Credit" item. Items 4-10 are all ratings on one Andrich scale or test-level partial-credit scale (Grouping 1), and items 11-20 are all ratings on another Andrich scale or test-level partial-credit scale (Grouping 2). These are grouped with ISGROUPS=.

Winsteps discovers from the data what the item structure is. Items 3-20 have response codes "A", "B", "C", "D", "E" or "a", "b", "c", "d", "e".

; This file is EXAM9.TXT
TITLE="Grouping and Modeling"
ITEM1=11 ; Item responses start in column 11
NI=20 ; There are 20 items
ISGROUPS=DD011111112222222222 ; The groupings in item order
IREFER=DDAAAAAAAAAAAAAAAA ; for recoding
CODES = 10ABCDEabcde ; Response codes to all items
IVALUED= 10********** ; for Items 1 & 2
IVALUEA= **1234512345 ; for Items 3-20
DATA=EXAM9DAT.TXT ; Location of data file
IWEIGHT=* ; Item weighting file list
 3 2.5 ; Item 3 has weight of 2.5
+
&END

RD Prompt 1 ; Item id's remind us of MODELS= and ISGROUPS=
RD Prompt 2 ; MODELS= and ISGROUPS= are shown in item measure Tables 10, 13, 14, 15

S0 Logic
R1 Grammar 1
R1 Grammar 2
R1 Grammar 3
R1 Grammar 4
R1 Grammar 5
R1 Grammar 6
R1 Grammar 7
R2 Meaning 1
R2 Meaning 2
R2 Meaning 3
R2 Meaning 4
R2 Meaning 5
R2 Meaning 6
R2 Meaning 7
R2 Meaning 8
R2 Meaning 9
R2 Meaning 10
END NAMES

The data is in file EXAM9DAT.TXT:

Richard M 00bCDCDddCDdddCDccE
Tracie F 00BcBABBccBbbBbBBBb
James M 00ccaBbabBAcbacbaBbb
Joe M 10BdBBBccBccbbccbcC

54. Example 10: Combining tests with common items

This uses MFORMS=, but it can also be done, more awkwardly, with FORMAT=

Test A, in file EXAM10A.TXT, and TEST B, in EXAM10B.TXT, are both 20 item tests. They have 5 items in common, but the distractors are not necessarily in the same order. The responses must be scored on an individual test basis. Also the validity of each test is to be examined separately. Then one combined analysis is wanted to equate the tests and obtain bankable item measures. For each file of original test responses, the person information is in columns 1-11, the item responses in 41-60.

The combined data file specified in EXAM10C.TXT, is to be in RFIELD= format. It contains

Person information of 11 characters: Columns 1-30 (always), but only 1-11 are active.
Item responses to 35 items: Columns 31-64

The identification of the common items is:
Test Item Number (=Location in item string)
Bank: 1 2 3 4 5 6-20 21-35
A: 3 1 7 8 9 2,4-6,10-20
B: 4 5 6 2 11 1,3,7-10,12-20
I. From Test A, make a response (RFILE=) file.

; This file is EXAM10A.TXT
TITLE="Analysis of Test A"
RFILE=EXAM10AR.TXT ; The constructed response file for Test A
NI=20 ; 20 items
ITEM1=41 ; Items start in column 41 of data record
NAME1=1 ; Start of person label
NAMELEN=11 ; Length of person label
CODES="ABCD# " ; Beware of blanks and # meaning wrong!
; Blanks are included in CODES=, but they are scored incorrect, because
never keyed correct
KEY1=CCBDACABDABCBCABCA ; the MCQ key
&END
BANK 2 TEST A 1 ; first item name
BANK 6 TEST A 2
BANK 1 TEST A 3
.
BANK 20 TEST A 20
END NAMES
Person 01 A                             BDABCDBDBABCACBCDBBA
.
Person 12 A                             BADCACDCABCABBCACB

The RFILE= file, EXAM10AR.TXT, is:

00000000110010001001
..........
00001110000001001011

II. From Test B, make a response (RFILE=) file.

; This file is EXAM10B.TXT
TITLE="Analysis of Test B"
RFILE=EXAM10BR.TXT ; The constructed response file for Test B
NI=20
ITEM1=41 ; Items start in column 26 of reformatted record
NAME1=1 ; Start of person label
NAMELEN=11 ; Length of person label
CODES="ABCD# " ; Beware of blanks meaning wrong!
KEY1=CDABCDBDBABCACBCDBCAD ; Key in data record format
&END
BANK 21 TEST B 1
BANK 4 TEST B 2
BANK 22 TEST B 3
.
BANK 35 TEST B 20
END NAMES
Person 01 B                             BDABDDCDBBBBBCCDAACBC
.
Person 12 B                             BADBAADCBADDBBB

The RFILE= file, EXAM10BR.TXT, is:

01110101011001000100
..........
00000001010000101000

III. Analyze Test A's and Test B's RFILE='s together:

; This file is EXAM10C.TXT
TITLE="Analysis of Tests A & B (already scored)"
NI=35 ; 35 items in total
After running EXAM10C, I want to see two ICCs: One for test A and another Test B. How do I do this?

This graph is not produced directly by Winsteps, but can be produced in Excel.
use the “Specification” pull-down menu to delete items not in Test A: Use IDELETE=21-35
Display the Test Characteristic Curve
Select "Copy data to clipboard".
Paste into an Excel worksheet

Use the “Specification” pull-down menu to reinstate all items: IDELETE=+1-35

Use the “Specification” pull-down menu to delete items not in Test B: Use IDELETE=6-20
Display the Test Characteristic Curve
Select "Copy data to clipboard".
Paste into an Excel worksheet

In Excel, scatterplot the pasted columns.

55. Example 11: Item responses two characters wide

The "Liking for Science" data (see RSA) is in file EXAM11.TXT. Each observation is on a rating scale where 0 means "dislike", 1 means "don't care/don't know" and 2 means "like". The data has been recorded in two columns as 00, 01 and 02. XWIDE= is used.

; This file is EXAM11.TXT
TITLE='LIKING FOR SCIENCE (Wright & Masters p.18)'
XWIDE=2 ; Responses are 2 columns wide
CODES=00102 ; Codes are 00 01 and 02
CLFILE=* ; Category label filelist
00 Dislike ; Category 00 in data file means "Dislike"
01 Neutral
02 Like
* ; End of category list
ITEM1=1 ; Items start in column 1
NI=25 ; 25 Items
NAME1=51 ; Person id starts in column 51
NAMLMP=20 ; Show 20 characters of id on maps
TABLES=11111111111111111111111111111111 ; All Tables
CURVES=111 ; Print all curves in Tables 2 and 21
IFILE = EXAM11IF.TXT ; Output item measure file
PFILE = EXAM11PF.TXT ; Output person measure file
SPIFILE = EXAM11SPIF.TXT ; Output structure calibration file
RFFILE = EXAM11RF.TXT ; Output reformatted response file
XFILE = EXAM11XF.TXT ; Output observation and residual file
UIMEAN = 455 ; User scaling: mean 455
USCALE = 94 ; 94 user units per logit
LINLEN = 0 ; Print with minimum of split lines
MAXPAG = 0 ; Print with no page breaks in long tables
&END

WATCH BIRDS
READ BOOKS ON ANIMALS
TALK W/FRIENDS ABOUT PLANTS

56. Example 12: Comparing high and low samples with rating scales

Rasch estimates are constructed to be as sample independent as is statistically possible, but you must still take care to maintain comparability of measures across analyses. For instance, if a rating scale or partial credit structure is used, and a high-low measure split is made, then the low rating scale (or partial credit) categories may
not appear in the data for the high measure sample and vice versa. To compare item calibrations for the two samples requires the response structure to be calibrated on both samples together, and then the response structure calibrations to be anchored for each sample separately. Comparison of patient measures from separate analyses requires both the response structure calibrations and the item calibrations to share anchor calibrations.

35 arthritis patients have been through rehabilitation therapy. Their admission to therapy and discharge from therapy measures are to be compared. They have been rated on the 13 mobility items of the Functional Independence Measure (FIM™). Each item has seven levels. At admission, the patients could not perform at the higher levels. At discharge, all patients had surpassed the lower levels (Data courtesy of C.V. Granger & B. Hamilton, ADS). A generic control file is in EXAM12.TXT. The admission ratings are in EXAM12LO.TXT and the discharge ratings in EXAM12HI.TXT. Three analyses are performed: 1) joint analysis of the admission (low) and discharge (high) data to obtain response structure calibrations, 2 & 3) separate runs for the admission (low) and discharge (high) data to obtain item calibrations. For a more complex situation, see Example 17.
The admission data is in file EXAM12LO.TXT:

```
21101 5523133322121 Patient number in cols 1-5, ratings in 7-19
21170 4433443454545
| 22618 4433255542141
| 22693 352432421111
```

The discharge data is in file EXAM12HI.TXT:

```
21101 5734366655453 Ratings generally higher than at admission
21170 6466677777676
| 22618 7667666666565
| 22693 7776677676677
```

The batch file to run this is (see BATCH=), under Windows XP, 2000, EXAM12.CMD:

```
REM  COMPARISON OF ITEM CALIBRATIONS FOR HIGH AND LOW SAMPLES
START /WAIT ..\WINSTEPS EXAM12.TXT EXAM12OU.TXT DATA=EXAM12LO.TXT+EXAM12HI.TXT
   TITLE=ADMIT+DISCHARGE SFILE=EXAM12SF.TXT BATCH=Y
START /WAIT ..\WINSTEPS EXAM12.TXT EXAM12LU.TXT DATA=EXAM12LO.TXT TITLE=ADMIT
   SFILE=EXAM12SF.TXT IFILE=EXAM12LIF.TXT BATCH=Y
START /WAIT ..\WINSTEPS EXAM12.TXT EXAM12HU.TXT DATA=EXAM12HI.TXT TITLE=DISCHARGE
   SFILE=EXAM12SF.TXT IFILE=EXAM12HIF.TXT BATCH=Y
```

under WINDOWS-95 or -98, EXAM12BAT.BAT:

```
REM  COMPARISON OF ITEM CALIBRATIONS FOR HIGH AND LOW SAMPLES
START /w ..\WINSTEPS EXAM12.TXT EXAM12OU.TXT DATA=EXAM12LO.TXT+EXAM12HI.TXT
   TITLE=ADMIT+DISCHARGE SFILE=EXAM12SF.TXT BATCH=Y
START /w ..\WINSTEPS EXAM12.TXT EXAM12LU.TXT DATA=EXAM12LO.TXT TITLE=ADMIT
   SFILE=EXAM12SF.TXT IFILE=EXAM12LIF.TXT BATCH=Y
START /w ..\WINSTEPS EXAM12.TXT EXAM12HU.TXT DATA=EXAM12HI.TXT TITLE=DISCHARGE
   SFILE=EXAM12SF.TXT IFILE=EXAM12HIF.TXT BATCH=Y
```

Under WINDOWS-NT (early versions), EXAM12NT.BAT:

```
REM  COMPARISON OF ITEM CALIBRATIONS FOR HIGH AND LOW SAMPLES
..\WINSTEPS EXAM12.TXT EXAM12OU.TXT DATA=EXAM12LO.TXT+EXAM12HI.TXT
   TITLE=ADMIT&DISCHARGE SFILE=EXAM12SF.TXT BATCH=Y
..\WINSTEPS EXAM12.TXT EXAM12LU.TXT DATA=EXAM12LO.TXT TITLE=ADMIT SAFILE=EXAM12SF.TXT
   IFILE=EXAM12LIF.TXT BATCH=Y
..\WINSTEPS EXAM12.TXT EXAM12HU.TXT DATA=EXAM12HI.TXT TITLE=DISCHARGE SAFILE=EXAM12SF.TXT
   IFILE=EXAM12HIF.TXT BATCH=Y
```

To run this, select "Run batch file" from "Batch" pull-down menu, and right-click on "Exam12bat.bat" or "Exam12cmd.cmd" in the dialog box, then left-click on "open".

The shared structure calibration anchor file is EXAM12SF.TXT:

```
; structure measure FILE FOR
; ADMIT&DISCHARGE
; May 23 13:56 1993
; CATEGORY  structure measure
1    .00
2  -2.11
3  -1.61
4  -1.25
5    .06
6   1.92
7   2.99
```
The item calibrations measures for admission and discharge are written into IFILE= files, with comma-separated values (CSV=Y), so that they can easily be imported into a spreadsheet.

57. Example 13: Paired comparisons as the basis for measurement

Paired comparisons can be modeled directly with the Facets computer program. For WINSTEPS a dummy facet of "occasion" must be introduced. On each occasion (in this example, each column), there is a winner '1', a loser '0', or a draw 'D' recorded for the two players. In column 1 of the response data in this example, Browne (1) defeated Mariotti (0). In column 2, Browne (D) drew with Tatai (D). Specifying PAIRED=YES adjusts the measures for the statistical bias introduced by this stratagem. Each player receives a measure and fit statistics. Occasion measures are the average of the two players participating. Misfitting occasions are unexpected outcomes. Point-biserial correlations have little meaning. Check the occasion summary statistics in Table 3 to verify that all occasions have the same raw score.

; This common control file is EXAM13.TXT
TITLE = 'Chess Matches at the Venice Tournament, 1971'
Name1 = 1 ; Player's name
Item1 = 11 ; First match results
PERSON = PLAYER
ITEM = MATCH ; Example of paired comparison
CODES = 0D1 ; 0 = loss, D = draw (non-numeric), 1 = win

; if you wish to just consider won-loss, and ignore the draws, omit the following line:
NEWSCORE = 012 ; 0 = loss, 1 = draw, 2 = win

CLFILE=*  
D Loss  
D Draw  
1 Win  
*

NI = 66  ; 66 matches (columns) in total
PAIRED = YES  ; specify the paired comparison adjustment
INUMBER = YES  ; number the matches in the Output
&END

Browne 1D011111
Mariotti 01D011111
Tatai D01111111
Hort 1D1D1111D
Kavalek 01DD1111D
Damjanovic 00DDDD11D1
Gligoric 00DDDDD1110
Radulov 00DDDDDD1D1
Bobotsov 00DDDDDDDD01
Cosulich DDDDDDDDD1
Westerinen 0D0DDDD101
Zichichi 00D1D01000

Part of the output is:

<table>
<thead>
<tr>
<th>ENTRY</th>
<th>RAW</th>
<th>INFIT</th>
<th>OUTFIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17</td>
<td>1.09</td>
<td>.35</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>.68</td>
<td>.32</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>.50</td>
<td>.31</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>.50</td>
<td>.31</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>.33</td>
<td>.31</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>.00</td>
<td>.30</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>-.17</td>
<td>.30</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>-.34</td>
<td>.31</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>-.51</td>
<td>.31</td>
</tr>
</tbody>
</table>

PLAYER STATISTICS: MEASURE ORDER

+---------------------------------------------------------------------+
<table>
<thead>
<tr>
<th>ENTRY</th>
<th>RAW</th>
<th>COUNT</th>
<th>MEASURE</th>
<th>ERROR</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>PLAYER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17</td>
<td>11</td>
<td>1.09</td>
<td>.35</td>
<td>.25</td>
<td>.10</td>
<td>Browne</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>11</td>
<td>.68</td>
<td>.32</td>
<td>.26</td>
<td>.12</td>
<td>Mariotti</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>11</td>
<td>.50</td>
<td>.31</td>
<td>.31</td>
<td>.15</td>
<td>Tatai</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>11</td>
<td>.50</td>
<td>.31</td>
<td>.31</td>
<td>.15</td>
<td>Hort</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>11</td>
<td>.33</td>
<td>.31</td>
<td>.31</td>
<td>.15</td>
<td>Kavalek</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>11</td>
<td>.00</td>
<td>.30</td>
<td>.35</td>
<td>.28</td>
<td>Damjanovic</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>11</td>
<td>-.17</td>
<td>.30</td>
<td>.35</td>
<td>.37</td>
<td>Gligoric</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>11</td>
<td>-.34</td>
<td>.31</td>
<td>.52</td>
<td>.52</td>
<td>Radulov</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>11</td>
<td>-.51</td>
<td>.31</td>
<td>1.00</td>
<td>.00</td>
<td>Bobotsov</td>
</tr>
</tbody>
</table>
58. 

Example 14: Multiple rescorings, response structures and widths

Introductory example: A test has 30 multiple choice questions (keyed A, B, C, D) and 5 essay questions, each has its own rating scale definition (scored 0, 1, or 2), i.e., they accord with the partial credit model. ISGROUPS= is used to identify the partial credit items. IREFER= and IVALUE= are used for scoring the questions

The control file is EXAM14.TXT

TITLE = "AN EXAMPLE OF MULTIPLE ITEM TYPES"

IREFER = ABCAAACBAAABABAAAAABABBABABABBABABABBABAB00000

IVALUEA= 0100000 ; MCQ ITEMS WITH A AS THE CORRECT ANSWER
IVALUEN= 0010000 ; B IS CORRECT
IVALUENC= 0001000 ; D IS CORRECT

MISSING-VALUES-SCORED = 0 ; SO ALL 9'S (OR ANYTHING ELSE NOT IN CODES=) ARE SCORED 0

EXAMPLE 14: A test comprises multiple-choice questions (MCQ), true-false, and two different response structure formats. These are combined in one data file. The MCQ scoring key, and also the True-False scoring key are used as item cluster references in the IREFER= and IVALUE= specifications to simplify key checking.

EXAM14DT.TXT data format is:
Cols 1-5  Person id.
Cols 6-10 5 MCQ items (A,B,C,D, # = missing, wrong. Some items may be miskeyed
Cols 11-15 5 True-False items, responses (S, N). For some of these S="True" is correct, for others "N"=False is correct
Cols 16-25 10 Rating scale items (N=0,P=1,M=2,S=3). Some items may be reversed.
Cols 26-29 2 Evaluation items (0-12). - (See exam14b.txt) - second analysis only.

First analysis: all responses one column wide

The control file, EXAM14.TXT, is:

; This file is EXAM14.TXT
TITLE="Multiple Response Formats, Same Response Width"
DATA=exam14dt.txt
;EXAM14DT.TXT data format is
;Cols 1-5  Person id.
;Cols 6-10 5 MCQ items (A,B,C,D, # = missing, wrong)
; Some items may be miskeyed
;Cols 11-15 5 True-False items, responses (S, N)
; For some of these S="True" is correct, for others "N"=False is correct
;Cols 16-25 10 Rating scale items (N=0,P=1,M=2,S=3)
; Some items may be reversed.
;Cols 26-29 2 Evaluation items (0-12). - (See exam14b.con)
NAME1=1
ITEM1=6
NI=20

; THESE CODES LINE UP WITH THE ITEM COLUMNS
; 0   1   2
; 12345678901234567890

; TO DEFINE RESPONSE STRUCTURE CLUSTERS
ISGROUPS = 11111222223333333333

; TO DEFINE RESCORING CLUSTERS
IREFER = BACDCSNSNSRSSSSSSSSSSS
; IREFER = X MATCHES IVALUEX=

; IVALUE?= MATCHES WITH CODES=
CODES = ABCD#SNPM ; CODES IN ORIGINAL DATA FILE
IVALUEA = 10000**** ; MCQ RESPONSE A IS CORRECT
IVALUEB = 01000**** ; MCQ RESPONSE B IS CORRECT
IVALUERC = 00100**** ; MCQ RESPONSE C IS CORRECT
IVALUED = 00010**** ; MCQ RESPONSE D IS CORRECT
IVALUES = *****10** ; "S" IS THE CORRECT ANSWER
IVALUEN = *****01** ; "N" IS THE CORRECT ANSWER
IVALUER = *****3012 ; "NPMS" RATING SCALE
STKEEP=YES  ; KEEP UNUSED INTERMEDIATE CATEGORIES IN RATING SCALES
INUMBER=YES ; NO ITEM INFORMATION AVAILABLE
&END

Second analysis: responses one and two columns wide

Including the last two items with long numeric response structures and 2-column format, the control file becomes EXAM14B.TXT. Since some responses are in 2 column format, XWIDE=2. FORMAT= is used to transform all responses into XWIDE=2 format.

EXAM14B.TXT is:

&INST
TITLE="Multiple Response Formats, Same Response Width"
DATA=EXAM14DT.TXT
; EXAM14DT.TXT data format is
; Cols 1-5 Person id.
; Cols 6-10 5 MCQ items (A,B,C,D, # = missing, wrong)
; Some items may be miskeyed
; Cols 11-15 5 True-False items, responses (S, N)
; For some of these S="True" is correct, for others "N"=False is correct
; Cols 16-25 10 Rating scale items (N=0,P=1,M=2,S=3)
; Some items may be reversed.
; Cols 26-29 2 Evaluation items (0-12)
NAME1=1
ITEM1=6
NI=22 ; 20-1 COLUMN + 2 2-COLUMN ITEMS
XWIDE=2 ; XWIDE FOR WIDEST FIELD
FORMAT=(5A1, 20A1,2A2) ; PERSON LABEL & FIRST 20 ITEMS 1 CHARACTER COLUMNS
; LAST 2 ITEMS ARE 2 CHARACTER COLUMNS
; THESE ARE SET UP FOR XWIDE=2

; FOR RESPONSE STRUCTURE DEFINITIONS
; TO DEFINE RATING SCALE CLUSTERS
ISGROUPS  = 1111122223333333344
IREFER  = BACDCSNSNSNSNSNSNSNSNSNS
; IREFER = X MATCHES IVALUEX=
; IVALUE?= MATCHES WITH CODES=
; CODES ARE 2 CHARACTERS WIDES
CODES   = "A B C D # S N P M 121110 9 8 7 6 5 4 3 2 1 0"; CODES IN DATA
IVALUEA = "1 0 0 0 0 * * * * * * * * * * * * * * * * * * * * * * * * *"; MCQ A IS CORRECT
IVALUEB = "0 1 0 0 0 * * * * * * * * * * * * * * * * * * * * * * * * *"; MCQ B IS CORRECT
IVALUEC = "0 0 1 0 0 * * * * * * * * * * * * * * * * * * * * * * * *"; MCQ C IS CORRECT
IVALUED = "0 0 0 1 0 * * * * * * * * * * * * * * * * * * * * * * * *"; MCQ D IS CORRECT
IVALUES = "; * * * * * 1 0 * * * * * * * * * * * * * * * * * * * * * * *"; "S" IS CORRECT
IVALUEN = "* * * * * 0 1 * * * * * * * * * * * * * * * * * * * * * * *"; "N" IS CORRECT
IVALUER = "* * * * * 3 0 1 2 * * * * * * * * * * * * * * * * * * *"; "NPMS" RATING SCALE
IVALUEE = "* * * * * * * * * 121110 9 8 7 6 5 4 3 2 1 0"; 0-12 RATING SCALE
STKEEP=YES ; KEEP UNUSED INTERMEDIATE CATEGORIES IN RATING SCALES
INUMBER=YES ; NO ITEM INFORMATION AVAILABLE

This can also be done with MFORMS=

EXAM14C.TXT is:

; This file is EXAM14C.TXT
&INST
TITLE="Multiple Response Formats, Same Response Width"
; EXAM14DT.TXT data format is
; Cols 1-5 Person id.
; Cols 6-10 5 MCQ items (A,B,C,D, # = missing, wrong)
; Some items may be miskeyed
; Cols 11-15 5 True-False items, responses (S, N)
; For some of these S="True" is correct, for others "N"=False is correct
; Cols 16-25 10 Rating scale items (N=0,P=1,M=2,S=3)
; Some items may be reversed.
; Cols 26-29 2 Evaluation items (0-12)

; Reformatted data record is:
; Cols 1-5 Person id
; Cols 6-7 Item 1 = original Col. 6
NAME1=1
ITEM1=6
NI=22 ; 20-1 COLUMN + 2 2-COLUMN ITEMS
XWIDE=2 ; XWIDE FOR WIDEST FIELD

mforms=
data=exam14dt.txt ; the name of an input data file
L = 1 ; the are 2 lines in input data file for each data record
P1-5 = 1 ; person label characters 1 through 5 start in column 1
; in the following "C" is used because "I" uses XWIDE=2
C6 = 6 ; original item 1 in column 6 goes in column 6
C8 = 7 ; original item 2 in column 7 goes in column 8
C10 = 8
C12 = 9
C14 = 10
C16 = 11
C18 = 12
C20 = 13
C22 = 14
C24 = 15
C26 = 16
C28 = 17
C30 = 18
C32 = 19
C34 = 20
C36 = 21
C38 = 22
C40 = 23
C42 = 24
C44 = 25 ; original item 20 in column 25 goes in column 44-45
I21-22 = 26 ; two-character items 21 and 22 start in column 26
*

; THESE ARE SET UP FOR XWIDE=2 FOR RESPONSE STRUCTURE DEFINITIONS
; TO DEFINE RATING SCALE CLUSTERS
ISGROUPS  = 1111222222333333333344
IREFER  = BACDCSNSNSRRRRRRRRREE
; IREFER = X MATCHES IVALUEX=
; IVALUE?= MATCHES WITH CODES=
; CODES ARE 2 CHARACTERS WIDES
CODES   = "A B C D # S N P M 121110 9 8 7 6 5 4 3 2 1 0"; CODES IN DATA
IVALUEA = "1 0 0 0 0 * * * * * * * * * * * * * * * * " ; MCQ A IS CORRECT
IVALUEB = "0 1 0 0 0 * * * * * * * * * * * * * * * * " ; MCQ B IS CORRECT
IVALUEC = "0 0 1 0 0 * * * * * * * * * * * * * * * * " ; MCQ C IS CORRECT
IVALUED = "0 0 0 1 0 * * * * * * * * * * * * * * * * " ; MCQ D IS CORRECT
IVALUEN = "0 0 0 0 1 * * * * * * * * * * * * * * * * " ; "N" IS CORRECT
IVALUER = "* * * * * 0 1 * * * * * * * * * * * * * * " ; "NPMS" RATING SCALE
IVALUEE = "% * * * * * * * * 121110 9 8 7 6 5 4 3 2 1 0" ; 0-12 RATING SCALE
STKEEP=YES ; KEEP UNUSED INTERMEDIATE CATEGORIES IN RATING SCALES
INumber=YES ; NO ITEM INFORMATION AVAILABLE
&END
59. Example 15: Figure skating: Multidimensionality, DIF or Bias

The Pairs Skating competition at the 2002 Winter Olympics in Salt Lake City was contentious. It resulted in the awarding of Gold Medals to both a Russian and a Canadian pair, after the French judge admitted to awarding biased scores. Multidimensionality, differential item functioning, and item bias are all manifestations of disparate subdimensions within the data. In judged competitions, judge behavior can introduce unwanted subdimensions.

For this analysis, each pair is allowed to have a different skill level, i.e., different measure, on each skill of each performance. The judges are modeled to maintain their leniencies across all performances. The control file and data are in exam15.txt.

```
; This control file is EXAM15.TXT
Title = "Pairs Skating: Winter Olympics, SLC 2002"
Item = Judge
Person = Pair
NI = 9 ; the judges
Item1 = 14 ; the leading blank of the first rating
Xwide = 3 ; Observations are 3 CHARACTERS WIDE for convenience
NAME1 = 1 ; start of person identification
NAMELENGTH = 13 ; 13 characters identifiers

; CODES NEXT LINE HAS ALL OBSERVED RATING SCORES
CODES=" 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44+
      + 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60"
STEPKEEP=YES ; maintain missing intermediate rating scores in the scoring structure
@order = 1-2 ; order number at finish of competition in person label columns 1-2
DIF = @order ; judge "DIF" across skating pairs
tfile=* ; produce Table 30 for judge "DIF"

&END
1 Rus ;Mrs. Marina SANAIA : RUSSIA
2 Chn ;Mr. Jiasheng YANG : CHINA
3 USA ;Mrs. Lucy BRENNAN : USA
4 Fra ;Miss Marie Reine LE GOUGNE : FRANCE
5 Pol ;Mrs. Anna SIEROCKA : POLAND
6 Can ;Mr. Benoit LAVOIE : CANADA
7 Ukr ;Mr. Vladislav PETUKHOV : UKRAINE
8 Ger ;Mrs. Sissy KRICK : GERMANY
9 Jap ;Mr. Hideo SUGITA : JAPAN

; Description of Person Identifiers
; Cols. Description
; 1-2 Order immediately after competition (@order)
; 4-5 Skaters' initials
; 7-9 Nationality
; 11 Program: S=Short  F=Free
; 13 Skill: T=Technical Merit, A=Artistic Impression
END LABELS
1 BS-Rus S T 58 57 58 58 58 58 58 57 ; 1 BEREZHNAIA Elena / SIKHARULIDZE Anton : RUS
2 BS-Rus A T 58 58 58 58 58 58 58 58 ; 2 BEREZHNAIA Elena / SIKHARULIDZE Anton : RUS
3 BS-Rus F T 58 58 58 58 58 58 58 57 ; 3 BEREZHNAIA Elena / SIKHARULIDZE Anton : RUS
4 BS-Rus A 59 59 59 59 59 59 59 59 ; 4 BEREZHNAIA Elena / SIKHARULIDZE Anton : RUS
2 SP-Can S T 57 57 57 57 58 58 58 57 ; 5 SALE Jamie / PELLETIER David : CAN
3 SP-Can A 58 58 58 58 58 58 58 58 ; 6 SALE Jamie / PELLETIER David : CAN
3 SP-Can F T 58 58 58 58 58 58 58 58 ; 7 SALE Jamie / PELLETIER David : CAN
4 SP-Can F A 58 58 58 58 58 58 58 58 ; 8 SALE Jamie / PELLETIER David : CAN
3 SZ-Chn S T 57 57 57 57 57 57 57 57 ; 9 SHEN Xue / ZHAO Hongbo : CHN

From this data file, estimate judge severity. In my run this took 738 iterations, because the data are so thin, and the rating scale is so long.

Here is some of the output of Table 30, for Judge DIF, i.e., Judge Bias by skater pair order number, @order = SS1W2.

| Pair DIF DIF DIF Pair DIF DIF DIF JOINT Judge |
| CLASS ADDED S.E. CLASS ADDED S.E. CONTRAST S.E. t Number Name |
The most significant statistical bias is by the Japanese judge on skater pairs 13 and 14 vs. 18. These pairs are low in the final order, and so of little interest.

Table 23, the principal components/contrast analysis of Judge residuals is more interesting. Note that Judge 4, the French judge, is at the top with the largest contrast loading. The actual distortion in the measurement framework is small, but crucial to the awarding of the Gold Medal!

60. Example 16: Disordered categories - thresholds and person anchoring

This example illustrates two types of disordering in rating scale structures: category disordering and Rasch-Andrich threshold disordering. For further discussion, see disordering. It also illustrates anchor values in the person labels and category labeling.

The control file and data are in exam16.txt.

; This control file is EXAM16.TXT
; title = "Attitude Survey illustrating two types of disordering"
; ni    = 3 ; three items
; item1 = 1 ; item responses start in column 1
; xwide = 2 ; data codes are two columns wide
; codes = "1 2 3 " ; valid two character codes
; namel = 9 ; person label starts in column 9
; namelength = 3 ; length of person labels

; pafie = $s1w3 ; person anchor values in columns 1-3 of person labels
; panchors = 1-3 ; person anchor field is in columns 1 to 3 of person labels
; pafie = @panchors ; person anchors in field @panchors

ISGROUPS = 0 ; allow each item to have its own rating scale (the partial credit model)
c1file = * ; category labels: item+category
1+1 Never ; item 1 category 1 is "Never"
1+2 Sometimes ; well-behaved rating scale
1+3 Often
2+1 Car ; categories as ordered in frequency in 1930
2+2 Ship ; now these categories are disordered
2+3 Plane ; ship travel now rarer than planes
3+1 No
3+2 Neutral ; very few in this narrow intermediate category
3+3 Yes
Smooth advance everywhere - probability curves a "range of hills"
Disordered categories - disordered empirical average measures for categories
Low middle frequency - high discrimination - disordered Rasch-Andrich thresholds

On the Diagnosis Menu: Empirical Item-Category Measures:

TABLE 2.5 Attitude Survey
OBSERVED AVERAGE MEASURES FOR PERSONS (BY OBSERVED CATEGORY)

<table>
<thead>
<tr>
<th>NUM</th>
<th>ITEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

On the Diagnosis Menu: Category Function:

TABLE 3.2 Attitude Survey
FOR GROUPING "0" ITEM NUMBER: 1 Smooth advance everywhere - probability curves a "range of hills"
ITEM ITEM DIFFICULTY MEASURE OF 2.00 ADDED TO MEASURES

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>OBSERVED</th>
<th>INFIT OUTFIT</th>
<th>STRUCTURE</th>
<th>CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>22</td>
<td>(  -.29)</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>5</td>
<td>56</td>
<td>2.00</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>22</td>
<td>3.00</td>
</tr>
</tbody>
</table>

TABLE 3.3 Attitude Survey
FOR GROUPING "0" ITEM NUMBER: 2 Disordered categories - disordered empirical average measures for categories
### TABLE 3.4 Attitude Survey

FOR GROUPING "0" ITEM NUMBER: 3  Low middle frequency - high discrimination - disordered Rasch-Andrich thresholds

<table>
<thead>
<tr>
<th>ITEM</th>
<th>ITEM DIFFICULTY MEASURE OF 2.00 ADDED TO MEASURES</th>
</tr>
</thead>
<tbody>
<tr>
<td>CATEGORY</td>
<td>OBSERVED</td>
</tr>
<tr>
<td>LABEL</td>
<td>SCORE COUNT %</td>
</tr>
<tr>
<td>-------</td>
<td>------------</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

---

![Diagram of Ordered Rasch-Andrich thresholds and disordered categories](image1)

![Diagram of Disordered Rasch-Andrich thresholds and ordered categories](image2)
Example 17: Rack, stack, principal components

Comparisons of measures at two time-points or in two situations can become complicated. Example 12 illustrates a straightforward situation. It uses the 13 motor items of the FIM®. This example uses the 18 items of the FIM, 13 motor and 5 cognitive, at two time points. There is a 7-level rating scale that is intended to operate in the same way across all items.

In exam17s.txt, the data have been stacked. The items are modeled to maintain their difficulties across the two time points, and the 52 patients are entered twice, once at admission to rehabilitation and once at discharge from it, so there 104 data records. Changes in patient independence can be identified by cross-plotting the admission and discharge measures for each patient, as in Example 12. This can be done by using the Plots menu to plot the measures against themselves, and then, in EXCEL, pasting the discharge measures over the top of the admission measures for the y-axis.

In exam17r.txt, the data have been racked. The persons are modeled to maintain their abilities across the two time points, and the 18 FIM items are entered twice, once at admission to rehabilitation and once at discharge from it, so there 36 items. Average changes in patient performance on individual items can be identified by cross-plotting the admission and discharge measures for each item. This can be done by using the Plots menu to plot the measures against themselves, and then, in EXCEL, pasting the discharge measures over the top of the admission measures for the y-axis.

A further feature is the contrasts in the Principal Components Analysis of Residuals, Tables 23 and 24. The 13 motor items and 5 cognitive items are probing different aspects of patient independence. Do they function as one variable in this sample? See Principal components/contrast. Patients also have different impairments. Are their measures comparable? See Differential Item Function.
variables in person label
@SEX=$S1W1 ; Gender: 1=Male, 2=Female
@IGC=$S3W2 ; Impairment Group Code: 8=Orthopedic, 9=Cardiac, 10=Pulmonary, 11=Burns, 12=Congenital,
13=Other
@ID=$S5W2 ; Patient sequ. number

variables in item label
@SETTING=$S1W1 ; setting: A=Admission, D=Discharge
@SUBTEST=$S3W1 ; subtest: M=Motor, C=Cognitive
@ILETTER=$S5W1 ; code letter of item

CODES=1234567 ; 7 level rating scale
CLFILE=* ; Defines the rating scale
  1  0% Independent
  2  25% Independent
  3  50% Independent
  4  75% Independent
  5 Supervision
  6 Device
  7 Independent
*
&END
A M A. EATING
....
A C R. MEMORY
 - Blank for ease of seeing
D M A. EATING
....
D C R. MEMORY
END NAMES
334412312331123112 554535546665345545 1  8  1
4323224223334444 777677777666567777 2  8  2
....

62. ALPHANUM alphabetic numbering

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

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

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

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

63. ASCII output only ASCII characters

MS-DOS Tables include graphic characters which some printers can't print. These graphics characters can be replaced by the ASCII characters \ and -.

ASCII=N use graphics characters
ASCII=Y replace graphics characters with ASCII characters (the standard).

Example:
ASCII=N produces what follows, or else accented letters, e.g., ààà:

OVERVIEW TABLES          ITEM CALIBRATIONS
1* PERSON AND ITEM DISTRIBUTION MAP  12* ITEM MAP BY NAME

69
64. **ASYMPTOTE** item upper and lower asymptotes

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

**ASYMPTOTE=Y** report the values of the Upper and Lower asymptotes in the Item Tables and **IFILE=**

**ASYMPTOTE=N** do not report values for the Upper and Lower asymptotes.

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

Run Exam1.txt

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

Enter: **DISCRIM = Yes** to report the Item Discrimination

Enter: **ASYMP = Yes** to report the asymptotes

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

### Estimation

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

A lower-asymptote model for dichotomies or polytomies is:

\[ T_{ni} = ci + \left( mi - ci \right) \left( En_i / mi \right) \]

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

\[ ci = \frac{mi \left( T_{ni} - En_i \right)}{mi - En_i} \]

This provides the basis for a model for estimating \( ci \). Since we are concerned about the lower asymptote, let us only consider \( B_{ni} = B_{ni-Di < B(En_i=0.5)} \) and weight the observations, \( X_{ni} \), with \( W_{ni} = B_{ni} - B(En_i=0.5) \).

\[ ci \approx \frac{\sum \left( W_{ni} \times mi \times (X_{ni} - En_i) \right)}{\sum \left( W_{ni} \times (mi - En_i) \right)} \]

Similarly, for \( di \), the upper asymptote,

\[ di \approx \frac{\sum \left( W_{ni} \times mi \times X_{ni} \right)}{\sum \left( W_{ni} \times En_i \right)} \]

for \( B_{ni} > B(En_i=mi=0.5) \)

But if the data are sparse in the asymptotic region, the estimates may not be good. This is a known problem in 3-PL estimation, leading many analysts to impute, rather than estimate, asymptotic values.


65. **BATCH Batch mode analysis**

If you want Winsteps to close itself as performing an analysis and writing out any output specified in the control file, e.g., by `TABLES=`, then specify `BATCH=YES` in the control file. You can launch batch files from the **Batch menu**.

If you want Winsteps to run in "background" with the minimum user interaction, then specify `BATCH=YES` in the Shortcut, DOS or Shell command which invokes Winsteps.

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

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

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

i) On the main WINSTEPS screen, click on the "Batch" menu item.

ii) On the pull-down menu, select "Edit batch file".

iii) In the dialog box, select **Winbatchcmd.cmd** and click on "Open"

iv) The following batch file is available to edit:

```
echo  This is the version for Windows-NT, 2000
echo  This is a batch file to run WINSTEPS in batch mode
echo  Edit the next lines and add more.
echo  Format of lines is:
echo  START /WAIT ..\WINSTEPS BATCH=YES Control-file Output-file
Extra=specifications
START /WAIT ..\WINSTEPS BATCH=YES EXAMPLE0.txt EXAMPLE0.OUT TABLES=111
START /WAIT ..\WINSTEPS BATCH=YES SF.txt SF.OUT TFILE=* 1 * PERSON=CASE
START /WAIT ..\WINSTEPS BATCH=YES KCT.txt KCT.OUT TFILE=* 3 20 * MRANGE=4
```

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

v) The lines starting with "echo" are comments.

vi) Lines starting "START /WAIT WINSTEPS BATCH=YES" execute WINSTEPS

vii) The format is `START /WAIT WINSTEPS BATCH=YES control-file output-file extra-specifications`

viii) Each new WINSTEPS line is an additional run of the WINSTEPS program

ix) Edit and save this file. You can save it with any name ending ".cmd"

x) From the "Batch" pull-down menu, select "Run batch file".

xi) **In the right-click menu, left-click on "open"**

xii) The batch file will run - if nothing happens, the batch file is incorrect.

xiii) Exit from the Winsteps dialog by clicking on "Cancel".

xiv) You can minimize the batch screen by clicking on the underline in the top right corner.

xv) You can cancel the batch run by right clicking on the Batch icon in the Task bar, usually at the bottom of the screen.

**Under early versions of Windows (e.g. -95, -98) except early Windows-NT**

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

i) On the main WINSTEPS screen, click on the "Batch" menu item.

ii) On the pull-down menu, select "Edit batch file".

iii) In the dialog box, select **Winbatchbat.bat** and click on "Open"

iv) The following batch file is available to edit:

```
echo  This is the version for WINDOWS-95, WINDOWS-98 and ME
```

71
echo This is a batch file to run WINSTEPS in batch mode
echo Edit the next lines and add more.

Format of lines is:

```
START /w ..\WINSTEPS BATCH=YES Control-file Output-file Extra=specifications
```

```
START /w ..\WINSTEPS BATCH=YES EXAMPLE0.txt EXAMPLE0.OUT.txt TABLES=111
```

```
START /w ..\WINSTEPS BATCH=YES SF.txt SF.OUT.txt TFILE=* 1 * PERSON=CASE
```

```
START /w ..\WINSTEPS BATCH=YES KCT.txt KCT.OUT.txt TFILE=* 3 20 * MRANGE=4
```

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

v) The lines starting with "echo" are comments.
v) Lines starting "START /w WINSTEPS BATCH=YES" execute WINSTEPS
vi) The format is

```
WINSTEPS BATCH=YES control-file output-file extra-specifications
```

vii) Each new WINSTEPS line is an additional run of the WINSTEPS program
viii) Edit and save this file. You can save it with any name ending ".bat"

ix) From the "Batch" pull-down menu, select "Run batch file".
x) Right-click on the desired batch file

xi) In the right-click menu, left-click on "open"
x) The batch file will run - if nothing happens, the batch file is incorrect.
xii) You can minimize the batch screen by clicking on the underline in the top right corner.

Under early Windows-NT

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

i) On the main WINSTEPS screen, click on the "Batch" menu item.

ii) On the pull-down menu, select "Edit batch file".

iii) In the dialog box, select **NTbatch.bat** and click on "Open"

iv) The following batch file is available to edit:

```
 echo This is for early versions of WINDOWS-NT
 echo For later versions use *.cmd files
 echo This is a batch file to run WINSTEPS in batch mode
 echo Edit the next lines and add more.
 echo Format of lines is:
 echo WINSTEPS BATCH=YES Control-file Output-file Extra=specifications
 ..\WINSTEPS BATCH=YES EXAMPLE0.txt EXAMPLE0.OUT.txt Tables=111
 ..\WINSTEPS BATCH=YES SF.txt SF.OUT.txt TFILE=* 1 * PERSON=CASE
 ..\WINSTEPS BATCH=YES KCT.txt KCT.OUT.txt TFILE=* 3 20 * MRANGE=4
```

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

v) The lines starting with "echo" are comments.
v) Lines starting "WINSTEPS BATCH=YES" execute WINSTEPS
vi) The format is

```
WINSTEPS BATCH=YES control-file output-file extra-specifications
```

vii) Each new WINSTEPS line is an additional run of the WINSTEPS program
viii) Edit and save this file. You can save it with any name ending ".bat"

ix) From the "Batch" pull-down menu, select "Run batch file".
x) Right-click on the desired batch file

xi) In the right-click menu, left-click on "open"
x) The batch file will run - if nothing happens, the batch file is incorrect.
xii) You can minimize the batch screen by clicking on the underline in the top right corner.
xiii) You can cancel the batch run by right clicking on the Batch icon in the Task bar, usually at the bottom of the screen.

Example: I want to automatically run multiple DIF reports for the same set of data. Since Winsteps can only perform one DIF analysis at a time in batch mode, you can use anchor files:

First line in batch file, produce measure files

\texttt{Winsteps \texttt{BATCH=YES} in\texttt{file} out\texttt{file} dif=\texttt{$s1w1$} ifile=ifile.txt pfile=pfile.txt sfile=sfile.txt}

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

\texttt{Winsteps \texttt{BATCH=YES} in\texttt{file} out\texttt{file2} dif=\texttt{$s2w1$} ifile=iafile.txt pfile=pafile.txt safile=sfile.txt tfile=* 30 *}
\texttt{Winsteps \texttt{BATCH=YES} in\texttt{file} out\texttt{file3} dif=\texttt{$s3w1$} ifile=iafile.txt pfile=pafile.txt safile=sfile.txt tfile=* 30 *}

A Windows-XP batch processor

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

a) Create a new subfolder of \texttt{c:\winsteps}, called \texttt{c:\winsteps\test}

b) Copy into folder "test" all the control and data files to be analyzed. For instance all the Winsteps example control and data files, which are found in \texttt{c:\winsteps\examples}

c) Use Notepad to create a file in \texttt{c:\winsteps\test} to do the analysis. This file is "saved as" \texttt{test.bat}

This file contains, for instance:

\texttt{start /w ..\winsteps \texttt{BATCH=YES} \texttt{c:\winsteps\examples\example0.txt} \texttt{example0.out.txt} \texttt{DISC=YES} \texttt{TABLES=111}}
\texttt{start /w ..\winsteps \texttt{BATCH=YES} \texttt{c:\winsteps\examples\example9.txt} \texttt{example9.out.txt} \texttt{DISC=YES} \texttt{TABLES=111}}
\texttt{start /w ..\winsteps \texttt{BATCH=YES} \texttt{c:\winsteps\examples\sf.txt} \texttt{sf.out.txt} \texttt{DISC=YES} \texttt{TABLES=111}}

You can replace \texttt{\winsteps} with the pathname to your copy of \texttt{winsteps.exe}

d) Double-click on \texttt{test.bat} in \texttt{c:\winsteps\test} to run this batch file.

e) Winsteps "flashes" on the task bar several times, and progress through the batch file is shown in a DOS-style window.

e) The .out files are written into \texttt{c:\winsteps\test}

Difficulty running Batch or Command files?

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

\textit{Common problems are solved at: www.winsteps.com/problems.htm}

i) Run Winsteps in standard mode from the DOS prompt.

ii) Have the full paths to everything in your batch or command file, e.g., called \texttt{mybatch.cmd}, START \texttt{/WAIT c:\winsteps\WINSTEPS BATCH=YES c:\winsteps\examples\example0.txt c:\winsteps\examples\example0.out.txt}

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

\texttt{DATA = c:\c:\winsteps\examples\mydata.txt}

Note: In this Batch command:

\texttt{START /WAIT c:\winsteps\WINSTEPS BATCH=YES c:\winsteps\examples\controlfile.txt outputfile.txt}

file "\texttt{outputfile.txt}" will be placed in directory "\texttt{c:\winsteps\examples\}".

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

\texttt{c:\winsteps\WINSTEPS c:\winsteps\examples\example0.txt c:\winsteps\examples\example0.out.txt table=1}
Does Winsteps start in the ordinary way? This tests the Windows command line interface.

iv) Windows "Start" menu. "Run". Copy and paste the following line into the Run box. Click OK:
c:\winsteps\WINSTEPS BATCH=YES  c:\winsteps\examples\exam15.txt c:\winsteps\examples\exam15.out.txt table=1

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

v) On your desktop, right-click, "New", "Text document". Double-click on icon. Paste in:
START /WAIT c:\winsteps\WINSTEPS c:\winsteps\examples\example0.txt c:\winsteps\examples\example0.out.txt table=1

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

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

vi) On your desktop, right-click, "New", "Text document". Double-click on icon. Paste in:
START /WAIT c:\winsteps\WINSTEPS BATCH=YES  c:\winsteps\examples\exam15.txt c:\winsteps\examples\exam15.out.txt table=1

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

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

vii) Now build your own .cmd batch file, using lines like:
START /WAIT c:\winsteps\WINSTEPS BATCH=YES  c:\winsteps\examples\example0.txt c:\winsteps\examples\example0.out.txt

Running Winsteps within other Software
Automating the standard version of Winsteps is straightforward using the control instruction BATCH=YES. Winsteps will run under Windows in background (as much as Windows permits).

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

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

This is being done routinely by users of SAS.

66. **BYITEM** display graphs for items

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

**BYITEM = Yes** show empirical curves at the item level.

**BYITEM = No** show empirical curves at the grouping level.

67. **CATREF** reference category for Table 2

*If a particular category corresponds to a criterion level of performance, choose that category for CATREF=.*
Table 2, "most probable responses/scores", maps the items vertically and the most probable responses, expected scores, and 50% cumulative probabilities horizontally. Generally, the vertical ordering is item difficulty measure. If, instead, a particular category is to be used as the reference for sorting, give its value as scored and recoded.

Special uses of CATREF= are:
- **CATREF=-3** for item entry order
- **CATREF=-2** for item measure order
- **CATREF=-1** for items measure order with ISGROUPS=
- **CATREF=0** for item measure order
- **CATREF=1...254** for item measure order based on this category.

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

```
CODES =ABCD  original responses
NEWSCORE=1234  rescored values
RESCORE=2  rescore all
```

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

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

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

```
CODES=1234
NI= 14
ISGROUPS= 11111122222222
TFILE=* 
 2,2 0 0 0 -1  -1 means CATREF=-1
 2,2 0 0 0 0  last 0 means CATREF=0
  *
```

68. **CFILE** scored category label file

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

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

Example 1: Identify the three LFS categories, 0=Dislike, 1=Don't know, 2=Like.
```
CODES=012
CFILE=* 
0 Dislike
1 Don't know
2 Like
 *
```

The labels are shown in Table 3 as:

```
CATEGORY OBSERVED AVGE  INFIT OUTFIT  STRUCTURE
LABEL    COUNT   MEASURE  MNSQ  MNSQ  MEASURE
```

75
Example 2: Items 1-10 (Grouping 1) are "Strong Disagree, Disagree, Agree, Strongly Agree". Items 11-20 (Grouping 2) are "Never, Sometimes, Often, Always".

NI=20
CODES=1234
ISGROUPS=1111111222222222
CFILE=*  
7+1 Strongly Disagree ; We could use any item number in Grouping 1, i.e., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
7+2 Disagree ; item 7 has been chosen
7+3 Agree
7+4 Strong Agree
13+1 Never ; We could use any item number in Grouping 2, i.e., 11, 12, 13, 14, 15, 16, 17, 18, 19, 20
13+2 Sometimes ; item 13 has been chosen
13+3 Often
13+4 Always
*

Example 3: To enter CFILE= information on the DOS Prompt or Extra Specifications lines, using commas instead of blanks as separators:
C:>WINSTEPS SF.TXT SFO.TXT CFILE=* 1,Dislike 2,Don't-know 3,Like *

Example 4: Some items have one rating scale definition, but most items have another rating scale definition. But each item is calibrated with its own structure: ISGROUPS=0
NI=20
CODES=1234
ISGROUPS=0
CFILE=*  
1 Strongly Disagree This scale is used by most items
2 Disagree
3 Agree
4 Strong Agree
16+1 Never 16 is one item using the other scale
16+2 Sometimes
16+3 Often
16+4 Always
17+1 Never 17 is another item using the other scale
17+2 Sometimes
17+3 Often
17+4 Always
.... for all the other items using the other scale
*

Example 5: Several categories are collapsed into one category. The original codes are A-H. After rescoring there is only a dichotomy: 0, 1.
NI=30
CODES =ABCDEFGH
NEWSCORE=00011110
CFILE=*  
0 Fail Specify the categories as recoded
1 Pass
*

69. CHART graphical plots in Tables 10, 13-15

CHART=N  Omit the graphical plots.

CHART=Y  Include graphical plots

<table>
<thead>
<tr>
<th>ENTRY</th>
<th>MEASURE</th>
<th>INFIT MEAN-SQUARE</th>
<th>OUTFIT MEAN-SQUARE</th>
<th>PUPIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUMBR</td>
<td>- +</td>
<td>0 7 1 1.3 2</td>
<td>0 7 1 1.3 2</td>
<td>PUPIL</td>
</tr>
<tr>
<td>53</td>
<td>*</td>
<td>: : * :</td>
<td>K</td>
<td>: : :</td>
</tr>
<tr>
<td>32</td>
<td>*</td>
<td>: . :</td>
<td>w</td>
<td>* : :</td>
</tr>
<tr>
<td>21</td>
<td>*</td>
<td>: . :</td>
<td>a</td>
<td>* : :</td>
</tr>
</tbody>
</table>

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

70.  **CLFILE**  codes label file

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

- XWIDE=2 ; observations are two columns wide
- CODES = "0 1 299" ; codes are 0, 1, 2, 99
- CLFILE=* 99  Strongly Agree ; original code of 99 has the label "Strongly Agree"
- 2  Agree ; original code of blank+2 (or 2+blank) has the label "Agree"
- 2+99  Heartily Agree ; for item 2, code 99 has the label "Heartily Agree"
- 3+0  Disagree  ; for item 3, code 0 means "Disagree"

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

<p>| CATEGORY OBSERVED AVGE  INFIT OUTFIT  STRUCTURE |</p>
<table>
<thead>
<tr>
<th>LABEL</th>
<th>COUNT</th>
<th>MEASURE</th>
<th>MNSQ</th>
<th>MNSQ</th>
<th>MEASURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>378</td>
<td>-.87</td>
<td>1.08</td>
<td>1.19</td>
<td>NONE</td>
</tr>
<tr>
<td>1</td>
<td>620</td>
<td>.13</td>
<td>.85</td>
<td>.69</td>
<td>-.85</td>
</tr>
<tr>
<td>2</td>
<td>852</td>
<td>2.23</td>
<td>1.00</td>
<td>1.46</td>
<td>.85</td>
</tr>
</tbody>
</table>

The labels are shown in Table 3 as:

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

<p>| CATEGORY OBSERVED AVGE  INFIT OUTFIT  STRUCTURE |</p>
<table>
<thead>
<tr>
<th>LABEL</th>
<th>COUNT</th>
<th>MEASURE</th>
<th>MNSQ</th>
<th>MNSQ</th>
<th>MEASURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>378</td>
<td>-.87</td>
<td>1.08</td>
<td>1.19</td>
<td>NONE</td>
</tr>
<tr>
<td>1</td>
<td>620</td>
<td>.13</td>
<td>.85</td>
<td>.69</td>
<td>-.85</td>
</tr>
<tr>
<td>2</td>
<td>852</td>
<td>2.23</td>
<td>1.00</td>
<td>1.46</td>
<td>.85</td>
</tr>
</tbody>
</table>

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

<p>| CATEGORY OBSERVED AVGE  INFIT OUTFIT  STRUCTURE |</p>
<table>
<thead>
<tr>
<th>LABEL</th>
<th>COUNT</th>
<th>MEASURE</th>
<th>MNSQ</th>
<th>MNSQ</th>
<th>MEASURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>378</td>
<td>-.87</td>
<td>1.08</td>
<td>1.19</td>
<td>NONE</td>
</tr>
<tr>
<td>1</td>
<td>620</td>
<td>.13</td>
<td>.85</td>
<td>.69</td>
<td>-.85</td>
</tr>
<tr>
<td>2</td>
<td>852</td>
<td>2.23</td>
<td>1.00</td>
<td>1.46</td>
<td>.85</td>
</tr>
</tbody>
</table>
Example 3: To enter CLFILE= information on the DOS Prompt or Extra Specifications lines, using commas instead of blanks as separators:

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

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

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

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

NI=30
CODES =ABCDEFGH
NEWSCORE=000111110
CLFILE=*  
0 Fail Specify the categories as recoded  
1 Pass  
*  
: or
CLFILE=*  
A Fail  
B Fail  
C Fail  
D Pass  
E Pass  
F Pass  
G Pass  
H Pass  

71. CODES  valid data codes

Says what characters to recognize as valid codes in your data file. If XWIDE=1 (the standard), use one column/character per legitimate code. If XWIDE=2, use two columns/characters per valid code. Characters in your data not included in CODES= are given the MISSCORE= value.

Example 1: A test has four response choices. These are "1", "2", "3", and "4". All other codes in the data file are to be treated as "item not administered". Each response uses 1 column in your data file. Data look like:
Example 2: There are four response choices. Each response takes up 2 columns in your data file and has leading 0's, so the codes are "01", "02", "03" and "04". Data look like: 0030240103020104040301

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

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

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

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

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

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

Example 9: The valid responses are percentages in the range 00 to 99.
Example 10: Codes are in the range 0-254 (the maximum possible).
XWIDE=3 ; 3 characters per response
CODES="  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23+
+ 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47+
+ 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71+
+ 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95+
+ 96 97 98 99100101102103104105106107108109110111112113114115116117118119+
+120121122123124125126127128129130131132133134135136137138139140141142143+
+144145146147148149150151152153154155156157158159160161162163164165166167+
+168169170171172173174175176177178179180181182183184185186187188189190191+
+192193194195196197198199200201202203204205206207208209210211212213214215+
+216217218219220221222223224225226227228229230231232233234235236237238239+
+240241242243244245246247248249250251252253254"

72. CONVERGE select convergence criteria

This selects which of LCONV= and RCONV= set the convergence criterion. See convergence considerations.

CONVERGE=L  LCONV= for "Logit change size" controls convergence.
Iteration stops when the biggest logit change is less or equal to LCONV=, or when the biggest logit change size increases (divergence).

CONVERGE=R  RCONV= for "Residual size" controls convergence.
Iteration stops when the biggest residual score is less or equal to RCONV=, or when the biggest residual size increases (divergence).

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

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

CONVERGE=F  Force both LCONV= for "Logit change size" and RCONV= for "Residual size" to control convergence.
Iteration stops when both the biggest logit change is less or equal to LCONV= and the biggest residual score is less or equal to RCONV=.

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

Example 2: We want to set the convergence criteria to match BIGSTEPS version 2.59
CONVERGE=B ; the rule was LCONV= and RCONV=
RCONV= 0.5 ; the BIGSTEPS standards or whatever value you used
LCOVN= .01

Example 3: We want to set the convergence criteria to match Winsteps version 3.20
CONVERGE=E ; the rule was LCONV or RCONV
RCONV= 0.5 ; the 3.20 standards or whatever value you used
LCOVN= .01

Example 4: We want the convergence criteria to match Winsteps version 2.85
CONVERGE= F; force both LCONV and RCONV to be met
RCONV= 0.5 ; the 2.85 standards or whatever value you used
LCONV=.01
You may also want:
WHEXACT=NO; centralized Wilson-Hilfterty was the default

Example 5: Question: With anchored analyses, iterations never stop!

<table>
<thead>
<tr>
<th>JMLE</th>
<th>MAX SCORE</th>
<th>MAX LOGIT</th>
<th>LEAST CONVERGED</th>
<th>CATEGORY</th>
<th>STEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITERATION</td>
<td>RESIDUAL*</td>
<td>CHANGE</td>
<td>EXID</td>
<td>BYCASE</td>
<td>CAT</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------</td>
<td>--------------</td>
<td>--------</td>
<td>--------</td>
<td>-----</td>
</tr>
<tr>
<td>1</td>
<td>-239.04</td>
<td>.5562</td>
<td>1993</td>
<td>392*</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>-105.65</td>
<td>-.1513</td>
<td>1993</td>
<td>392*</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>-5.35</td>
<td>-.0027</td>
<td>2228</td>
<td>352*</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>-5.16</td>
<td>.0029</td>
<td>2228</td>
<td>352*</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>-5.05</td>
<td>.0025</td>
<td>2228</td>
<td>352*</td>
<td>3</td>
</tr>
<tr>
<td>21</td>
<td>-5.00</td>
<td>.0010</td>
<td>2228</td>
<td>352*</td>
<td>3</td>
</tr>
<tr>
<td>22</td>
<td>-4.99</td>
<td>-.0008</td>
<td>2228</td>
<td>352*</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>170</td>
<td>-5.00</td>
<td>-.0011</td>
<td>1377</td>
<td>352*</td>
<td>3</td>
</tr>
<tr>
<td>171</td>
<td>-5.00</td>
<td>.0018</td>
<td>187</td>
<td>352*</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

To have the current Winsteps do this automatically, set
CONVERGE=L
LCONV=.005 ; set to stop at iteration 22 - to be on the safe side.

73. **CSV comma-separated values in output files**

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

CSV=N Use fixed field length format (the standard)
CSV=Y or CSV=S Separate values by commas (or their international replacements) with character fields in " " marks.
CSV=T Separate values by tab characters with character fields in " " marks (convenient for EXCEL).
CSV=S SPSS format.

Examples:

**Fixed space:**

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

**Tab-delimited:**

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

**Comma-separated:**

```
"MATCH Chess Matches at the Venice Tournament, 1971  Feb 11 0:47 2004"
";","ENTRY","MEASURE","STATUS","COUNT","SCORE","ERROR","IN.MSQ","IN.ZSTD",.....
```

81
SPSS format: This is the SPSS .sav file format.

74. CURVES  probability curves for Table 21 and Table 2

Interpreting rating (or partial credit) scale structure is an art, and sometimes counter-intuitive. See the examples in RSA.

Table 21 provides three curves for each rating (or partial credit) scale definition. Table 2 provides the equivalent response locations for each item. The first curve shows the probability of response in each category. The second curve shows the expected score ogive. The third curve shows the cumulative probability of response in each category or below. The 50% cumulative probability medians are at the intersections of these curves with the .5 probability line. The control indicators of "1" and "0", in the 1st, 2nd or 3rd position of the CURVES= variable, select or omit output of the corresponding curve.

?? Probability curves
?? Expected score ogives, model inter characteristic curves
??1 Cumulative probability curves, showing .5 probability median category boundaries.

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

CURVES=111  draw all 3 curves in Table 21 (and 3 versions of Table 2)

CURVES=001  draw only the 3rd, cumulative probability score, curve.

75. CUTHI cut off responses with high probability of success

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

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

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

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

CUTHI= 3

This produces a scalogram with eliminated responses blanked out:

RESPONSES SORTED BY MEASURE:
KID   TAP
     11111111
123745698013245678
------------------------
15   111  111000000 observations for extreme scores remain
14   111  111000000
28   111  111010000000
30   111  1111000000000
27  111111100000000000
76. **CUTLO** cut off responses with low probability of success

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

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

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

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

```
CUTLO= 2
```

RESPONSES SORTED BY MEASURE:

<table>
<thead>
<tr>
<th>KID</th>
<th>TAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>111111111</td>
<td>123745698013245678</td>
</tr>
</tbody>
</table>

--------------------------

15 1110101010111000000 observations for extreme scores remain
14 1110110011100000000
28 1110101110 00000
30 11110111 00000
27 111111 00000

77. **DATA** name of data file

Your data can be the last thing in the control file (which is convenient if you only have a small amount of data), but if you have a large amount of data, you can place it in a separate file, and then use **DATA=** to say where it is. **FORMAT=** reformats these records. **MFORMS=** enables multiple reformating.

Example 1: Read the observations from file "A:\PROJECT\RESPONSE.TXT".

```
DATA=A:\PROJECT\RESPONSE.TXT
```

Example 2: Read scanned MCQ data from file DATAFILE in the current directory.

```
DATA=DATAFILE
```

You may specify that several data files be analyzed together in one run, by listing their file names, separated by "+" signs. The list, e.g., FILE1.TXT+MORE.TXT+YOURS.D, can be up to 200 characters long. The layout of all data files must be identical.

Example 3: A math test has been scanned in three batches into files "BATCH.1", "BATCH.2" and "BATCH.3". They are to be analyzed together.

```
DATA=BATCH.1+BATCH.2+BATCH.3
```

78. **DELMITER** data field delimiters

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

To check that your data file has decoded properly, look at **RFILE=**

To do this, specify the following command **DELMITER=** value (or **SEPARATOR=** value). This value is the separator.
Examples:  DELIMITER=" "  fixed-field values
DELIMITER=" ", comma-separated values. The \ must be ","
DELIMITER=BLANK  blank-separated values
or  DELIMITER=SPACE  space-separated values
DELIMITER=TAB  tab-separated values
DELIMITER=";"  semi-colon separated values. The ; must be ";;", otherwise it is treated as a comment.

When decoding delimited values, leading and trailing blanks, and leading and trailing quotation marks, " " and '" in each value field are ignored. Responses are left-aligned, and sized according to XWIDE=.

For NAME1= and ITEM1=, specify the value number in the data line, starting with 1 as the leftmost value. FORMAT= does not apply to this data design.

Combine your person name and demographic information into one field that is to be referenced by NAME1=.

Example 1 of a data line:

; the following is ONE data line:

"01"; 02; "01"; "01"; 00; 02; 00; "01"; 02; 02; 02; 00; 02; "01"; 02; 02 ; 00; 02; "01"; 00; 02; 00; ROSSNER, MARC DANIEL

; which decodes as:

010201010001200102020020101202000201000200ROSSNER, MARC DANIEL

ITEM1=1  ; item responses start in first field
NI=25  ; there are 25 responses, i.e., 25 response fields
NAME1=26  ; the person name is in the 26th field
DELIMITER = ";"  ; the field delimiters are semi-colons
XWIDE=2  ; values are right-aligned, 2 characters wide.
CODES=000102  ; the valid codes.
NAMLEN=20  ; override standard person name length of 30 characters.

Example 2 of a data line:

; the following is ONE data line:

ROSSNER - MARC DANIEL, "01", 02 , "01", "01", "01", 00, 02, 00, "01", 02, 02, 02, 00, 02, "01", "01", 02, 02, 00, 02, "01", 00, 02, 00

; which decodes as:

010201010001200102020020101202000201000200ROSSNER - MARC DANIEL

ITEM1=2  ; item responses start in second field
NI=25  ; there are 25 responses, i.e., 25 response fields
NAME1=1  ; the person name is in the 1st field
DELIMITER = ";"  ; the field delimiters are commas (so no commas in names)
XWIDE=2  ; values are right-aligned, 2 characters wide.
CODES=000102  ; the valid codes
NAMLEN=20  ; override standard person name length of 30 characters.

Example:  Here is the data file, "Book1.txt"

fred,1,0,1,0
george,0,1,0,1

Here is the control file:

name1=1  ; first field
item1=2  ; second field
ni=4  ; 4 fields
data=book1.txt
codes=01
delimiter = ",,"
&END
looking
viewing
peeking
seaking
END LABELS

Here is the reformatted file from the Edit Pull-Down menu: View Delimiter File:
1010fred
0101george

79. **DIF columns within person label for Table 30**

DIF= specifies the part of the person label which is to be used for classifying persons in order to identify Differential Item Function (DIF) - uniform or non-uniform - using the column selection rules. See also DIF Table and DIF and DPF considerations.

DIF= location is usually column number **within person label** field. DIF=1 means “DIF selection character is first character of person label.”

Example 0: I have tab-separated data and my DIF indicator is in a separate field from the Person label. Solution: for the DIF analysis, specify the DIF field as the person label field using NAME1=, then $S1W1

Example 1: Columns 18-20 of the person label (in columns 118-120 of the data record) contain a district code:
NAME1=101 ; person label starts in column 101
DIF = $S18W3 ; district starts in column 18 of person label with a width of 3
or
@district = 18W3 ; district starts in column 18 of person label with a width of 3
DIF = @district ; DIF classifier

tfile=* 30 Table 30 for the DIF report (or use Output Tables menu) *

Example 2: Columns 18-20 of the person label (in columns 118-120 of the data record) contain a district code. Column 121 has a gender code. Two independent DIF analyses are needed:
NAME1=101 ; person label starts in column 101
DIF = * $S18W3 ; start in person label column 18 with a width of 3 - district
$S21W1 ; start in person label column 21 with a width of 1 - gender
*
tfile=* 30 Table 30 for the DIF report (or use Output Tables menu) *

Example 3: An investigation of non-uniform DIF with high-low ability classification for the KCT data.
: action the following with the Specification pull-down menu
@SEX = $S9W1 ; the sex of participants is in column 9 of the person label
DIF = @SEX + MA2; look for non-uniform DIF (gender + two ability strata)
PSUBTOT = @SEX + MA2; summary statistics by gender and ability strata
Tfile=* ; This is more easily actioned through the Output Tables Menu
30 ; Table 30 - DIF report
28 ; Table 28 - Person subtotals for DIF classifications
*

<table>
<thead>
<tr>
<th>Table 30: DIF specification is: DIF=SEX+MA2</th>
</tr>
</thead>
</table>

Table 30: DIF specification is: DIF=SEX+MA2
<table>
<thead>
<tr>
<th>KID</th>
<th>DIF</th>
<th>CLASS MEASURE S.E.</th>
<th>JOINT S.E.</th>
<th>TAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>KID</td>
<td>DIF</td>
<td>CLASS MEASURE S.E.</td>
<td>CONTRAST S.E.</td>
<td>t d.f. Number Name</td>
</tr>
</tbody>
</table>

85
Example 4: With Example0.txt (the Liking for Science rating scale data) you want to see if any items were biased against names starting with any letter of the alphabet, then:

run example0.txt
request the DIF Table (Table 30) from the Output Tables menu
specify: $S1W1
a DIF table is produced.

The equivalent DIF specification is: DIF=$S1W1

Positive DIF size is higher ACT difficulty measure

Rasch models assert that items exhibit the model-specified item discrimination. Empirically, however, item discriminations vary. During the estimation phase of Winsteps, all item discriminations are asserted to be equal, of value 1.0, and to fit the Rasch model. But empirical item discriminations never are exactly equal, so Winsteps can also report an estimate of those discriminations post-hoc (as a type of fit statistic). The amount of the departure of a discrimination from 1.0 is an indication of the degree to which that item misfits the Rasch model.

DISCRIM=NO Do not report an estimate of the empirical item discrimination.

DISCRIM=YES Report an estimate of the empirical item discrimination in the IFILE= and Tables 6.1, 10.1, etc.

An estimated discrimination of 1.0 accords with Rasch model expectations for an item of this difficulty. A value greater than 1 means that the item discriminates between high and low performers more than expected for an item of this difficulty. A value less than 1 means that the item discriminates between high and low performers less than expected for an item of this difficulty. In general, the geometric mean of the estimated discriminations approximates 1.0, the Rasch item discrimination.

Rasch analysis requires items which provide indication of relative performance along the latent variable. It is this information which is used to construct measures. From a Rasch perspective, over-discriminating items are tending to act like switches, not measuring devices. Under-discriminating items are tending neither to stratify nor to measure.

Over-discrimination is thought to be beneficial in many raw-score and IRT item analyses. High discrimination usually corresponds to low MNSQ values, and low discrimination with high MNSQ values. In Classical Test Theory, Guttman Analysis and much of Item Response Theory, the ideal item acts like a switch. High performers pass, low performers fail. This is perfect discrimination, and is ideal for sample stratification, but such an item provides no information about the relative performance of low performers, or the relative performers of high performers.

Winsteps reports an approximation to what the discrimination parameter value would have been in a 2-PL IRT
program, e.g., BILOG for MCQ, or PARSCALE for partial credit items. IRT programs artificially constrain discrimination values in order to make them estimable, so Winsteps discrimination estimates tend to be wider than 2-PL estimates. For the lower asymptote, see ASYMPTOTE=.

A Rasch-Andrich threshold discrimination is also reported, see Table 3.2.

With DISCRIM=YES,

<table>
<thead>
<tr>
<th>ENTRY</th>
<th>RAW</th>
<th>COUNT</th>
<th>MEASURE</th>
<th>ERROR</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>DISCR</th>
<th>ACTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>40</td>
<td>74</td>
<td>2.19</td>
<td>.21</td>
<td>2.42</td>
<td>6.3</td>
<td>4.13</td>
<td>.89</td>
</tr>
<tr>
<td>17</td>
<td>93</td>
<td>74</td>
<td>.16</td>
<td>.19</td>
<td>.65</td>
<td>-2.7</td>
<td>.59</td>
<td>-2.5</td>
</tr>
</tbody>
</table>

81. **DISFILE** category/distractor/option count file

DISFILE=filename produces an output file containing the counts for each distractor or option or category of each item. This file contains 1 heading lines (unless HLINES=N), followed by one line for each CODES= of each item containing:

Columns:

- Start
- End
- Description
- Item entry number
- Response code in CODES= or *** = missing
- Scored value of code. (MISSSCORE=-1)
- Count of response code in data set (excludes missing except for "MISSING" lines)
- % of responses to this item
- Count of response codes used for measurement (non-extreme persons and items)
- % of responses to this item
- Average person measure for these used responses
- Standard Error of average person measure
- Outfit Mean-square of responses in this category.
- Point-biserial PTBIS=Y or point-measure correlation for this response (distractor, option)
- Item label

Since the DISFILE= has the same number of CODES= and MISSING entries for every item, the repeated fields are filled out with "0" for any unobserved response codes.

When CSV=Y, commas separate the values with quotation marks around the "Item label", response in CODES=, and MISSING. When CSV=T, the commas are replaced by tab characters.

Example: You wish to write a file on disk called "DISCOUNT.TXT" containing the item distractor counts from Table 14.3, for use in constructing your own tables:

```
DISFILE=DISCOUNT.TXT
```

82. **DISTRACTOR** output option counts in Tables 10, 13-15

This variable controls the reporting of counts of option, distractor or category usage in Table 10.3 etc. The standard is DISTRT=Y, if more than two values are specified in CODES=.

- DISTRACTOR=N Omit the option or distractor information.
- DISTRACTOR=Y Include counts, for each item, for each of the values in CODES=, and for the number of responses counted as MISSCORE=
83. **DPF columns within item label for Table 31**

DPF= specifies the part of the item label which is to be used for classifying items in order to identify Differential Person Function (DPF) - uniform or non-uniform - using the column selection rules. See also DPF Table and DIF and DPF considerations.

See [ISUBTOTAL](#) for format.

See [DPF Table](#).

**Example 1:** Columns 3 and 4 of the item label (between &END and END LABELS) contains content-area code:

```
DPF = $S3E4  ; start in column 3 and end in column 4 of item label
```

```
tfile=*  
```

```
31       ; Table 31 is DPF Table (or use Output Tables menu)
```

**Example 2:** Columns 3 of the item label contains a content code. Column 5-6 have a complexity code. Two independent DIF analyses are needed:

```
DPF =  
```

```
$S3W1     ; content analysis
$S5W2     ; complexity
```

```
tfile=*  
```

```
31       ; Table 31 is DPF Table (or use Output Tables menu)
```

84. **EDFILE edit data file**

This permits the replacement of data values in your data file with other values, without altering the data file. Data values are in the original data file format, specified in CODES=. If specified as decimals, they are rounded to the nearest integers.

Its format is:
```
person entry number   item entry number   replacement data value
```

Ranges are permitted: first-last.

**Example 1:** In your MCQ test, you wish to correct a data-entry error. Person 23 responded to item 17 with a D, not whatever is in the data file.

```
EDFILE=*  
```

```
23 17 D   ; person 23, item 17, data value of D
```

**Example 2:** Person 43 failed to read the attitude survey instructions correctly for items 32-56. Mark these missing.

```
43 32-56 " " ; person 43, items 32 to 56, blanks are missing data.
```

**Example 3:** Persons 47-84 are to be given a rating of 4 on item 16.

```
47-84 16 4   ; persons 47 to 84, item 16, data value of 4
```

**Example 4:** Items 1-10 are all to be assigned a datum of 1 for the control subsample, persons 345-682.

```
345-682 1-10 1 ; persons 345-682, items 1 to 10, data value 1.
```

**Example 5:** Missing data values are to be imputed with the values nearest to their expectations.
a. Produce PFILE=, IFILE= and SFILE= from the original data (with missing).
b. Use those as PAFILE=, IAFILE=, SAFILE= anchor files with a data set in which all the original non-missing
data are made missing, and vice-versa - it doesn't matter what non-missing value is used.
c. Produce XFILE= to obtain a list of the expected values of the originally missing data.
d. Use the EDFILE= command to impute those values back into the data file. It will round expected values to
the nearest integer, for us as a category value.

17 6 2.6 ; persons 17, item 6, expected value 2.6, imputed as category "3".

85. END LABELS or END NAMES

The first section of a control file contains control variables, one per line, and ends with &END. This is followed by
the second section of item labels, one per line, matching the items in the analysis. This sections ends with END
LABELS or END NAMES, which mean the same thing. The data can follow as a third section, or the data can be
in a separate file specified by the control variable DATA=.

TITLE = "5 item test"
ITEM1 = 1
NI = 5
.....
&END
Addition ; label for item 1
Subtraction ; label for item 2
Multiplication ; label for item 3
Division ; label for item 4
Geometry ; label for item 5
END LABELS
..... ; data here

86. EQFILE code equivalences

This specifies that different demographic or item-type codes are to be reported as one code. This is useful for
Tables 27, 28, 30, 31, 33 and Use EQFILE=filename or EQFILE=*, followed by a list, followed by a *. These
values can be overwritten from the equivalence boxes when invoking the Tables from the Output Tables menu.

The format is
@Field name = $S1W1 ; user defined field name and location, see selection rules.
EQFILE=* ; start of list
@Field name ; field to be referred to
Base Code Code Code Code ..... ; code list
Base Code Code Code Code ..... 
Base Code Code Code Code ..... 
@Field name ; field to be referred to
Base Code Code Code Code ..... ; code list
Base Code Code Code Code ..... 
Base Code Code Code Code ..... 
* ; end of list

where @Field name is the name of field in the person or item label, such as
@GENDER = $S1W1 ; M or F
@STRAND = $S10W2 ; 01 to 99

Base is the demographic or item-type code to be reported. It need not be present in a label
Code is a demographic or item-type code to be included with the Base code.

87. EXTRSC extreme score correction for extreme measures

EXTRSCORE= is the fractional score point value to subtract from perfect scores, and to add to zero scores, in
order to estimate finite values for extreme scores (formerly MMADJ=). Look at the location of the E's in the tails of the test ogive in Table 20. If they look too far away, increase EXTRSC= by 0.1. If they look too bunched up, reduce EXTRSC= by 0.1.

The measure corresponding to an extreme (perfect or zero) score is not estimable, but the measure corresponding to a score a little less than perfect, or a little more than zero, is estimable, and is often a useful measure to report.

Rasch programs differ in the way they estimate measures for extreme scores. Adjustment to the value of EXTRSC= can enable a close match to be made to the results produced by other programs.

There is no "correct" answer to the question: “How large should EXTRSC= be?” The most conservative value, and that recommended by Joseph Berkson, is 0.5. Some work by John Tukey indicates that 0.167 is a reasonable value. The smaller you set EXTRSC=, the further away measures corresponding to extreme scores will be located from the other measures. The technique used here is Strategy 1 in www.rasch.org/rmt/rmt122h.htm.

Example 1: You wish to estimate conservative finite measures for extreme scores by subtracting 0.4 score points from each perfect score and adding 0.4 score points to each zero person score.

<table>
<thead>
<tr>
<th>ENTRY</th>
<th>RAW</th>
<th>COUNT</th>
<th>MEASURE</th>
<th>ERROR</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>PTBIS</th>
<th>PERSON</th>
</tr>
</thead>
<tbody>
<tr>
<td>46</td>
<td>60</td>
<td>20</td>
<td>7.23</td>
<td>1.88</td>
<td>MAXIMUM ESTIMATED MEASURE</td>
<td>XPQ003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>62</td>
<td>21</td>
<td>5.83</td>
<td>1.12</td>
<td>.44</td>
<td>-.9</td>
<td>.08</td>
<td>.62</td>
</tr>
<tr>
<td>86</td>
<td>18</td>
<td>6</td>
<td>5.11</td>
<td>1.90</td>
<td>MAXIMUM ESTIMATED MEASURE</td>
<td>XPQ009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>50</td>
<td>17</td>
<td>4.94</td>
<td>1.09</td>
<td>.53</td>
<td>-.7</td>
<td>.13</td>
<td>.60</td>
</tr>
</tbody>
</table>

Here, the 5.11 corresponds to a perfect score on 6 easier items. The 5.83 was obtained on 21 harder items (perhaps including the 6 easier items.) To adjust the "MAXIMUM ESTIMATED" to higher measures, lower the value of EXTRSCORE=, e.g., to EXTRSCORE=0.2

Example 2: With the standard value of EXTRSCORE=, this Table is produced:

FITHIGH higher bar in charts

Use FITHIGH= to position the higher acceptance bars in Tables like 6.2. Use FITLOW= to position the lower bar.

FITHIGH=0 cancels the instruction.

Example: We want the lower mean-square acceptance bar to be shown at 1.4

FITHIGH=1.4 ; show higher fit bar
CHART=YES ; produce Tables like 6.2
89. **FITI** item misfit criterion

Specifies the minimum t standardized fit value at which items are selected for reporting as misfits. For Table 10, the table of item calibrations in fit order, an item is omitted only if the absolute values of both t standardized fit statistics are less than FITI=, both mean-square statistics are closer to 1 than (FITI=)/10, and the item point-biserial correlation is positive.

For Table 11, the diagnosis of misfitting items, all items with a t standardized fit greater than FITI= are reported. Selection is based on the OUTFIT statistic, unless you set OUTFIT=N in which case the INFIT statistic is used. If MNSQ=YES, then selection is based on the mean-square value: 1 + FITI=/10.

Example 1: You wish to focus on grossly "noisy" items in Tables 10 and 11.

FITI=4  *an extreme positive value*

Example 2: You wish to include all items in Tables 10 and 11.

FITI=-10  *a value more negative than any fit statistic*

90. **FITLOW** lower bar in charts

Use FITLOW= to position the lower acceptance bars in Tables like 6.2. Use FITHIGH= to position the higher bar.

FITLOW=0 cancels the instruction.

Example: We want the lower mean-square acceptance bar to be shown at 0.6

FITLOW=0.6  ; show lower fit bar

CHART=YES  ; produce Tables like 6.2

91. **FITP** person misfit criterion

Specifies the minimum t standardized fit value at which persons are selected for reporting as misfits. For Table 6, person measures in fit order, a person is omitted only if the absolute values of both t standardized fit statistics are less than FITP=, both mean-square statistics are closer to 1 than (FITP=)/10, and the person point-biserial correlation is positive.

For Table 7, the diagnosis of misfitting persons, persons with a t standardized fit greater than FITP= are reported. Selection is based on the OUTFIT statistic, unless you set OUTFIT=N in which case the INFIT statistic is used. If MNSQ=YES, then selection is based on the mean-square value: 1 + FITP=/10.

Example 1: You wish to examine wildly guessing persons in Tables 6 and 7.

FITP=3  *an extreme positive value*

Example 2: You wish to include all persons in Tables 6 and 7.

FITP=-10  *a value more negative than any fit statistic*

92. **FORMAT** reformat data

**Enables you to process awkwardly formatted data! But MFORMS= is easier**

FORMAT= is rarely needed when there is one data line per person.
Place the data in a separate file, then the Winsteps screen file will show the first record before and after
FORMAT=

Control instructions to pick out every other character for 25 two-character responses, then a blank, and then the
person label:
XNIDE=1
data=datafile.txt
format=(T2,25(1A,1X),T90,1A,Tl1,30A)

This displays on the Winsteps screen:
Opening: datafile.txt
Input Data Record before FORMAT=:
1         2         3         4         5         6         7
123456789012345678901234567890123456789012345678901234567890
-------------------------------------------------------------------
01xx 1x1 1000200010202020002010200002000201000200ROSSNER, MARC DANIEL
Input Data Record after FORMAT=:
1x1110202222021122021020 L
^I                      ^N^P
^N is the last item according to NI=
^P is Name1= column

FORMAT= enables you to reformat one or more data record lines into one new line in which all the component
parts of the person information are in one person-id field, and all the responses are put together into one
continuous item-response string. A FORMAT= statement is required if
1) each person's responses take up several lines in your data file.
2) if the length of a single line in your data file is more than 10000 characters.
3) the person-id field or the item responses are not in one continuous string of characters.
4) you want to rearrange the order of your items in your data record, to pick out sub-tests, or to move a set of
connected forms into one complete matrix.
5) you only want to analyze the responses of every second, or nth, person.

FORMAT= contains up to 512 characters of reformatting instructions, contained within (..), which follow special
rules. Instructions are:

nA read in n characters starting with the current column, and then advance to the next column after them.
Processing starts from column 1 of the first line, so that 5A reads in 5 characters and advances to the sixth
column.

nX means skip over n columns. E.g. 5X means bypass this column and the next 4 columns.

Tc go to column c. T20 means get the next character from column 20.
T55 means "tab" to column 55, not "tab" passed 55 columns (which is TR55).

TLc go c columns to the left. TL20 means get the next character the column which is 20 columns to the left of the
current position.

TRc go c columns to the right. TR20 means get the next character the column which is 20 columns to the right
of the current position.

/ means go to column 1 of the next line in your data file.

n(..) repeat the string of instructions within the ( ) exactly n times.

, a comma is used to separate the instructions.
Set XWIDE=2 and you can reformat your data from original 1 or 2 column entries. Your data will all be analyzed as XWIDE=2. Then:

nA2 read in n pairs of characters starting with the current column into n 2-character fields of the formatted record. (For responses with a width of 2 columns.)

A1 read in n 1-character columns, starting with the current column, into n 2-character fields of the formatted record.

Always use nA1 for person-id information. Use nA1 for responses entered with a width of 1-character when there are also 2-character responses to be analyzed. When responses in 1-character format are converted into 2-character field format (compatible with XWIDE=2), the 1-character response is placed in the first, left, character position of the 2-character field, and the second, right, character position of the field is left blank. For example, the 1-character code of "A" becomes the 2-character field "A ". Valid 1-character responses of "A", "B", "C", "D" must be indicated by CODES="A B C D " with a blank following each letter.

**ITEM1=** must be the column number of the first item response in the formatted record created by the **FORMAT=** statement. **NAME1=** must be the column number of the first character of the person-id in the formatted record.

**Example 1:** Each person's data record file is 80 characters long and takes up one line in your data file. The person-id is in columns 61-80. The 56 item responses are in columns 5-60. Codes are "A", "B", "C", "D". No **FORMAT=** is needed. Data look like:

```
xxxx DCBD ABC D BC DB A D CB A D CD BA D CD AD CB A D C B A CZarahrustra-Xerxes
```

**Without FORMAT=**

- XWIDE=1 response width (the standard)
- ITEM1=5 start of item responses
- NI=56 number of items
- NAME1=61 start of name
- NAMLEN=20 length of name
- CODES=ABCD valid response codes

**With FORMAT=**

Reformatted record will look like:

```
DCBD ABC D BC DB A D CB A D CD BA D CD AD CB A D C B A CZarahrustra-Xerxes
```

- XWIDE=1 response width (the standard)
- FORMAT=(4X,56A,20A) skip unused characters
- ITEM1=1 start of item responses
- NI=56 number of items
- NAME1=57 start of name
- NAMLEN=20 length of name
- CODES=ABCD valid response codes

**Example 2:** Each data record is one line of 80 characters. The person-id is in columns 61-80. The 28 item responses are in columns 5-60, each 2 characters wide. Codes are "A", "B", "C", "D". No **FORMAT=** is necessary. Data look like:

```
```

**Without FORMAT=**

- XWIDE=2 response width
- ITEM1=5 start of item responses
- NI=28 number of items
- NAME1=61 start of name
- NAMLEN=20 length of name
- CODES="A B C D" valid response codes

**With FORMAT=**

Columns of reformatted record:

```
1-2-3-4-5-6-7-8-9-0-1-2-3-4-5-6-7-8-9-0-1-2-3-4-5-6-7-8-90123456789012345678
```

- XWIDE=2 response width
- FORMAT=(4X,29A2,20A1) skip unused characters
Example 3: Each person's data record is 80 characters long and takes one line in your data file. Person-id is in columns 61-80. 30 1-character item responses, "A", "B", "C" or "D", are in columns 5-34, 13 2-character item responses, "01", "02" or "99", are in 35-60.

Example 4: The person-id is 10 columns wide in columns 15-24 and the 50 1-column item responses, "A", "B", "C", "D", are in columns 4000-4019, then in 4021-50. Data look like:

Example 5: There are five records or lines in your data file per person. There are 100 items. Items 1-20 are in columns 25-44 of first record; items 21-40 are in columns 25-44 of second record, etc. The 10 character person-id is in columns 51-60 of the last (fifth) record. Codes are "A", "B", "C", "D". Data look like:

Example 6: There are three lines per person. In the first line from columns 31 to 50 are 10 item responses, each 2 columns wide. Person-id is in the second line in columns 5 to 17. The third line is to be skipped. Codes are "A", "B", "C", "D". Data look like:
Columns:
1-2-3-4-5-6-7-8-9-0-1234567890123
A C B D A D C B A DJoseph-Carlos

FORMAT=(T31,10A2,/,T5,13A1,/)  
ITEM1=1  start of item responses  
NI=10  number of items  
XWIDE=2  2 columns per response  
NAME1=11  starting "A" of person name  
NAMLEN=13  length of person name  
CODES='A B C D '  valid response codes

If the third line isn't skipped, format a redundant extra column in the skipped last line. Replace the first control variable in this with:
FORMAT=(T31,10A2,/,T5,13A1,/,A1) last A1 unused

Example 7: Pseudo-random data selection
You have a file with 1,000 person records. This time you want to analyze every 10th record, beginning with the 3rd person in the file, i.e., skip two records, analyze one record, skip seven records, and so on. The data records are 500 characters long.

XWIDE = 1
FORMAT = (/,/,500A,/,/,/,/,/,/,/)  
or
XWIDE = 2
FORMAT = (/,/,100A2,300A1,/,/,/,/,/,/,/) ; 100 2-character responses, 300 other columns

Example 8: Test A, in file EXAM10A.TXT, and TEST B, in EXAM10B.TXT, are both 20 item tests. They have 5 items in common, but the distractors are not necessarily in the same order. The responses must be scored on an individual test basis. Also the validity of each test is to be examined separately. Then one combined analysis is wanted to equate the tests and obtain bankable item difficulties. For each file of original test responses, the person information is in columns 1-25, the item responses in 41-60.

The combined data file specified in EXAM10C.TXT, is to be in RFILe= format. It contains

Person information  30  characters (always)
Item responses       Columns 31-64

The identification of the common items is:
Test  Item Number (=Location in item string)
Both: 1 2 3 4 5 6-20 21-35
A:  3 1 7 8 9 2,4-6,10-20
B:  4 5 6 2 11 1,3,7-10,12-20

I. From Test A, make a response (RFILe=) file rearranging the items with FORMAT=.

; This file is EXAM10A.TXT
&INST
TITLE="Analysis of Test A"
RFILe=EXAM10AR.TXT ; The constructed response file for Test A
NI=20
ITEM1=26  ; Items start in column 26 of reformatted record
CODES=ABCD#  ; Beware of blanks meaning wrong!
; Use your editor to convert all "wrong" blanks into another code,
; e.g., #, so that they will be scored wrong and not ignored as missing.
KEYFRM=1  ; Key in data record format
&END
Key 1 Record                            CCBDACADBACABCDABCA
BANK 1 TEST A 3  ; first item name
BANK 20 TEST A 20
END NAMES
The RFILE= file, EXAM10AR.TXT, is:

Person 01 A  00001000010010001001
Person 02 A  00000100001110100111
Person 12 A  00100001100001001011

II. From Test B, make a response (RFILE=) file rearranging the items with FORMAT=. Responses unique to Test A are filled with 15 blank responses to dummy items.

; This file is EXAM10B.TXT
&INST
TITLE=“Analysis of Test B”
RFILE=EXAM10BR.TXT ; The constructed response file for Test B
NI=35

ITEM1=26 ; Items start in column 26 of reformatted record
CODES=ABCD# ; Beware of blanks meaning wrong!
KEYFRM=1 ; Key in data record format
&END

Key 1 Record CDABCDBABCADCDABDBCAD

BANK 1   TEST B 4
BANK 5   TEST B 11
BANK 6   TEST A 2
BANK 20  TEST A 20
BANK 21  TEST B 1

BANK 35  TEST B 20

END NAMES

Person 01 B  BDABDDCDBBCCCDAAACBC
Person 12 B  BADABBADCBADBDBBBBBB

The RFILE= file, EXAM10BR.TXT, is:

Person 01 B  10111  0101010010001001
Person 02 B  00000  0100000000010000
Person 11 B  00010  0010000000001000
Person 12 B  00000  0001010001010000

III. Analyze Test A's and Test B's RFILE='s together:

; This file is EXAM10C.TXT
&INST
TITLE=“Analysis of Tests A & B (already scored)”
NI=35
ITEM1=31 ; Items start in column 31 of RFILE=
CODES=01 ; Blanks mean "not in this test"
DATA=EXAM10AR.TXT+EXAM10BR.TXT ; Combine data files

; or, first, at the DOS prompt,
; C:> COPY EXAM10AR.TXT+EXAM10BR.TXT EXAM10AB.TXT(Enter)
; then, in EXAM10C.TXT,
; DATA=EXAM10AB.TXT

96
SHORTENING FORMAT= STATEMENTS
If the required FORMAT= statement exceeds 512 characters, consider using this technique:

Relocate an entire item response string, but use an IDFILE= to delete the duplicate items, i.e., replace them by blanks. E.g., for Test B, instead of
\begin{verbatim}
NI=35
\end{verbatim}
Put Test 2 as items 21-40 in columns 51 through 70:
\begin{verbatim}
NI=40
\end{verbatim}
Blank out (delete) the 5 duplicated items with an IDFILE= containing:
24-26
22
31

93. FORMFD the form feed character
Do not change FORMFD= unless you have problems printing the tables or importing them into some other program.

The form feed character indicates the start of a new page of print-out. The DOS standard is Ctrl+L (ASCII 12) which is what represented by ^ (Shift+6). The DOS standard is understood by most word-processing software and PC printers as the instruction to skip to the top of a new page, i.e., form feed. The ASA (FORTRAN) form feed character is 1.

WordPad does not have a "form feed" or page advance feature. You must put extra blank lines in the output files.

Example 1: You want your EPSON LQ-500 printer to form feed automatically at each new page of output. (You have already set the printer to use compressed print, at 15 cpi, because output lines contain up to 132 characters):
\begin{verbatim}
FORMFD=^ (the standard)
\end{verbatim}
Example 2: Your main-frame software understands a "1" in the first position of a line of print-out to indicate the top of a new page:
\begin{verbatim}
FORMFD=1
\end{verbatim}

94. FRANGE half-range of fit statistics on plots
Specifies the t standardized fit Y-axis half-range, (i.e. range away from the origin), for the t standardized fit plots. FRANGE= is in units of t standardized fit (i.e., expected mean = 0, standard deviation = 1).

Example: You want the fit plots to display from -3 to +3 units of t standardized fit:
\begin{verbatim}
FRANGE=3
\end{verbatim}
95. **GRFILE probability curve coordinate output file**

If GRFILE=filename is specified, a file is output which contains a list of measures (x-axis coordinates) and corresponding expected scores and category probabilities (y-axis coordinates) to enable you to use your own plotting program to produce item-category plots like those in Table 21.

These are usually relative to the item difficulty, so
If you want the item information function relative to the item difficulty, use the GRFILE=
If you want the item information function relative to the latent variable, add the item difficulty to the GRFILE=

The plotted range is at least MRANGE= away from its center.

This file contains:
1. An example item number from the response-structure grouping (I5) - see ISGROUPS=
2. The measure (F7.2) (user-rescaled by USCALE=)
3. Expected score (F7.2)
4. Statistical information (F7.2)
5. Probability of observing lowest category (F7.2)
6 etc. Probability of observing higher categories (F7.2).

If CSV=Y, values are separated by commas. When CSV=T, values are separated by tab characters.

Example: You wish to write a file on disk called "MYDATA.GR" containing x- and y-coordinates for plotting your own category response curves.

```
GRFILE=MYDATA.GR
```

With CSV=Y

```
"PROBABILITY CURVES FOR LIKING FOR SCIENCE (Wright & Masters p.18) Jul 4 16:03 2000"
;ITEM,MEAS,SCOR,INFO,0,1,2
1,-3.00,.11,.10,.89,.10,.00
1,-2.94,.12,.11,.89,.11,.00
1,-2.88,.12,.11,.88,.12,.00
1,-2.82,.13,.12,.87,.12,.00
1,-2.76,.14,.12,.87,.13,.00
1,-2.70,.14,.13,.86,.14,.00
```

With CSV=N (fixed spacing)

```
; PROBABILITY CURVES FOR LIKING FOR SCIENCE (Wright & Masters p.18) Jul 4 16:03 2000
; ITEM MEAS SCOR INFO 0 1 2
1 -3.00 .11 .10 .89 .10 .00
1 -2.94 .12 .11 .89 .11 .00
1 -2.88 .12 .11 .88 .12 .00
1 -2.82 .13 .12 .87 .12 .00
1 -2.76 .14 .12 .87 .13 .00
1 -2.70 .14 .13 .86 .14 .00
```

96. **GROUPS or ISGROUPS assigns items to rating scale groupings**

Items in the same "grouping" share the same dichotomous, rating scale or partial credit response structure. For tests comprising only dichotomous items, or for tests in which all items share the same rating (or partial credit) scale definition, all items belong to one grouping, i.e., they accord with the simple dichotomous Rasch model or the Andrich "Rating Scale" model. For tests using the "Masters' Partial Credit" model, each item comprises its own grouping (dichotomous or polytomous). For tests in which some items share one polytomous response-structure definition, and other items another response-structure definition, there can be two or more item groupings. Groups are called "blocks" in the PARSCALE software.
where $P$ is a probability, and the Rasch parameters are $B_n$, the ability of person, $D_{gi}$, the difficulty of item $i$ of grouping $g$, and $F_{gj}$, the Rasch-Andrich threshold between categories $j-1$ and $j$ of grouping $g$. If there is only one grouping, this is the Andrich "rating scale" model. If each item forms a grouping of its own, i.e., $g=i$, this is the Masters' "partial credit" model. When several items share the rating scale, then this could be called an item-grouping-level Andrich rating-scale model, or an item-grouping-level Masters' partial-credit model. They are the same thing.

ISGROUPS= also acts as IREFER=, when IVALUE= is specified, but IREFER= is omitted.

ISGROUPS= has three forms: ISGROUPS=1101110 and ISGROUPS=* list * and ISGROUPS=*filename

ISGROUPS=" " (standard if only one model specified with MODELS=)
   All items belong to one grouping. This is sometimes called "Andrich's Rating Scale Model"

ISGROUPS=0 (standard if MODELS= specifies multiple models)
   Each item has a grouping of its own, so that a different response structure is defined for each item, as in the "Masters' Partial Credit model". This is also used for rank order data.

ISGROUPS= some combination of numbers and letters: 0's, 1's, 2's, A's, B's, etc. ("a" is the same as "A"), also #, @, !,& etc.

Use only one letter or number per grouping, regardless of the XWIDE= setting.
   Items are assigned to the grouping label whose location in the ISGROUPS= string matches the item sequence number. Each item assigned to grouping 0 is allocated to a "partial credit" grouping by itself. Items in groupings labeled "1", "2", "A", etc. share the response-structure definition with other items in the same labeled grouping.

Valid one-character ISGROUPS= codes include: !#$%&-./123456789<>@ABCDEFGHIJKLMNOPQRSTUVWXYZ^_|~

A-Z are the same as a-z.

For the ISGROUPS= specification, "0" has the special meaning: "this is a one item grouping" - and can be used for every 1 item grouping.

Characters with ASCII codes from 129-255 can also be used, but display peculiarly: ÏÎÎÎÎÎÎÎ×ØUUÝÝÝßåååå etc.

When XWIDE=2 or more, then either (a) Use one character per XWIDE= and blanks,
   NI=8
   XWIDE=2
   ISGROUPS='1 0 1 0 1 0 1 1'
   or (b) Use one character per item with no blanks
   NI=8
   XWIDE=2
   ISGROUPS='10101011'

ISGROUPS=*
   item number grouping code
   item number-item number grouping code
*   Each line has an item number, e.g., 23, or an item number range, e.g., 24-35, followed by a space and then a grouping code, e.g., 3. The items can be out of order. If an item is listed twice, the last entry dominates.

ISGROUPS="filename"
This has a file with the format of the ISGROUPS= list.

Particularly with ISGROUPS=0, some extreme categories may only be observed for persons extreme scores. To reinstate them into the measurement analysis, see Extreme Categories: Rescuing the Dropped.

### Groupings vs. Separate Analyses

ISGROUPS= is very flexible for analyzing together different item response structures in one analysis. Suppose that an attitude survey has 20 Yes-No items, followed by 20 5-category Likert (Strongly disagree - disagree - neutral - agree- strongly agree) items, followed by 20 3-category frequency (never - sometimes - often) items. When possible, we analyze these together using ISGROUPS=. But sometimes the measurement characteristics are too different. When this happens, the fit statistics stratify by item type: so that, say, all the Yes-No items overfit, and all the Frequency items underfit. Then analyze the test in 3 pieces, and equate them together - usually into the measurement framework of the response structure that is easiest to explain. In this case, the Yes-No framework, because probabilistic interpretation of polytomous logits is always difficult to explain or perform.

The "equation" would be done by cross-plotting the person measures for different item types, and getting the slope and intercept of the conversion from that. Drop out of the "equation" any noticeably off-trend-line measures. These are person exhibiting differential performance on the different item types.

**Example 1:** Responses to all items are on the same 4-point rating scale, an Andrich "Rating Scale" model, ISGROUPS=" "

**Example 2:** An arithmetic test in which partial credit is given for an intermediate level of success on some items. There is no reason to assert that the intermediate level is equivalent for all items. 0=No success, 1=Intermediate success (or complete success on items with no intermediate level), 2=Complete success on intermediate level items.

```plaintext
CODES=012 valid codes
ISGROUPS=0 each item has own response structure, i.e., Masters' Partial Credit model
```

or

```plaintext
ISGROUPS=*  
1 0 ; item 1 is in Grouping 0, no other items mentioned, so all assumed to be in Grouping 0  
*  
```

**Example 3:** An attitude survey consists of four questions on a 0,1,2 rating scale (grouping 1), an Andrich "Rating Scale" model, followed by three 0,1 items (grouping 2), an other Andrich "Rating Scale" model, and ends with one 0,1,2,3,4,5 question (grouped by itself, 0), a Masters' "Partial Credit" model.

```plaintext
NI=8 number of items
CODES=012345 valid codes for all items
ISGROUPS=11112220 the item groupings
```

or

```plaintext
ISGROUPS=*  
1-4 1  
5-7 2  
8 0 ; this line is optional, 0 is the standard.  
*  
```

When XWIDE=2, use two columns for each ISGROUPS= code. Each ISGROUPS= code must be one character, a letter or number, specified once in the two columns, e.g. "1" or "1" mean "1", and "0" or "0" mean "0".

**Example 4:** You wish most items on the "Liking for Science" Survey to share the same rating scale, an Andrich "Rating Scale" model, (in Grouping A). Items about birds (1, 10, 21) are to share a separate response structure, another Andrich "Rating Scale" model, (in Grouping B). Items 5 (cans) and 18 (picnic) each has its own response structure, i.e., the "Masters' Partial Credit" model, (Grouping 0).

```plaintext
NI=25 number of items
XWIDE=2
CODES=000102 valid codes for all items
ISGROUPS=' B A A A 0 A A A A B A A A A A A A A A A A A A 0 A A B A A A A'  
```
item groupings - use only one letter/number codes.

or

ISGROUPS=*; XWIDE=2 is not a worry here, but use one letter/number codes.
1-25 A; most items in grouping A
1 B; item 1 transferred to grouping B
10 B
21 B
5 0; grouping 0 means item is by itself
18 0

Example 5: Four separate tests of patient performance (some on a 4-point rating scale, some on a 5-point rating scale) are to be Rasch-analyzed. All 500 persons were given all 4 tests. I analyzed each separately, to get an idea of good-fitting and bad-fitting items, etc. Now, I'd like to run all 4 tests together using a partial credit model. There is no problem running all four tests together. Put them all in one file, or use MFORMS=. If you intend every item of every test to have its own rating scale (i.e., a strict partial-credit model), use ISGROUPS=0. But if you intend items on test 1 to share the same rating scale, similarly test 2 etc. (i.e., a test-level partial-credit model), then specify ISGROUPS=11111112222233334444.... matching the grouping number indicators to the count of items in each test.

Example 6: Items are to be rescored in 3 different ways, and then the items are to be divided into 4 rating scale structures.

ISGROUPS=11112223332444 ; 4 RATING SCALE GROUPINGS
IREFER =AAAAABBCCBABBB ; 3 RECODINGS
CODES =01234 ; ORIGINAL CODES IN DATA
IVALUEA =01234 ; ORIGINAL CODING MAINTAINED - THIS LINE CAN BE OMITTED
IVALUEB =43210 ; CODING IS REVERSED
IVALUEC =*112* ; DICHTOMIZED WITH EXTREME CATEGORIES MARKED MISSING

Example 7: A five-item test.
Item 1 Dichotomy already scored 0-1; let's call this a "D" (for dichotomy) group item
Item 2 Dichotomy already scored 0-1; this is another "D" (for dichotomy) group item. Under the Rasch model, all dichotomies have the same response structure.
Item 3 Partial credit polytomy already scored 0-1-2; this is an "0" type item. "0" means "this item has its own response structure".
Item 4 Rated polytomy already scored 1-2-3-4; let's call this an "R" group item.
Item 5 Rated polytomy already scored 1-2-3-4 with the same rating scale as item 4, so this is another "R" group item.
CODES =01234 ; this is all possible valid codes in the data
ISGROUPS = DD0RR ; Winsteps detects from the data which are the responses for each item-group and what they mean.

97. GRPFROM location of ISGROUPS

Only use this if you have too many items to put conveniently on one line of the ISGROUPS= control variable.

Instructs where to find the ISGROUPS= information.

GRPFROM=N
ISGROUPS= is a control variable before &END (the standard).

GRPFROM=Y
ISGROUPS= information follows just after &END, but before the item names. It is formatted exactly like a data record. It is helpful to enter "ISGROUPS=", for reference, where the person name would go.

Example: An attitude survey of 10 items with 3 rating scale definitions. Items 1 through 3 on Rating Scale 1, items 4 through 6 on Rating Scale 2 and items 7 through 10 on Rating Scale 3. The ISGROUPS= information is formatted like a data record and entered after &END and before the item names. The responses are in columns 1-10, and the person-id in column 11 onwards.

NAME1=11 start of person-id
ITEM1=1 start of responses
NI=10  number of items
CODES=12345  valid responses
GRPFRM=Y  ISGROUPS= formatted like data
&END
1112223333  ISGROUPS= information
Item name 1  item names
| Item name 10
END NAMES
2213243223 John Smith  first data record
|

98.  **GUFILE (G0ZONE, G1ZONE) Guttmanized response file**

This writes out the response file edited to more closely match an ideal Guttman scalogram. It is in a format close to the original data file, with items and person in entry order.

Outlying 1's are converted to 0's according to G0ZONE=
Outlying 0's are converted to 1's according to G1ZONE=

This removes unlikely 1's in the G0ZONE (e.g., lucky guesses) and unlikely 0's in the G1ZONE (e.g. careless mistakes)

It is also useful for imputing theories about item hierarchies.

G0ZONE= sets the % of observed 0's, starting from the "difficult" side of the Guttman scalogram, among which all 1's are turned to 0's. (The item hierarchy is constructed with the current data, but can be enforced through anchoring.) Standard value is 50.

G1ZONE= sets the % of observed 1's, starting from the "easy" side of the Guttman scalogram, among which all 0's are turned to 1's. Standard value is 50.

Example:  GUFILE= guttman.txt
G0ZONE = 20%
G1ZONE = 40%

Original data (Guttman ordered)
11100110011001001010
becomes
11111110011001001000

The file format matches the input data file if both are in fixed-field format.
When GUFILE= is written with CSV=Y, comma-separated or CSV=T, tab-separated, the item responses precede the person label.

Example:  KCT.txt Guttmanized with fixed field format:
Richard M 111111100000000000
Tracie F 111111111000000000
Walter M 111111111000000000

KCT.txt Guttmanized with comma-separated, CSV=Y, format:
1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,Richard M
1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,Tracie F
1,1,1,1,1,1,1,1,1,0,0,1,0,0,0,0,0,0,Walter M

99.  **HEADER** display or suppress subtable headings

Subtables are usually displayed with two heading lines, showing Table number, Title, date, etc.
To display these (the standard), specify HEADER=YES.
To suppress these, specify HEADER=NO.

Example: In Table 7.2, the person misfits. Heading lines are not wanted between persons.
  HEADER=NO

100. **HIADJ** correction for top rating scale categories

The Rasch model models the measure corresponding to a top rating (or partial credit) scale category as infinite.
This is difficult to think about and impossible to plot. Consequently, graphically in Table 2.2 and numerically in
Table 3.1 a measure is reported corresponding to a top category. This is the measure corresponding to an
imaginary rating HIADJ= rating points below the top category. The corresponding instruction for the bottom
category is **LOWADJ=**.

Example: The standard spread in Table 2.2 is based on HIADJ=0.25. You wish the top category number to be
printed more to the right, further away from the other categories.
  HIADJ=0.1

101. **HLINES** heading lines in output files

To facilitate importing the **IFILE=** **FILE=** **SFILE=** and **XFILE=** files into spreadsheet and database programs, the
heading lines can be omitted from the output files.

**HLINES=Y** Include heading lines in the output files (the standard)
In **IFILE=** and **PFILE=**, specifying **HLINES=Y** also puts ";" at the start of missing, deleted and extreme
lines.
**HLINES=N** Omit heading lines.

Example: I want a tab-separated score-to-measure file, without the column headings:
  SCOREFILE=mysc.txt
  HLINES=NO
  CSV=TAB

<table>
<thead>
<tr>
<th></th>
<th>-6.46</th>
<th>1.83</th>
<th>.28</th>
<th>217</th>
<th>85</th>
<th>1</th>
<th>2.9</th>
<th>1</th>
<th>2.9</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-5.14</td>
<td>1.08</td>
<td>.81</td>
<td>278</td>
<td>50</td>
<td>0</td>
<td>.0</td>
<td>1</td>
<td>2.9</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>-4.22</td>
<td>.86</td>
<td>1.29</td>
<td>321</td>
<td>40</td>
<td>1</td>
<td>2.9</td>
<td>2</td>
<td>5.7</td>
<td>4</td>
</tr>
</tbody>
</table>

with column headings, **HLINES=**YES, the standard:

";" "KID" "SCORE FILE FOR"
"; TABLE OF SAMPLE NORMS (500/100) AND FREQUENCIES CORRESPONDING TO COMPLETE TEST"
"; SCORE" "MEASURE" "S.E." "INFO" "NORMED" "S.E." "FREQUENCY" "%" "CUM.FREQ." "%"
"PERCENTILE"
|   | -6.46 | 1.83 | .28 | 217 | 85 | 1 | 2.9 | 1 | 2.9 |

102. **IAFILE** item anchor file

*The IFILE= from one analysis can be used unedited as the item anchor file, IAFILE=, of another.*

The item parameter values (deltas) can be anchored (fixed) using IAFILE= . Anchoring facilitates equating test
forms and building item banks. The items common to two test forms, or in the item bank and also in the current
form, can be anchored at their other form or bank calibrations. Then the measures constructed from the current
data will be equated to the measures of the other form or bank. Other measures are estimated in the frame of
reference defined by the anchor values.

In order to anchor items, a data file must be created of the following form:
1. Use one line per item-to-be-anchored.
2. Type the sequence number of the item in the current analysis, a blank, and the measure-value at which to anchor the item (in logits if **USCALE=**1, or in your user-rescaled units otherwise).
Further values in each line are ignored. An IFILE= works well as an IAFILE=.

Anything after ";" is treated as a comment.

IAFILE = filename
  Item anchor information is in a file containing lines of format
  item entry number   anchor value
  item entry number   anchor value

IAFILE=*  
  Item anchor information is in the control file in the format
  IAFILE=*  
  item entry number   anchor value
  item entry number   anchor value
  *

IAFILE=$SnnEnn or IAFILE=$SnnWnn or @Field
  Item anchor information is in the item labels using the column selection rules. Blanks or non-numeric values indicate no anchor value.

Example 1: The third item is to be anchored at 1.5 logits, and the fourth at 2.3 logits.
  1. Create a file named, say, "ANC.FIL"
  2. Enter the line "3 1.5" into this file, which means "item 3 in this test is to be fixed at 1.5 logits".
  3. Enter a second line "4 2.3" into this file, which means "item 4 in this test is to be fixed at 2.3 logits".
  3. Specify, in the control file,
     IAFILE=ANC.FIL
     CONVERGE=L  ; only logit change is used for convergence
     LCONV=0.005  ; logit change too small to appear on any report.

or place directly in the control file:
  IAFILE=*  
  3 1.5  
  4 2.3  
  *  
  CONVERGE=L  ; only logit change is used for convergence
  LCONV=0.005  ; logit change too small to appear on any report.

or in with the item labels:
  IAFILE=$S10W4 ; location of anchor value in item label
  CONVERGE=L  ; only logit change is used for convergence
  LCONV=0.005  ; logit change too small to appear on any report.
  &END
  Zoo
  House 1.5 ; item label and anchor value
  Garden 2.3
  Park
  END LABELS

To check: "A" after the measure means "anchored"

Example 2: The calibrations from one run are to be used to anchor subsequent runs. The items have the same numbers in both runs. This is convenient for generating tables not previously requested.
  1. Perform the calibration run, say,
     C:>
     WINSTEPS SF.TXT SOMEQ.TXT IFILE=ANCHORS.SF TABLES=111
2. Perform the anchored runs, say,
   C:> WINSTEPS SF.TXT MOREO.TXT IAFILE=ANCHORS.SF TABLES=0001111
   C:> WINSTEPS SF.TXT CURVESO.TXT IAFILE=ANCHORS.SF CURVES=111

Example 3: Score-to-measure Table 20 is to be produced from known item and rating scale structure difficulties.
   Specify:
   IAFILE= ; the item anchor file
   SAFILE= ; the structure/step anchor file (if not dichotomies)
   CONVERGE=L ; only logit change is used for convergence
   LCONV=0.005 ; logit change too small to appear on any report.
   STBIAS=NO ; anchor values do not need estimation bias correction.
   The data file comprises two dummy data records, so that every item has a non extreme score, e.g.,
   For dichotomies:
   Record 1: 10101010101
   Record 2: 01010101010
   For a rating scale from 1 to 5:
   Record 1: 15151515151
   Record 2: 51515151515

103. IANCHQU anchor items interactively

Items to be anchored can be entered interactively by setting IANCHQ=Y. If you specify this, you are asked if you want to anchor any items. If you respond "yes", it will ask if you want to read these anchored items from a file; if you answer "yes" it will ask for the file name and process that file in the same manner as if IAFILE= had been specified. If you answer "no", you will be asked to enter the sequence number of each item to be anchored, one at a time, along with its logit (or user-rescaled by USCALE=, UMEAN=) value. When you are finished, enter a zero.

Example: You are doing a number of analyses, anchoring a few, but different, items each analysis. You don't want to create a lot of small anchor files, but rather just enter the numbers at the terminal, so specify:
   IANCHQ=Y
   CONVERGE=L ; only logit change is used for convergence
   LCONV=0.005 ; logit change too small to appear on any report.
   You want to anchor items 4 and 8.
   WINSTEPS asks you:
   DO YOU WANT TO ANCHOR ANY ITEMS?
   respond YES(Enter)
   DO YOU WISH TO READ THE ANCHORED ITEMS FROM A FILE?
   respond NO(Enter)
   INPUT ITEM TO ANCHOR (0 TO END):
   respond 4(Enter) (the first item to be anchored)
   INPUT VALUE AT WHICH TO ANCHOR ITEM:
   respond 1.45(Enter) (the first anchor value)
   INPUT ITEM TO ANCHOR (0 TO END): 8(Enter)
   INPUT VALUE AT WHICH TO ANCHOR ITEM:-0.23(Enter)
   INPUT ITEM TO ANCHOR (0 TO END): 0(Enter) (to end anchoring)

104. ICORFILE item residual correlation file

This writes out the Table of inter-item correlations which is the basis of the principal components analysis of residuals. Missing data: for these Winsteps substitutes their expectations when possible. For residuals and standardized residuals, these are 0. Persons with extreme scores (minimum possible or maximum possible): Winsteps drops these from the correlation computation. The reason for these choices is to make the principal components analysis of residuals as meaningful as possible.

   ICORFILE= file name

Example 1: Write out the Table of inter-item residual correlations. ICORFILE=file.txt - Then file.txt contains, for
Example 2: When ICORFILE= is selected on the Output Files menu or MATRIX=YES, the Data Format: Matrix option can be selected:

This produces:

```
1.0000  -.0451   .0447   .0095  ....
-.0451  1.0000  -.0448  -.2024  ....
.0447  -.0448  1.0000  -.0437  ....
......  ......  ......  ......  ....
```

105. **IDELETE**  item one-line item deletion

A one-line list of items to be deleted or reinstated can be conveniently specified with IDELETE=. This is designed to be used in the post-analysis Specification pull-down menu box.

The formats are:
- IDELETE= 3 ; an entry number: delete item 3
- IDELETE= 6 1 ; delete items 6 and 1
- IDELETE= 2-5 ; delete items 2, 3, 4, 5
- IDELETE= +3-10 ; delete all items, then reinstate items 3 to 10.
- IDELETE= 4-20 +8 ; delete items 4-20 then reinstate item 8
- IDELETE= 3,7,4,10 ; delete items 3, 7, 4, 10. Commas, blanks and tabs are separators. At the "Extra information" prompt, use commas.
- IDELETE= (blank) ; resets temporary item deletions

Example 1: After an analysis is completed, delete all items except for one subtest in order to produce a score-to-measure Table for the subtest.

In the Specifications pull-down box:
IDELETE = +11-26 ; the subtest is items 11-26

Screen displays: CURRENTLY REPORTABLE ITEMS = 16

In the Output Tables menu (or SCOREFILE=)
Table 20. Measures for all possible scores on items 11-26.

Example 2: 9 common items. 3 items on Form A. 4 items on Form B. Score-to-measure tables for the Forms.

For Form A: in the Specifications pull-down box:
IDELETE = 13-16 ; deletes Form B items

In the Output Tables menu:
Table 20. Measures for all possible scores on items in Form A.

For Form B: in the Specifications pull-down box:
IDELETE= ; to reset all deletions
then
IDELETE = 10-12 ; deletes Form A items

In the Output Tables menu:
Table 20. Measures for all possible scores on items in Form B.

106. **IDELQU** delete items interactively

Use this if you have one or two items to delete or will be running repeatedly with different deletion and selection patterns, otherwise use IDFILE=.
If your system is interactive, items to be deleted or selected can be entered interactively by setting IDELQU=Y. If you specify this, you will be asked if you want to delete any items. If you respond "yes", it will ask if you want to read these deleted items from a file; if you answer "yes" it will ask for the file name and process that file in the same manner as if IDFILE= had been specified. If you answer "no", you will be asked to enter the sequence number or numbers of items to be deleted or selected one line at a time, following the rules specified for IDFILE=. When you are finished, enter a zero.

Example: You are doing a number of analyses, deleting a few, but different, items each analysis. You don't want to create a lot of small delete files, but rather just enter the numbers directly into the program using:

\[
\begin{align*}
&NI=60 \\
&ITEM1=30 \\
&IDELQU=Y \\
&\&END
\end{align*}
\]

You want to delete items 23 and 50 through 59.

WINSTEPS asks you:

DO YOU WANT TO DELETE ANY ITEMS?
respond YES(Enter)

DO YOU WISH TO READ THE DELETED ITEMS FROM A FILE?
respond NO(Enter)

INPUT ITEM TO DELETE (0 TO END):
respond 23(Enter) (the first item to be deleted)

INPUT ITEM TO DELETE (0 TO END): 50-59(Enter)

INPUT ITEM TO DELETE (0 TO END): 0(Enter) (to end deletion)

If you make a mistake, it is simple to start again, reinstate all items with

INPUT ITEM TO DELETE (0 TO END): +1-999(Enter)
where 999 is the length of your test or more, and start selection again.

107. **IDFILE** item deletion file

Deletion or selection of items from a test for an analysis, but without removing the responses from your data file, is easily accomplished by creating a file in which each line contains the sequence number or numbers of items to be deleted or selected. Specify this file by means of the control variable, IDFILE=, or enter the deletion list in the control file using IDFILE=*. Your control file should include item labels for all items, including the ones you are deleting.

a) Delete an item: enter the item number. E.g., to delete item 5, enter

5

b) Delete a range of items: enter the starting and ending item number on the same line separated by a *blank or dash*. E.g., to delete items 13 through 24

13-24

or

13  24

c) Select an item for analysis: enter a plus sign then the number.

E.g., to select item 19 from a previously deleted range

+19

d) Select a range of items for analysis: enter a plus sign, the starting number, a blank or dash, then the ending number. E.g., to select items 17 through 22

+17-22

or

+17  22

e) If a + selection is the *first entry* in the deletion file, then all items are deleted before the first selection is undertaken, so that the items analyzed will be limited to those selected, e.g.,

if +10-20 is the only line in the item deletion file for a 250 item test, it means
1-250 ; delete all 250 items
+10-20 ; reinstate items 10 through 20.

1) You may specify an item deletion list in your control with
   IDFILE=* (List)
   e.g.,
     IDFILE=*  
        17 ; delete item 17
        2 ; delete item 2
     *

Example 1: You wish to delete the fifth and tenth through seventeenth items from an analysis, but then keep item fourteen.
1. Create a file named, say, ITEM.DEL
2. Enter into the file, the lines:
   5
   10-17
   +14
3. Specify, in the control file,
   NI=50
   ITEM1=63
   IDFILE=ITEM.DEL
   TABLES=1110111
   &END
   or, specify in the control file,
   NI=50
   ITEM1=63
   IDFILE=*  
         5
         10-17
         +14
         *
         TABLES=1110111
         &END

Example 2: The analyst wants to delete the most misfitting items reported in Table 10.
1. Set up a standard control file.
2. Specify
   IDFILE=*  
   *
3. Copy the target portion of Table 10.
4. Paste it between the "**"
5. Delete characters before the entry numbers.
6. Type ; after the entry numbers to make further numbers into comments.

TITLE = 'Example of item deletion list from Table 10'
IDFILE = *

Delete the border character before the entry number
; ENTRY RAW  INFIT  OUTFIT
; NUM SCORE  COUNT  MEASURE  ERROR MNSQ ZSTD MNSQ ZSTD PTBIS ACTS G
5 ; 2  4 .00  1.03  1.48  1.8 1.50  1.8 A-.83  FIND BOTTLES AND CANS  0
8 ; 2  4 .00  1.03  1.40  1.6 1.43  1.6 B-.71  LOOK IN SIDEWALK CRACKS  0
4 ; 3  4 .00  1.62  1.33  .7 1.49  .9  C-.21  WATCH GRASS CHANGE  0
9 ; 4  4 .00  .74  1.51  .8 1.57  .9  D-.59  LEARN WEED NAMES  0
20 ; 1  4 .00  1.03  1.12  .5 1.14  .6  E-.05  WATCH BUGS  0
24 ; 6  4 .30  1.03  1.15  .6 1.13  .5  F-.15  FIND OUT WHAT FLOWERS LIVE ON  0

Enter the ; to make details to right of entry numbers into comments
Example 3: The analyst wants to delete item 4 and items 18 to 23 on the DOS control (or Extra Specifications) line:

Extra specifications? IDFILE=* 4 18-23 * (Enter)

or

C:>WINSTEPS CONTROL.FIL OUTPUT.FIL  IDFILE=* 4 18-23 *

108. **IDROPEXTREME drop items with extreme scores**

Unanchored items with extreme (zero or perfect, minimum possible or maximum possible) scores provide no information for estimating person measures, but they are reported and included in summary statistics. To remove them:

IDROPEXTREME = No ; do not drop extreme items (standard)

IDROPEXTREME = Yes or All ; drop items with zero and perfect scores

IDROPEXTREME = Zero or Low or Bottom or Minimum ; drop items with zero or minimum-possible scores

IDROPEXTREME = Perfect or High or Top or Maximum ; drop items with perfect or maximum-possible scores

Example: The instrument contains items asking about very rare conditions (scored "0" - not observed). These are skewing the survey summary statistics:

IDROPEXTREME = Minimum ; items about conditions never observed in the sample are dropped.

109. **IFILE item output file**

IFILE=filename produces an output file containing the information for each item. This file contains 4 heading lines (unless HLINES=N), followed by one line for each item containing:

Columns:

<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>Format</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>A1</td>
<td>Blank or &quot;;&quot; if HLINES=Y and there are no responses or deleted or extreme (status =0,-1, -2, -3)</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>I5</td>
<td>1. The item sequence number (ENTRY)</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>F8.2</td>
<td>2. Item's calibration (user-rescaled by UMEAN=, USCALE=, UDECIM) (MEASURE)</td>
</tr>
<tr>
<td>15</td>
<td>17</td>
<td>I3</td>
<td>3. The item's status (STATUS)</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>I3</td>
<td>3. The item's status (STATUS)</td>
</tr>
<tr>
<td>18</td>
<td>25</td>
<td>F8.1</td>
<td>4. The number of responses used in calibrating (COUNT) or the observed count (TOTAL=Y)</td>
</tr>
<tr>
<td>26</td>
<td>34</td>
<td>F8.1</td>
<td>5. The raw score used in calibrating (SCORE) or the observed score (TOTAL=Y)</td>
</tr>
<tr>
<td>35</td>
<td>41</td>
<td>F7.2</td>
<td>6. Item calibration's standard error (user-rescaled by USCALE=, UDECIM=) (ERROR)</td>
</tr>
<tr>
<td>42</td>
<td>48</td>
<td>F7.2</td>
<td>7. Item infit: mean square infit (IN.MSQ)</td>
</tr>
<tr>
<td>49</td>
<td>55</td>
<td>F7.2</td>
<td>8. Item infit: t standardized (ZSTD), locally t standardized (ZEMP) or log-scaled (LOG)</td>
</tr>
<tr>
<td>56</td>
<td>62</td>
<td>F7.2</td>
<td>9. Item outfit: mean square outfit (OUT.MS)</td>
</tr>
<tr>
<td>63</td>
<td>69</td>
<td>F7.2</td>
<td>10. Item outfit: standardized (ZSTD), locally t standardized (ZEMP) or log-scaled (LOG)</td>
</tr>
<tr>
<td>70</td>
<td>76</td>
<td>F7.2</td>
<td>11. Item displacement (user-rescaled by USCALE=, UDECIM=) (DISPLACE)</td>
</tr>
<tr>
<td>77</td>
<td>83</td>
<td>F7.2</td>
<td>12. Item by test-score correlation: point-biserial (PTBS) or point-measure (PTME)</td>
</tr>
<tr>
<td>84</td>
<td>90</td>
<td>F7.2</td>
<td>13. Item weight (WEIGHT)</td>
</tr>
<tr>
<td>91</td>
<td>96</td>
<td>F6.1</td>
<td>14. Observed percent of observations matching prediction (OBSMA)</td>
</tr>
<tr>
<td>97</td>
<td>102</td>
<td>F6.1</td>
<td>15. Expected percent of observations matching prediction (EXPMA)</td>
</tr>
<tr>
<td>103</td>
<td>109</td>
<td>F7.2</td>
<td>16. Item discrimination</td>
</tr>
</tbody>
</table>

If ASYMPTOTE=Yes:

110 115 F6.2 17. Item lower asymptote (LOWER)
If PVALUE=Yes:

Item upper asymptote (UPPER)  

... add 12 to the next column locations

Item p-value or average rating (PVALUE)  

... add 6 to the next column locations

The format descriptors are:

In = Integer field width n columns
Fn.m = Numeric field, n columns wide including n-m-1 integral places, a decimal point and m decimal places
An = Alphabetic field, n columns wide
nX = n blank columns.

When CSV=Y, commas separate the values, which are squeezed together without spaces between. Quotation marks surround the "Item name", e.g., 1,2,3,4,"Name". When CSV=T, the commas are replaced by tab characters.

Example:  You wish to write a file on disk called "ITEM.CAL" containing the item statistics for use in updating your item bank, with values separated by commas:

IFILE=ITEM.CAL
CSV=Y

When W300=Yes, then this is produced in Winsteps 3.00, 1/1/2000, format:

Columns:
Start  End  Format  Description
1   1   A1  Blank or ";" if HLINES=Y and there are no responses or deleted (status = -2, -3)
2   6   I5  1. The item sequence number (ENTRY)
7  14   F8.2  2. Item's calibration (user-rescaled by UMEAN=, USCALE=, UDECIM) (MEASURE)
15  17   I3  3. The item's status (STATUS)
18  23   I6  4. The number of responses used in calibrating (COUNT) or the observed count (TOTAL=Y)
24  30   I6  5. The raw score used in calibrating (SCORE) or the observed score (TOTAL=Y)
31  37   F7.2  6. Item calibration’s standard error (user-rescaled by USCALE=, UDECIM=) (ERROR)
38  44   F7.2  7. Item mean square infit (IN.MSQ)
45  51   F7.2  8. Item infit: t standardized (ZSTD), locally t standardized (ZEMP) or log-scaled (LOG)
52  58   F7.2  9. Item mean square outfit (OUT.MS)
59  65   F7.2  10. Item outfit: t standardized (ZSTD), locally t standardized (ZEMP) or log-scaled (LOG)
66  72   F7.2  11. Item displacement (user-rescaled by USCALE=, UDECIM=) (DISPLACE)
73  79   F7.2  12. Item by test-score correlation: point-biserial (PTBS) or point-measure (PTME)
80  80   1X  15. Blank
81  81   A1  16. Grouping to which item belongs (G)
82  82   1X  17. Blank
83  83   A1  18. Model used for analysis (R=Rating, S=Success, F=Failure) (M)
84  84   1X  19. Blank
85 132+   A30+  18. Item name (NAME)
ILFILE  item label file

Useful item identification greatly enhances the output tables.

You usually specify items labels between &END and END LABELS in the control file. You may also use ILFILE= for the initial or additional sets of item labels.

ILFILE= commands also add 'Edit ILFILE=' lines to the File pull-down menu.

Example: You wish to use abbreviated item labels for the initial analysis, but then switch to longer item labels to produce the final reports.

In your control file specify the shorter labels, one per line,
(a) between &END and END LABELS
or (b) between ILFILE=* and * in the control file
or (c) in a separate file identified by ILFILE=*.

You can switch to the longer labels, in a file called "Longer.txt" by using the "Specification" menu item, and entering ILFILE=Longer.txt.

If you have ILFILE= in your control file and your data is also in your control file, be sure that there is an "END LABELS" before your data (or that you specify INUMB=YES).

Example: 4 item arithmetic test.

\[ \begin{align*}
\text{NI} & = 4 \\
\text{ILFILE} & = * \\
\text{Addition} & ; \text{ labels for the 4 items} \\
\text{Subtraction} & \\
\text{Multiplication} & \\
\text{Division} & *
\end{align*} \]

&End
END LABELS

IMAP  item label on item maps Tables 1, 12

This specifies what part of the item label is to be used on the item map. The length of IMAP= overrides NAMLMP=.

It’s format is IMAP = $S..W.. or $S..E. etc. using the column selection rules.

Example: Item type is in columns 3 and 4 of the item label. Item content area is in columns 7 and 8.

\[ \begin{align*}
\text{IMAP} & = \$S3W2+$/+$S7W2 \\
\text{title} & = *
\end{align*} \]

12 ; Item maps in Table 12 (or use Output Tables menu)

* If the item label is "KH323MXTR", the item label on the map will be "32/XT"

INUMB  label items by sequence numbers

Are item names provided, or are they the entry sequence numbers?
INUMB=Y
a name is given to each item based on its sequence number in your data records. The names are "I0001", "I0002", ..., and so on for the NI= items. This is a poor choice as it produces noninformative output.

INUMB=N, the standard
Your item names are entered (by you) after the "&END" at the end of the control variables. Entering detailed item names makes your output much more meaningful to you.

The rules for supplying your own item names are:
1. Item names are entered, one per line, generally directly after &END.
2. Item names begin in column 1.
3. Up to 300 characters (or ITLEN=) in each item name line will be used, but you may enter longer names in the control file for your own record keeping.
4. The item names must be listed in exactly the same order as their responses appear in your data records.
5. There should be the same number of item names as there are items specified by NI=. If there are too many or too few names, a message will warn you and sequence numbers will be used for the names of any unnamed items. You can still proceed with the analysis.
6. Type END NAMES starting in column 1 of the line after the last item name.

Example: An analysis of 4 items for which you supply identifying labels.

; these lines can start at any column
NI=4  four items
ITEM1=10  responses start in column 10
INUMB=N  item names supplied (the standard)
&END
My first item name ; must start at column 1.
My second item label
My third item identifier
My fourth and last item name
END NAMES ; must start at column 1, in capital letters
Person A 1100 data records
Person Z 1001

113. IPMATRIX response-level matrix

IPMATRIX= is only available from the Output Files pull-down menu. It constructs a rectangular matrix in which there is one row for each person and one column for each item, or vice-versa. The entries in the matrix are selected from the following screen:

The first rows and columns can be the entry numbers, measures, labels and/or fields from the labels.

The matrix must contain one of
3. Original response value (after keying/scoring) (I4) (OBS)
4. Observed response value (after recounting) (I4) (ORD)
5. Expected response value (F7.3) (EXPECT)
6. modeled variance of observed values around the expected value (F7.3) (VAR)
   This is also the statistical information in the observation.
   Square root(modeled variance) is the observation's raw score standard deviation.
7. Standardized residual: (Observed - Expected)/Square root Variance (F7.3) (ZSCORE)
8. Score residual: (Observed - Expected) (F7.3) (RESID)
11. Measure difference (Person measure - Item measure) (F7.3) (MEASURE)
12. Log-Probability of observed response (F7.3) (LOGe(PROB))
13. Predicted person measure from this response (F7.3) (PMEASURE)
14. Predicted item measure from this response (F7.3) (IMEASURE)
15. Response code in data file (A) (CODE)

Field numbers shown here are those for XFILE=.

Depending on CSV=, data values are separated by "tab" or comma characters. In fixed field format, all fields are 7 characters long, separated by a blank. Missing data codes are "." as standard, but can be any character, or nothing.

Example: I want a table of probabilities with items as the columns and possible scores as the rows, like the one on page 166 of Doug Cizek's book, Setting Performance Standards, based on work by Mark Reckase.

From your main analysis, write out an IFILE=itemanc.txt
Create a data set with one record for each possible score (it doesn't matter what the actual pattern of 1's and 0's is).
Enter the intended raw score as the person label.
In the control file for the new data set, put IAFILE=itemanc.txt
Analyze this second data set. Ignore any "subset" warning messages.
Check that the items are anchored in Table 14.
Check that the reported raw score for each person match that in the person label in Table 18.

Here it is from Exam1.txt (using the non-extreme items):

TITLE='KNOX CUBE TEST' ; Report title
NAME1=1 ; First column of person label in data file
ITEM1=11 ; First column of responses in data file
NI=18 ; Number of items
CODES=01 ; Valid response codes in the data file
iafile = itemanc.txt ; item calibrations from the original analysis
&END

1 10000000000000
2 11000000000000
3 11100000000000
4 11110000000000
5 11111000000000
6 11111100000000
7 11111110000000
8 11111111000000
9 11111111100000
10 11111111110000
11 11111111111000
12 11111111111100
13 11111111111110

Here's the array in EXCEL, after some editing. Add across the columns to confirm that the probabilities add up to the scores.
114. IREFER identifying items for recoding

Responses are revalued according to the matching codes in IVALUE=. If this implies that the items may have different rating (or partial credit) scale structures, so ISGROUPS= may also be required.

IREFER= has three forms: IREFER=AABBCDAAD and IREFER=* list * and IREFER=*filename

Valid one-character IREFER= codes include: !#$%&-/123456789<>@ABCDEFGHIJKLMNOPQRSTUVWXYZ^_|~

A-Z are the same as a-z.

Characters with ASCII codes from 129-255 can also be used, but display peculiarly: ÌÎÐÑÒÓÔÕÖרÙÚÛÜÝÞßàáâã etc.

When XWIDE=2 or more, then either (a) Use one character per XWIDE and blanks,

NI=8
XWIDE=2
IREFER=' A B C D D C B A'

or (b) Use one character per item with no blanks

NI=8
XWIDE=2
RESCORE='ABCDDCBA'

Item identifying codes can be letters or numbers. "A" is the same as "a", etc.

Example 1. There are 3 item types. Items are to rescored according to Type A and Type B. Other items to keep original scoring.

CODES = 1234
IREFER = AAAAAAAABBBBbbbb***** 3 item types: ("a" is the same as "A" in these codes)
IVALUEA = 1223  Recode Type A items
IVALUEB = 1123  Recode Type B items
IVALUE* = 1234  Recode Type * item. Can be omitted

or

IREFER=*
1-8  A
9-16  B
17-23 *
*

or

IREFER=*filename.txt

in filename.txt:
1-8  A
9-16  B
17-23 *
Example 2. There are 3 item types. Responses are 2 characters wide. Items are to be rescored according to Type A and Type B. Other items to keep original scoring.

XWIDE=2
CODES = '1 2 3 4 '  
IREFER = AAAAAAABBBBBB** 3 item types  
IVALUEA = '1 2 2 3 '  Recode Type A items  
IVALUEB = '1 1 2 3 '  Recode Type B items  
IVALUE* = 1234  Recode Type * item. Can be omitted  

Example 3: All items are to be rescored the same way
NI = 100  100 ITEMS  
IREFER=*
1-100  X  FOR ALL 100 ITEMS, reference is X  
*  
Codes  = 12345  rescore 12345  
IVALUEx  = 12223  into 12223  

Example 4: Items are to be rescored in 3 different ways, and then the items are to be divided into 4 rating scale structures.

ISGROUPS=1111222332444  ; 4 RATING SCALE GROUPINGS  
IREFER  =AAAAABBBCCABBBB  ; 3 RECODINGS  
CODES  =01234  ; ORIGINAL CODES IN DATA  
IVALUEA =01234  ; ORIGINAL CODING MAINTAINED - THIS LINE CAN BE OMITTED  
IVALUEx =43210  ; CODING IS REVERSED  
IVALUEC =*112*  ; DICHOTOMIZED WITH EXTREME CATEGORIES MARKED MISSING  

Example 5: Multiple-choice test with 4 options, ABCD

IREFER=ABCDDABADCD  ; SCORING KEY  
CODES  =ABCD  ; VALID OPTIONS  
IVALUEA=1000  ; A SCORED 1  
IVALUEx =0100  ; B SCORED 1  
IVALUEC=0010  ; C SCORED 1  
IVALUEx =0001  ; D SCORED 1  
MISSCORVE=0  ; EVERYTHING ELSE IN THE DATA SCORED 0  

115.  ISELECT  item selection criterion

Items to be selected may be specified by using the ISELECT= instruction to match characters within the item name. Items deleted by IDFILE= or similar are never selected by ISELECT=.

This can be done before analysis in the control file or with "Extra specifications". It can also be done after the analysis using the "Specification" pull-down menu.

Control characters to match item name:
?  matches any character
{..}  braces characters which can match a single character:  {ABC} matches A or B or C.
{.. - ..}  matches single characters in a range.  {0-9} matches digits in the range 0 to 9.
{.. --..}  matches a single "-"  {AB--} matches A or B or "-".
*  matches any string of characters - must be last selection character.
Other alphanumeric characters match only those characters.

Each ISELECT= performed using the "Specification" pull-down menu selects from all those analyzed. For incremental selections, i.e., selecting from those already selected, specify +ISELECT=

Example 1: Select for analysis only items with M in the 5th column of item name.

ISELECT=?????M*  M in column  means Math items  
0001M 2x4  selected  
0002R the cat  omitted  
END NAMES

Example 2: Select for analysis all items with code "A 4" in columns 2-4 of their names.
ISELECT="?A 4*" quotes because a blank is included. A is in col. 2 etc.

**Example 3:** Select all Math (M in column 2) items which are Addition or Division (A or D in column 4):

ISELECT="?M?[AD]*"

<table>
<thead>
<tr>
<th>ISELECT</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1M3A456</td>
<td>2+2 selected</td>
</tr>
<tr>
<td>1M5D689</td>
<td>23/7 selected</td>
</tr>
<tr>
<td>1H2A123 George</td>
<td>omitted (History, American)</td>
</tr>
</tbody>
</table>

**Example 3:** Select codes A, 1,2,3,4,5,6 in column 3:

ISELECT=??(A1-6)*

**Example 4:** Select "- " in columns 2 and 3:

ISELECT="?- "

**Example 5:** Select "-" or "x" in column 2 with " " in column 3:

ISELECT="?(-x)"

**Example 6:** Analyze only math (column 4 or person-id). Then report only items in Topic C (column 1). Then only report items in Strands 4 and 5 (column 2) in Topic C.

ISELECT=??M* in the Control file or at the Extra Specifications prompt.

ISELECT=C* using the Specification pull-down menu, after the analysis

+ISELECT=?(45)*

### ISFILE item structure output file

**Do not use this file for anchoring.** ISFILE=filename produces an output file containing the category structure measure information for each item. All measures are added to the corresponding item's calibration and rescaled by USCALE= and UDECIMALS=. This file contains 4 heading lines (unless HLINES=N), followed by one line for each item containing:

| Columns: |
|----------|-----------------------------------------------------|
| Start    | End | Format | Description |
| 1        | 1   | A1     | Blank or ";" if no responses or deleted (status = -2, -3) |
| 2        | 6   | I5     | 1. The item sequence number (ENTRY) |
| 7        | 11  | I5     | 2. The item's status (STATUS) |
| 12       | 5   | I5     | 3. Number of active categories (MAXIMUM) |
| 17       | 5   | I5     | 4. Lowest active category number (CAT) |
| 22       | 29  | F8.2   | 5. Measure for an expected score of LOWADJ= (E.G., CAT+.25) |
| 30       | 34  | I5     | 6. Active category number (CAT) |
| 35       | 39  | I5     | 7. Ordered category number in structure (STRUCTURE) = "Step counting from zero" |
| 40       | 47  | F8.2   | 8. Structure measure (MEASURE) = Rasch-Andrich threshold + item measure = Dij, the Rasch-Andrich threshold. The number of decimal places is set by UDECIMAL= |
| 48       | 55  | F8.2   | 9. Rasch-Andrich threshold's standard error (ERROR) - reported if only one item in the ISGROUPS= | **Do not use for anchoring. Use the SFILE= and IFILE=** |
| 56       | 63  | F8.2   | 10. Measure for an expected score of category — 0.5 score points (CAT-0.5). This is the Rasch-half-point threshold, the boundary between categories when conceptualized as average performances. It is not a model parameter. |
| 64       | 71  | F8.2   | 11. Measure for an expected score of category score points (AT CAT). This is the measure |
corresponding to a category when predicting for an individual or sample about which nothing else is known. For the top category this value corresponds to the top category value less \( \text{HIADJ}= \) (e.g., \( \text{CAT}-0.25 \)), the measure for an expected score of HIADJ= score points less than the top category value.

72 79 F8.2 12. Measure at the 50% cumulative probability (50%PRB). This is the Rasch-Thurstone threshold.

The "AT CAT" for the extreme categories are actually at -infinity and +infinity. So the value \( \text{LOWADJ}= \) above minimum score on the item substitutes for -infinity. For the extreme high category, the value \( \text{HIADJ}= \) less than the maximum score for that item substitutes for +infinity. It is these values that are plotted on Table 2.2 for the extreme categories.

Since the ISFILE= has the same number of category entries for every item, the repeated fields are filled out with "0" for any further categories up to the maximum categories for any item.

The format descriptors are:

- \( \text{In} = \) Integer field width \( n \) columns
- \( \text{Fn.m} = \) Numeric field, \( n \) columns wide including \( n-m-1 \) integral places, a decimal point and \( m \) decimal places
- \( \text{An} = \) Alphabetic field, \( n \) columns wide

When \( \text{CSV}= \)Y, commas separate the values with quotation marks around the "Item name". When \( \text{CSV}= \)T, the commas are replaced by tab characters.

When \( \text{STKEEP}= \)YES and there are intermediate null categories, i.e., with no observations, then the structure calibration into the category is set 40 logits above the previous calibration. The structure calibration out of the category, and into the next category, is set 40 logits above. Thus:

<table>
<thead>
<tr>
<th>Category</th>
<th>structure Calibraton</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>NULL 0.00</td>
</tr>
<tr>
<td>1</td>
<td>-1.00 -1.00</td>
</tr>
<tr>
<td>2</td>
<td>NULL 39.00</td>
</tr>
<tr>
<td>3</td>
<td>1.00 -38.00</td>
</tr>
<tr>
<td>TOTAL:</td>
<td>0.00 0.00</td>
</tr>
</tbody>
</table>

**Meanings of the columns**

There are several ways of conceptualizing the category boundaries or thresholds of a rating (or partial credit) scale item. Imagine a rating (or partial credit) scale with categories, 1, 2, 3:

From the "expected score ogive", also called the "model item characteristic curve"

- Average rating: Measure (must be ordered)
- 1.25 Measure for an expected score of .25 (CAT+.25) when \( \text{LOWADJ}=0.25 \)
- 1.5 Measure for an expected score of category — .5 score points (CAT-0.5)
- 2 Measure for an expected score of category score points (AT CAT)
- 2.5 Measure for an expected score of category — .5 score points (CAT-0.5)
- 2.75 Measure for an expected score of category score points (AT CAT)

since this is the top extreme category the reported values is for CAT-0.25 when HIADJ=0.25

From the "category probability curves" relative to the origin of the measurement framework (need not be ordered)

- 1-2 equal probability Structure measure = Rasch-Andrich threshold + item measure = \( \text{Dij (MEASURE)} \)
  standard error Rasch-Andrich threshold's standard error (ERROR)
- 2 maximum probability Measure for an expected score of category score points (AT CAT) - (yes, same as for the ogive)
- 2-3 equal probability Structure measure = Rasch-Andrich threshold + item measure = \( \text{Dij (MEASURE)} \)
  standard error Rasch-Andrich threshold's standard error (ERROR)

From the "cumulative probability curves" (preferred by Thurstone) (must be ordered)

Category 1 at .5 probability Measure at the 50% cumulative probability (50%PRB)
Category 1+2 at .5 probability  Measure at the 50% cumulative probability (50%PRB)

Example 1: You wish to write a file on disk called "ITEMST.FIL" containing the item statistics reported in Table 2.2, for use in constructing your own tables:

```
ISFILE = ITEMST.FIL
ISGROUPS = 0  ; each item has its own "partial credit" scale
LOWADJ = 0.25  ; the standard for the low end of the rating scale
HIADJ = 0.25   ; the standard for the high end of the rating scale
```

<table>
<thead>
<tr>
<th>ENTRY</th>
<th>STAT</th>
<th>MAX</th>
<th>CAT</th>
<th>CAT+.25</th>
<th>CAT</th>
<th>STRU</th>
<th>MEASURE</th>
<th>ERROR</th>
<th>CAT-0.5</th>
<th>CAT-.25</th>
<th>50%PRB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>-2.47</td>
<td>1</td>
<td>1</td>
<td>-1.25</td>
<td>.00</td>
<td>-1.58</td>
<td>-.40</td>
<td>-1.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.46</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>-2.78</td>
<td>1</td>
<td>1</td>
<td>-1.57</td>
<td>.00</td>
<td>-1.89</td>
<td>-.71</td>
<td>-1.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.15</td>
</tr>
</tbody>
</table>

*ENTRY* is the item entry number

*STAT* is the item status, see *IFILE=

*MAX* is the highest category

*CAT* is the current category

*CAT+0.25* is the measure corresponding to an expected score of the lowest category+0.25 score points on the item

*STRU* (structure calibration) or step measure is a Rasch model parameter estimate (Rasch-Andrich thresholds), also the point at which adjacent categories are equally probable. See "Category probability curves" graph.

*MEASURE* is the item difficulty + structure calibration.

*ERROR* is an estimate of the standard error. It is reported as .00 if it is not known.

*CAT-0.5* is the location where the expected score on the item is the category half-point value, e.g., for a scale for 0,1,2 the "CAT-0.5" values correspond to expected scores of 0.5 and 1.5. See the "Expected score ICC" graph.

*AT CAT* is the location where the expected score on the item is the category point value, e.g., for a scale for 0,1,2 the "AT step" values correspond to expected scores of 0.25, 1, and 1.75. Since the "at step" values 0 and 2 are infinite they are reported for 0.25 and 1.75. See the "Expected score ICC" graph.

*50%PRB* is the location of the Rasch-Thurstone threshold, the point at which the probability of all categories below = the probability of all categories at or above. See the "Cumulative probabilities" graph.

Example 2: To produce a Table of expected measures per item-category similar to Pesudovs, K., E. Garamendi, et al. (2004). "The quality of life impact of refractive correction (QIRC) questionnaire: Development and validation." Optometry and Vision Science 81(10): 769-777, write the ISFILE= to Excel. Then delete or hide unwanted columns.

```
ISFILE = EFILE.XLS
```

117. **ISORT column within item name for alphabetical sort in Table 15**

Table 15 lists items alphabetically. Table 1 and Table 12 list them alphabetically within lines. As standard, the
whole item name is used. Select the sorting columns in the item labels with ISORT= using the column selection rules, e.g., starting in column Snn and ending in column Enn or of width Wnn.

Example 1: The item name is entered in the specification file as sequence number, followed by identification in column 6. Sort by identification for Table 15.
NI=4
TABLES=1111111111111111111111111
ISORT=5-30 ; sort the items reported in Table 15 by item descriptions
&END
0001 Addition Item
0002 Subtraction Item
0003 Multiplication item
0004 Division item
  sort column
END NAMES

Example 2: The item name contains several important classifiers. Table 15 is required for each one:
TFILE=* 
  15 - - - 1  sort starts with column 1 of item name
  15 - - - 6  sort starts with column 6
  15 - - - 13 sort starts with column 13 of the item name and goes to the end of the item name
    - entered as place-holders, see TFILE=
    *
&END
MCQU Geogr 1995-0234
  sort column
    sort column
    sort column
| END NAMES

Example 3: A version of Table 15, sorted on item name column 13, is to be specified on the DOS command line or on the Extra Specifications line. Commas are used as separators, and "-" as place-holders:
TFILE=*  15,-,-,-,13  *

118. ISUBTOTAL columns within item label for subtotals in Table 27

This specifies what part of the data record is to be used to classify items for subtotal in Table 27.

$S..W.. e.g., $S2W13 means that the label to be shown on the map starts in column 2 of the item label and is 13 columns wide.
$S..E.. e.g., $S3E6 means that the label to be shown on the map starts in column 3 of the item label and ends in column 6.

These can be combined, and constants introduced, e.g,
ISUBTOTAL=$S3W2+"\"+$S7W2

If the item label is "KH323MXTR", the subgrouping will be shown as "32/XT"

Format 2: ISUBTOTAL=* This is followed by a list of subgroupings, each on a new line using the column selection rules:

ISUBTOTAL=* 
  $S1W1+$S7W2 ; Subtotals reported for item classifications according to these columns
  $S3E5 ; Subtotals reported for item classifications according to these columns

Example: Subtotal by first letter of item name:

```
ISUBTOTAL=$S1W1
TFILE=*  ; produce the subtotal report
```

Here is a subtotal report (Tables 27) for items beginning with "R"

"R" SUBTOTAL FOR 8 NON-EXTREME ITEMS

<table>
<thead>
<tr>
<th>RAW</th>
<th>MODEL</th>
<th>INFIT</th>
<th>OUTFIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCORE</td>
<td>COUNT</td>
<td>MEASURE</td>
<td>ERROR</td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>MEAN</td>
<td>28.1</td>
<td>25.0</td>
<td>4.04</td>
</tr>
<tr>
<td>S.D.</td>
<td>5.5</td>
<td>.0</td>
<td>6.63</td>
</tr>
<tr>
<td>MAX.</td>
<td>38.0</td>
<td>25.0</td>
<td>16.30</td>
</tr>
<tr>
<td>MIN.</td>
<td>19.0</td>
<td>25.0</td>
<td>-6.69</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>REAL RMSE</th>
<th>ADJ.SD</th>
<th>SEPARATION</th>
<th>PUPIL RELIABILITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.63</td>
<td>5.54</td>
<td>1.52</td>
<td>.70</td>
</tr>
<tr>
<td>MODEL RMSE</td>
<td>ADJ.SD</td>
<td>SEPARATION</td>
<td>PUPIL RELIABILITY</td>
</tr>
<tr>
<td>3.48</td>
<td>5.64</td>
<td>1.62</td>
<td>.72</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>S.E. OF PUPIL MEAN = 2.50</th>
</tr>
</thead>
</table>

MAXIMUM EXTREME SCORE: 1 PUPILS
MINIMUM EXTREME SCORE: 1 PUPILS
LACKING RESPONSES: 1 PUPILS
DELETED: 1 PUPILS

119. **ITEM** title for item labels

Up to 12 characters to use in table headings to describe the kind of items, e.g.
ITEM=MCQ.

Choose a word which makes its plural with an "s", e.g. MCQS, since an S is added to whatever you specify. If you say ITEM=mcq, then the plural will be "mcqs".

120. **ITEM1** column number of first response

Specifies the column position where the response-string begins in your data file record, or the column where the response-string begins in the new record formatted by FORMAT=.

If you have the choice, put the person-identifiers first in each record, and then the item responses with each response taking one column.

**Error messages** regarding **ITEM1=** may be because your control file is not in "Text with line breaks" format.

It is easy to miscount the **ITEM1=** column. Scroll to the top of the Winsteps screen and check column positions:

```
Input in process..
Input Data Record:
  1  2
123456789012345678901234567
Richard M 111111000000000000
^P ^I ^N
35 KID  Records Input.
```

^P marks the Name1=1 column position with ^.
^I marks the Item1=11 column position with ^.
^N marks the NI=18 column position with ^.

Example 1: The string of 56 items is contained in columns 20 to 75 of your data file.

| ITEM1=20 | response to item 1 in column 20 |
| NI=56    | for 56 items                     |
| XWIDE=1  | one column wide (the standard)   |

Example 2: The string of 100 items is contained in columns 30 to 69 of the first record, 11 to 70 of the second record, followed by 10 character person i.d.

| XWIDE=1  | one column wide (the standard)   |
| FORMAT=(T30,40A,/,T11,60A,10A) | two records per person |
| ITEM1=1  | item 1 in column 1 of reformatted record |
| NI=100   | for 100 items                     |
| NAME1=101| person id starts in column 101   |
| NAMLEN=10| person id starts in 10 columns wide |

121. ITLEN  maximum length of item label

ITLEN= specifies the maximum number of columns in the control file that are to be used as item names. The maximum possible is 300 characters.

Example 1: You only wish to use the first five columns of item information to identify the items on the output:

| NI=4     |
| ITLEN=5  |
| AX123    | This part is not shown on the output |
| BY246    | Trial item |
| AZ476    | This item may be biased |
| ZZ234    | Hard item at end of test |
| END NAMES |

Example 2: Your item names may be up to 50 characters long:

| NI=4     |
| ITLEN=50 |
| END      |

122. IVALUEx recoding of data

Responses are revalued according to the matching codes in IREFER= (or ISGROUPS= if IREFER= is omitted). Items in IREFER= not referenced by an IVALUEx= are not recoded.

IVALUEa= is the same as IVALUEA=

The recoded values in IVALUEx= line up vertically with the response codes in CODES=, if a data value does not match any value in CODES= it will be treated as missing.

Valid one-character IVALUE= codes include: !#$%&-./123456789<>@ABCDEFGHIJKLMNOPQRSTUVWXYZ^_|~

A-Z are the same as a-z.

Characters with ASCII codes from 129-255 can also be used, but display peculiarly: ÛÜÖÖÔØ×ØÜÜÜÝ¥§âââ etc.

When XWIDE=2 or more, then either (a) Use one character per XWIDE and blanks, 

| NI=8     |
or (b) Use one character per item with no blanks

```
NI=8
XWIDE=2
RESCORE='ABCDABCD'
```

Layout is:

```
NI = 17
IREFER = AABCADBCDEDEEAABC ; the recoding type designators for the 17 items ; see the vertical line up here
CODES = 0123456 ; valid codes across all items
IVALUEA = 012*** ; recodes for Grouping A
IVALUEB = *1224** ; recodes for Grouping B: "2" and "3" recoded to "2"
IVALUEC = *122226 ; 1-2-6 acts as 1-2-3 because STKEEP=NO
IVALUED = 01233*
IVALUEE = 00123**
STKEEP=YES ; missing intermediate codes are squeezed out
```

Example 1: Items identified by Y and Z in IREFER= are to be recoded.
Y-type items are 1-3, 7-8. Z-type items are 4-6, 9-10. All items have their own rating (or partial credit) scales,
```
NI = 10
IREFER = YYYYZZZZZZ ; items identified by type: item 1 is Y, item 4 is Z etc.
CODES = ABCD ; original codes in the data
IVALUEY = 1234
IVALUEZ = 4321
STKEEP=NO  ; allow each item to have its own rating (or partial credit) scale structure
```

Example 2: Items identified by 1, 2, 3 in ISGROUPS= are to be recoded and given there own rating (or partial credit) scales
Y-type items are 1-3, 7-8. Z-type items are 4-6, 9-10.
```
NI = 10
ISGROUPS = YYYZZZYYZZ
CODES = ABCD ; original codes in the data
IVALUEY = 1234
IVALUEZ = 4321
```
or
```
NI = 10
ISGROUPS = YYYZZZYYZZ
IREFER = YYYYZZZZZZ
CODES = ABCD ; original codes in the data
IVALUEY = 1234
```

Example 3: All items are to be recoded the same way.
```
NI = 100   100 ITEMS
IREFER=* 1-100 X FOR ALL 100 ITEMS, reference is X
Codes = 12345 rescore 12345
IVALUEX = 12223 into 12223
```

123. **IWEIGHT**  item (variable) weighting

IWEIGHT= allows for differential weighting of items. The standard weights are 1 for all items. To change the weighting of items, specify IWEIGHT=

Raw score, count, and standard error of measurement reflect the absolute size of weights as well as their relative
sizes. Measure, inﬁt and outfit and correlations are sensitive only to relative weights.

Weighting is treated for estimation as that many independent observations. So, if you weight all items by two, you will divide the S.E. by the square-root of 2, but will not change the measures or ﬁt statistics.

If you want to do different weighting at different stages of an analysis, one approach is to use weighting to estimate all the measures. Then anchor them all (IFILE= and IAFILE= etc.) and adjust the weighting to meet your “independent observation” S.E. and reporting requirements.

If you want the standard error of the ﬁnal weight-based measure to approximate the S.E. of the unweighted measure, then ratio-adjust case weights so that the total of the weights is equal to the total number of independent observations.

Formats are:

IWEIGHT= ﬁle name the weights are in a ﬁle of format:

<table>
<thead>
<tr>
<th>item number</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td>2-20</td>
<td>1</td>
</tr>
</tbody>
</table>

A better weighting, which would make the reported person standard errors more realistic by maintaining the original total sum of weights at 20, is:

IWEIGHT=*

<table>
<thead>
<tr>
<th>item number</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.33</td>
</tr>
<tr>
<td>2-20</td>
<td>0.93</td>
</tr>
</tbody>
</table>

The sum of all weights is 20.0

or adjust the weights to keep the sample-based “test” separation and reliability about the same - so that the reported statistics are still reasonable:

e.g., original sample “test” reliability = .9, separation = 3, but separation with weighting = 4

Multiply all weights by (3/4)^2 to return separation to about 3.

Example 2:

The item labels contain the weights in columns 16-18.

IWEIGHT= $S16W3 ; or $S16E18

Item 1 Hello   0.5
Item 2 Goodbye 0.7

Example 3:

Item 4 is a pilot or variant item, to be given weight 0, so that item statistics are computed, but this item does not affect person measurement.

IWEIGHT=*

4 0 ; Item 4 has weight 0, other items have standard weight of 1.
124. **KEYn scoring key**

*Usually only KEY1= is needed for an MCQ scoring key.*

Up to 99 keys can be provided for scoring the response choices, with control variables KEY1= through KEY99=. Usually KEY1= is a character string of "correct" response choices. The standard is one column per correct response, or two columns if XWIDE=2.

As standard, responses matching the characters in KEY1= are scored 1. Other valid responses are scored 0. KEY2= through KEY99= are character strings of successively "more correct" response choices to be used when more than one level of correct response choice is possible for one or more items. The standard score value for KEY2= is 2, and so on up to the standard score value for KEY99= which is 99. The values assigned to these keys can be changed by means of KEYSCR=. If XWIDE=1, only the values assigned to KEY1= through KEY9= can be changed, KEY10= through KEY99= retain their standard values of 10 through 99. If XWIDE=2, the all KEYn= values can be changed.

**Example 1:** A key for a 20-item multiple choice exam, in which the choices are coded "1", "2", "3" and "4", with one correct choice per item.

```
CODES=1234   valid codes
KEY1 =31432432143142314324 correct answers
```

**Example 2:** A 20-item MCQ test with responses entered as "a", "b", "c", "d".

```
CODES=abcd   valid responses
KEY1 =cadcbdcbdacbdacbd correct answers
```

**Example 3:** A 20 item multiple choice exam with two somewhat correct response choices per item. One of the correct choices is "more" correct than the other choice for each item, so the "less correct" choice will get a score of "1" (using KEY1=) and the "more correct" choice will get a score of "2" (using KEY2=). All other response choices will be scored "0":

```
CODES=1234   valid responses
KEY1=233131413241342113 assigns 1 to these responses
KEY2=31432432143142314324 assigns 2 to these responses
0 is assigned to other valid responses
```

**Example 4:** A 100 item multiple choice test key.

```
CODES= ABCD
KEY1 = BCADDDCBBADCDACBCDADDCCDBADCDADDCA+
       DBADBCDACCDADDCBDADACBDADDDBADDCD+
       +ACBCDADDCCBDADCBC continuation lines
```

**Example 5:** Multiple score key for items 1 to 10. Items 11 to 15 are on a rating scale of 1 to 5

```
CODES = abcd12345
KEY1 = bacdbadcd*****
RESCORE= 11111111100000 ; RESCORE= signals when to apply KEY1=
```

**Example 6:** A 10 item test. 5 MCQ items have responses entered as "ABCD", with one of those being correct: Item 1, correct response is B. Item 2 is C. 3 is D. 4 is A. 5 is C. Then 5 partial-credit performance items rated 0-5.

```
CODES =ABCD012345
ISGROUPS=1111100000
KEY1 = BCDAC11111 ; Key1= automatically has the value "1", etc.
KEY2 = *****22222 ; * can be any character not in CODES=.
KEY3 = *****33333
KEY4 = *****44444
KEY5 = *****55555
```
125. **KEYFROM** location of KEYn

*Only use this if you have the scoring Key conveniently in data-record format.*

Instructs where to find the KEYn= information.

- **KEYFROM=0**  
  KEY1= through KEY99=, if used, are before &END.

- **KEYFROM=1**  
  KEY1 information follows after &END, but before the item names. The key is formatted exactly like a data record. It is helpful to place the name of the key, e.g. "KEY1=", where the person name would usually go, for reference.

- **KEYFROM=n**  
  KEY1=, then KEY2=, and so on up to KEYn= (where n is a number up to 99) follow &END, but placed before the item names. Each key is formatted exactly like a data record. It is helpful to place the name of the key, e.g. "KEY2=", where the person name would usually go.

Example: KEY1 and KEY2 information are to follow directly after &END

| NAME1=1 | start of person-id (the standard) |
| ITEM1=10 | start of response string |
| NI=7 | number of items |
| CODES=abcd | valid codes |
| KEYFROM=2 | two keys in data record format |

&END

KEY1=****bacddba keys formatted as data

KEY2=****cddbaac

| Item 1 name | item names |
| Item 7 name |

END NAMES

Mantovanibbacdba first data record

| subsequent data records |

126. **KEYSCR** reassign scoring keys

*This is only needed for complicated rescoring.*

Specifies the score values assigned to response choices which match KEY1= etc. To assign responses matching key to the "missing" value of -1, make the corresponding KEYSCR= entry blank or some other non-numeric character.

When XWIDE=1, each value can only take one position, so that only KEY1= through KEY9= can be reassigned. KEY10= through KEY99= can also be used but keep their standard values of 10 through 99.

When XWIDE=2, each value takes two positions, and the values corresponding to all keys, KEY1= through KEY99=, can be reassigned.

Example 1: Three keys are used, and XWIDE=1.

Response categories in KEY1= will be coded "1"

Response categories in KEY2= will be coded "2"

Response categories in KEY3= will be coded "3"

KEYSCR=123 (standard)

Example 2: Three keys are used, and XWIDE=1.

Response categories in KEY1= will be coded "2"

Response categories in KEY2= will be coded "1"

Response categories in KEY3= will be coded "1"

KEYSCR=211
Example 3: Three keys are used, and XWIDE=2
Response categories in KEY1= will be coded "3"
Response categories in KEY2= will be coded "2"
Response categories in KEY3= will be coded "1"

KEYSCR=030201

or

KEYSCR=" 3 2 1"

Example 4: Three keys are used, and XWIDE=1
Response categories in KEY3= will be coded "1"
Response categories in KEY6= will be coded "missing"
Response categories in KEY9= will be coded "3"

KEY3=BACDCACDBA response keys
KEY6=ABDADCDCA
KEY9=CCBCBBBBCC
KEYSCR=xx1xxXxx3 scores for keys

The "x"'s correspond to unused keys, and so will be ignored.
The "X" corresponds to specified KEY6=, but is non-numeric and so will cause responses matching KEY6= to be ignored, i.e. treated as missing.

Example 5: Some items in a test have two correct answers, so two keys are used. Since both answers are equally good, KEY1= and KEY2= have the same value, specified by KEYSCR=. But some items have only one correct answer so in one key "*", a character not in CODES=, is used to prevent a match.

CODES=1234
KEY1 =2331314132413242113
KEY2 =31*324321*3142314**** * is not in CODES=
KEYSCR=11 both KEYs scored 1

Example 6: More than 9 KEYn= lines, together with KEYSCR=, are required for a complex scoring model for 20 items, but the original data are only one character wide.

Original data: Person name: columns 1-10
20 Item responses: columns 21-40

Looks like: M. Stewart..........1321233212321232134

Solution: reformat from XWIDE=1 to XWIDE=2

```
TITLE="FORMAT= from XWIDE=1 to =2"
FORMAT=(10A1,10X,20A1) 10 of Name, skip 10, 20 of responses
N=20
NAME=1
ITEM=11 Responses in column 11 of reformatted record
XWIDE=2
CODES="1 2 3 4 " Original response now "response blank"
KEY1 ="1 2 1 3 2 1 2 3 1 4 3 2 1 1 1 1 1 2 1 1 " Keying 20 items
KEY2 ="2 1 2 1 1 2 1 1 2 1 1 2 3 3 2 2 * 2 1 "
....
KEY10="3 3 3 2 3 3 2 3 4 2 3 * * * 4 4 4 4 4 4 4 4"
KEYSCR="1 2 3 2 2 2 3 4 1 4 " Renumbering 10 KEYn=
&END
```

127. **LCONV** logit change at convergence

*Measures are only reported to 2 decimal places, so a change of less than .005 logits will probably make no visible difference.*

Specifies what value the largest change in any logit estimate for a person measure or item calibration or rating (or partial credit) scale structure calibration must be less than, in the iteration just completed, for iteration to cease. The current largest value is listed in Table 0 and displayed on your screen. See convergence considerations.
The standard setting is CONVERGE="E", so that iteration stops when either LCONV= or RCONV= is satisfied. (Note: this depends on Winsteps version - and may explain differences in converged values.)

Example: To set the maximum change at convergence to be less or equal to .001 logits:
LCONV=.001
RCONV=0; set to zero, so does not affect convergence decision
CONVERGE=Logit

128. **LINLEN** length of printed lines in Tables 7, 10-16, 22

The misfitting responses, name maps, scalogram, and option frequency tables can be output in any convenient width. Specify **LINLEN=0** for the maximum page width (132 characters).

Example: You want to print the map of item names with up to 100 characters per line.
LINLEN=100 set line length to 100 characters

129. **LOCAL** locally restandardize fit statistics

**LOCAL=N** accords with large-sample statistical theory.

Standardized fit statistics test report on the hypothesis test: "Do these data fit the model (perfectly)?" With large sample sizes and consequently high statistical power, the hypothesis can never be accepted, because all empirical data exhibit some degree of misfit to the model. This can make t standardized statistics meaninglessly large. t standardized statistics are reported as unit normal deviates. Thus ZSTD=2.0 is as unlikely to be observed as a value of 2.0 or greater is for a random selection from a normal distribution of mean 0.0, standard deviation, 1.0. ZSTD (standardized as a z-score) is used of a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's t-distribution value has been adjusted to a unit normal value.

**LOCAL=N**  t standardized fit statistics are computed in their standard form. Even the slightest item misfit in tests taken by many persons will be reported as very significant misfit of the data to the model. Columns reported with this option are headed "ZSTD" for model-exact standardization. This is a "significance test" report on "How unexpected are these data if the data fit the model perfectly?"

**LOCAL=L** Instead of t standardized statistics, the natural logarithm of the mean-square fit statistic is reported. This is a linearized form of the ratio-scale mean-square. Columns reporting this option are headed "LOG", for mean-square logarithm.

**LOCAL=Y**  t standardized fit statistics are transformed to reflect their level of unexpectedness in the context of the amount of disturbance in the data being analyzed. The model-exact t standardized fit statistics are divided by their local sample standard deviation. Thus their transformed sample standard deviation becomes 1.0. Columns reported with this option are headed "ZEMP" for empirically restandardized. The effect of the local-rescaling is to make the fit statistics more useful for interpretation. The meaning of ZEMP statistics is an "acceptance test" report on "How unlikely is this amount of misfit in the context of the overall pattern of misfit in these data?"

Ronald A. Fisher ("Statistical Methods and Scientific Inference" New York: Hafner Press, 1973 p.81) differentiates between "tests of significance" and "tests of acceptance". "Tests of significance" answer hypothetical questions: "how unexpected are the data in the light of a theoretical model for its construction?" "Tests of acceptance" are concerned with whether what is observed meets empirical requirements. Instead of a theoretical distribution, local experience provides the empirical distribution. The "test" question is not "how unlikely are these data in the light of a theory?", but "how acceptable are they in the light of their location in the empirical distribution?"

130. **LOGFILE** accumulates control files

Specifying **LOGFILE=file name** causes the current control file to be appended to the log file, enabling an audit trail of the Winsteps analysis. The contents of Table 0.3 are saved.

Example: An audit trail of Winsteps analyses is to be maintained at c:\winsteps.log.txt
131. **LOWADJ** correction for bottom rating scale categories

The Rasch model models the measure corresponding to a bottom rating (or partial credit) scale category as infinite. This is difficult to think about and impossible to plot. Consequently, graphically in Table 2.2 and numerically in Table 3.1 a measure is reported corresponding to a bottom category. This is the measure corresponding to an imaginary rating LOWADJ= rating points above the bottom category. HIADJ= is the corresponding instruction for top categories.

Example: The standard spread in Table 2.2 is based on LOWADJ=0.25. You wish the bottom category number to be printed more to the right, close to the other categories.

LOWADJ=0.4

132. **MAKEKEY** construct MCQ key

For multiple-choice and True-False questions, the analyst is usually provided with the answer key. When an answer key is not available, MAKEKEY=YES constructs one out of the most frequent responses to each item. The answer key is used in the analysis and reported at the end of Table 0.3 in the Report Output File. Inspect the Item Tables, particularly the "CATEGORY/OPTION/Distractor FREQUENCIES", to identify items for which this scoring key is probably incorrect. The correct answer is expected to have the highest measure.

If you have no KEY1= at all, put in a dummy key, e.g., all A's or whatever, to get Winsteps to run.

Example: The scoring key for Example5.con is lost.

```
MAKEKEY=YES
KEY1 = aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa ;
dummy key
```

Constructed key is:

```
KEY1 = dabbbadbdacacaadbababaaaaacbdaddabccbcaccbccccacbbcbdbdbacaccbcdbdb
```

Original key was:

```
KEY1 = dcbbbadbdacacaddbabaaaccbdddcaadccccdbcccbccbadccccdbcccbdbacaccbcdbdb
```

The keys match for 48 of the 69 items. Item fit Tables suggest up to 29 items whose keys may not be correct.

The key is reported on the Iteration screenon the and after Table 0.3 in the Report Output file accessed by the Edit File pull-down menu.

133. **MATRIX** correlation output format

The correlation matrix ICORFILE= or PCORFILE= can be produced in list or matrix format.

**MATRIX = NO** is the list format

```
<table>
<thead>
<tr>
<th>Item</th>
<th>Item</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-.04</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>.05</td>
</tr>
</tbody>
</table>
```

**MATRIX = YES** is the matrix format

```
1.0000  -.0451  .0447  .0095  ......  
-.0451  1.0000  -.0448  -.2024  ......  
.0447  -.0448  1.0000  -.0437  ......  
......  ......  ......  ......  ......  
```
134. **MAXPAGE**  the maximum number of lines per page

*For no page breaks inside Tables, leave MAXPAG=0*

If you prefer a different number of lines per page of your output file, to better match the requirements of your printer or word processor, give that value (see *Using a Word Processor or Text Editor* in Section 2). If you prefer to have no page breaks, and all plots at their maximum size, leave MAXPAG=0.

On Table 1 and similar Tables, MAXPAG= controls the length of the Table.

Example: You plan to print your output file on standard paper with 60 lines per page (pages are 11 inches long, less 1 inch for top and bottom margins, at 6 lpi):

```
MAXPAG=60  (set 60 lines per page)
FORMFD=^ (standard: Word Processor form feed)
```

135. **MFORMS** reformat input data

MFORMS= supports the reformatting of input data records, and also equating multiple input files in different formats, such as alternate forms of the same test. Data after END NAMES or END LABELS is processed first, as is data specified by DATA= in the core control file.

Data reformatted by MFORMS= can be accessed, viewed, edited and "saved as" permanently using the "Edit" pull-down menu. It has a file name of the form: ZMF.....txt

Here is the layout:

```
mforms=*  
data=forma.txt   ; the name of an input data file  
L=2   ; there are 2 lines in input data file for each data record  
I1 = 20  ; response to item 1 of the test is in column 20 of the input data file  
I3-5 = 21  ; items 3, 4, 5 are in columns 21, 22, 23 of the input data file  
I16-20=11  ; items 16, 17,18, 19, 20 are in columns 11, 12, 13, 14, 15  
P1=9  ; the first character of person label is in column 9 of the input data file  
P3-8=1  ; information for columns 3-7 of person label starts in column 1  
C20-24="FORMA"  ; put in columns 20-24 the letters FORMA  
C40-90 = 2:1  ; put in columns 40-90 the characters in the second line of the data record  
#  ; end of definition - start of next file reformat  
DATA=formb.txt  ; name of input data file  
P3-7=1  ; information for columns 3-7 of person label starts in column 1 of data record  
.....  
*  ; end of mforms= command
```

Details:

- `mforms=*` instructions follow in control file, and end with another *.
- `mforms=filename` instructions are in a file but in the same format as above.

- `data=filename` name of input file name to be reformatteed.
  - The reformatted records are placed in a temporary work file. This may be accessed from the Edit pull-down menu, and saved into a permanent file.
  - This temporary file is processed after any Data Files specified with the master `Data=` instruction and in the same way, e.g., any `FORMAT=` command will be applied also to the temporary work file.

- `L=nnn` nnn is the count of lines in each input data record
  - If `L=1` this can be omitted
  - `L=4` means that 4 input data lines are processed for each data record output.

- `Cnnn=` nnn is the column number in the formatted data record.
XWIDE= does not apply. C10-12 means columns 10, 11, 12 of the formatted record. C1= refers to column 1 of the formatted data record. This can also be used to move item and person information.

Innn=....  nnn is the starting item number in the formatted data record
nnn-mmm are the starting to ending item numbers in the formatted data record

XWIDE= is applied, so that I13-5= with XWIDE=2 means 6 characters. I1= points to column Item1= in the formatted data record.

Pnnn=....  nnn is the starting column number in the person label in the formatted person label.

XWIDE= is not applied. P6-8= always means 3 columns starting in column 6. P1= points to column Name1= in the formatted data record.

......=nnn  nnn is the starting column of the only, or the first, line in the input data record.

......=m:nnn  m is the line number in each data record
nnn is the starting column number of that line

......="xxxx"  "xxxx" is a character constant to be placed in the formatted data record.

Note: for I18-20="abc" with XWIDE=2, then response to Item 18 is "ab", 19 is "c", 20 is " ".

#  end of processing of one file, start of the next

*  end of Mforms= processing

Example 1:  See Exam10c.txt

Example 2:  Three data files with common items and one MCQ scoring key.

Datafile1.tx: (Items 1-6)
TOMY ABCDAB
BILL BCDADD

Datafile2.txt (Items 1-3 and 7-9)
TOTO BBABAB
MOULA BADADD

Datafile3.txt (Items 1-3 and 10-12)
IHSANI ACCDAB
MALIK CBDDCD

TITLE="Multiple MCQ forms with one scoring key"
NI=12  ; 12 ITEMS IN TOTAL
ITEM1=11
NAME1=1
CODES="ABCD"
KEY1=BACCADACADDA
mforms=*data=datafile1.txt  ; name of data file
L=1  ; one line per person
P1-10=1  ; person label in columns 1-10
I1-3=11  ; items 1-3 in columns 11-13
I4-6=14  ; items 4-6 in columns 14-16
#
data=datafile2.txt
L=1
P1-10=1
I1-3=11
I7-9=14  ; items 8-9 in columns 14-16
#
data=datafile3.txt
L=1
Here is how the data appear to Winsteps for analysis:

Example 3: Test 1 is a 4-item survey. Test 2 is a 4-item survey with two items in common with Test 1 which are to be anchored to their Test 1 values.

Test 1 has 4 rating scale items. Each item has its own partial-credit structure:

```plaintext
title = "Test 1"
item1 = 1 ; items start in column 1
ni = 4 ; 4 items
name1 = 5 ; person label starts in column 5
namlen = 14 ; length of person name
codes = 01234 ; rating scale
ISGROUPS = 0 ; each item has its own rating scale structure
stkeep = YES ; this is probably what you want for these type of data
data = data1.txt
ifile = items1if.txt ; item calibrations from Test 1 for Test 2 (output)
sfile = items1sf.txt ; structure calibrations from Test 1 for Test 2 (output)
&END
Test 1 item 1
Test 1 item 2
Test 1 item 3
Test 1 item 4
END NAMES
```

data 1.txt is:

```
1234Person 1-1
3212Person 1-2
......
```

Test 2 has 4 items. 1 and 4 are new - we will call these items 5 and 6 of the combined Test 1 and 2. Item 2 is Test 1 item 4, and item 3 is Test 1 item 2.

```plaintext
title = "Test 2 (formatted to match Test 1)"
item1 = 1 ; items start in column 1
ni = 6 ; 4 items in Test 1 + 2 more in Test 2
name1 = 7 ; person label starts in column 7
namlen = 14 ; length of person name
codes = 01234 ; rating scale
stkeep = YES ; this is probably what you want for these type of data
ISGROUPS = 0 ; each item has its own rating scale structure
iafile = items1if.txt ; item calibrations from Test 1 (input - unchanged)
safie = items1sf.txt ; structure calibrations from Test 1 (input - unchanged)
MFORMS = * ; reformat the Test 2 data to align with Test 1
data = data2.txt ; the name of an input data file
&END
```
L = 1 ; there is 1 line in input data file for each data record
I2 = 3 ; response to item 2 of Test 1 is in 3 of the data2.txt file
I4 = 2 ; response to item 4 of Test 1 is in 2 of the data2.txt file
I5 = 1 ; item 5 is in column 1 of data2.txt
I6 = 4 ; item 6 is in column 4 of data2.txt
P1-14 = 5 ; the first character of person label is in column 5 of data2.txt for 14 columns.
* ; end of mforms= command
&END
END NAMES

data2.txt is:
  5426Person 2-1
  1234Person 2-2
  ....

The formatted file (see Edit pull-down menu MFORMS==) is
  2 456Person 2-1
  3 214Person 2-2
  ....

136. MHSlice Mantel-Haenszel slice width

Differential item functioning (DIF) can be investigated using log-odds estimators, Mantel-Haenszel (1959) for dichotomies or Mantel (1963) for polytomies. The sample is divided into difference classes (also called reference groups and focal groups). These are produced for Table 30 specified with DIF=. In principle, when the data fit the Rasch model, these estimators should concur with the DIF contrast measures. When DIF estimates disagree, it indicates that the DIF in the data is non-uniform with ability level. The DIF contrast weights each observation equally. Mantel-Haenszel weights each slice equally.

The simplest way to do a direct comparison of the MH and DIF Contrast methods is to anchor all persons at the same ability:
PAFILE=*  
1-NN 0
*

This will put everyone in one MH cross-tab.

MHSlice= specifies the width of the slice (in logits) of the latent variable be included in each cross-tab. The lower end of the first slice is always the lowest observed person measure.

MHSlice = 0 bypasses Mantel-Haenszel or Mantel computation.

MHSlice = .1 logits and smaller. The latent variable is stratified into thin slices.

MHSlices = 1 logit and larger. The latent variable is stratified into thick slices.

For each slice, a cross-tabulation is constructed for each pair of person classifications against each scored response level. An odds-ratio is computed from the cross-tab. Zero and infinite ratios are ignored. A homogeneity chi-square is also computed when possible.

Thin slices are more sensitive to small changes in item difficulty across person classifications, but more persons
are ignored in inestimable cross-tabs. Thick slices are more robust because fewer persons are ignored. Use the Specification pull-down menu to set different values of MHSLICE= and then produce the corresponding Table 30.

Person classifications are A, B, ... They are compared pairwise. Starting from the lowest person measure, each slice is MHSLICE= logits wide. There are K slices up through the highest person measure. For the target item, in the kth slice and comparing classifications A and B: 

ACk and BCk are the counts of persons in classifications A and B in slice k. 

ABk = ACk + BCk.

ASk and BSk are the summed ratings or scores of persons in classifications A and B in slice k on the target item. 

ABSk = ASk + BSk.

ABQk are the summed squared ratings or scores of persons in both classifications A and B in slice k on the target item.

Then the Mantel or Mantel-Haenszel DIF chi-square for the target item is:

\[
\chi^2_1 = \sum_{k=1}^{K} \frac{AC_k BC_k}{AB C_k} \left( \frac{AB C_k}{AB C_k - 1} \right) \left( \frac{AB C_k A B Q_k - A B S_k^2}{A B C_k} \right)
\]

For dichotomous items, the Mantel-Haenszel logit DIF size estimate for a dichotomous item is summed across estimable slices:

\[
\ln(\hat{\alpha}_{MH}) = \log \left( \sum_{k=1}^{K} \frac{A S_k}{B S_k} (B C_k - B S_k) \right) \left/ \sum_{k=1}^{K} \frac{B S_k}{A C_k - A S_k} \right.
\]

For polytomous items using adjacent, transitional, sequential odds, the logit DIF size estimate becomes:

\[
\ln(\hat{\alpha}_{MH}) = \log \left( \sum_{j=1}^{m} \sum_{k=1}^{K} \frac{A S_{jk}}{B S_{j-1,k}} \right) \left/ \sum_{j=1}^{m} \sum_{k=1}^{K} \frac{B S_{jk}}{A C_{jk} - A S_{jk}} \right.
\]

where ASjk is the count of responses by Classification A in category j of slice k.


Example:

<table>
<thead>
<tr>
<th>PERSON</th>
<th>DIF</th>
<th>DIF</th>
<th>PERSON</th>
<th>DIF</th>
<th>DIF</th>
<th>JOINT</th>
<th>MantelHanzl ITEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASS</td>
<td>MEASURE S.E.</td>
<td>CLASS</td>
<td>MEASURE S.E.</td>
<td>CONTRAST S.E.</td>
<td>t</td>
<td>d.f.</td>
<td>Prob.</td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>-------</td>
<td>--------</td>
<td>-------</td>
<td>-------</td>
<td>---------</td>
<td>-----------------</td>
</tr>
<tr>
<td>A</td>
<td>1.47</td>
<td>.28</td>
<td>P</td>
<td>2.75</td>
<td>.34</td>
<td>-1.28</td>
<td>.44 -2.94 104 0.0041 .0040 -1.20 1 Response</td>
</tr>
</tbody>
</table>

Size of Mantel-Haenszel slice = .100 logits
title="MH computation"
; d.f.=1 chi=8.3052 p=0.0040
; log-odds = 1.198
codes=01
clfile=*
1 Better
0 Same
* item1=1
name1=1
NI=1
pweight=$s9w2 ; weighting substitutes for entering multiple records
PAFILE=$S6W1 ; anchoring forces stratification
DIF = $4W1 ; cross-tab by Gender, F or M
&end
Response
;234567890
END LABELS
1 FA 1 16
0 FA 1 11
1 FP 1 5
0 FP 1 20
1 MA 2 12
0 MA 2 16
1 MP 2 7

133
MISSCORE scoring of missing data codes

This is NOT the missing-value code in your data. All codes NOT in CODES= are missing value codes. Use this control specification when you want missing data to be treated as valid responses. Wnsteps and Missing Data: No Problem!

Winsteps processes one observation at a time. For each observation, Xni by person n on item i, it computes an expectation Eni, based on the current person measure estimate Bn and the current item measure Di and, if relevant, the current rating (or partial credit) scale structure (calibrations) {Fk}. Pnik is the probability of observing category k for person n on item i.

In this computation it skips over, omits, ignores “missing” data.

It then compares sum(Xni) with sum(Eni) for each person n, and adjusts Bn.
It then compares sum(Xni) with sum(Eni) for each item i, and adjusts Di
It then compares the count of (Xni=k) with the sum (Pnik) for each k, and adjusts Fk

These sums and counts are only over the observed data. There is no need to impute missing data.

There are no pairwise, listwise or casewise deletions associated with missing data.

MISSCORE= says what to do with characters that are not valid response codes, e.g. blanks and data entry errors. Usually any characters not in CODES= are treated as missing data, and assigned a value of -1 which means “ignore this response.” This is usually what you want when such responses mean “not administered”. If they mean “I don’t know the answer”, you may wish to assign missing data a value of 0 meaning “wrong”, or, on a typical attitude survey, 3, meaning “neutral” or “don’t know”.

MISSING=0 is the same as MISSCORE=0 meaning that all codes in the data not listed in CODES= are to be scored 0.

Non-numeric codes included in CODES= (without rescoring/recoding) or in NEWSCORE= or IVALUE= are always assigned a value of “not administered”, -1.

Example 0a: In my data file, missing data are entered as 9. I want to score them 0, wrong answers. Valid codes are 0 and 1.
CODES = 01 do not specify a 9 as valid
MISSCORE = 0 specifies that all codes not listed in CODES=, e.g., 9's, are to be scored 0.

Example 0b: In my data file, missing data are entered as 9. I want to ignore them in my analysis. Valid codes are 0 and 1.
CODES = 01 do not specify a 9 as valid
; the following line is the standard, it can be omitted.
MISSCORE = -1 specifies that all codes not listed in CODES=, e.g., 9’s.
are to be treated as “not administered”

Example 1: Assign a code of "0" to any responses not in CODES=
MISSCORE=0 missing responses are scored 0.

Example 2: In an attitude rating scale with three categories (0, 1, 2), you want to assign a middle code of "1" to missing values
MISSCORE=1 missing responses scored 1

Example 3: You want blanks to be treated as “wrong” answers, but other unwanted codes to be ignored. In a questionnaire with responses "Y" and "N".
CODES="YN " blank included as valid response
NEWSCORE=100 new response values
RESCORE=2 rescore all items
MISSCORE=-1  ignore missing responses (standard)

Example 4: Your optical scanner outputs an "@" if two bubbles are marked for the same response. You want to ignore these for the analysis, but you also want to treat blanks as wrong answers:
CODES ="1234 "  blank is the fifth valid code
KEY1 =31432432143142314324  correct answers
MISSCORE=-1  applies to @ (standard)

Example 5: Unexpected codes are scored "wrong", but 2's to mean "not administered".
CODES = 012
NEWSCORE= 01X X is non-numeric, matching 2's ignored
MISSCORE= 0  all non-CODES= responses scored 0

Example 6: You have a long 4-option MCQ test with data codes ABCD. Most students do not have the time to complete all the items. This requires a two-stage item analysis:
Stage 1.  Item calibration:
Deliberately skipped responses are coded "S" and scored incorrect. The student could not answer the question.
Not-items are coded "R" and scored "not administered". This prevents easy items at the end of the test being calibrated as "very difficult".
CODES="ABCDS"
KEY1="CDBAD. . . . ."
MISSCORE=-1
IFILE=ITEMCAL.TXT  ; write out the item calibrations

Stage 2.  Person measurement:
The convention with MCQ tests is that all missing responses are scored incorrect when measuring the persons.
IAFILE=ITEMCAL.TXT  ; anchor on the Stage 1 item calibrations
CODES="ABCDS"
KEY1="CDBAD. . . . ."
MISSCORE=0  ; all missing data are scored incorrect

138.  MJMLE  maximum number of JMLE iterations

JMLE iterations may take a long time for big data sets, so initially set this to -1 for no JMLE iterations. Then set MJMLE= to 10 or 15 until you know that more precise measures will be useful. The number of PROX iterations, MPROX=, affects the number of JMLE iterations but does not affect the final estimates.

MJMLE= specifies the maximum number of JMLE iterations to be performed. Iteration will always cease when both LCONV= and RCONV= criteria have been met, see CONVERGE=. To specify no maximum number limitation, set MJMLE=0. Iteration always be stopped by Ctrl with F, see "Stopping Winsteps".

Example 1: To allow up to 4 iterations in order to obtain rough estimates of a complex rating (or partial credit) scale:
MJMLE=4  4 JMLE iterations maximum

Example 2: To allow up to as many iterations as needed to meet the other convergence criteria:
MJMLE=0  Unlimited JMLE iterations

Example 3: Perform no JMLE iterations, since the PROX estimates are good enough.
MJMLE=-1  No JMLE iteration

Example 4: Run as quick estimation as possible to check out control options.
MPROX=-1 ; minimal prox estimation iterations
MJMLE=-1 ; no JMLE iterations

139.  MNSQ  show mean-square or standardized fit statistics

The mean-square or t standardized fit statistics are shown in Tables 7, 11 to quantify the unexpectedness in the
response strings, and in Tables 4, 5, 8, 9 for the fit plots.

MNSQ=N  Show standardized (ZSTD) fit statistics. ZSTD (standardized as a z-score) is used of a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's t-distribution value has been adjusted to a unit normal value.

MNSQ=Y  Show mean-square fit statistics. Use LOCAL=L for log scaling.

### TABLE 7.1 TABLE OF POORLY FITTING PERSONS   (ITEMS IN ENTRY ORDER)
NUMBER - NAME -- POSITION ------ MEASURE - INFIT (MNSQ) OUTFIT

<table>
<thead>
<tr>
<th>NUMBER</th>
<th>NAME</th>
<th>POSITION</th>
<th>INFIT</th>
<th>MEASURE</th>
<th>MNSQ</th>
<th>OUTFIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>Rod</td>
<td>M</td>
<td>-1.41</td>
<td>2.4</td>
<td>A</td>
<td>2.2</td>
</tr>
</tbody>
</table>

**Z-RESIDUAL:**

-2 2 -2 -2 2 -2

**Mean-square:**

### TABLE 9.1

<table>
<thead>
<tr>
<th>ITEM</th>
<th>MEASURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>-5</td>
<td>-4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ITEM</th>
<th>MEASURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>-5</td>
<td>-4</td>
</tr>
</tbody>
</table>

**t standardized ZSTD:**

<table>
<thead>
<tr>
<th>ITEM</th>
<th>MEASURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>-5</td>
<td>-4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ITEM</th>
<th>MEASURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>-5</td>
<td>-4</td>
</tr>
</tbody>
</table>

140. MODELS assigns model types to items

Winsteps estimates calibrations for four different ordered response category structures. *Dichotomies* are always analyzed using the Rasch dichotomous model, regardless of what model is specified for polytomies.
MODELS=R (standard) is the default option, specifying standard Rasch analyses using the Rasch dichotomous model, Andrich "Rating Scale" model and Masters' "Partial Credit" model, see ISGROUPS=

MODELS=S uses the Rasch dichotomous model and the Glas-Verhelst "Success" (growth) model, also called the "Steps" Model (Verhelst, Glas, de Vries, 1997). If and only if the person succeeds on the first category, another category is offered until the person fails, or the categories are exhausted, e.g. an arithmetic item, on which a person is first rated on success on addition, then, if successful, on multiplication, then, if successful, on division etc. "Scaffolded" items can function this way. This is a continuation ratio model parameterized as a Rasch model with missing data on unreached categories. Verhelst N.D., Glas C.A.W. & De Vries H.H. (1997) A Steps model to analyze partial credit. In W.J. van der Linden & R.K. Hambleton (Eds.), Handbook of modern item response theory (pp. 123 - 138) New York: Springer.

MODELS=F uses the Rasch dichotomous model and the Linacre "Failure" (mastery) model. If a person succeeds on the first category, top rating is given and no further categories are offered. On failure, the next lower category is administered until success is achieved, or categories are exhausted. This is a continuation ratio model parameterized as a Rasch model with missing data on unreached categories. The Success and Failure model computations were revised at Winsteps version 3.36, August 2002.

MODELS= has three forms: MODELS=RRSSFR and MODELS=* list * and MODELS=*filename. When only one letter is specified with MODELS=, e.g., MODELS=R, all items are analyzed using that model. Otherwise MODELS=some combination of R's, F's, S's, and G's, e.g., MODELS=RRSF

Items are assigned to the model for which the serial location in the MODELS string matches the item sequence number. The item grouping default becomes each item with its own rating scale, ISGROUPS=0.

When XWIDE=2 or more, then either (a) Use one character per XWIDE and blanks,
   NI=8
   XWIDE=2
   MODELS=' R S R F R S R R' ; this also forces ISGROUPS=0 to be the default
or (b) Use one character per item with no blanks
   NI=8
   XWIDE=2
   RESCORE='RSRFRSRR' ; this also forces ISGROUPS=0 to be the default

Example 1: All items are to be modeled with the "Success" model. MODELS=S the Success model

Example 2: A competency test consists of 3 success items followed by 2 failure items and then 10 dichotomies. The dichotomies are to be reported as one grouping.
   NI=15 fifteen items
   MODELS=SSSFFFFFFRRRRRRRRRR ; matching models: ; forces ISGROUPS=0 to be the default
   ISGROUPS=0000011111111111 ; dichotomies grouped: overriding the default ISGROUPS=0
or MODELS=* 1-3 S
   4 F
   5 F
   6-15 R
   *

141. MODFROM location of MODELS

This command has not proved productive. It is maintained for backwards compatibility.

Only use this if you have too many items to put conveniently on one line of the MODELS= control variable. It is easier to use "+" continuation lines
Instructs where to find the MODELS= information.
MODFROM=N MODELS= is a control variable before &END (the standard).

MODFROM=Y

MODELS= information follows just after &END but before the item names. It is formatted exactly like a data record. It is helpful to enter "MODELS=" where the person name would go.

Example: A test consists of 10 three-category items. The highest level answer is scored with KEY2=. The next level with KEY1=. Some items have the “Success” structure, where the higher level is administered only after success has been achieved on the lower level. Some items have the “Failure” structure, where the lower level is administered only after failure at the higher level. The MODELS=, KEY1=, KEY2= are formatted exactly like data records. The data records are in a separate file.

| NAME1 = 5          | start of person-id |
| ITEM1 = 20        | start of responses |
| NI = 10           | ten items         |
| CODES = ABCDE     | valid codes       |
| MODFRM = Y        | MODELS= in data format |
| KEYFRM = 2        | two keys in data format |
| DATA = DATAFILE   | location of data  |
| ;23456789          | 1 column          |
| ;345678901234567890| 2 columns         |

&END

MODELS= SSSFFFSSSS data format
KEY1= BCDABCDABC starts in column ITEM1 = 20
KEY2= ABCDBCBAAA

142.  MPROX  maximum number of PROX iterations

Specifies the maximum number of PROX iterations to be performed. PROX iterations will always be performed so long as inestimable parameters have been detected in the previous iteration, because inestimable parameters are always dropped before the next iteration. At least 2 PROX iterations will be performed. PROX iteration ceases when the spread of the persons and items no longer increases noticeably (0.5 logits). The spread is the logit distance between the top 5 and the bottom 5 persons or items.

If you wish to continue PROX iterations until you intervene with Ctrl and S, set MPROX=0. JMLE iterations will then commence.

Example: To set the maximum number of PROX iterations to 20, in order to speed up the final JMLE estimation of a symmetrically-distributed set of parameters,

MPROX=20

143.  MRANGE  half-range of measures on plots

Specifies the measure (X-axis on most plots) half-range, (i.e., range away from the origin or UMEAN=), of the maps, plots and graphs. This is in logits, unless USCALE= is specified, in which case it must be specified in the new units defined by USCALE=. To customize particular tables, use the Specification pull-down menu, or see TFILE=.

Example 1: You want to see the category probability curves in the range -3 to +3 logits:

MRANGE=3

Example 2: With UMEAN=500 and USCALE=100, you want the category probability curves to range from 250 to 750:

UMEAN=500  new item mean calibration
USCALE=100  value of 1 logit
MRANGE=250 to be plotted each way from UMEAN=
144. **NAME1** first column of person label

**NAME1=** gives the column position where the person label information starts in your data file or in the new record formatted by **FORMAT=**.

It is easy to miscount the **NAME1=** column. Scroll to the top of the Winsteps screen and check column positions:

```
Input in process.
Input Data Record:
  1  2
1234567890123456789012345678
^P        ^I               ^N
35 KID     Records Input.
```

^P marks the **Name1=** column position with ^.
^I marks the **Item1=** column position with ^.
^N marks the **NI=** column position with ^.

Example 1: The person-id starts in column 10, data responses are 1 column wide, in columns 1-8:

```
NAME1=10       starting column of person-id
XWIDE=1        width of response
NI=8           number of responses
```

Example 2: The person-id in column 10, there are 4 data responses are 2 columns wide, in columns 1-8:

```
NAME1=10       starting column of person-id
XWIDE=2        width of response
NI=4           number of responses
```

Example 3: The person id starts in column 23 of the second record.

```
FORMAT=(80A,/,80A)       concatenate two 80 character records
NAME1=103          starts in column 103 of combined record
```

Example 4: The person id starts in column 27 of a record with **XWIDE=**2 and **FORMAT=**.
This becomes complicated, see **FORMAT=**

145. **NAMLEN** length of person label

*Use this if too little or too much person-id information is printed in your output tables.*

**NAMLEN=** allows you define the length of the person-id name with a value in the range of 1 to 30 characters. This value overrides the value obtained according to the rules which are used to calculate the length of the person-id. These rules are:

- 1) Maximum person-id length is 300 characters
- 2) Person-id starts at column **NAME1=**
- 3) Person-id ends at **ITEM1=** or end of data record.
- 4) If **NAME1=** equals **ITEM1=** then length is 30 characters.

Example 1: The 9 characters including and following **NAME1=** are the person's Social Security number, and are to be used as the person-id.

```
NAMLEN=9
```

Example 2: We want to show the responses in Example0.txt as the person label to help diagnose the fit statistics:

```
ITEM1 = 1
NI = 25
NAME1 = 1
NAMLEN = 25
```

```
<table>
<thead>
<tr>
<th>ENTRY</th>
<th>RAW</th>
<th>MODEL</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>PTMEA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NUMBER</td>
<td>SCORE</td>
<td>COUNT</td>
<td>MEASURE</td>
<td>S.E.</td>
<td>MNSQ</td>
<td>ZSTD</td>
</tr>
</tbody>
</table>
```

139
146. **NAMLMP** name length on map for Tables 1, 12, 16

The id fields are truncated for Tables 12 and 16. The name-length for maps variable, NAMLMP=, overrides the calculated truncation.

This is ignored when IMAP= or PMAP= is specified.

Example: The 9 characters including and following NAME1= are the person’s Social Security number, and are to be used as the person-id on the maps.

NAMLMP=9

147. **NEWCORE** recoding values

NEWCORE= says which values must replace the original codes when RESCORE= is used. If XWIDE=1 (the standard), use one column per code. If XWIDE=2, use two columns per code. The length of the NEWSCORE= string must match the length of the CODES= string. For examples, see RESCORE=. NEWSCORE= is ignored when KEYn= is specified.

The responses in your data file may not be coded as you desire. The responses to some or all of the items can be rescored or keyed using RESCORE=. RESCORE= and NEWSCORE= are ignored when KEYn= is specified, except as below.

**RESCORE=** " or 2 or is omitted

All items are recoded using NEWSCORE=. RESCORE=2 is the standard when NEWSCORE= is specified.

**RESCORE=** some combination of 1’s and 0’s

Only items corresponding to 1’s are recoded with NEWSCORE= or scored with KEYn=. When KEYn= is specified, NEWSCORE= is ignored.

If some, but not all, items are to be recoded or keyed, assign a character string to RESCORE= in which “1” means “recode (key) the item”, and “0” (or blank) means “do not recode (key) the item”. The position of the “0” or “1” in the RESCORE= string must match the position of the item-response in the item-string.

**Example 1:** The original codes are “0” and “1”. You want to reverse these codes, i.e., 1 0 and 0 1, for all items.

```
XWIDE=1  one character wide responses (the standard)
CODES =01  valid response codes are 0 and 1 (the standard)
NEWCORE=10  desired response scoring
RESCORE=2  rescore all items - this line can be omitted
```

or

```
NI = 100  100 ITEMS
IREFER=*  FOR ALL 100 ITEMS, reference is X
*. Codes = 01  recode 01
IVALUEX = 10  into 10
```

**Example 2:** Your data is coded “0” and “1”. This is correct for all 10 items except for items 1 and 7 which have the reverse meaning, i.e. 1 0 and 0 1.

```
NI=10  ten items
CODES =01  the standard, shown here for clarity
```

**(a) old method - which still works:**

```
NEWCORE=10  revised scoring
RESCORE=1000001000 only for items 1 and 7
```
(b) new method - recommended:

IVALUE1 =10    revised scoring
IVALUEO =01    scoring unchanged, so this line can be omitted.
IREFER =1000001000 only for items 1 and 7

If XWIDE=2, use one or two columns per RESCORE= code, e.g., "1" or "1 " mean recode (key). " 0" or "0 "
mean do not recode (key).

Example 3: The original codes are " 0" and " 1". You want to reverse these codes, i.e., 1 0 and 0 1, for items 1
and 7 of a ten item test.

NI    =10    ten items
XWIDE =2    two characters wide
CODES = " 0 1" original codes
NEWSCORE=" 1 0" new values
RESCORE =" 1 0 0 0 0 0 1 0 0 0" rescore items 1 & 7

Example 4: The original codes are "0", "1", and "2". You want 0 0, 1 1, and 2 2 for all items

XWIDE=1 one character wide (standard)
CODES =012 valid codes
NEWSCORE=011 desired scoring

Example 5: The original codes are "0", "1", and "2". You want to make 0 2, 1 1, and 2 0, for even-numbered
items in a twenty item test.

NI=20 twenty items
CODES =012 three valid codes
NEWSCORE=210 desired scoring
RESCORE=01010101010101010101 rescore "even" items

Example 6: The original codes are "0", "1", "2", "3" and some others. You want to make all non-specified codes
into "0", but to treat codes of "2" as missing.

CODES = 0123 four valid codes
NEWSCORE= 01X3 response code 2 will be ignored
MISSCORE=0 treat all invalid codes as 0

Example 7: The original codes are "0", "1", "2", "3". You want to rescore some items selectively using KEY1= and
KEY2= and to leave the others unchanged - their data codes will be their rating values. For items 5 and 6, 0 0, 1
0, 2 1, 3 2; for item 7, 0 0, 1 0, 2 0, 3 1. Responses to other items are already entered correctly as 0, 1, 2, or 3.

CODES =0123 valid codes
RESCORE=0001110000 rescore items 5,6,7
KEY1 =****223*** keyed for selected items
KEY2 =****33X*** the X will be ignored
^ read these columns vertically

148. NI number of items

The total number of items to be read in (including those to be deleted by IDFILE= etc.). NI= is limited to about
30000 for one column responses or about 15000 for two column responses in the standard program. NI= is
usually the length of your test (or the total number of items in all test forms to be combined into one analysis).

Example: If there are 230 items in your test, enter

NI=230 ; 230 items

It is easy to miscount the NI= column. Scroll to the top of the Winsteps screen and check column
positions:

Input in process..
Input Data Record:
  1   2
1234567890123456789012345678
Richard M 111111100000000000
^P  ^I  ^N
35 KID Records Input.
149. **NORMAL** normal distribution for standardizing fit

The standard generally matches the statistics used in BTD and RSA.

Specifies whether distribution of squared residuals is hypothesized to accord with the chi-square or the normal distribution. Values of $t$ standardized fit statistics are obtained from squared residuals by means of one of these distributions.

NORMAL=N $t$ standardized fit statistics are obtained from the squared residuals by means of the chi-square distribution and the Wilson-Hilferty transformation (the standard).


150. **OSORT** option/distractor sort

Specifies the order in which categories, options and distractors are listed within items in the Item Category/Option/Distractor Tables, such as Table 13.3

OSORT = D Options within items are listed in their Data order in CODES = (the standard).

OSORT = S or V or " Score or Value: Options are listed in their Score Value order.

OSORT = A or M Average or Measure: Options are listed in their Average Measure order.

Example: List distractors in original CODES = order:

OSORT=D
CODES=000102
XWIDE=2

<table>
<thead>
<tr>
<th>ENTRY DATA CODE VALUE</th>
<th>DATA COUNT % AVERAGE S.E. OUTFIT</th>
<th>ACT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>23 A 00 0</td>
<td>44 59</td>
</tr>
<tr>
<td></td>
<td>01 1</td>
<td>20 27</td>
</tr>
<tr>
<td></td>
<td>02 2</td>
<td>11 15</td>
</tr>
<tr>
<td></td>
<td>MISSING ***</td>
<td>1 1</td>
</tr>
<tr>
<td></td>
<td>5 B 00 0</td>
<td>47 63</td>
</tr>
<tr>
<td></td>
<td>01 1</td>
<td>19 25</td>
</tr>
<tr>
<td></td>
<td>02 2</td>
<td>9 12</td>
</tr>
</tbody>
</table>

151. **OUTFIT** sort misfits on infit or outfit

Other Rasch programs may use infit, outfit or some other fit statistic. There is no one "correct" statistic. Use the one you find most useful.

Specifies whether mean-square infit or mean-square outfit is used as your output sorting and selection criterion for the diagnosis of misfits.

OUTFIT=Y For each person, the greater of the outfit mean-square and infit mean-square is used as the fit statistic for sorting and selection (the standard).

OUTFIT=N Infit mean-square only is used as the fit statistic for sorting and selection.
152. **PAFILE person anchor file**

The `PAFILE=` from one run can be used unedited as the person anchor file, `PAFILE=`, of another.

The person parameter values (thetas) can be anchored (fixed) using `PAFILE=`. Person anchoring can also facilitate test form equating. The persons common to two test forms can be anchored at the values for one form. Then the measures constructed from the second form will be equated to the measures of the first form. Other measures are estimated in the frame of reference defined by the anchor values.

In order to anchor persons, an anchor file must be created of the following form:
1. Use one line per person-to-be-anchored.
2. Type the sequence number of the person, a blank, and the measure value (in logits if USCALE=1, otherwise your user-rescaled units) at which to anchor the person.

Anything after ";" is treated as a comment.

**PAFILE = filename**

Person anchor information is in a file containing lines of format

```
person entry number       anchor value
person entry number       anchor value
```

**PAFILE=*

Person anchor information is in the control file in the format

```
PAFILE=*
person entry number       anchor value
person entry number       anchor value
```

**PAFILE=$SnnEnn or PAFILE=$SnWnn**

Person anchor information is in the person data records using the column selection rules, e.g., starting in column Snn and ending in column Enn or of width Wnn. Blanks of non-numeric values indicate no anchor value. `PAFILE=$S10E12` or `PAFILE=$S10W2` means anchor information starts in column 10 and ends in column 12 of the person's data record (not person label). This can be expanded, e.g., `PAFILE = $S23W1+"."+$S25W2` places the columns next to each other (not added to each other).

Example: The third person in the test is to be anchored at 1.5 logits, and the eighth at -2.7.
1. Create a file named, say, "PERSON.ANC"
2. Enter the line "3 1.5" into this file, meaning "person 3 is fixed at 1.5 logits".
3. Enter the line "8 -2.7", meaning "person 8 is fixed at -2.7 logits".
4. Specify, in the control file,
```
PAFILE=PERSON.ANC
CONVERGE=L  ; only logit change is used for convergence
LCONV=0.005  ; logit change too small to appear on any report.
```

or, enter directly into the control file
```
PAFILE=*  
3 1.5
8 -2.7
*
CONVERGE=L  ; only logit change is used for convergence
LCONV=0.005  ; logit change too small to appear on any report.
```

or, include the anchor information in the data record
```
PAFILE=$S1E4
NAME1=5
ITEM1=11
NI=12
CONVERGE=L  ; only logit change is used for convergence
LCONV=0.005  ; logit change too small to appear on any report.
```
To check: "A" after the measure means "anchored"

<table>
<thead>
<tr>
<th>ENTRY</th>
<th>RAW</th>
<th>COUNT</th>
<th>MEASURE</th>
<th>ERROR</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>PTMEA</th>
<th>PERSONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>32</td>
<td>35</td>
<td>1.5</td>
<td>.05</td>
<td>.80</td>
<td>-.3</td>
<td>.32</td>
<td>.6</td>
</tr>
</tbody>
</table>

153. **PAIRED** correction for paired comparison data

Paired comparison data is entered as only two observations in each row (or each column). The raw score of every row (or column) is identical. In the simplest case, the "winner" receives a '1', the "loser" a '0', and all other column (or rows) are left blank, indicating missing data.

Example: Data for a chess tournament is entered. Each row is a player. Each column a match. The winner is scored '2', the loser '0' for each match. For draws, each player receives a ‘1’.

Example: Data for a chess tournament is entered. Each row is a player. Each column a match. The winner is scored '2', the loser '0' for each match. For draws, each player receives a ‘1’.

**PAIRED=YES** ; paired comparisons

CODES=012 ; valid outcomes

NI=56 ; number of matches

154. **PANCHQU** anchor persons interactively

If your system is interactive, persons to be anchored can be entered interactively by setting PANCHQ=Y before the &END line. If you specify this, you will be asked if you want to anchor any persons. If you respond "yes", it will ask if you want to read these anchored persons from a file; if you answer "yes" it will ask for the file name and process that file in the same manner as if PAFILE= had been specified. If you answer "no", you will be asked to enter the sequence number of each person to be anchored, one at a time, along with the logit (or user-rescaled by UANCHOR=, USCALE=, UMEAN=) calibration. When you are finished, enter a zero.

Example: You are doing a number of analyses, anchoring a few, but different, persons each analysis. This time, you want to anchor person 4.

Enter on the DOS control line, or in the control file:

```
PANCHQ=Y
CONVERGE=L ; only logit change is used for convergence
LCONV=0.005 ; logit change too small to appear on any report.
```

You want to anchor person 4:

```
WINSTEPS asks you:
DO YOU WANT TO ANCHOR ANY PERSONS?
respond YES(Enter)
DO YOU WISH TO READ THE ANCHORED PERSONS FROM A FILE?
respond NO(Enter)
INPUT PERSON TO ANCHOR (0 TO END): 
respond 4(Enter) (the first person to be anchored)
INPUT VALUE AT WHICH TO ANCHOR PERSON:
respond 1.45(Enter) (the first anchor value)
```

155. **PCORFIL** person residual correlation file

This writes out the Table of inter-person correlations which is the basis of the principal components analysis of
residuals. Missing data: for these Winsteps substitutes their expectations when possible. For residuals and standardized residuals, these are 0. Items with extreme scores (minimum possible or maximum possible): Winsteps drops these from the correlation computation. The reason for these choices is to make the principal components analysis of residuals as meaningful as possible.

**PCORFILE** = file name

Example: Write out the Table of inter-person residual correlations. **PCORFIL=file.txt** - Then file.txt contains, for SF.txt,

<table>
<thead>
<tr>
<th>Person</th>
<th>Person</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>.23</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>-.31</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>.13</td>
</tr>
<tr>
<td>....</td>
<td>....</td>
<td>....</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>-.10</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>-.02</td>
</tr>
<tr>
<td>....</td>
<td>....</td>
<td>....</td>
</tr>
</tbody>
</table>

Example 2: When **PCORFILE=** is selected on the **Output Files** menu or **MATRIX=** **YES**, the Data Format: Matrix option can be selected:

<table>
<thead>
<tr>
<th>Data Format:</th>
<th>Matrix</th>
<th>List</th>
</tr>
</thead>
</table>
|              | ![Temporary file: automatic file name](image)

This produces:

```
1.0000  .2265  -.3147  .1306  ....
 .2265  1.0000  -.1048  -.0222  ....
-.3147  -.1048  1.0000  .0403  ....
......  ......  ......  ......  ....
```

156. **PDELETE** person one-line item deletion

A one-line list of persons to be deleted or reinstated can be conveniently specified with **PDELETE=**. This is designed to be used in the post-analysis Specification pull-down menu box.

The formats are:

- **PDELETE= 3** ; an entry number: delete person 3
- **PDELETE= 6 1** ; delete persons 6 and 1
- **PDELETE= 2-5** ; delete persons 2, 3, 4, 5
- **PDELETE= +3-10** ; delete all persons, then reinstate persons 3 to 10.
- **PDELETE= 4-20 +8** ; delete persons 4-20 then reinstate person 8
- **PDELETE=3,7,4,10** ; delete persons 3, 7, 4, 10. Commas, blanks and tabs are valid separators. Commas are useful at the "Extra information prompt.

Example 1: After an analysis is completed, delete criterion-performance synthetic cases from the reporting. In the **Specification pull-down box**:

```
PDELETE=16 25 87
```

Example 2: Delete all except persons 5-10 and report

```
Specification menu box: PDELETE=+5-10
```

Output Tables Menu

```
Now reinstate item 11 and report items 5-11.
```

Specification menu box: **PDELETE= 1 +11** ; item 1 is already deleted, but prevents deletion of all except +11.

Output Tables Menu

157. **PDELQU** delete persons interactively

Persons to be deleted or selected can be entered interactively by setting **PDELQU=Y**. If you specify this, you will
be asked if you want to delete any persons. If you respond "yes", it will ask if you want to read these deleted persons from a file; if you answer "yes" it will ask for the file name and process that file in the same manner as if PDFILE= had been specified. If you answer "no", you will be asked to enter the sequence number or numbers of persons to be deleted or selected one line at a time, following the rules specified for IDFILE=. When you are finished, enter a zero.

Example: You are doing a number of analyses, deleting a few, but different, persons each analysis. You don’t want to create a lot of small delete files, but rather just enter the numbers at the terminal, so specify:

PDELQU=Y

You want to delete persons 23 and 50. WINSTEPS asks you:

DO YOU WANT TO DELETE ANY PERSONS?
respond YES(Enter)
DO YOU WISH TO READ THE DELETED PERSONS FROM A FILE?
respond NO(Enter)
INPUT PERSON TO DELETE (0 TO END):
respond 23(Enter) (the first person to be deleted)
INPUT PERSON TO DELETE (0 TO END): 50(Enter)
INPUT PERSON TO DELETE (0 TO END): 0(Enter) (to end deletion)

158. PDFILE person deletion file

Deletion or selection of persons from a test to be analyzed, but without removing their responses from your data file, is easily accomplished by creating a file in which each line contains the sequence number of a person or persons to be deleted or selected (according to the same rules given under IDFILE=), and then specifying this file by means of the control variable, PDFILE=, or enter the deletion list in the control file using PDFILE=*.

Example 1: You wish to delete the fifth and tenth persons from this analysis.
1. Create a file named, say, “PERSON.DEL”
2. Enter into the file, the lines:
   5
   10
3. Specify, in the control file,
   PDFILE=PERSON.DEL
   or, enter directly into the control file,
   PDFILE=*
   5
   10
   *

Example 2: The analyst wants to delete the most misfitting persons reported in Table 6.
1. Set up a standard control file.
2. Specify
   PDFILE=*  
   *
3. Copy the target portion of Table 6.
4. Paste it between the "*
5. Delete characters before the entry numbers.
6. Type ; after the entry numbers to make further numbers into comments.

TITLE = 'Example of person deletion list from Table 6'
IDFILE = *
Delete the border character before the entry number
; ENTRY RAW INFIT OUTFIT
; NUM SCORE COUNT MEASURE ERROR MNSQ ZSTD MNSQ ZSTD PTBIS PUP
73 ; 21 22 .14 .37 .95 -.3 1.03 .2 B-.19 SAN
75 ; 16 22 -.56 .39 .95 -.3 1.03 .2 C-.19 PAU
Enter the ; to make other numbers into comments
159. **PDROPEXTREME** drop persons with extreme scores

Unanchored persons with extreme (zero or perfect, minimum possible or maximum possible) scores provide no information for estimating item measures, but they are reported and included in summary statistics. To remove them:

PDROPEXTREME = No ; do not drop extreme persons (standard)

PDROPEXTREME = Yes or All ; drop zero and perfect scores

PDROPEXTREME = Zero or Low or Bottom or Minimum ; drop zero or minimum-possible scores

PDROPEXTREME = Perfect or High or Top or Maximum ; drop perfect or maximum-possible scores

Example: The data file contains many data records of persons who did not attempt this test and so were scored 0. They are skewing the test statistics:

PDROPEXTREME = Zero

160. **PERSON** title for person labels

Up to 12 characters to use in table headings to describe the persons, e.g.

PERSON=KID

Choose a word which makes its plural with an "s", e.g. KIDS.

If you specify, PERSON=kid, then the plural will be "kids"

161. **PFILE** person output file

PFILE=filename produces an output file containing the information for each person. This file contains 4 heading lines (unless HLINES=N), followed by one line for each person containing:

Columns:

<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>Format</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>A1</td>
<td>Blank or &quot;;&quot; if HLINES=Y and there are no responses or deleted or extreme (status =0,-1, -2, -3)</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>I5</td>
<td>1. The person sequence number (ENTRY)</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>F8.2</td>
<td>2. Person's measure (user-rescaled by UMEAN=, USCALE=, UDECIM=) (MEASURE)</td>
</tr>
<tr>
<td>15</td>
<td>17</td>
<td>I3</td>
<td>3. The person's status (STATUS)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td>2 = Anchored (fixed) measure</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td>1 = Estimated measure</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td></td>
<td>0 = Extreme maximum (estimated using EXTRSC=)</td>
</tr>
<tr>
<td></td>
<td>-1</td>
<td></td>
<td>-1 = Extreme minimum (estimated using EXTRSC=)</td>
</tr>
<tr>
<td></td>
<td>-2</td>
<td></td>
<td>-2 = No responses available for measure</td>
</tr>
<tr>
<td></td>
<td>-3</td>
<td></td>
<td>-3 = Deleted by user</td>
</tr>
<tr>
<td>18</td>
<td>25</td>
<td>F7.1</td>
<td>4. The number of responses used in measuring (COUNT) or the observed count (TOTAL=Y)</td>
</tr>
<tr>
<td>26</td>
<td>34</td>
<td>F8.1</td>
<td>5. The raw score used in measuring (SCORE) or the observed score (TOTAL=Y)</td>
</tr>
</tbody>
</table>
The format descriptors are:

In = Integer field width n columns
Fn.m = Numeric field, n columns wide including n-m-1 integral places, a decimal point and m decimal places
An = Alphabetic field, n columns wide
Nx = n blank columns.

When CSV=Y, commas separate the values with quotation marks around the "Person name". When CSV=T, the commas are replaced by tab characters.

Example: You wish to write a file on disk called "STUDENT.MES" containing the person statistics for import later into a student information database:

PFILE=STUDENT.MES

When W300=Yes, then this is produced in Winsteps 3.00, 1/1/2000, format:

Columns:
Start   End   Description
---     ---   -------------------------------------
1 1 A1    Blank or ";" if HLINES=Y and there are no responses or deleted (status = -2, -3)
2 6 I5    1. The Person sequence number (ENTRY)
7 14 F8.2 2. Person's calibration (user-rescaled by UMEAN=, USCALE=, UDECIM=) (MEASURE)
15 17 I3    3. The Person's status (STATUS)
18 23 I6    4. The number of responses used in calibrating (COUNT) or the observed count (TOTAL=Y)
24 30 I6    5. The raw score used in calibrating (SCORE) or the observed score (TOTAL=Y)
31 37 F7.2 6. Person calibration's standard error (user-rescaled by USCALE=, UDECIM=) (ERROR)
38 44 F7.2 7. Person mean square infit (IN.MSQ)
45 51 F7.2 8. Person infit: t standardized (ZSTD), locally t standardized (ZEMP) or log-scaled (LOG)
52 58 F7.2 9. Person mean square outfit (OUT.MSQ)
59 65 F7.2 10. Person outfit: t standardized (ZSTD), locally t standardized (ZEMP) or log-scaled (LOG)
66 72 F7.2 11. Person displacement (user-rescaled by USCALE=, UDECIM=) (DISPLACE)
73 79 F7.2 12. Person by test-score correlation: point-biserial (PTBS) or point-measure (PTME)
80 80 1X   15. Blank
81 132+ A30+ 16. Person name (NAME)

Example of standard PFILE=

; PERSON Knox Cube Test (Best Test Design p.31) Nov 10 15:40 2005
;ENTRY MEASURE STTS COUNT SCORE ERROR IN.MSQ IN.ZSTD OUT.MS OUT.ZSTD DISPL PTME WEIGHT OBSMA EXPMA
NAME
1 -3.08 1 14.0 4.0 .83 .61 -1.31 .29 -.14 .00 .80 1.00 92.9 84.8
Richard M
35 -6.78 -1 14.0 .0 1.88 1.00 .00 1.00 .00 .00 .01 1.00 .0 0.0

148
162. **PMA\_P** person label on person map: Tables 1, 16

This specifies what part of the data record is to be used on the person map.

It's format is \texttt{PMA\_P} = \$S..W.. or \$S..E.. using the column selection rules.

\$S..W.. e.g., \$S2W13 means that the person label to be shown on the map starts in column 2 of the person label and is 13 columns wide.

\$S..E.. e.g., \$S3E6 means that the person label to be shown on the map starts in column 3 of the person label and ends in column 6.

These can be combined, and constants introduced, e.g,
\texttt{PMA\_P} = \$\texttt{S3W2+}/"\texttt{+S7W2}

If the person label is "KH323MXTR", the person label on the map will be "32/XT"

The length of \texttt{PMA\_P} overrides \texttt{NAMLMP}=

163. **PRCOMP** residual type for principal components analyses in Tables 23, 24

Principal components analysis of item-response or person-response residuals can help identify structure in the misfit patterns across items or persons. The measures have been extracted from the residuals, so only uncorrelated noise would remain, if the data fit the Rasch model.

\texttt{PRCOMP=S} or \texttt{Y} Analyze the standardized residuals, (observed - expected)/(model standard error).
Simulation studies indicate that \texttt{PRCOMP=S} gives the most accurate reflection of secondary dimensions in the items.

\texttt{PRCOMP=R} Analyze the raw score residuals, (observed - expected) for each observation.

\texttt{PRCOMP=L} Analyze the logit residuals, (observed - expected)/(model variance).

\texttt{PRCOMP=O} Analyze the observations themselves.

Example 1: Perform a Rasch analysis, and then see if there is any meaningful other dimensions in the residuals:
\texttt{PRCOMP=S} Standardized residuals

Example 2: Analysis of the observations themselves is more familiar to statisticians.
\texttt{PRCOMP=O} Observations

164. **PSELECT** person selection criterion

Persons to be selected may be specified by using the \texttt{PSELECT=} instruction to match characters within the person name. Persons deleted by \texttt{PDFILE=} etc. are never selected by \texttt{PSELECT=}.

This can be done \textbf{before} analysis in the control file or with "Extra specifications". It can also be done \textbf{after} the analysis using the "Specification" pull-down menu.

Control characters to match person name:

\texttt{?} matches any character

\texttt{(. .)} braces characters which can match a single character: \texttt{\{ABC\}} matches A or B or C.

\texttt{(. . .)} matches single characters in a range. \texttt{\{0-9\}} matches digits in the range 0 to 9.

\texttt{(. . . .)} matches a single "-" \texttt{\{AB-\}} matches A or B or ".".

\texttt{*} matches any string of characters - must be last selection character.

Other alphanumeric characters match only those characters.

Each \texttt{PSELECT=} performed using the "Specification" pull-down menu selects from all those analyzed. \textbf{For...}
incremental selections, i.e., selecting from those already selected, specify +PSELECT=

PSELECT= works best for single column indicators. It is usually possible to convert 2 columns (such as age) to one column using the rectangular copy feature in Word or in EXCEL =code(age+65) converts "age" to letters, starting with A as 0.

Example 1: Select for analysis only persons with M in the 5th column of person name. Person name starts in column 6 of the data record:

```
NAME1=6  Person name field starts in col. 6
NAMLEN=8  Person name field is 8 characters long
PSELECT=????M*  Column 5 of person name is sex
```

```
END NAMES
```

```
xxxxxBPL M J 01101000101001 selected
xxxxxMEL F S 01001000111100 omitted
```

Example 2: Select for analysis all persons with code "A 4" in columns 2-4 of their names. Person name starts in column 23, so target characters starts in column 24:

```
NAME1=23  person name starts in column 23
PSELECT="A 4"*  quotes because a blank is included. A is in col. 2 etc.
ZA 4PQRS selected
```

Example 3: Select all Male (M in column 2) persons who are Asian or Hispanic (A or H in column 4):

```
PSELECT=??M?{AH}*  
```

```
1M3A456 MPQRS  selected
1M6H689 ABCDE  selected
1X2A123 QWERT omitted
```

Example 4: Select Males (M in column 8) in School 23 (023 in column 14-16):

```
PSELECT=???????M?????023*  
Selects: 1234567
```

Example 5: Select codes 1,2,3,4,5,6,7, in column 2:

```
PSELECT=?{1-7}*
```

Example 6: Analyze only males (column 4 or person-id). Then report only School C (column 1). Then only report Grades 4 and 5 (column 2) in School C.

```
PSELECT=???M*  in the Control file or at the Extra Specifications prompt.
PSELECT=C*  using the Specification pull-down menu, after the analysis
+PSELECT=?{45}*  using the Specification pull-down menu.
```

165. **PSORT** column within person label for alphabetical sort in Table 19

Table 19 lists persons alphabetically. Table 1 and Table 16 list them alphabetically within lines. Ordinarily, the whole person name is used. Select the sorting columns in the person labels using the column selection rules, e.g., starting in column Snn and ending in column Enn or of width Wnn.

Example 1: The person name is entered in the data file starting in column 20. It is a 6-digit student number followed by a blank and then gender identification in column 8 of the person name. Sort by gender identification for Table 19, then by student number.

```
NAME1=20
NAMLEN=8  ; student number + gender
PSORT=8+1-6  ; same as $S8W1+$S8W6 alphabetical sort on gender
TABLES=1111111111111111111111
&END
```

```
END NAMES
```
Example 2: The person name contains several important classifiers. Table 19 is needed for each one:

NAME1=14  Person name starts in column 14
ITEM1=24  Response start in column 24
TFILE=* 
19 - - - 1  sort starts with column 1 of person name 
19 - - - 8  sort starts with column 8 of person name 
19 - - - 6  sort starts with column 6 of person name upto the end of the person name 
   - entered as place-holders, see TFILE= 
   * 
&END 
END NAMES

Example 3: A version of Table 19, sorted on person name column 6, is to be specified on the DOS command line or on the Extra Specifications line. Commas are used as separators, and ".-" as place-holders: 
TFILE=*  19,-,-,-,6  *

166.  PSUBTOTAL  columns within person label for subtotals in Table 28

This specifies what part of the data record is to be used to classify persons for subtotal in Table 28.

With tab-separated data and the subtotal indicator in a separate field from the Person label, specify the subtotal field as the person label field using NAME1=, then PSUBTOTAL=$S1W1

$S..W.. e.g., $S2W13 means that the label to be shown on the map starts in column 2 of the person label and is 13 columns wide.
$S..E.. e.g., $S3E6 means that the label to be shown on the map starts in column 3 of the person label and ends in column 6.

These can be combined, and constants introduced, e.g,
PSUBTOTAL=$S3W2+"/"+$S7W2

If the person label is "KH323MXTR", the subgrouping will be shown as "32/XT"

Format 2: PSUBTOTAL=* 
This is followed by a list of subgroupings, each on a new line:

   PSUBTOTAL=* 
   $S1W1+$S7W2  ; Subtotals reported for person classifications according to these columns 
   $S3E5  ; Subtotals reported for person classifications according to these columns 
   *

Example: Subtotal by first letter of person name:

   PSUBTOTAL=$S1W1 
   TFILE=* 
   27 ; produce the subtotal report 
   *

Here is a subtotal report (Table 28) for person beginning with "R"
"R" SUBTOTAL FOR 8 NON-EXTREME PUPILS

<table>
<thead>
<tr>
<th>RAW</th>
<th>MODEL</th>
<th>INFIT</th>
<th>OUTFIT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SCORE</td>
<td>COUNT</td>
<td>MEASURE</td>
</tr>
<tr>
<td>MEAN</td>
<td>28.1</td>
<td>25.0</td>
<td>4.04</td>
</tr>
<tr>
<td>S.D.</td>
<td>5.5</td>
<td>.0</td>
<td>6.63</td>
</tr>
<tr>
<td>MAX.</td>
<td>38.0</td>
<td>25.0</td>
<td>16.30</td>
</tr>
<tr>
<td>MIN.</td>
<td>19.0</td>
<td>25.0</td>
<td>-6.69</td>
</tr>
</tbody>
</table>

+-----------------------------------------------------------------------------+
| REAL RMSE | 3.63 | ADJ.SD | 5.54 | SEPARATION | 1.52 | PUPIL RELIABILITY | .70 |
| MODEL RMSE | 3.48 | ADJ.SD | 5.64 | SEPARATION | 1.62 | PUPIL RELIABILITY | .72 |
| S.E. OF PUPIL MEAN | 2.50 |

+-----------------------------------------------------------------------------+
| WITH 2 EXTREME = TOTAL 10 PUPILS | MEAN = 3.05, S.D. = 28.19 |
| REAL RMSE | 8.88 | ADJ.SD | 26.75 | SEPARATION | 3.01 | PUPIL RELIABILITY | .90 |
| MODEL RMSE | 8.83 | ADJ.SD | 26.77 | SEPARATION | 3.03 | PUPIL RELIABILITY | .90 |
| S.E. OF PUPIL MEAN | 9.40 |

MAXIMUM EXTREME SCORE: 1 PUPILS
MINIMUM EXTREME SCORE: 1 PUPILS
LACKING RESPONSES: 1 PUPILS
DELETED: 1 PUPILS

167. PTBIS compute point-biserial correlation coefficients

PTBIS=Y Compute and report conventional point bi-serial correlation coefficients, r_pbis. These are reported not only for items but also for persons. Extreme measures are included in the computation, but missing observations are omitted. In Rasch analysis, r_pbis is a useful diagnostic indicator of data miscoding or item miskeying: negative or zero values indicate items or persons with response strings that contradict the variable. The Biserial correlation can be computed from the Point-biserial.

PTBIS=N (or PTBIS=RPM). Compute and report point-measure correlation coefficients, r_pm or RPM, shown as PTMEA. These are reported in Tables 14 and 18 for items and persons. They correlate an item's (or person's) responses with the measures of the encountered persons (or items). r_pm maintains its meaning better than r_pbis in the presence of missing observations. Extreme measures are included in the computation. Negative or zero values indicate response strings that contradict the variable.

The formula for this product-moment correlation coefficient is:

\[ r_{pbis} = \frac{\left( \sum (x-x \bar{bar})(y-y \bar{bar}) \right)}{\sqrt{\left( \sum (x-x \bar{bar})^2 \right) \cdot \left( \sum (y-y \bar{bar})^2 \right)}} \]

where \( x \) = observation for this item (or person), \( y \) = total score for this person **omitting this item** (or for this item **omitting this person**).

Conventional computation of \( r_{pbis} \) includes persons with extreme scores. These correlation can be obtained by forcing observations for extreme measures into the analysis. The procedure is:

1) Perform a standard analysis but specify `PFILE=`, `IFILE=` and, if a polytomous analysis, `SFILE=.`

2) Perform a second analysis setting `PAFILE=` to the name of the `PFILE=` of the first analysis, `IAFILE=` to the name of the `IFILE=`, and `SAFILE=` to the name of the `SFILE=.` The PTBIS of this second analysis is the conventional `r_pbis`, which includes extreme scores.

RPM (PTMEA) is reported instead of PTBIS when PTBIS=N or PTBIS=RPM is specified. RPM is the point-measure correlation, r_pm. It is computed in the same way as the point bi-serial, except that Rasch measures replace total scores.

Example: For rank-order data, point-biserials are all -1. So specify Point-measure correlations.

PTBIS=NO
168. **PVALUE proportion correct or average rating**

An important statistic in classical item analysis is the item p-value, the proportion of the sample succeeding on a dichotomous item. In Winsteps, this is interpreted as the average of the responses to the item whether dichotomous or polytomous.

PVALUE = NO  Do not report the item p-value.

PVALUE = YES  Report item p-values in the IFILE= and Tables 6, 10, etc.

Example:  To parallel a classical analysis of an MCQ data set, it is desired to report the raw scores (including extreme persons and items), the point-biserials and the p-values.

TOTALSCORE= YES ; report original total raw scores
PTBIS = YES ; report point-biserial, not point-measure correlations
PVALUE = YES ; report p-values

+----------------------------------------------------------------------------------------+
|ENTRY   TOTAL                       |   INFIT  |  OUTFIT  |PTBIS| P-  |                 |
|NUMBER  SCORE  COUNT  MEASURE  ERROR|MNSQ  ZSTD|MNSQ  ZSTD|CORR.|VALUE| ITEM            |
|------------------------------------+----------+----------+-----+-----+-----------------|
|     1     35     35   -6.59    1.85| MINIMUM ESTIMATED MEASURE | 1.00| 1= 1-4          |
|     4     32     35   -4.40     .81| .90    .0| .35    .8|  .48|  .91| 4= 1-3-4        |
|   12      6     35    2.24     .55|1.16    .6|1.06    .5|  .26|  .17| 12=1-3-2-4-3    |
|   18      0     35    6.13    1.84| MAXIMUM ESTIMATED MEASURE |  .00| 18=4-1-3-4-2-1-4|
+----------------------------------------------------------------------------------------+

169. **PWEIGHT person (case) weighting**

PWEIGHT= allows for differential weighting of persons. The standard weights are 1 for all persons. To change the weighting of items, specify PWEIGHT=*

Raw score, count, and standard error of measurement reflect the absolute size of weights as well as their relative sizes.
Measure, infit and outfit and correlations are sensitive only to relative weights.

If you want the standard error of the final weight-based measure to approximate the S.E. of the unweighted measure, then ratio-adjust case weights so that the total of the weights is equal to the total number of independent observations.

Formats are:

PWEIGHT= file name
the weights are in a file of format:
person entry number   weight

PWEIGHT=*
person entry number   weight

* *

PWEIGHT= $S...$W... or $S...$E...
weights are in the data records using the column selection rules, e.g., starting in column S... with a width of W... or starting in column S and ending in column E. This can be expanded, e.g., PWEIGHT = $S23W1+"."+$S25W2 places the columns next to each other (not added to each other)

Example 1:
In a sample of 20 respondents, person 1 is to be given a weight of 2.5, all other persons have a weight of 1.

PWEIGHT=*
  1  2.5
  2-20   1

*
A better weighting, which would make the reported item standard errors more realistic by maintaining the original total sum of weights at 20, is:

\[ \text{PWEIGHT= } * \]

\[ 1 \quad 2.33 \quad ; \quad 2.5 * 0.93 \]

\[ 2-20 \quad 0.93 \quad ; \quad \text{the sum of all weights is 20.0} \]

<table>
<thead>
<tr>
<th>ENTRY</th>
<th>RAW</th>
<th>MODEL</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>PTMEA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NUMBER</td>
<td>SCORE</td>
<td>COUNT</td>
<td>MEASURE</td>
<td>S.E.</td>
<td>MNSQ</td>
<td>ZSTD</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>10</td>
<td>-1.20</td>
<td>0.85</td>
<td>.65</td>
<td>-.9</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>10</td>
<td>1.70</td>
<td>1.17</td>
<td>.18</td>
<td>-1.2</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>10</td>
<td>1.70</td>
<td>1.17</td>
<td>2.04</td>
<td>1.2</td>
</tr>
</tbody>
</table>

**Example 2:**

The data records contain the weights in columns 16-18 of the record. The person label starts in column 15, so this is also column 2 of the person label

\[ \text{PWEIGHT= } $C16W3 \quad ; \quad \text{or } $C16E18 \quad ; \quad \text{column in data record} \]

or

\[ \text{NAME1 = 15} \quad ; \quad \text{start of person label} \]

\[ \text{NAMELEN = 20} \quad ; \quad \text{length of person label} \]

\[ \text{PWEIGHT= } $S2W3 \quad ; \quad \text{location in person label} \]

&END

... END NAMES

10110110011001 0.5 Person A
01001100011011 0.7 Person B

......

**Example 3:**

Person 4 is a dummy case, to be given weight 0.

\[ \text{PWEIGHT= } * \]

\[ 4 \quad 0 \quad ; \quad \text{Person 4 has weight 0, other persons have standard weight of 1.} \]

170. QUOTED quote-marks around labels

Non-numeric values in the output files can be placed within quote-marks. This is required by some software in order to decode internal blanks within labels correctly. These apply to comma-separated and tab-separated output files.

**QUOTED=Y** "non-numeric values within quotation marks"

**QUOTED=N** non-numeric values without quotation marks.

Example:  Produce an SFILE=

\[ \text{CSV=Y} \quad ; \quad \text{produce a comma-separated output file} \]

\[ \text{QUOTED=Y} \quad ; \quad \text{with labels in quotation marks} \]

\[ ; \text{STRUCTURE MEASURE ANCHOR FILE} \]

\[ ; \text{CATEGORY},"Rasch-Andrich Threshold" \]

\[ 0,.00 \]

\[ 1,-.86 \]

\[ 2,.86 \]

\[ \text{QUOTED=N} \quad ; \quad \text{labels without quotation marks} \]

\[ ; \text{STRUCTURE MEASURE ANCHOR FILE} \]

\[ ; \text{CATEGORY},\text{Rasch-Andrich Threshold} \]

\[ 0,.00 \]

\[ 1,-.86 \]

\[ 2,.86 \]
171. **RCONV** score residual at convergence

*Scores increment in integers so that 0.1 is about as precise a recovery of observed data as can be hoped for.*

Specifies what value the largest score residual, corresponding to any person measure or item calibration, must be less than in the iteration just completed for iteration to cease. The current largest value is listed in Table 0, and displayed on your screen. In large data sets, the smallest meaningful logit change in estimates may correspond to score residuals of several score points. See convergence considerations.

The standard setting is **CONVERGE=**"E", so that iteration stops when either **LCONV=** or **RCONV=** is satisfied. (Note: this depends on Winsteps version - and may explain differences in converged values.)

Example: To set the maximum score residual, when convergence will be accepted, at 5 score points and maximum logit change in estimates of .01 logits. Your data consists of the responses of 5,000 students to a test of 250 items.

```
NI=250 ; 250 items
RCONV=5 ; score residual convergence at 5 score points
LCONV=.01 ; this is the standard.
CONVERGE=Both :convergence when both RCONV= and LCONV= are met
```

172. **REALSE** inflate S.E. for misfit

*The modeled, **REALSE=N**, standard errors of measure estimates (abilities and difficulties) are the smallest possible errors. These always overstate the measurement precision.*

Controls the reporting of standard errors of measures in all tables.

- **REALSE=N**
  Report modeled, asymptotic, standard errors (the standard).

- **REALSE=Y**
  Report the modeled standard errors inflated by the square root of the infit mean square, when it is greater than 1.0. This inflates the standard error to include uncertainty due to overall lack of fit of data to model.

See [Standard Errors: Model and Real](#) for more details.

173. **RESCORE** response recoding

The responses in your data file may not be coded as you desire. The responses to some or all of the items can be rescored or keyed using **RESCORE=**. **RESCORE=** and **NEWSCORE=** are ignored when **KEYn=** is specified, except as below. If rescoring implies that the items have different rating (or partial credit) scale structures, **ISGROUPS=** may also be required.

**RESCORE=** has three forms: **RESCORE=1101110** and **RESCORE=* list * and RESCORE=*filename**

- **RESCORE=**" " or 2 or is omitted
  All items are recoded using **NEWSCORE=**. **RESCORE=2** is the standard when **NEWSCORE=** is specified.

- **RESCORE=** some combination of 1's and 0's
  Only items corresponding to 1's are recoded with **NEWSCORE=** or scored with **KEYn=**. When **KEYn** is specified, **NEWSCORE=** is ignored.

If some, but not all, items are to be recoded or keyed, assign a character string to **RESCORE=** in which "1" means "recode (key) the item", and "0" (or blank) means "do not recode (key) the item". The position of the "0" or "1" in the **RESCORE=** string must match the position of the item-response in the item-string.

When **XWIDE=2** or more, then either
- (a) Use one character per **XWIDE** and blanks,
Example 1: The original codes are "0" and "1". You want to reverse these codes, i.e., 1 0 and 0 1, for all items.

```
NI=8
XWIDE=2
RESCORE='1 0 1 0 1 0 1 1'
```

```
or (b) Use one character per item with no blanks
NI=8
XWIDE=2
RESCORE='10101011'
```

Example 2: Your data is coded "0" and "1". This is correct for all 10 items except for items 1 and 7 which have the reverse meaning, i.e. 1 0 and 0 1.

```
NI=10
ten items
CODES =01
valid response codes are 0 and 1 (the standard)
NEWSCORE=10
desired response scoring
RESCORE=2
rescore all items - this line can be omitted
```

(a) old method - which still works:

```
NEWSCORE=10
revised scoring
RESCORE=100001000
only for items 1 and 7
```

```
or
NEWSCORE=20
RESCORE=*  
1 1 item 1 is to be rescored  
7 1 item 7 is to be rescored  
*
```

(b) new method - recommended:

```
IVALUE1 =10   revised scoring
IVALUE0 =01   scoring unchanged, so this line can be omitted.
IREFER   =100001000 only for items 1 and 7
```

If XWIDE=2, use one or two columns per RESCORE= code, e.g., " 1" or "1 " mean recode (key). " 0" or "0" mean do not recode (key).

Example 3: The original codes are " 0" and " 1". You want to reverse these codes, i.e., 1 0 and 0 1, for items 1 and 7 of a ten item test.

```
NI  =10   ; ten items
XWIDE =2   ; two characters wide
CODES  =" 0 1"  ; original codes
NEWSCORE=" 1 0"  ; new values
RESCORE =" 1000010001000" ; rescore items 1 & 7
```

Example 4: The original codes are "0", "1", and "2". You want 0 0, 1 1, and 2 1 for all items

```
XWIDE=1   one character wide (standard)
CODES  =012   valid codes
NEWSCORE=011   desired scoring
```

Example 5: The original codes are "0", "1", and "2". You want to make 0 2, 1 1, and 2 0, for even-numbered items in a twenty item test.

```
NI=20   twenty items
CODES =012   three valid codes
NEWSCORE=210   desired scoring
RESCORE=010101010101010101010101  rescore "even" items
```

Example 6: The original codes are "0", "1", "2", "3" and some others. You want to make all non-specified codes into "0", but to treat codes of "2" as missing.
CODES = 0123 four valid codes
NEWSCORE= 01X3 response code 2 will be ignored
MISSCORE=0 treat all invalid codes as 0

Example 7: The original codes are "0", "1", "2", "3". You want to rescore some items selectively using KEY1= and KEY2= and to leave the others unchanged - their data codes will be their rating values. For items 5 and 6, 0 0, 1 0, 2 1, 3 2; for item 7, 0 0, 1 0, 2 0, 3 1. Responses to other items are already entered correctly as 0, 1, 2, or 3.
CODES =0123 valid codes
RESCORE=0000110000 rescore items 5,6,7
KEY1=****223*** keyed for selected items
KEY2=****33X*** the X will be ignored
^ read these columns vertically

Example 8: Multiple score key for items 1 to 10. Items 11 to 15 are on a rating scale of 1 to 5
CODES = abcd12345
KEY1 = bacdbaddcd*****
RESCORE= 11111111100000 ; RESCORE= signals when to apply KEY1=

174. RESFROM location of RESCORE

Only use this if you have too many items to put conveniently on one line of the RESCORE= control variable. Instructs where to find the RESCORE= information.
RESFRM=N
RESCORE= is a control variable between before &END (the standard).

RESFRM=Y
RESCORE= information follows after &END but before the item names, if any, and is formatted exactly like a data record. It is helpful, for reference, to enter the label "RESCORE=" where the person name would go in your data record.

Example: KEY1= and KEY2= information follows the RESCORE= information, all are formatted like your data. No item names are provided,
NAME1 = 1 start of person-id
ITEM1 = 10 start of responses
NI = 10 ten items
INUMB = Y use item sequence numbers as names
CODES = ABCDE valid codes
RESFRM = Y rescore information in data record format
KEYFRM = 2 two keys in data record format
&END

RESCORE= 0000110000 RESCORE= looks like data
KEY1= ****AB***** KEY1= looks like data
KEY2= ****CA***** KEY2= looks like data record
George ABCDABCDAB first data record
| subsequent data records

175. RFILE scored response file

Useful for reformatting data from a family of test forms, linked by a network of common items, into a single common structure suitable for one-step item banking.

If this parameter is specified in the control file with RFILE=filename, a file is output which contains a scored/keyed copy of the input data. This file can be used as input for later analyses. Items and persons deleted by PDFILE= or the like are replaced by blank rows or columns in the scored response file. The file format is:
1. Person-id A30 or maximum person-id length
2. Responses one per item:
   A1 if largest scored response is less than or equal to 9
A2 if largest scored response is more than 9.
The width of the responses is not determined by XWIDE=.

176. SAFILE item structure anchor file

The SAFILE of one analysis may be used unedited as the SAFILE of another.

The rating-scale structure parameter values (taus, Rasch-Andrich thresholds, steps) can be anchored (fixed) using SAFILE=. The anchoring option facilitates test form equating. The structure in the rating (or partial credit) scales of two test forms, or in the item bank and in the current form, can be anchored at their other form or bank values. Then the common rating (or partial credit) scale calibrations are maintained. Other measures are estimated in the frame of reference defined by the anchor values.

In order to anchor category structures, an anchor file must be created of the following form:

1. Use one line per category Rasch-Andrich threshold to be anchored.
2. If all items use the same rating scale (i.e. ISGROUPS="", the standard, or you assign all items to the same grouping, e.g ISGROUPS=222222..), then type the category number, a blank, and the "structure measure" value (in logits or your user-rescaled units) at which to anchor the Rasch-Andrich threshold measure corresponding to that category (see Table 3). If you wish to force category 0 to stay in an analysis, anchors its calibration at 0. Specify SAITEM=Yes to use the multiple ISGROUP= format or

    If items use different rating (or partial credit) scales (i.e. ISGROUPS=0, or items are assigned to different groupings, e.g ISGROUPS=122113..), then type the sequence number of any item belonging to the grouping, a blank, the category number, a blank, and the "structure measure" value (in logits if USCALE=1, otherwise your user-rescaled units) at which to anchor the Rasch-Andrich threshold up to that category for that grouping. If you wish to force category 0 to stay in an analysis, anchors its calibration at 0.

    This information may be entered directly in the control file using SAFILE=*

Anything after ";" is treated as a comment.

Example 1: A rating scale, common to all items, of three categories numbered 2, 4, and 6, is to be anchored at pre-set calibrations. The calibration of the Rasch-Andrich threshold from category 2 to category 4 is -1.5, and of the Rasch-Andrich threshold to category 6 is +1.5.

1. Create a file named, say, "STANC.FIL"
2. Enter the lines

   2 0   place holder for bottom category of this rating scale
   4 -1.5 Rasch-Andrich threshold from category 2 to category 4, anchor at -1.5 logits
   6 1.5 Rasch-Andrich threshold from category 4 to category 6, anchor at +1.5 logits

Note: categories are calibrated pair-wise, so the Rasch-Andrich threshold values do not have to advance.

3. Specify, in the control file,

   ISGROUPS=" " (the standard)
   SAFILE=STANC.FIL structure anchor file

   or, enter directly in the control file,

   SAFILE=* 4 -1.5
   6 1.5 *

   If you wish to use the multiple grouping format, i.e., specify an example item, e.g., 13

   SAITEM=YES
   SAFILE=* 13 4 -1.5
   13 6 1.5 *

To check this: "A" after the structure measure
Example 2: A partial credit analysis (ISGROUPS=0) has a different rating scale for each item. Item 15 has four categories, 0, 1, 2, 3 and this particular response structure is to be anchored at pre-set calibrations.
1. Create a file named, say, "PC.15"
2. Enter the lines

```
15 0 0  Bottom categories are always at logit 0
15 1 -2.0  item 15, Rasch-Andrich threshold to category 1, anchor at -2 logit
15 2 0.5
15 3 1.5
```
3. Specify, in the control file,
   ISGROUPS=0
   SAFILE=PC.15

Example 3: A grouped rating scale analysis (ISGROUPS=21134..) has a different rating scale for each grouping of items. Item 26 belongs to grouping 5 for which the response structure is three categories, 1, 2, 3 and this structure is to be anchored at pre-set calibrations.
1. Create a file named, say, "GROUPING.ANC"
2. Enter the lines

```
26 2 -3.3  for item 26, representing grouping 5, Rasch-Andrich threshold to category 2, anchored at -3.3
26 3 3.3
```
3. Specify, in the control file,
   ISGROUPS=21134..
   SAFILE=GROUPING.ANC

Example 4: A partial-credit scale has an unobserved category last time, but we want to use those anchor values where possible.
We have two choices.

a) Treat the unobserved category as a structural zero, i.e., unobservable. If so...
Rescore the item using IVALUE=, removing the unobserved category from the category hierarchy, and use a matching SAFILE=.

In the run generating the anchor values, which had STKEEP=NO,

```
<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>OBSERVED</th>
<th>OBSVD SAMPLE</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>STRUCTURE</th>
<th>CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LABEL</td>
<td>SCORE</td>
<td>COUNT %</td>
<td>AVERAGE</td>
<td>EXPECT</td>
<td>MNSQ</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>-------</td>
<td>---------</td>
<td>---------</td>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>1</td>
<td>1 1</td>
<td>33</td>
<td>0</td>
<td>-.23</td>
<td>-.15</td>
<td>.91</td>
</tr>
<tr>
<td>2</td>
<td>2 2</td>
<td>23</td>
<td>0</td>
<td>.15</td>
<td>.05</td>
<td>.88</td>
</tr>
<tr>
<td>4</td>
<td>3 2</td>
<td>2</td>
<td>0</td>
<td>.29</td>
<td>.17</td>
<td>.95</td>
</tr>
</tbody>
</table>
```

In the anchored run:

```
IREFER=A....... ; item 1 is an "A" type item
CODES=1234       ; valid categories
IVALUEA=12*3     ; rescore "A" items from 1,2,4 to 1,2,3
SAFILE=*         
1 1 .00
1 2 -1.12
1 3 1.12
*                 
```

If the structural zeroes in the original and anchored runs are the same then, the same measures would result from:

```
STKEEP=NO
SAFILE=*     
```
b) Treat the unobserved category as an incidental zero, i.e., very unlikely to be observed.

Here is Table 3.2 from the original run which produced the anchor values. The NULL indicates an incidental or sampling zero.

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>OBSERVED</th>
<th>OBSVD SAMPLE</th>
<th>INFIT  OUTFIT</th>
<th>STRUCTURE</th>
<th>CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LABEL</td>
<td>SCORE</td>
<td>COUNT %</td>
<td>AVERAGE EXPECT MNSQ MNSQ</td>
<td>MEASURE</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>33</td>
<td>0</td>
<td>-.27 -.20</td>
<td>.91 .95</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>23</td>
<td>0</td>
<td>.08 -.02</td>
<td>.84 .68</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>.00 .00</td>
<td>NULL</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>.22 .16</td>
<td>.98 .87</td>
</tr>
</tbody>
</table>

Here is the matching SAFILE=

SAFILE=*  
1 1 .00  
1 2 -.69  
1 3 46.71 ; flag category 3 with a large positive value, i.e., unlikely to be observed.  
1 4 -46.02 ; maintain sum of structure measures (step calibrations) at zero.

Example 5: Score-to-measure Table 20 is to be produced from known item and rating scale structure difficulties.

Specify:

IAFILE= ; the item anchor file
SAFILE= ; the structure/step anchor file (if not dichotomies)
CONVERGE=L ; only logit change is used for convergence
LCONV=0.005 ; logit change too small to appear on any report.
STBIAS=NO ; anchor values do not need estimation bias correction.

The data file comprises two dummy data records, so that every item has a non extreme score, e.g.,
For dichotomies:
Record 1: 10101010101  
Record 2: 01010101010

For a rating scale from 1 to 5:
Record 1: 15151515151  
Record 2: 51515151515

Pivot Anchoring

Pivots are the locations in the dichotomy, rating (or partial credit) scale at which the categories would be dichotomized, i.e., the place that indicates the transition from "bad" to "good", "unhealthy" to "healthy". Ordinarily the pivot is placed at the point where the highest and lowest categories of the response structure are equally probable. Pivot anchoring redefines the item measures.

Dichotomies (MCQ, etc.):

Example 1: To set mastery levels at 75% on dichotomous items (so that maps line up at 75%, rather than 50%) set

    SAFILE=*  
    1 -1.1 ; set the Rasch-Andrich threshold point 1.1 logits down, so that the item ability appears at 75% success.

    ; If you are using USCALE=, then the value is -1.1 * USCALE=

*
Polytomies (rating scales, partial credit, etc.):
When a variety of rating (or partial credit) scales are used in an instrument, their different formats perturb the item hierarchy. This can be remedied by choosing a point along each rating (or partial credit) scale that dichotomizes its meaning (not its scoring) in an equivalent manner. This is the pivot point. The effect of pivoting is to move the structure calibrations such that the item measure is defined at the pivot point on the rating (or partial credit) scale, rather than the standard point (at which the highest and lowest categories are equally probable).

Example 2: Pivoting with ISGROUPS=. Positive (P) items pivot at an expected score of 2.5. Negative (N) items at an expected score of 2.0

```
ISGROUPS=PPPPPPNNNNNN
SAFILE=* 1 2 0.7 ; put in the values necessary to move the center to the desired spot
5 2 0.5 ; e.g., the "structure calibration" - "score-to-measure of pivot point"
```

Example 3: To set a rating (or partial credit) scale turning point: In the Liking for Science, with 0=Dislike, 1=Neutral, 2=Like, anything less than an expected score of 1.5 indicates some degree of lack of liking:

```
SAFILE=* 1 -2.22 ; put in the step calibration necessary to move expected rating of 1.5 to the desired spot
```

```
RATING SCALE PIVOTED AT 1.50
|CATEGORY   OBSERVED|OBSVD SAMPLE|INFIT OUTFIT||STRUCTURE|CATEGORY|
|LABEL SCORE COUNT %|AVRGE EXPECT|  MNSQ  MNSQ||CALIBRATN| MEASURE|
|-------------------+------------+------------++---------+--------|
|  0   0     197  22| -2.29 -2.42|  1.05   .99||  NONE   |( -3.42)| dislike
|  1   1     322 36| -1.17  -.99|   .90   .79||   -2.22 |  -1.25 | neutral
|  2   2     368  41|   .89   .80|   .98  1.29||    -.28 |(   .92)| like
|-------------------+------------+------------++---------+--------|
|MISSING       1   0|   .04      |            ||         |        |
```

AVERAGE MEASURE is mean of measures in category.

```
+----------------------------------------------------------+
|CATEGORY STRUCTURE | SCORE-TO-MEASURE | CUMULATIV| COHERENCE| | 
| LABEL  CALIBRATN S.E. | AT CAT. ----ZONE----| PROBABLTY| M->C C->M| |
|------------------------+---------------------+---------+----------|
|   0      NONE          |( -3.42) -INF   -2.50|         |  63%  44%| dislike
|   1       -2.22    .10 |  -1.25  -2.50    .00|   -2.34 |  55%  72%| neutral
|   2        -.28    .09 |(   .92)   .00  +INF |    -.16 |  84%  76%| like
+----------------------------------------------------------+

Values of .00 for scores of 1.5 show effect of pivot anchoring on the rating (or partial credit) scale. The structure calibrations are offset.

```
TABLE 21.2 LIKING FOR SCIENCE (Wright & Masters p.18) sf.out Aug 1 21:31 2000
EXPECTED SCORE OGIVE: MEANS
```

<table>
<thead>
<tr>
<th>EXPECTED SCORE OGIVE: MEANS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 +</td>
</tr>
<tr>
<td>E</td>
</tr>
<tr>
<td>K 1.5</td>
</tr>
<tr>
<td>P</td>
</tr>
<tr>
<td>E</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>T</td>
</tr>
<tr>
<td>E 1</td>
</tr>
<tr>
<td>D</td>
</tr>
<tr>
<td>S</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>O .5</td>
</tr>
<tr>
<td>R</td>
</tr>
<tr>
<td>E</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
```

161
Example 4: A questionnaire includes several rating (or partial credit) scales, each with a pivotal transition-structure between two categories. The item measures are to be centered on those pivots.
1. Use ISGROUPS= to identify the item response-structure groupings.
2. Look at the response structures and identify the pivot point:
   e.g., here are categories for "grouping A" items, after rescoring, etc.
   Strongly Disagree 1
   Disagree 2
   Neutral 3
   Agree 4
   Strongly Agree 5
   If agreement is wanted, pivot between 3 and 4, identified as transition 4.
   If no disagreement is wanted, pivot between 2 and 3, identified as transition 3.

3. Anchor the transition corresponding to the pivot point at 0, e.g., for agreement:
   e.g., for
   ISGROUPS=AAAAAAABBBABAACCC
   SAFILE=*
   6 4 0 6 is an item in grouping A, pivoted at agreement (Rasch-Andrich threshold from category 3 into category 4)
   8 2 0 8 is an item in grouping B, pivoted at Rasch-Andrich threshold from category 2 into category 3
   ; no pivoting for grouping C, as these are dichotomous items

Example 5: Anchor files for dichotomous and partial credit items. Use the IAFILE= for anchoring the item difficulties, and SAFILE= to anchor partial credit structures. Winsteps decomposes the Dij of partial credit items into Di + Fij.
The Di for the partial credit and dichotomous items are in the IAFILE=
The Fij for the partial credit files are in the SAFILE=

Suppose the data are A,B,C,D, and there are two partial credit items, scored 0,1,2, and two merely right-wrong.
0,1 then:

   CODES=ABCD
   KEY1=BCBC ; SCORE OF 1 ON THE 4 ITEMS
   KEY2=DA** ; SCORE OF 2 ON THE PARTIAL CREDIT ITEMS
   ISGROUPS=0

If the right-wrong MCQ items are to be scored 0,2, then
   CODES=ABCD
   KEY1=BC** ; SCORE OF 1 ON THE 4 ITEMS
   KEY2=DABC ; SCORE OF 2 ON THE PARTIAL CREDIT ITEMS
   ISGROUPS=0

but better psychometrically is:

   CODES=ABCD
   KEY1=BCBC ; SCORE OF 1 ON THE 4 ITEMS
   KEY2=DA** ; SCORE OF 2 ON THE PARTIAL CREDIT ITEMS
   IWEIGHT=*
   3-4 2 ; items 3 and 4 have a weight of 2.
   *
   ISGROUPS=0

Then write out the item and partial credit structures
   IFILE= items.txt
   SFILE=pc.txt

In the anchored run:
   CODES= ... etc.
   IAFILE=items.txt
   SAFILE=pc.txt
   CONVERGE=L ; only logit change is used for convergence
Anchored values are marked by "A" in the Item Tables, and also Table 3.2

177. **SAITEM item numbers in SAFILE with one grouping**

Step Files *SFILE=* and Step Anchor Files *SAFILE=* (Rating (or partial credit) scale structure anchoring files) have the format:

For only one active grouping in *ISGROUPS=*, or no *ISGROUPS=*

<table>
<thead>
<tr>
<th>step number</th>
<th>step calibration</th>
</tr>
</thead>
</table>

For multiple groupings in *ISGROUPS=*, or no *ISGROUPS=*

<table>
<thead>
<tr>
<th>example item number in grouping</th>
<th>step number</th>
<th>step calibration</th>
</tr>
</thead>
</table>

To specify that the multiple grouping format be used when there is only one active grouping, specify *SAITEM=Yes*

Example: "Liking for Science" step anchor file:

Standard format for the Rating Scale model:

<table>
<thead>
<tr>
<th>ISGROUPS=</th>
<th>no groupings specified</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAFILE=*</td>
<td>1 -1.8; Rasch-Andrich threshold from category 0 to category 1</td>
</tr>
<tr>
<td>2 1.8; Rasch-Andrich threshold from category 1 to category 2</td>
<td>*</td>
</tr>
</tbody>
</table>

Alternative format allowing an example item number to identify the grouping, say item 10

<table>
<thead>
<tr>
<th>SAITEM=Yes</th>
<th>SAFILE=*</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 1 -1.8</td>
<td>10 2 1.8</td>
</tr>
</tbody>
</table>

178. **SANCHQ anchor category structure interactively**

If your system is interactive, steps to be anchored can be entered interactively by setting *SANCHQ=Y* before the &END line. If you specify this, you will be asked if you want to anchor any steps. If you respond "yes", it will ask if you want to read these anchored items from a file; if you answer "yes" it will ask for the file name and process that file in the same manner as if *SAFILE=* had been specified. If you answer "no", you will be asked to enter the step measures (found in Table 3).

If there is only one rating (or partial credit) scale, enter the category numbers for which the Rasch-Andrich thresholds are to be anchored, one at a time, along with their logit (or user-rescaled by *USCALE=* ) structure measure calibrations. Bypass categories without measures. Enter 0 where there is a measure of "NONE". When you are finished, enter -1 in place of the category number.

If there are several rating (or partial credit) scales, enter one of the item numbers for each rating (or partial credit) scale, then the structure measures corresponding to its categories. Repeat this for each category of an item for each rating (or partial credit) scale. Enter 0 where there is a structure measure for a category of "NONE". Entering 0 as the item number completes anchoring.

Example 1: You are doing a number of analyses, anchoring the common rating (or partial credit) scale to different values each time. You want to enter the numbers at your PC:

```
SANCHQ=Y
```

You want to anchor items 4 and 8.

WINSTEPS asks you:

**DO YOU WANT TO ANCHOR ANY STRUCTURES?** respond YES(Enter)
**DO YOU WISH TO READ THE ANCHORED STRUCTURES FROM A FILE?**
respond NO(Enter)
INPUT STRUCTURE TO ANCHOR (-1 TO END):
respond 2(Enter) (the first category Rasch-Andrich threshold to be anchored)
INPUT VALUE AT WHICH TO ANCHOR Rasch-Andrich threshold:
respond 0(Enter) (the first anchor value)
INPUT Rasch-Andrich threshold TO ANCHOR (-1 TO END): 4(Enter)
INPUT VALUE AT WHICH TO ANCHOR Rasch-Andrich threshold: -1.5(Enter)
INPUT Rasch-Andrich threshold TO ANCHOR (-1 TO END): 6(Enter)
INPUT VALUE AT WHICH TO ANCHOR Rasch-Andrich threshold: 1.5(Enter)
INPUT Rasch-Andrich threshold TO ANCHOR (-1 TO END): -1(Enter) to end anchoring

Example 2: You wish to enter the structure measures for several rating scales, each comprising a grouping of items: SANCHQ=Y

WINSTEPS asks you:
DO YOU WANT TO ANCHOR ANY Rasch-Andrich thresholds? YES(Enter)
DO YOU WANT TO READ THE ANCHORED Rasch-Andrich thresholds FROM A FILE? NO
Item 1 represents the first grouping of items, sharing a common rating scale:
INPUT AN ITEM, REPRESENTING A GROUPING (0 TO END): 1
INPUT Rasch-Andrich threshold TO ANCHOR (-1 TO END): 0 bottom category
INPUT VALUE AT WHICH TO ANCHOR Rasch-Andrich threshold: 0 "NONE"
INPUT AN ITEM, REPRESENTING A GROUPING (0 TO END): 1
INPUT Rasch-Andrich threshold TO ANCHOR (-1 TO END): 1
INPUT VALUE AT WHICH TO ANCHOR Rasch-Andrich threshold: -0.5
INPUT AN ITEM, REPRESENTING A GROUPING (0 TO END): 1
INPUT Rasch-Andrich threshold TO ANCHOR (-1 TO END): 2
INPUT VALUE AT WHICH TO ANCHOR Rasch-Andrich threshold: 0.5
Item 8 represents the second grouping of items, sharing a common rating scale:
INPUT AN ITEM, REPRESENTING A GROUPING (0 TO END): 8
When all are anchored, enter 0 to end:
INPUT AN ITEM, REPRESENTING A GROUPING (0 TO END): 0

179. **SCOREFILE** person score file

If SCOREFILE=filename is specified, a file is output which contains the measure and model standard error corresponding to every possible score on a test consisting of all the items. This is also shown in Table 20. It has 3 heading lines (unless HLINES=N), and has the format:

```
; PERSON SCORE FILE FOR 
; EXAMPLE ANALYSES 
; Apr 28 22:37 2005  UMEAN=0.000  USCALE=1.000 
; TABLE OF SAMPLE NORMS (500/100) AND FREQUENCIES CORRESPONDING TO COMPLETE TEST 
; SCORE MEASURE S.E. INFO NORMED S.E. FREQUENCY % CUM.FREQ. % PERCENTILE 
0 5.2891 1.4230 .49 -76 125 0 .0 0 .0 0 
... 58 1.8144 .3351 8.90 546 29 15 2.9 363 71.2 70 
59 1.9303 .3461 8.35 556 30 20 3.9 383 75.1 73 
60 2.0544 .3588 7.77 567 31 21 4.1 404 79.2 77 
... 70 5.3232 1.4245 .49 853 125 3 .6 510 100.0 99 
```

1. **SCORE**: Score on test of all items
The score file shows integer raw scores, unless there are decimal weights for IWEIGHT=. In which case, scores to 1 decimal place are shown. To obtain other decimal raw scores for short tests, go to the Graphs pull-down menu. Select "Test Characteristic Curve”. This displays the score-to-measure ogive. Click on "Copy data to clipboard”. Open Excel. Paste. There will be to three columns. The second column is the measure, the third column is the raw score.
2. **MEASURE**: Measure (user-scaled bu USCALE=)
3. **S.E.**: Standard error (user scaled by USCALE=) - model, because empirical future misfit is unknown.
4. **INFO**: Statistical information in measure (=1/Logit S.E.²)

Measures locally-rescaled, so that sample mean=500, sample S.D.=100
5. **NORMED**: Measure (rescaled)
6. **S.E.**: Standard error (rescaled)

Sample distribution:
7. **FREQUENCY**: Count of sample at this measure
8. **%**: Percent of sample at this measure
9. **CUM.FREQ.**: Count of sample at or below this measure
10. **%**: Percent of sample at or below this measure
11. **PERCENTILE**: Percentile of this sample lower than the current measure (range is 0-99).

If CSV=\texttt{Y}, these values are separated by commas. When CSV=\texttt{T}, the commas are replaced by tab characters.

Example 1: You wish to write a file on disk called "MYDATA.SCF.txt" containing a score-to-measure table for the complete test.

\texttt{SCOREFILE=MYDATA.SCF.txt}

Example 2: You want a score-to-measure file for items with known difficulties.

\texttt{ITEM1=1  ; start of response string}
\texttt{NI=10   ; number of items}
\texttt{CODES=01  ; valid item codes}
\texttt{IAFILE=*}
\texttt{; known item difficulties here}
\texttt{0 0.53}
\texttt{*}
\texttt{SAFILE=*}
\texttt{; known structure "step" calibrations here, if rating scale or partial credit items}
\texttt{*}
\texttt{SCOREFILE=sm.txt  ; the score-to-measure file - also see Table 20}

&END

END LABELS

0101010101  ; two dummy data records
1010101010  ; give every item a non-extreme score

\textbf{180. SDELQU delete category structure interactively}

This is better performed with \texttt{IREFER=} and \texttt{IVALUE=}. \texttt{SDELQU=} is not supported in this version of Winsteps.

If your system is interactive, categories to be deleted can be entered interactively by setting \texttt{SDELQU=Y} before the \&END line. If you specify this, you will be asked if you want to delete any category structure.

If you respond "yes", it will ask if you want to read these deleted categories from a file; if you answer "yes" it will ask for the file name and process that file in the same manner as if \texttt{SDFILE=} had been specified.

If you answer "no", you will be asked to enter

a) the sequence number of each item (representing a grouping, as described under \texttt{SDFILE=}). This question is omitted if all items are in one grouping.

b) the score value of one category to be deleted from that item and its grouping.

Enter these deletions one at a time. When you are finished, enter a zero.

\textbf{181. SDFILE category structure deletion file}

This is better performed with \texttt{IREFER=} and \texttt{IVALUE=}. \texttt{SDFILE=} is not supported in this version of Winsteps.

Deletion of categories from a test analysis (i.e. conversion of responses in these categories to "missing data"), but without removing these responses from your data file, is easily accomplished by creating a file in which each line contains the number of the category to be deleted from that item and its grouping. If there is more than one grouping, the sequence number of an item representing a grouping followed by a blank.
Specify this file by means of the control variable, SDFILE=, or this information may be specified in the control file using SDFILE=*.

Since rating (or partial credit) scales may be shared by groupings of items, deletion of categories is performed by grouping:

a) If no ISGROUPS= control variable is specified, no item need be entered, since specifying deletion of a category deletes that category for all items.

b) If a ISGROUPS= control variables is specified, then specifying deletion of a category for one item deletes that category for all items in the same grouping.

c) If ISGROUPS=0 is specified, only the specified category for the specified item is deleted.

Example: You wish to delete particular categories for the fifth and tenth “partial credit” items for this analysis.

1. Create a file named, say, "CAT.DEL"

2. Enter into the file, the lines:
   5 3  (item 5 category 3)
   10 2 (item 10 category 2)

3. Specify, in the control file,
   SDFILE=CAT.DEL
   ISGROUPS=0

or, enter in the control file,
   SDFILE=*
   5 3
   10 2
   *
   ISGROUPS=0

182. SFILE  category structure output file

If SFILE=filename is specified, a file is output which contains the item and category information needed for anchoring structures. It has 4 heading lines (unless HLINES=N), and has the format:

1. The item sequence number (I6)
   Only if there are multiple groupings in ISGROUPS= or SAITEM=Yes, otherwise this is omitted.

2. The category value (I3) (STRU) - the Rasch-Andrich threshold.

3. structure calibration (F7.2) (user-rescaled by USCALE=) (number of decimals by UDECIM=) (MEASURE)
   If the category is an intermediate category with no observations, kept in the analysis by STKEEP=YES, then its structure calibration is shown as very large. The next higher calibration is shown as very low.
   **Use this for anchoring, together with IFILE=**

   If CSV=Y, these values are separated by commas. When CSV=T, the commas are replaced by tab characters.

    Anchoring with Unobserved Categories

When STKEEP=YES and there are intermediate null categories, i.e., with no observations, then the Rasch-Andrich threshold into the category is set 40 logits above the highest calibration. The Rasch-Andrich threshold out of the category, and into the next category, is set the same amount down. Thus:

<table>
<thead>
<tr>
<th>Category</th>
<th>structure Calibration</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>NULL</td>
</tr>
<tr>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
This is an attempt to model unobserved categories for the next time, when they may be observed.

If categories will not be observed next time, then please specify STKEEP=NO, which automatically drops the category from the rating (or partial credit) scale.

If categories may be observed next time, then it is better to include a dummy data record in your data file which includes an observation of the missing category, and reasonable values for all the other item responses that accord with that missing category. This one data record will have minimal impact on the rest of the analysis.

183. SIFILE simulated data file

This uses the estimated (or anchored) person, item and structure measures to simulate a data file equivalent to the raw data. This can be used to investigate the stability of measures, distribution of fit statistics and amount of statistical bias. Each time SIFILE= is run, or selected from the Output Files pull-down menu, a different simulated data file is produced. Do simulated analyses with several simulated datasets to verify their overall pattern. When SIFILE= is specified, missing data values are filled in, where possible.

Example 1: It is desired to investigate the stability of the "Liking for Science" measures.
(1) Estimate measures from SF.txt
(2) Choose SIFILE= from the Output Files menu. SIFILE=SFSIMUL.TXT
(3) Rerun Winsteps with DATA=SFSIMUL.TXT on the "Extra Specifications" line.
(4) Compare person, item and structure measures.

The file format matches the input data file if both are in fixed-field format. When SIFILE= is written with CSV=Y, comma-separated or CSV=T, tab-separated, the item responses precede the person label.

Example: KCT.txt simulated with fixed field format:
Richard M 111101101000000000
Tracie F 111111110000000000
Walter M 111111110111000000

KCT.txt simulated with comma-separated, CSV=Y, format:
1,1,1,0,0,1,1,0,0,0,0,0,0,0,0,0,Richard M
1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,Tracie F
1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,Walter M

Example 2. To estimate the measure standard errors in a linked equating design.
Do a concurrent calibration with Winsteps and then simulate data files SIFILE= from the Output Files menu. Specify "missing data" as "missing" to maintain the same data pattern. Save 10 simulated sets as S1.txt S2.txt ......
Then rerun your Winsteps analysis 10 times specifying in Extra Specifications "DATA=S1.txt PFILE=P1.txt CSV=TAB" etc. This will produce 10 PFILE=s, P1.txt P2.txt ..... in tab-separated format. These are easy to import into Excel. So you will have 10 measures for every person. Compute the standard deviation of the measures for each person based on the 10 person measures - this is their model standard error for the equating design for each measure. Inflate these values by 20%, say, to allow for systematic equating errors, misfit, etc.
184. SPFILE supplementary control file

There is often a set of control instructions that you wish to make a permanent or temporary part of every control file. Such files can be specified with SPFILE=. Multiple SPFILE= specifications can be included in one control file. Supplemental files called with SPFILE= can also include SPFILE= specifications.

Example 1: The analyst has a standard set of convergence criteria and other control instructions to include in every control file.

a) Enter these into a standard DOS TEXT/ASCII file, e.g., SUPPL.TXT

The analyst's SUPPL.TXT contains:

LCONV=.01
ITEM=TASK
PERSON=CLIENT
TABLES=10110011

b) Specify this supplemental file in the main control file, say, MAIN.TXT

```
TITLE='NEW YORK CLIENTS'
SPFILE=SUPPL.TXT
ITEM1=37
NI=100
```  

Example 2: The analyst has a set of control instructions that are used only for the final run. These are coded in a separate DOS TEXT file called FINAL.SPC

```
C:>WINSTEPS CONTROL.FIL OUTPUT.FIL SPFILE=FINAL.SPC
```  

Example 3: Keyn= is a particularly useful application of SPFILE=.

Put the KEY1= instruction for each test form in its own DOS TEXT file, then reference that file rather than including the key directly in the control file.

Here is FORMA.KEY:

```
NI=23
CODES=ABCD
KEY1=ABCDDADBCDADDABBCCAABBB
```  

Here is the control file:

```
TITLE='FORM A READING RASCH ANALYSIS'
ITEM1=20
SPFILE=FORMA.KEY
TABLES=111011011
```  

185. STBIAS correct for estimation bias

STBIAS=Y causes an approximate correction for estimation bias in JMLE estimates to be applied to measures and calibrations. This is only relevant if an exact probabilistic interpretation of logit differences is required for short tests or small samples. Set STBIAS=NO when using IWEIGHT=, PWEIGHT=, anchoring or artificially lengthened tests or augmented samples, e.g., by replicating item or person response strings.

Fit statistics are computed without this estimation-bias correction. Estimation-bias correction makes the measures more central, generally giving a slight overfit condition to Outfit and Infit. Correct "unbiased" computation of INFIT and OUTFIT needs not only unbiased measures, but also probabilities adjusted for the possibility of extreme score vectors (which is the cause of the estimation bias). The XMLE algorithm (implemented experimentally) attempts to do this - but it has other drawbacks that often make it impractical.

Example 1: You have a well-behaved test of only a few items, for which you judge the statistical bias correction to be useful because you are planning to make exact probabilistic inferences based on differences between logit measures:

```
STBIAS=Y
```
STKEEP keep non-observed intermediate categories in structure

Unobserved categories can be dropped from rating (or partial credit) scales and the remaining category recounted during estimation. For intermediate categories only, recounting can be prevented and unobserved categories retained in the analysis. This is useful when the unobserved categories are important to the rating (or partial credit) scale logic or are usually observed, even though they happen to have been unused this time. Category Rasch-Andrich thresholds for which anchor calibrations are supplied are always maintained wherever computationally possible, even when there are no observations of a category in the current data set.

Use STKEEP=YES when there may be intermediate categories in your rating (or partial credit) scale that aren’t observed in this data set, i.e., incidental zeroes.

Use STKEEP=NO when your category numbering deliberately skips over intermediate categories, i.e., structural zeroes.

STKEEP=N Eliminate unused categories and close up the observed categories.

STKEEP=Y Retain unused non-extreme categories in the ordinal categorization.

When STKEEP=Y, missing categories are retained in the rating (or partial credit) scale, so maintaining the raw score ordering. But missing categories require locally indefinite structure calibrations. If these are to be used for anchoring later runs, compare these calibrations with the calibrations obtained by an unanchored analysis of the new data. This will assist you in determining what adjustments need to be made to the original calibrations in order to establish a set of anchor calibrations that maintain the same rating (or partial credit) scale structure.

To remind yourself, STKEEP=YES can be written as STRUCTUREKEEP=YES, STRKEEP=YES or STEPKEEP=YES and other similar abbreviations starting STK, STR and STEPK.

Example 1: Incidental unobserved categories. Keep the developmentally important rating (or partial credit) scale categories, observed or not. Your small Piaget rating scale goes from 1 to 6. But some levels may not have been observed in this data set.

STKEEP=Y

Example 2: Structural unobserved categories. Responses have been coded as "10", "20", "30", "40", but they really mean 1,2,3,4

CODES = "10203040"
XWIDE = 2
STKEEP=NO
; if STKEEP=YES, then data are analyzed as though categories 11, 12, 13, 14, etc. could exist, which would distort the measures.
; for reporting purposes, multiply Winsteps SCOREs by 10 to return to the original 10, 20, 30 categorization.

Example 3: Some unobserved categories are structural and some incidental. Rescore the data and use STKEEP=YES. Possible categories are 2,4,6,8 but only 2,6,8 are observed this time.

(a) Rescore 2,4,6,8 to 1,2,3,4 using IVALUE= or NEWSCORE=
(b) Set STKEEP=YES, so that the observed 1,3,4 and unobserved 2 are treated as 1,2,3,4
(c) For reporting purposes, multiply the Winsteps SCORE by 2 using Excel or similar software.

CODES=2468
NEWSCORE=1234
STKEEP=YES

Incidental and Structural Zeroes: Extreme and Intermediate

For missing intermediate categories, there are two options.

If the categories are missing because they cannot be observed, then they are "structural zeroes". Specify "STKEEP=NO".
This effectively recounts the observed categories starting from the bottom category, so that 1,3,5,7 becomes
1,2,3,4.

If they are missing because they just do not happen to have been observed this time, then they are "incidental or sampling zeros". Specify "STKEEP=YES". Then 1,3,5,7 is treated as 1,2,3,4,5,6,7.

Categories outside the observed range are always treated as structural zeroes.

When STKEEP=Y, unobserved intermediate categories are imputed using a mathematical device noticed by Mark Wilson. This device can be extended to runs of unobserved categories.

187. **STEPT3** include category structure summary in Table 3 or 21

The structure summary statistics usually appear in Table 3. For grouped analysis this part of Table 3 can become long, in which case it can be moved to Table 21.

Example: Don't output partial credit structure summaries in Table 3. Move them to Table 21:

```
ISGROUPS=0 each item has own rating scale
STEPT3=N report scale statistics in Table 21
```

188. **T1I#** number of items summarized by "#" symbol in Table 1

For ease in comparing the outputs from multiple runs, force consistent x-axis scaling by using MRANGE=, T1I#= and T1P#. Choose T1I# to be the largest number of items summarized by one "#" from any of the separate runs.

Example: In one run, the bottom of Table 1 states that

```
EACH "#" IN THE ITEM COLUMN IS 20 ITEMS
```

In another run:

```
EACH "#" IN THE ITEM COLUMN IS 15 ITEMS
```

To make the runs visually comparable, specify the bigger value:

```
T1I#=20
```

189. **T1P#** number of persons summarized by "#" symbol in Table 1

For ease in comparing the outputs from multiple runs, force consistent x-axis scaling by using MRANGE=, T1I#= and T1P#. Choose T1P# to be the largest number of persons summarized by one "#" from any of the separate runs.

Example: In one run, the bottom of Table 1 states that

```
EACH "#" IN THE PERSON COLUMN IS 250 PERSONS
```

In another run:

```
EACH "#" IN THE PERSON COLUMN IS 300 PERSON
```

To make the runs visually comparable, specify the bigger value:

```
T1P#=300
```

190. **TABLES** output tables

For more flexibility, use the **Output Tables** pull-down menu.

```
Tables= causes the specified Tables to be written into the report output file. It is has no effect on the pull-down menus.

A character string that tells WINSTEPS which output tables to prepare for printing. The sequence number of the "1" or "0" in the TABLES= string matches the table number. For more elaborate table selection, use **TFILE**=.

"1" means prepare the corresponding table.
"0" or anything else means do not prepare the corresponding table.

Example 1: You want only Tables 2,4,6,8,10 and 20 to be prepared

```
TABLES=01010101000000000001000
```
This is the same as specifying:
TFILE=*  
2  
4  
6  
8  
10  
20  
*

Example 2: You want only Tables 1-4.  
TABLES=1111

191.  **TARGET** estimate using information-weighting

TARGET=Y lessens the effect of guessing on the measure estimates, but increases reported misfit. A big discrepancy between the measures produced by TARGET=N and TARGET=Y indicates much anomalous behavior disturbing the measurement process.

Unwanted behavior (e.g. guessing, carelessness) can cause unexpected responses to off-target items. The effect of responses on off-target items is lessened by specifying TARGET=Y. This weights each response by its statistical information during estimation. Fit statistics are calculated as though the estimates were made in the usual manner. Reported displacements show how much difference targeting has made in the estimates.

Example: Some low achievers have guessed wildly on a MCQ test. You want to reduce the effect of their lucky guesses on their measures and on item calibrations.  
TARGET=Y

**How Targeting works:**
a) for each observation:
   - calculate probability of each category (0,1 for dichotomies)
   - calculate expected score (= probability of 1 for dichotomy)
   - calculate variance = information
     = probability of 1 * probability of 0 for dichotomies,
     so maximum value is 0.25 when person ability measure = item difficulty measure

b) for targeting:
   - weighted observation = variance * observation
   - weighted expected score = variance * expected score

c) sum these across persons and items (and structures)

d) required "targeted" estimates are obtained when, for each person, item, structure sum (weighted observations) = sum (weighted expected scores)

e) for calculation of fit statistics and displacement, weights of 1 are used but with the targeted parameter estimates. Displacement size and excessive misfit indicate how much "off-target" aberrant behavior exists in the data.

For targeting, there are many patterns of responses that can cause infinite measures, e.g. all items correct except for the easiest one. The convergence criteria limit how extreme the reported measures will be.

192. **TFILE** input file listing tables to be output

Omit TFILE= and use the pull-down menu.

TFILE= causes the specified Tables to be written into the report output file. It has no effect on the pull-down menus.
TABLES= selects the tables in a fixed sequence, and prints only one copy. TFILE= allows the analyst to print multiple copies of tables to different specifications. TFILE= specifies the name of an input ASCII file. Each line of the file contains a table number or table number.sub-table and other control parameters, separated by blanks or commas. Unused control values are specified with "-". The list may be entered directly into the control file with TFILE=" (see Example 2).

TFILE= Parameters: (enter unused parameters with ",")

Example 1: The analyst wishes to select and print several tables:

TFILE=TABLES.TF

TABLES.TF is a DOS (ASCII) file with the following lines:

<table>
<thead>
<tr>
<th>; Table</th>
<th>Low</th>
<th>High</th>
<th>Columns</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>10.2</td>
<td>0.5</td>
<td>1.5</td>
</tr>
<tr>
<td>8</td>
<td>-5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>-2</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>-5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>15</td>
<td>-</td>
<td>-</td>
<td>4</td>
</tr>
</tbody>
</table>

or enter directly into the control file,

TFILE="

Example 2: Analyst wishes to specify on the DOS control line, Table 15 sorted on item name column 4. Values are separated by commas, because blanks act as end-of-line separators.

C:\WINSTEPS SF.TXT SFO.TXT TFILE="* 15,-,-,-,4 *"

193. TITLE title for output listing

Use this option to label output distinctly and uniquely.

Up to 60 characters of title. This title will be printed at the top of each page of output.

Example: You want the title to be: Analysis of Math Test

TITLE="Analysis of Math Test"

Quote marks " " or "" are required if the title contains any blanks.
194. TOTALSCORE show total scores with extreme observations

TOTALSCORE=N, the standard
Winsteps uses an adjusted raw score for estimation, from which observations that form part of extreme scores have been dropped. This is displayed in Table 13, PFILE=, IFILE=, etc.

TOTALSCORE=Y
The total raw score from the data file, after any recoding and weighting, is shown. This usually matches the numbers used in raw-score analysis.

This can be changed from the Specification pull-down menu.

Example: KCT.txt with TOTALSCORE=N, the standard. This shows scores on the 15 measurable items.

<table>
<thead>
<tr>
<th>ENTRY</th>
<th>RAW</th>
<th>COUNT</th>
<th>MEASURE</th>
<th>ERROR</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>SCORE</th>
<th>WEIGH</th>
<th>KID</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>12</td>
<td>15</td>
<td>3.50</td>
<td>.91</td>
<td>1.81</td>
<td>1.4</td>
<td>.79</td>
<td>.0</td>
<td>1.00</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>15</td>
<td>2.68</td>
<td>.90</td>
<td>1.64</td>
<td>1.1</td>
<td>1.49</td>
<td>.1</td>
<td>1.00</td>
</tr>
<tr>
<td>15</td>
<td>11</td>
<td>15</td>
<td>2.68</td>
<td>.90</td>
<td>1.35</td>
<td>-1.7</td>
<td>-4</td>
<td>.76</td>
<td>1.00</td>
</tr>
</tbody>
</table>

With TOTALSCORE=Y. This shows scores on all 18 items.

<table>
<thead>
<tr>
<th>ENTRY</th>
<th>TOTAL</th>
<th>COUNT</th>
<th>MEASURE</th>
<th>ERROR</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>SCORE</th>
<th>WEIGH</th>
<th>KID</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>15</td>
<td>18</td>
<td>3.50</td>
<td>.91</td>
<td>1.81</td>
<td>1.4</td>
<td>.79</td>
<td>.0</td>
<td>1.00</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>18</td>
<td>2.68</td>
<td>.90</td>
<td>1.64</td>
<td>1.1</td>
<td>1.49</td>
<td>.1</td>
<td>1.00</td>
</tr>
<tr>
<td>15</td>
<td>14</td>
<td>18</td>
<td>2.68</td>
<td>.90</td>
<td>1.35</td>
<td>-1.7</td>
<td>-4</td>
<td>.76</td>
<td>1.00</td>
</tr>
</tbody>
</table>

195. UANCHOR anchor values supplied in user-scaled units

This simplifies conversion from previously computed logit measures to user-scaled measures.

UANCHOR=A or N or L specifies that the anchor values are in UASCALE= units per logit. Reported measures, however, will be user-rescaled by UMEAN= (or UIMEAN= or UPMEAN=) and USCALE=.

UANCHOR=Y specifies that anchor values are in USCALE= units per logit.

If UASCALE<>1.0 then UANCHOR=A is forced.

Example 1: Your item bank calibrations are user-scaled with 10 units per logits, but you want to report person measures in CHIPS (BTD p.201):
UASCALE=10 ; user-scaling of anchor values
UANCHOR=A
UMEAN=50 ; user-scaling of reported values
USCALE=4.55

Example 2: Your previous test was in logits, but now you want the sample mean to be 500, with user-scaling 100 per logit.
UASCALE=1 ; user-scaling of anchor values: logits
UANCHOR=Logits
UPMEAN=500 ; user-scaling of reported values
USCALE=100
196. **UASCALE** the anchor user-scale value of 1 logit

Specifies the number of units per logit of the anchor values.
If UASCALE=1, then specify UANCHOR=A to confirm that this is the anchor user-scaling.

Example 1: The anchor values are user-scaled such that 1 logit is 45.5 units, so that differences of -100, -50, 0, +50, +100 correspond to success rates of 10%, 25%, 50%, 75%, 90%:

UASCALE = 45.5

Example 2: The anchor values are on one scaling, but you want the reported values to be on another scaling.

- **UASCALE**= 5 ; there are 5 units per logit in the scaled values
- **USCALE**=10 ; you want 10 units per logit on the reported values
- **UPMEAN**=50 ; you want the reported person mean to be 50

197. **UCOUNT** number of unexpected responses: Tables 6, 10

This sets the maximum number of "most unexpected responses" to report in Tables 6.6, 10.6. Also the maximum number of persons and items to report in the anti-Guttman matrices in Tables 6.4, 6.5, 10.4, 10.5

**MOST UNEXPECTED RESPONSES**

<table>
<thead>
<tr>
<th>DATA</th>
<th>OBSERVED</th>
<th>EXPECTED</th>
<th>RESIDUAL</th>
<th>ST. RES.</th>
<th>ACT</th>
<th>KID</th>
<th>ACT</th>
<th>KID</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1.93</td>
<td>-1.93</td>
<td>-7.66</td>
<td>18</td>
<td>73</td>
<td>GO ON PICNIC</td>
<td>SANDBERG, RYNE</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>.07</td>
<td>1.93</td>
<td>7.57</td>
<td>23</td>
<td>72</td>
<td>WATCH A RAT</td>
<td>JACKSON, SOLOMON</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>.07</td>
<td>1.93</td>
<td>7.57</td>
<td>23</td>
<td>29</td>
<td>WATCH A RAT</td>
<td>LANDMAN, ALAN</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1.93</td>
<td>-1.93</td>
<td>-7.40</td>
<td>19</td>
<td>71</td>
<td>GO TO ZOO</td>
<td>STOLLER, DAVE</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1.93</td>
<td>-1.93</td>
<td>-7.40</td>
<td>19</td>
<td>71</td>
<td>GO TO ZOO</td>
<td>STOLLER, DAVE</td>
</tr>
</tbody>
</table>

Example: Show 100 "Most Unexpected Responses":

UCOUNT=100

198. **UDECIMALS** number of decimal places reported

This is useful for presenting your output measures and calibrations in a clear manner by removing meaningless decimal places from the output. Range is 0 (12345.) to 4 (1.2345).

How small is meaningless? Look at the Standard Error columns. Any value clearly less than a standard error has little statistical meaning.

Use the "Specification" pull-down menu to alter the value of UDECIMALS= for individual reports.

Example 1: You want to report measures and calibrations to the nearest integer:

UDECIMALS = 0

Example 2: You want to report measures and calibrations to 4 decimal places because of a highly precise, though arbitrary, pass-fail criterion level:

UDECIMALS = 4

199. **UIMEAN** the mean or center of the item difficulties

Assigns your chosen numerical value to the average measure of the non-extreme items, i.e., a criterion-referenced measure. Previous **UPMEAN=** values are ignored.

UMEAN= and UIMEAN= are the same specification. Anchor values are treated according to UANCHOR=

Table 20 gives the UMEAN= and USCALE= values for a conversion that gives the measures a range of 0-100.

Example 1: You want to recenter the item measures at 10 logits, and so add 10 logits to all reported measures, to avoid reporting negative measures for low achievers:

UIMEAN = 10
Example 2: You want to recenter and user-rescale the item measures, so that the range of observable measures goes from 0 to 100.

Look at Table 20.1. Beneath the Table are shown the requisite values, e.g.,
UMEAN = 48.3 ; this is the same as UIMEAN=48.3
USCALE = 9.7

For more examples, and how to compute this by hand, see User-friendly rescaling

200.  **UP**MEAN  the mean or center of the person abilities

Assigns your chosen numerical value to the average of the non-extreme abilities for persons, i.e., this provides a norm-referenced user-scaling. Previous **U**MEAN= values are ignored. Anchor values are treated according to **U**ANCHOR=  

Example 1: You want to used conventional IRT norm-referenced user-scaling with person mean of 0 and person S.D. of 1.

UPMEAN = 0
USCALE = 1 / (person S.D. in logits); find this from Table 3.1
If there are extreme person scores, see User-friendly rescaling

Example 2: I want to compare the mean performance of random samples of examinees from my database. Will UPMEAN= help in this?

UPMEAN=0 sets the local mean of the persons to zero (excluding extreme scores) regardless of the sampling.
If you wish to investigate the behavior of the person mean for different person samples, then
(1) analyze all persons and items: set UPMEAN=0, for convenience, and write IFILE=.
For better comparison, set STBIAS=NO.
(2) anchor the items using IAFILE=
(3) analyze samples of persons with the anchored items.
The person means reported in Table 3.1 now show person means (with or without extreme scores) in the one frame of reference across all analyses defined by the anchored items.

For more examples, and how to compute this by hand, see User-friendly rescaling

201.  **US**CALE  the user-scaled value of 1 logit

Specifies the number of reported user-scaled units per logit. When USCALE=1 (or USCALE= is omitted) then all measures are in logits.

Table 20 gives the **U**MEAN= and **U**SCALE= values for a conversion that gives the measures a range of 0-100.

Example 1: You want to user-rescale 1 logit into 45.5 units, so that differences of -100, -50, 0, +50, +100 correspond to success rates of 10%, 25%, 50%, 75%, 90%:

USCALE = 45.5

Example 2: You want to reverse the measurement directions, since the data matrix is transposed so that the 'items' are examinees and the 'persons' are test questions:

USCALE = -1
KEY=, RESCORE=, ISGROUPS= will still apply to the columns, not the rows, of the data matrix.
Centering will still be on the column measures.

Example 3: You want to approximate the "probit" measures used in many statistical procedures.

UPMEAN = 0  ; set the person sample mean to zero
USCALE = 0.59  ; probits * 1.7 = logits

Example 4: You want to report measures in "log-odds units to the base 10" instead of the standard "logits to the base e".

USCALE=0.434294482 ; the conversion between "natural" and base-10 logarithms.
For more examples, and how to compute this by hand, see User-friendly rescaling

202. **W300 Output files in Winsteps 3.00 format**

New or revised statistics are added as extra or changed columns in IFILE= and PFILE= when they are introduced into Winsteps.

To revert to an earlier format of IFILE= and PFILE=, specify W300=Yes. This produces the PFILE= and IFILE= in the format of Winsteps 3.00 1/1/2000.

Example: IFILE= in current format:

```
;ACT      FILE FOR
;ENTRY MEASURE STTS COUNT  SCORE  ERROR  IN.MSQ IN.ZSTD OUT.MS OUT.ZSTD DISPL PTME WEIGHT DISCR NAME
1  -.89 1  75.0  109.0  .23    .74 -1.97 .67 -1.89 .00 .64 1.00 1.06 R WATCH BIRDS
2  -.61 1  75.0  116.0  .20    .76 -1.54 .56 -1.55 .00 .58 1.00 1.07 R READ BOOKS ON ANIMALS
```

IFILE= in W300=Yes format:

```
;ACT      FILE FOR
;ENTRY MEASURE ST COUNT  SCORE  ERROR  IN.MSQ IN.ZSTD OUT.MS OUT.ZSTD DISPL CORR NAME
1  -.89 1  75  109    .23    .74 -1.97 .67 -1.89 .00 .64 R WATCH BIRDS
2  -.61 1  75  116    .20    .76 -1.54 .56 -1.55 .00 .58 R READ BOOKS ON ANIMALS
```

PFILE= in current format:

```
;PUPIL    FILE FOR
;ENTRY MEASURE STTS COUNT  SCORE  ERROR  IN.MSQ IN.ZSTD OUT.MS OUT.ZSTD DISPL PTME WEIGHT DISCR NAME
1  .49 1  25.0  30.0  .35    .96 -.15 .84 -.43 .00 .69 ROSSNER, MARC DANIEL
2  5.99 0  25.0  50.0  1.84  1.00 .00 .00 1.00 .00 ROSSNER, LAWRENCE F.
```

PFILE= in W300=Yes format:

```
;PUPIL    FILE FOR
;ENTRY MEASURE ST COUNT  SCORE  ERROR  IN.MSQ IN.ZSTD OUT.MS OUT.ZSTD DISPL CORR NAME
1  .49 1  25  30    .35    .96 -.15 .84 -.43 .00 .69 ROSSNER, MARC DANIEL
2  5.99 0  25  50  1.84  1.00 .00 .00 .00 ROSSNER, LAWRENCE F.
```

Notes:
TOTAL=YES is active for both current and old formats.
```` shown for extreme scores, such as person 2, in current format, but not in old format.
COUNT and SCORE are shown rounded to nearest integer in old format.

203. **WHEXACT Wilson-Hilferty exact normalization**

Some versions of Winsteps have the standard WHEXACT=NO.

ZSTD INFIT is the "t standardized Weighted Mean Square" shown at the bottom of RSA p. 100. ZSTD (standardized as a z-score) is used of a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's t-distribution value has been adjusted to a unit normal value.

ZSTD OUTFIT is the "t standardized Unweighted Mean Square" based on the terms on RSA p. 100.

The Wilson-Hilferty transformation converts mean-square values to their equivalent "t standardized" normal deviates. See RSA p. 101
Under certain circumstances, it can correctly report the paradoxical finding that the mean-square apparently reports an overfit condition, but the normal deviate an underfit.

To allow this possibility, specify WHEXACT=Y

To suppress it, specify WHEXACT=N

Example: A person takes a test of 20 dichotomous items and obtains an unweighted chi-square value of 19.5.

WHEXACT=Y
The OUTFIT mean-square is 0.975, i.e., apparently slightly overfitting. The exact normal deviate is .03, i.e., very slightly underfitting.

WHEXACT=N
The OUTFIT mean-square is 0.975, i.e., apparently slightly overfitting. The reported normal deviate is -.08, i.e., slightly overfitting.

204. XFILE analyzed response file

If XFILE=filename is specified in the control file, a file is output which enables a detailed analysis of individual response anomalies. This file contains 4 heading lines (unless HLINES=N) followed by one line for each person-by-item response used in the estimation. Each line contains:
1. Person number (I7) (PERSON)
2. Item number (I7) (ITEM)
3. Original response value (after keying/scoring) (I4) (OBS)
4. Observed response value (after recounting) (I4) (ORD)
5. Expected response value (F7.3) (EXPECT)
6. Modeled variance of observed values around the expected value (F7.3) (VARIAN)
   This is also the statistical information in the observation.
   Square root(modeled variance) is the observation's raw score standard deviation.
7. Standardized residual: (Observed - Expected)/Square root (Variance) (F7.3) (ZSCORE)
8. Score residual: (Observed - Expected) (F7.3) (RESIDL)
9. Person measure in USCALE= units (F7.2*) (PERMEA)
10. Item measure in USCALE= units (F7.2*) (ITMMEA)
11. Measure difference (Person measure - Item measure) in USCALE= units (F7.2*) (MEASDF)
12. Log-Probability of observed response (F7.3) (L-PROB)
13. Predicted person measure from this response alone in USCALE= units (F7.2*) (PPMEAS)
14. Predicted item measure from this response alone in USCALE= units (F7.2*) (PIMEAS)
15. Response code in data file (A) (CODE)
16. Person label (A) (PLABEL)
17. Item label (A) (ILABEL)

2* means decimal places set by UDECIM=.

Fields can be selected interactively, see below.

If CSV=Y, the values are separated by commas. When CSV=T, the commas are replaced by tab characters. For "non-numeric values in quotation marks", specify QUOTED=Y.

This file enables a detailed analysis of individual response anomalies. The response residual can be analyzed in three forms:
1) in response-level score units, from [(observed value - expected value)].
2) in logits, from [(observed value - expected value)/variance].
3) in standard units, [(observed value - expected value)/(square root of variance)].
The log-probabilities can be summed to construct log-likelihood and chi-square tests. Asymptotically, "chi-square = -2*log-likelihood".

**Predicted person measure:** Imagine that this observation was the only observation made for the person ... this value is the measure we would predict for that person given the item measure.

**Predicted item measure:** Imagine that this observation is the only observation made for this item ... this value is the measure we would predict for that item given the person measure.

The formulas are the same as for a response string of more than 1 observation. For dichotomies, see [www.rasch.org/rmt/rmt102t.htm](http://www.rasch.org/rmt/rmt102t.htm) and for polytomies [www.rasch.org/rmt/rmt122q.htm](http://www.rasch.org/rmt/rmt122q.htm)

Example: You wish to write a file on disk called "MYDATA.XF" containing response-level information for use in examining particularly response patterns:

```
XFILE=MYDATA.XF
```

Example: You wish to compute differential item functioning, DIF, for a specific subset of people:
If Table 30 is not suitable, here is a simple approximation:
Since one item does not have enough information to measure a person, for item bias we have to do it on the basis of a subset of people.
From the XFILE,
add the "score residuals" (not standardized) for everyone in subset "A" on a particular item.
Add the "modelled variance" for everyone in the subset.
Divide the residual sum by the variance sum. This gives an estimate of the DIF for subset "A" relative to the grand mean measure.
Do the same for subset "B" on the same item.
To contrast subset "A" with subset "B" then
DIF size "AB" =DIF estimate for "A" - DIF estimate for "B"
A significance t-test is \[ t = \frac{DIF \text{ size } "AB"}{\sqrt{\frac{1}{\text{variance sum for subset A}} + \frac{1}{\text{variance sum for subset B}}}} \]

When called from the "Output Files" menu, you can select what fields to include in the file. And you can also specify particular persons and/or items. If this is too big for your screen see Display too big.

![Fields in analyzed Response File: XFILE=](image)

**205. XMLEX consistent, almost unbiased, estimation**

**Experimental!** This implements Linacre's (1989) XCON algorithm as XMLE "Exclusory Maximum Likelihood Estimation".

The reason that JMLE is statistically inconsistent under some conditions, and noticeably estimation-biased for
short tests or small samples, is that it includes the possibility of extreme scores in the estimation space, but
cannot actually estimate them. The XMLE algorithm essentially removes the possibility of extreme response
vectors from the estimation space. This makes XMLE consistent, and much less biased than JMLE. In fact it is
even less biased than CMLE for small samples, this is because CMLE only eliminates the possibility of extreme
person response vectors, not the possibility of extreme item response vectors.

Considerations with XMLE=YES include:
(1) Anchoring values changes the XMLE probabilities. Consequently, measures from, say, a Table 20 score table
do not match measures from the estimation run. Consequently, it may be necessary to estimate item calibrations
with XMLE=YES. Then anchor the items and perform XMLE=NO.
(2) Items and persons with extreme (zero and perfect) scores are deleted from the analysis.
(3) For particular data structures, measures for finite scores may not be calculable.

Selecting XMLE=YES, automatically makes STBIAS=NO and PAIRED=NO, because XMLE is a more powerful
bias correction technique.

Example: Produce XMLE estimates, to compare with JMLE estimates, and so investigate the size of the JMLE
estimation bias.
XMLE=YES

206. **XWIDE columns per response**

The number of columns taken up by each response in your data file (1 or 2). If possible enter your data one
column per response. If there are two columns per response, make XWIDE=2. If your data includes responses
entered in both 1 and 2 character-width formats, use FORMAT= to convert all to XWIDE=2 format. When
XWIDE=2, these control variables require two columns per item or per response code: CODES=, KEYn=,
KEYSCR=, NEWSCORE=, IVALUE=. Either 1 or XWIDE= columns can be used for RESCORE=, ISGROUPS=,
RESCORE= and IREFER=.

Example 1: The responses are scanned into adjacent columns in the data records,
```
XWIDE=1 Observations 1 column wide
```

Example 2: Each response is a rating on a rating scale from 1 to 10, and so requires two columns in the date
record,
```
XWIDE=2 2 columns per datum
```

Example 3: Some responses take one column, and some two columns in the data record. Five items of 1-
character width, code "a", "b", "c", or "d", then ten items of 2-character width, coded "AA", "BB", "CC", "DD".
These are preceded by person-id of 30 characters.
```
XWIDE=2 Format to two columns per response
FORMAT=(30A1,5A1,10A2) Name 30 characters, 5 1-chars, 10 2-chars
CODES ="a b c d AABBCDDD" "a" becomes "a "
NEWSCORE="1 2 3 4 1 2 3 4 
" response values
RESCORE=2 rescore all items
NAME=1 person id starts in column 1
ITEM=31 item responses start in column 31
NI=15 15 items all now XWIDE=2
```

207. **@Field name for location in label**

@Fieldname= allows for user-defined names for locations with the person or item labels to be specified with the
column selection rules.

@Fieldname = value

Field name: a user-specified name which can include letters and numbers, but not = signs. Field names are
converted to capital letters, and must be referenced in full.

Value: a user-specified values which must accord with the column selection rules.
Example 1: The gender of persons is in column 14 of the person label. A DIF report on gender is wanted.

```plaintext
@GENDER = 14 ; gender indicator in column 14 of the person label
DIF = @GENDER ; DIF classification is by Gender column
or
DIF = @gender ; lower case letters in field names are allowed
but not
DIF = @GEN ; abbreviations of field names are not allowed
```

TFILE=* ; produce the DIF Table
*

This can also be done by the pull-down menus
Specification menu box: @GENDER = 14
Output Tables menu: 30. Items: DIF
Right-click on DIF selection box: @GENDER
Click on OK box

208. &END end of control variables

The first section in a control file contains the control variables, one per line. Its end is indicated by &END.

```plaintext
TITLE = "Example control file"
ITEM1 = 1
NI = 10
NAME1 = 12
&END
......; Item labels here
END LABELS
```

209. &INST start of control instructions

&INST is ignored by current versions of Winsteps. It is maintained for backward compatibility with earlier versions, where it was required to be the first control instruction. It is still present in some example files, again for backwards compatibility.

```plaintext
&INST ; this is allowed for compatibility
TITLE = "Old control file"
....
```

210. The Iteration Screen

While WINSTEPS is running, information about the analysis is displayed on the screen. The iterative estimation process is by logistic curve-fitting. Here is an example based on the "Liking for Science" data. The analysis was initiated with:

```plaintext
C:> WINSTEPS SF.TXT SFO.TXT (Enter)
```

The "====" is a horizontal bar-chart which moves from left to right to show progress through the work file during each phase of the analysis.

The screen display includes:

```plaintext
WINSTEPS  Version: 2.58  Program running - shows version number
Reading Control Variables .. Processing your control variables
Reading keys, groupings etc.. Processing special scoring instructions
Input in process .. Reading in your data: each . is 1,000 persons
1 2 3 4 5 6 7
1234567890123456789012345678901234567890123456789012345678901234567890 column in data record
```

180
Writing response file if RFILE= specified

To stop iterations: Press Ctrl with S
press Ctrl with F to cancel program

CONVERGENCE TABLE for more details see Table 0.2

Control: sf.TXT                         Output: sfO.TXT
PROX          ACTIVE COUNT       EXTREME 5 RANGE      MAX LOGIT CHANGE
ITERATION   PERSONS ITEMS   CATS     PERSONS ITEMS      MEASURES  STRUCTURES
1        76      25     3       3.78    3.20        3.8918     .0740
2        74      25     3       4.53    3.67         .7628    -.6167

DROPPING OUT OF RANGE OBSERVATIONS
This is reported for CUTLO= and CUTHI=
3        74      25     3       4.73    3.85         .2143    -.0991
4        74      25     3       4.82    3.90         .0846    -.0326

WARNING: DATA MAY BE AMBIGUOUSLY CONNECTED INTO 6 SUBSETS, see Connection Ambiguities

Control: sf.TXT                         Output: sfO.TXT
JMLE     MAX SCORE   MAX LOGIT    LEAST CONVERGED     CATEGORY STRUCTURE
ITERATION   RESIDUAL*    CHANGE    PERSON  ITEM    CAT   RESIDUAL   CHANGE
1         3.01      -.4155      60     24*     2      27.64   -.0184
2          .50      -.0258      53     24*     1       6.88    .0198
3          -.37       .0292      53      5*     1       3.10    .0091
4          .26       .0206      53      21*     1       2.74    .0079
5          .20       .0154      53      21*     0       1.90    .0056
6          .15       .0113      53     21*     0       1.42    .0042
7          .11       .0083      53     21*     0       1.05    .0030

Check that values diminish, and are near 0.

Calculating Fit Statistics an extra pass to calculate fit statistics

If Tables= is specified:
Processing Misfitting PERSONS for Table 7
Calculating Correlations for Table 10
Calculating Principal Components for Table 10
...............*...............*: one . per iteration, one * per contrast.
Processing Misfitting ITEMS for Table 11
Sorting ITEMS for Table 12
Sorting ITEMS for Table 15
Sorting PERSONS for Table 16
Sorting PERSONS for Table 19
Calculating Scores for Table 20
Writing Sorted Responses in Table 22 Guttman Scalogram
Analysis completed of SF.TXT  name of control file

LIKING FOR SCIENCE-(Wright-&-Masters-p.18)---------------

+-----------------------------------------------------------------------------+
|   76 KIDS    IN     74 KIDS    MEASURED             INFIT         OUTFIT    |
| SCORE      COUNT   MEASURED ERROR IMNSQ ZSTD OMNSQ ZSTD | INFIT OUTFIT |
| MEAN      26.4      16.8       56.99    5.74      1.01    -.2 .82    -.3 |
| S.D.      11.9      5.7        23.67    1.33       .65      1.4 .78    1.2 |
| REAL RMSE   5.90  ADJ.SD   22.93  SEPARATION  3.89  KID RELIABILITY .94 |
+-----------------------------------------------------------------------------+

| 25 TAPS IN  25 TAPS MEASURED INFIT OUTFIT |
| SCORE      COUNT   MEASURED ERROR IMNSQ ZSTD OMNSQ ZSTD | INFIT OUTFIT |
| MEAN      78.2      49.7       50.00    3.48      1.06   .0 .89    -.1 |
| S.D.      43.0      22.5       27.56    1.04       .36    1.3 .43    .7 |
| REAL RMSE   3.63  ADJ.SD   27.32  SEPARATION  7.53  TAP RELIABILITY .98 |
+-----------------------------------------------------------------------------+
TABLE 1.0 is identifies the current Table and sub-table.

LIKING FOR SCIENCE (Wright & Masters is set by TITLE=

ZOU214ws.txt is the name of the disk file containing this Table.

Feb 1 14:38 2005 is the date and time of this analysis.

INPUT: 75 KIDS, 25 ACTS

75 KIDS gives the number of cases in the data file(s) and the case (row) identification PERSON=

25 ACTS gives the number of items specified by NI= and the column (item) identification ITEM=

MEASURED: 75 KIDS, 25 ACTS, 3 CATS

The number of categories is determined by ISGROUPS= and the data structure. For details, see Table 3.1 and Table 3.2

Previously, ANALYZED was used to mean "used for item analysis", omitting extreme scores.

WINSTEPS 3.55.2 is the program name and version number performing this analysis.

212. Table 1.0, 1.2, 1.3, 1.10, 1.12 Distribution maps

(controlled by MRANGE=, MAXPAG=, NAMLMP=, ISORT=, PSORT=)

Table 1.2 is printed if the item map can be squeezed into one page. Table 1.3 is printed if the person map can be squeezed into one page. If person and item maps can be squeezed into one page, Table 1.0 is printed. You can use NAMLMP= to limit the number of characters of each name reported.

"TSMST" summarize the left-hand and right-hand distributions. An "M" marker represents the location of the mean measure. "S" markers are placed one sample standard deviation away from the mean. "T" markers are placed two sample standard deviations away.

MAXPAG= controls the length of the Table. MRANGE= controls the displayed range of measures. ISORT= and PSORT= control the sort order within rows.

In subtables 10 and above, the items are arranged by easiness. The item hierarchy is reversed.

Items arranged by measure: Look for the hierarchy of item names to spell out a meaningful construct from
The easiest items are at the bottom and the hardest items are at the top.

### PUPILS MAP OF ACTS

[Diagram with levels and items arranged by easiness]

**Subtables .10 and above:**

**Items arranged by easiness:** Look for the hierarchy of item names to spell out a meaningful construct from easiest at the top to hardest at the bottom.

The double line || indicates the two sides have opposite orientations. This is useful if the items and persons are being compared to the response structures.
LEARN WEED NAMES
WATCH ANIMAL MOVE
LOOK IN SIDEWALK  WATCH GRASS CHANGE
WATCH RATS
FIND BOTTLES AND
WATCH A RAT
FIND BOTTLES AND
WATCH A RAT
FIND BOTTLES AND

213. Table 1.1, 1.4 Distribution maps

These tables show the distribution of the persons and items. The variable is laid out vertically with the most able persons, and most difficult items at the top.

TABLE 1.1 LIKING FOR SCIENCE (Wright & Masters p.18)  SFO.TXT Dec 1 18:33 1996

76 PUPILS 25 ACTS ANALYZED: 74 PUPILS 25 ACTS 3 CATEGS v2.67

MAP OF KIDS AND TAPS

<table>
<thead>
<tr>
<th>MEASURE</th>
<th>KIDS</th>
<th>TAPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.0</td>
<td>XXXX</td>
<td>+</td>
</tr>
<tr>
<td>4.0</td>
<td>XXXX</td>
<td>+</td>
</tr>
<tr>
<td>3.0</td>
<td>XXXX</td>
<td>+</td>
</tr>
<tr>
<td>2.0</td>
<td>XXXX</td>
<td>+</td>
</tr>
<tr>
<td>1.0</td>
<td>XXXX</td>
<td>+</td>
</tr>
<tr>
<td>.0</td>
<td>XXXX</td>
<td>+</td>
</tr>
</tbody>
</table>

76 PUPILS 25 ACTS Number of persons in data file and items in NI= specification.
ANALYZED: 74 PUPILS 25 ACTS 3 CATEGS Number of persons, items and categories that with non-extreme scores.
v2.67 Winsteps version number.
In Table 1, each person or item is indicated by an "X", or, when there are too many "X"s to display on one line, several persons or items are represented by a "#". Less than that number by ".". So that if "#" represents 4, then "." represents 1 to 3.

The left-hand column locates the person ability measures along the variable. For dichotomous items, the right-hand column locates the item difficulty measures along the variable. Look for an even spread of items along the variable (the Y-axis) with no gaps, indicating poorly defined or tested regions of the variable. The persons often show a normal distribution. Good tests usually have the items targeted (lined up with) the persons.

For rating (or partial credit) scales, each item is shown three times in Table 1.4. In the center item column, each item is placed at its mean calibration, i.e., this is the location of the center of the rating (or partial credit) scale - the location at which being ratings in the top and bottom category are equally probable. In the left-hand item column, the item is shown at the measure level corresponding to a probability of .5 of exceeding (or being rated in) the bottom rating (or partial credit) scale category. In the right-hand item column, the item is shown at the measure level corresponding to a probability of .5 of being rated in (or falling below) the top rating (or partial credit) scale category. These locations are also shown in Table 2.3. Dichotomous items, "D", have only one location.

<table>
<thead>
<tr>
<th>MEASURE</th>
<th>MAP OF PUPILS</th>
<th>AND ACTS</th>
<th>P=50%</th>
<th>MEASURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.0</td>
<td>X +</td>
<td>+</td>
<td>+</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.0</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3.0</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>XX</td>
<td></td>
<td></td>
<td>XX</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>2.0</td>
<td>XXX +</td>
<td>+</td>
<td>+</td>
<td>2.0</td>
</tr>
<tr>
<td>XX</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>XX</td>
<td>XX</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>XX</td>
<td>X</td>
<td>XXX</td>
<td></td>
<td>XXX</td>
</tr>
<tr>
<td>XXX</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>1.0</td>
<td>XXXXXX +</td>
<td>+</td>
<td>X</td>
<td>1.0</td>
</tr>
<tr>
<td>XXX</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>XXXXXX</td>
<td>XX</td>
<td>X</td>
<td>XX</td>
<td>XXX</td>
</tr>
<tr>
<td>XXXXXX</td>
<td>XX</td>
<td>X</td>
<td>XX</td>
<td>X</td>
</tr>
<tr>
<td>XXXXXX</td>
<td>X</td>
<td></td>
<td></td>
<td>XXX</td>
</tr>
<tr>
<td>XXXXXX</td>
<td>X</td>
<td></td>
<td></td>
<td>XX</td>
</tr>
<tr>
<td>XXXXXX</td>
<td>+</td>
<td>+</td>
<td>X</td>
<td>0.0</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>XX</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>XX</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>-1.0</td>
<td>X +</td>
<td>+</td>
<td>+</td>
<td>-1.0</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>XX</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Observe that the top pupil (left column) is well above the top category of the most difficult act item (right-most column), but that all pupils are above the top category of the easiest item (bottom X in right-most column). "Above" means with greater than 50% chance of exceeding.

214. Table 2 Multiple-choice distractor plot

Here are the Tables produced by selecting "2.0 Measure Forms (All)" on the Output Tables pull-down menu. They reflect different conceptualizations of the category structure.

The codes for the response options (distractors) are located according to the measures corresponding to them. Each subtable is presented two ways: with the response code itself (or one of them if several would be in the same place), e.g., Table 2.1, and with the score corresponding to the option, e.g. Table 2.11 (numbered 10 subtables higher).

Table 2.1: shows the most probable response on the latent variable. In this example, for item "al07", "a" (or any other incorrect option) is most probable up to 3.2 logits, when "d", the correct response, becomes most probable according to the Rasch model.

Table 2.11 is the same as Table 1, but the options are shown by their scored values, not by their codes in the data.

Table 2.2: shows the predicted average response on the latent variable. In this example, for item "al07", "a" (or any other incorrect option) is the predicted average response up to 3.2 logits, then "d", the correct response, becomes the average predictions. The ":" is at the transition from an average expected wrong response to an
average expected "right" response, i.e., where the predicted average score on the item is 0.5, the Rasch-half-point thresholds. The "a" below "2" is positions where the expected average score on the item is 0.25. Similarly "d" would be repeated where the expected average score on the item is 0.75, according to the Rasch model.

TABLE 2.2 EXPECTED SCORE: MEAN (":" INDICATES HALF-POINT THRESHOLD) (ILLUSTRATED BY AN OBSERVED CATEGORY)
-4 -3 -2 -1 0 1 2 3 4
|-----------------------------| NUM TOPIC
a a : d 55 al07 newspaper
a

Table 2.12 is the same as Table 2, but the options are shown by their scored values, not by their codes in the data.

TABLE 2.12 EXPECTED SCORE: MEAN (":" INDICATES HALF-POINT THRESHOLD) (BY CATEGORY SCORE)
-4 -3 -2 -1 0 1 2 3 4
|-----------------------------| NUM TOPIC
0 0 : 1 55 al07 newspaper
0 0 : 1 64 sa01 magazine

Table 2.3 shows the 50% cumulative probability points, the Rasch-Thurstone thresholds. The lower category ("a" and other wrong answers) has a greater than than 50% probability of being observed up to 3.2 logits, when "d", the correct answer, has a higher than 50% probability.

TABLE 2.3 50% CUMULATIVE PROBABILITY: MEDIAN (ILLUSTRATED BY AN OBSERVED CATEGORY)
-4 -3 -2 -1 0 1 2 3 4
|-----------------------------| NUM TOPIC
d d 55 al07 newspaper

Table 2.13 is the same as Table 3, but the options are shown by their scored values, not by their codes in the data.

TABLE 2.13 50% CUMULATIVE PROBABILITY: MEDIAN (BY CATEGORY SCORE)
-4 -3 -2 -1 0 1 2 3 4
|-----------------------------| NUM TOPIC
1 1 55 al07 newspaper
1 1 64 sa01 magazine

Table 2.4 shows the item difficulties (or more generally the Rasch-Andrich thresholds) coded by the option of the higher category. For item "al07" this is "d", the correct option.

TABLE 2.4 STRUCTURE MEASURES (Rasch model parameters: equal-adjacent-probability thresholds) (ILLUSTRATED BY AN OBSERVED CATEGORY)
-4 -3 -2 -1 0 1 2 3 4
|-----------------------------| NUM TOPIC
d 55 al07 newspaper
c

Table 2.14 is the same as Table 4, the Rasch-Andrich thresholds, but the options are shown by their scored values, not by their codes in the data.

TABLE 2.14 STRUCTURE MEASURES (Rasch model parameters: equal-adjacent-probability thresholds) (BY CATEGORY SCORE)
-4 -3 -2 -1 0 1 2 3 4
|-----------------------------| NUM TOPIC
1 55 al07 newspaper
1 64 sa01 magazine

Table 2.5 shows the average measures of persons choosing wrong distractors (illustrated by one of the wrong distractors, "a") and the average measures or persons choosing a correct distractor (illustrated by one of the correct distractors, "d").

TABLE 2.5 OBSERVED AVERAGE MEASURES FOR STUDENTS (scored) (ILLUSTRATED BY AN OBSERVED
Table 2.15 is the same as Table 5, but the options are shown by their scored values, not by their codes in the data.

TABLE 2.15 OBSERVED AVERAGE MEASURES FOR STUDENTS (scored) (BY CATEGORY SCORE)

<table>
<thead>
<tr>
<th>NUM</th>
<th>TOPIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.6, shown first from the Diagnosis menu, shows the average measures of the persons choosing each distractor. "m" usually indicates the average measure of persons with missing data.

TABLE 2.6 OBSERVED AVERAGE MEASURES FOR STUDENTS (unscored) (BY OBSERVED CATEGORY)

<table>
<thead>
<tr>
<th>NUM</th>
<th>TOPIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.16 is the same as Table 6, but the options are shown by their scored values, not by their codes in the data.

TABLE 2.16 OBSERVED AVERAGE MEASURES FOR STUDENTS (unscored) (BY CATEGORY SCORE)

<table>
<thead>
<tr>
<th>NUM</th>
<th>TOPIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.7 shows the measures that would be predicted to be observed for incorrect and correct responses if the persons responded exactly as the Rasch model predicts. "a" (an incorrect distractor) shows the average measure for persons in the sample who would be predicted to fail the item, and "d" (a correct distractor) shows the average measure for persons in the sample who would be predicted to succeed on the item.

TABLE 2.7 EXPECTED AVERAGE MEASURES FOR STUDENTS (scored) (ILLUSTRATED BY AN OBSERVED CATEGORY)

<table>
<thead>
<tr>
<th>NUM</th>
<th>TOPIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.17 is the same as Table 7, but the options are shown by their scored values, not by their codes in the data.

TABLE 2.17 EXPECTED AVERAGE MEASURES FOR STUDENTS (scored) (BY CATEGORY SCORE)

<table>
<thead>
<tr>
<th>NUM</th>
<th>TOPIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

215. **Table 2** Most probable, expected, cumulative, structure, average measures

(controlled by MRANGE=, CATREF=, CURVES=)

Each plot answers a different question:

- What category is most likely? The maximum probability (mode) plot.
- What is the average or expected category value? The expected score (mean) plot.
- What part of the variable corresponds to the category? The cumulative probability (median) plot.
Which Table should be used for a standard setting procedure?

Most standard setting is based on "average" or "frequency" considerations. For instance,

"If we observed 1000 candidates whose measures are known to be exactly at the pass-fail point, ...

..., we would expect their average score to be the pass-fail score." If this is how you think, then the Table you want is Table 2.2 (matches Table 12.5)

..., we would expect 50% to pass and 50% to fail the pass-fail score." If this is how you think, then the Table you want is Table 2.3 (matches 12.6)

..., we would expect more to be in the criterion pass-fail category of each item than any other category." If this is how you think, then the Table you want is Table 2.1 (no matching 12.)

"Our current sample is definitive, ...

..., we would expect the next sample to behave in exactly the same way this sample did." If this is how you think, then the Table you want is Table 2.5 (or Table 2.6, if the responses have been rescored.)

..., we would expect the next sample to behave the way this sample should have behaved, if this sample had conformed to the Rasch model." If this is how you think, then the Table you want is Table 2.7.

The left-side of this table lists the items in descending order of measure. Anchored items are indicated by an * between the sequence number and name. A particular category can be used as the reference for sorting the items by specifying the CATREF= variable.

Across the bottom is the logit (or user-rescaled) variable with the distribution of the person measures shown beneath it. An "M" marker represents the location of the mean person measure. "S" markers are placed one sample standard deviation away from the mean. "T" markers are placed two sample standard deviations away. An "M" inside a plot indicates the measure corresponding to missing data.

To produce all subtables of Table 2, request Table 2.0

Tables 2.1 & 2.11: The "Most Probable Response" Table, selected with CURVES=001, answers the question "which category is a person of a particular measure most likely to choose?" This is the most likely category with which the persons of logit (or user-rescaled) measure shown below would respond to the item shown on the left. The area to the extreme left is all "0"; the area to the extreme right is at the top category. Each category number is shown to the left of its modal area. If a category is not shown, it is never a most likely response. An item with an extreme, perfect or zero, score is not strictly estimable, and is omitted here. Blank lines are used to indicate large gaps between items along the variable.

This table presents in one picture the results of this analysis in a form suitable for inference. We can predict for people of any particular measure measure what responses they would probably make. "M" depicts an "average" person. The left "T" a low performer. The right "T" a high performer. Look straight up from those letters to read off the expected response profiles.

Table 2.1 to 2.7 reports with observed categories, i.e., those in the CODES= statement.
Table 2.11 to 2.17 report with scored categories, i.e., after IVALUE=, RESCORE=, KEY1=, etc., but only if different from Table 2.1 to 2.7.

<table>
<thead>
<tr>
<th>MOST PROBABLE RESPONSE: MODE (BETWEEN &quot;0&quot; AND &quot;1&quot; IS &quot;0&quot;, ETC.)</th>
<th>NUM</th>
<th>ITEM</th>
<th>-5</th>
<th>-4</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 FIND BOTTLES 0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 WATCH BUGS 0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 LOOK IN SIDE 0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 WATCH ANIMAL 0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 WATCH WHAT A 0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

189
Tables 2.2 & 2.12: In the "Expected Score" Table, the standard output (or selected with CURVES=010) answers the question "what is the average rating that we expect to observer for persons of a particular measure?" This rating information is expressed in terms of expected scores (with ":" at the half-point thresholds). Extreme scores are located at expected scores .25 score points away from the extremes.

EXPECTED SCORE: MEAN (":" INDICATES HALF-POINT THRESHOLDS)
-5 -4 -3 -2 -1 0 1 2 3 4 5

<table>
<thead>
<tr>
<th>5 FIND BOTTLES 0</th>
<th>23 WATCH A RAT 0</th>
<th>9 LEARN WEED N 0</th>
<th>21 WATCH BIRD M 0</th>
<th>11 FIND WHERE A 0</th>
<th>19 GO TO ZOO 0</th>
<th>18 GO ON PICNIC 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>: 1</td>
<td>: 1</td>
<td>: 1</td>
<td>: 1</td>
<td>: 1</td>
<td>: 1</td>
<td>: 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tables 2.2 & 2.13: The "Cumulative Probability" Table: Rasch-Thurstone thresholds, selected with CURVES=001, answers the question "whereabouts in the category ordering is a person of a particular measure located?" This information is expressed in terms of median cumulative probabilities (the point at which the probability of scoring in or above the category is .5).

CUMULATIVE PROBABILITY: MEDIAN
-6 -4 -2 0 2 4 6

<table>
<thead>
<tr>
<th>5 FIND BOTTLES 0</th>
<th>23 WATCH A RAT 0</th>
<th>9 LEARN WEED N 0</th>
<th>21 WATCH BIRD M 0</th>
<th>11 FIND WHERE A 0</th>
<th>19 GO TO ZOO 0</th>
<th>18 GO ON PICNIC 0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tables 2.4 & 2.14 show Rasch structure calibrations: Rasch-Andrich thresholds (step parameters, step measures, step difficulties, rating (or partial credit) scale calibrations). These are the relationships between adjacent categories, and correspond to the points where adjacent category probability curves cross, i.e., are equally probable of being observed according to a Rasch model.

STRUCTURE MEASURES (Rasch Parameters)

<table>
<thead>
<tr>
<th>-40 -30 -20 -10 0 10 20 30 40 50</th>
<th>NUM ACT</th>
<th>5 FIND BOTTLES AND CANS</th>
<th>1 WATCH BIRDS</th>
<th>13 GROW GARDEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td></td>
<td>13 GROW GARDEN</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>12 GO TO MUSEUM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>11 FIND WHERE ANIMAL LIVES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>10 LISTEN TO BIRD SING</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>19 GO TO ZOO</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Tables 2.5 & 2.15 plot the observed average person measures for each scored category. It reflects how this sample used these categories. The plotted values cannot fall outside the range of the sample.

OBSERVED AVERAGE MEASURES BY SCORED CATEGORY FOR PUPILS

| NUM | ACT |
|-----+-----|
| 00102 | 5 FIND BOTTLES AND CANS |
| 000102 | 23 WATCH A RAT |
| 00 01 02 | 20 WATCH BUGS |
| 00 01 02 | 4 WATCH GRASS CHANGE |
| 0001 02 | 8 LOOK IN SIDEWALK CRACKS |

Tables 2.6 & Table 2.16 plot the observed average person measures for each observed category. It reflects how this sample used these categories. The plotted values cannot fall outside the range of the sample. "m" in the plot indicates the average measure of those for whom their observation is missing on this item. This Table is shown first from the Diagnosis pull-down menu.

OBSERVED AVERAGE MEASURES BY OBSERVED CATEGORY FOR PUPILS

| NUM | ACT |
|-----+-----|
| 00102 m | 5 FIND BOTTLES AND CANS |
| 00 01 02 | 20 WATCH BUGS |
| 00 01 02 | 4 WATCH GRASS CHANGE |
| 0001 02 | 8 LOOK IN SIDEWALK CRACKS |

Tables 2.7 & 2.17 plot the expected average person measures for each category score. It reflects how this sample were expected to use these categories. The plotted values cannot fall outside the range of the sample. This Table applies the empirical person distribution to Table 2.2.

EXPECTED AVERAGE MEASURES BY CATEGORY FOR PUPILS

| NUM | ACT |
|-----+-----|
| 00 01 02 | 5 FIND BOTTLES AND CANS |
| 00 01 02 | 23 WATCH A RAT |
| 00 01 02 | 20 WATCH BUGS |
| 00 01 02 | 4 WATCH GRASS CHANGE |
| 00 01 02 | 8 LOOK IN SIDEWALK CRACKS |

216. Table 3.1, 27.3, 28.3 Summaries of persons and items

This table summarizes the person, item and structure information. Extreme scores (zero and perfect scores) have no exact measure under Rasch model conditions, so they are dropped from the main summary statistics.
Using a Bayesian technique, however, reasonable measures are reported for each extreme score, see EXTRSC=. Totals including extreme scores are also reported, but are necessarily less inferentially secure than those totals only for non-extreme scores.

Table 3: Gives summaries for all persons and items.
Table 27: Gives subtotal summaries for items, controlled by ISUBTOT=
Table 28: Gives subtotal summaries for persons, controlled by PSUBTOT=

<table>
<thead>
<tr>
<th>SUMMARY OF 34 MEASURED (NON-EXTREME) KIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAW</td>
</tr>
<tr>
<td>SCORE COUNT MEASURE ERROR MNSQ ZSTD MNSQ ZSTD</td>
</tr>
<tr>
<td>MEAN 6.9 14.0 -.20 1.03 1.03 -.3 .73 -.3</td>
</tr>
<tr>
<td>S.D. 2.1 .0 2.07 .11 1.01 1.2 1.45 .5</td>
</tr>
<tr>
<td>MAX. 11.0 14.0 3.89 1.15 4.43 2.3 6.86 1.3</td>
</tr>
<tr>
<td>MIN. 2.0 14.0 -4.48 .82 .17 -1.6 .08 -.8</td>
</tr>
<tr>
<td>REAL RMSE 1.23 ADJ.SD 1.66 SEPARATION 1.35 KID RELIABILITY .65</td>
</tr>
<tr>
<td>MODEL RMSE 1.03 ADJ.SD 1.79 SEPARATION 1.73 KID RELIABILITY .75</td>
</tr>
<tr>
<td>S.E. OF KID MEAN = .36</td>
</tr>
<tr>
<td>MINIMUM EXTREME SCORE: 1 KIDS</td>
</tr>
<tr>
<td>MINIMUM EXTREME SCORE: 46 PUPILS</td>
</tr>
<tr>
<td>LACKING RESPONSES: 8 PUPILS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SUMMARY OF 35 MEASURED (EXTREME AND NON-EXTREME) KIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAW</td>
</tr>
<tr>
<td>SCORE COUNT MEASURE ERROR MNSQ ZSTD MNSQ ZSTD</td>
</tr>
<tr>
<td>MEAN 6.7 14.0 -.38 1.05</td>
</tr>
<tr>
<td>S.D. 2.4 .0 2.31 .18</td>
</tr>
<tr>
<td>MAX. 11.0 14.0 3.89 1.88</td>
</tr>
<tr>
<td>MIN. 0 14.0 -6.79 .82</td>
</tr>
<tr>
<td>REAL RMSE 1.25 ADJ.SD 1.95 SEPARATION 1.56 KID RELIABILITY .71</td>
</tr>
<tr>
<td>MODEL RMSE 1.07 ADJ.SD 2.05 SEPARATION 1.92 KID RELIABILITY .79</td>
</tr>
<tr>
<td>S.E. OF KID MEAN = .40</td>
</tr>
</tbody>
</table>

KID RAW SCORE-TO-MEASURE CORRELATION = 1.00 
CRONBACH ALPHA (KR-20) KID RAW SCORE RELIABILITY = .73  
UMEAN=.000 USCALE=1.000  
476 DATA POINTS. APPROXIMATE LOG-LIKELIHOOD CHI-SQUARE: 221.61  

For valid observations used in the estimation, 
SCORE is the raw score (number of correct responses).  
COUNT is the number of responses made.  
MEASURE is the estimated measure (for persons) or calibration (for items).  
ERROR is the standard error of the estimate.  

RAW SCORE-TO-MEASURE CORRELATION is the Pearson correlation between raw scores and measures, including extreme scores. When data are complete, this correlation is expected to be near 1.0 for persons and near -1.0 for items.  

INFIT is an information-weighted fit statistic, which is more sensitive to unexpected behavior affecting responses to items near the person's measure level.  

MNSQ is the mean-square infit statistic with expectation 1. Values substantially below 1 indicate dependency in your data; values substantially above 1 indicate noise.  

ZSTD is the infit mean-square fit statistic t standardized to approximate a theoretical mean 0 and variance 1 distribution. ZSTD (standardized as a z-score) is used of a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's t-distribution value has been adjusted to a unit normal value. When LOCAL=Y, then EMP is shown, indicating a local (0,1) standardization. When LOCAL=L, then LOG is shown, and the natural logarithms of the mean-squares are reported.
OUTFIT is an outlier-sensitive fit statistic, more sensitive to unexpected behavior by persons on items far from the person's measure level.

MNSQ is the mean-square outfit statistic with expectation 1. Values substantially less than 1 indicate dependency in your data; values substantially greater than 1 indicate the presence of unexpected outliers.

ZSTD is the outfit mean-square fit statistic standardized to approximate a theoretical mean 0 and variance 1 distribution. ZSTD (standardized as a z-score) is used of a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's t-distribution value has been adjusted to a unit normal value. When LOCAL=Y, then EMP is shown, indicating a local (0,1) standardization. When LOCAL=L, then LOG is shown, and the natural logarithms of the mean-squares are reported.

MEAN is the average value of the statistic.
S.D. is its sample standard deviation.
MAX. is its maximum value.
MIN. is its minimum value.

RMSE is the square-root of the average error variance. It is the Root Mean Square standard Error computed over the persons or over the items.

MODEL RMSE is computed on the basis that the data fit the model, and that all misfit in the data is merely a reflection of the stochastic nature of the model. This is a "best case" reliability, which reports an upper limit to the reliability of measures based on this set of items for this sample.

REAL RMSE is computed on the basis that misfit in the data is due to departures in the data from model specifications. This is a "worst case" reliability, which reports a lower limit to the reliability of measures based on this set of items for this sample.

ADJ. S.D. is the "adjusted" standard deviation, i.e., the "true" standard deviation. This is the sample standard deviation of the estimates after subtracting the error variance (attributable to their standard errors of measurement) from their observed variance.

\[(\text{ADJ.S.D.})^2 = (\text{S.D. of MEASURE})^2 - (\text{RMSE})^2\]

The ADJ.S.D. is an estimate of the "true" sample standard deviation from which the bias caused by measurement error has been removed.

SEPARATION is the ratio of the PERSON (or ITEM) ADJ.S.D., the "true" standard deviation, to RMSE, the error standard deviation. It provides a ratio measure of separation in RMSE units, which is easier to interpret than the reliability correlation. SEPARATION\(^2\) is the signal-to-noise ratio, the ratio of "true" variance to error variance.

RELIABILITY is a separation reliability. The PERSON (or ITEM) reliability is equivalent to KR-20, Cronbach Alpha, and the Generalizability Coefficient. See much more at Reliability.

S.E. OF MEAN is the standard error of the mean of the person (or item) measures for this sample.

WITH 1 EXTREME KIDS = 75 KIDS
MEAN is the mean of the measures including measures corresponding to extreme scores
S.D. is the sample standard deviation of those measures.
MEDIAN is the median measure of the sample including extreme scores (in Tables 27, 28).
The separation and reliability computations are repeated, but including any elements with extreme measures. Since the measures for extreme scores are imprecise, these statistics are often lower than their non-extreme equivalents. Conventional computation of a reliability coefficient (KR-20, Cronbach Alpha) includes persons with extreme scores. The PERSON SEP REL. of this second analysis is the conventional reliability, and is usually between the MODEL and REAL values, closer to the MODEL.

KID RAW SCORE-TO-MEASURE CORRELATION is the correlation between the marginal scores (person raw scores and item scores) and the corresponding measures. The item correlation is expected to be negative because higher measure implies lower probability of success and so lower item scores.

CRONBACH ALPHA (KR-20) KID RAW SCORE RELIABILITY is the conventional "test" reliability index. It reports an approximate test reliability based on the raw scores of this sample. It is only reported for complete data. See more at Reliability.
UMEAN=.000 USCALE=1.000 are the current settings of UMEAN= and USCALE=.

476 DATA POINTS is the number of observations that are used for standard estimation, and so are not missing and not in extreme scores.

APPROXIMATE LOG-LIKELIHOOD CHI-SQUARE: 221.61 is the approximate value of the global fit statistic. The accuracy of the approximation depends on how close the reported estimated measures are to their “true” maximum likelihood estimates. The degrees of freedom, d.f., of the chi-square are the number of data points less the number of free parameters, where number of free parameters for complete data = the minimum of ((number of different non-extreme person raw scores) or (number of different non-extreme item raw scores)) - 1 (for identifiability of local origin) + the sum, across rating scales, of the number of categories in each scale - 2 for each scale. Increase this by 1 for each for response string with a different missing data pattern.

217. Table 3.2 Summary of rating scale category structure

The average measures and category fit statistics are how the response structure worked "for this sample" (which might have high or low performers etc.). For each observation in category k, there is a person of measure Bn and an item of measure Di. Then:

average measure = sum( Bn - Di ) / count of observations in category. These are not estimates of parameters.

The probability curves are how the response structure is predicted to work for any future sample, provided it worked satisfactorily for this sample.

Our logic is that if the average measures and fit statistics don't look reasonable for this sample, why should they in any future sample? If they look OK for this sample, then the probability curves tell us about future samples. If they don’t look right now, then we can anticipate problems in the future.

a) For dichotomies,

SUMMARY OF MEASURED STRUCTURE
FOR GROUPING "0", MODEL "R", ACT NUMBER: 12 GO TO MUSEUM

ACT MEASURE OF -1.07 ADDED TO MEASURES
+-----------------------------------------------+-----+
|CATEGORY OBSERVED|OBSVD SAMPLE|INFIT OUTFIT| COHERENCE|ESTIM|
|LABEL SCORE COUNT|%AVRGE EXPECT| MNSQ | MNSQ | M->C C->M DISCR|
|-------------------+------------+------------+----------|-----|
| 1  1  13  18 | -.38 .01| .83 .52| 75%  23% |     |01 neutral
| 2  2  61  82 | 1.12 1.03| .78 .85| 85%  98% | 1.23|02 like
+-----------------------------------------------+-----+

AVERAGE MEASURE is mean of measures in category.
M->C = Does Measure imply Category?
C->M = Does Category imply Measure?

CATEGORY LABEL is the number of the category in your data set after scoring/keying.
CATEGORY SCORE is the ordinal value of the category used in computing raw scores - and in Table 20.
OBSERVED COUNT and % is the count of occurrences of this category used in the estimation (i.e., for non-extreme persons and items). Counts of all occurrences of categories are given in the distractor Tables, e.g., Table 14.3.

OBSVD AVERAGE is the average of the measures that are modelled to produce the responses observed in the category. The average measure is expected to increase with category value. Disordering is marked by **. This is a description of the sample, not a Rasch parameter. For each observation in category k, there is a person of measure Bn and an item of measure Di. Then:

average measure = sum( Bn - Di ) / count of observations in category.

SAMPLE EXPECT is the expected value of the average measure for this sample. These values always advance
with category. This is a description of the sample, not a Rasch parameter.

INFIT MNSQ is the average of the INFIT mean-squares associated with the responses in each category. The expected values for all categories are 1.0.

OUTFIT MNSQ is the average of the OUTFIT mean-squares associated with the responses in each category. The expected values for all categories are 1.0. This statistic is sensitive to grossly unexpected responses.

Note: Winsteps always reports the MNSQ values in Table 3.2. An approximation to their standardized values can be obtained by using the number of observations in the category as the degrees of freedom, and then looking at the plot below.

COHERENCE
M->C shows what percentage of the measures that were expected to produce observations in this category actually did. Do the measures imply the category?

Guttman's Coefficient of Reproducibility is the count-weighted average of the M->C, i.e.,
Reproducibility = sum (COUNT * M->C) / sum(COUNT * 100)

C->M shows what percentage of the observations in this category were produced by measures corresponding to the category. Does the category imply the measures?

ESTIM DISCR is an estimate of the local discrimination when the model is parameterized in the form: log-odds = aj (Bn - Dn - Fj)

RESIDUAL (when shown) is the residual difference between the observed and expected counts of observations in the category. Shown as % of expected, unless observed count is zero. Then residual count is shown. Only shown if residual count is >= 1.0. Indicates lack of convergence, structure anchoring, or large data set.

CATEGORY CODES and LABELS are shown to the right based on CODES=, CFILE= and CLFILE=.

Measures corresponding to the dichotomous categories are not shown, but can be computed using the Table at "What is a Logit?" and LOWADJ= and HIADJ=.

b) For rating (or partial credit) scales, the structure calibration table lists:

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>OBSERVED</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>STRUCTURE</th>
<th>CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>5</td>
<td>14</td>
<td>-.51</td>
<td>.89</td>
<td>.68</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>26</td>
<td>.39</td>
<td>1.45</td>
<td>1.63</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>60</td>
<td>.73</td>
<td>1.34</td>
<td>1.32</td>
</tr>
</tbody>
</table>

AVERAGE MEASURE is mean of measures in category.

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>MEASURE</th>
<th>S.E.</th>
<th>AT CAT.</th>
<th>PROBABILITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 Supervision</td>
<td>NONE</td>
<td>(-.22)</td>
<td>INF</td>
<td>-1.50</td>
</tr>
<tr>
<td>6 Device</td>
<td>-.79</td>
<td>.52</td>
<td>-.61</td>
<td>-1.50</td>
</tr>
<tr>
<td>7 Independent</td>
<td>-.43</td>
<td>.39</td>
<td>1.00</td>
<td>.28</td>
</tr>
</tbody>
</table>

M->C = Does Measure imply Category?
C->M = Does Category imply Measure?
ITEM MEASURE OF -.64 ADDED TO MEASURES
When there is only one item in a grouping (the Partial Credit model), the item measure is added to the reported measures.

CATEGORY LABEL, the number of the category in your data set after scoring/keying.
CATEGORY SCORE is the value of the category in computing raw scores - and in Table 20.
OBSERVED COUNT and %, the count of occurrences of this category used in the estimation.

OBSVD AVERAGE is the average of the measures that are modelled to produce the responses observed in the category. The average measure is expected to increase with category value. Disordering is marked by "***". This is a description of the sample, not the estimate of a parameter. For each observation in category k, there is a person of measure Bn and an item of measure Di. Then: average measure = \( \frac{\text{sum}(Bn - Di)}{\text{count of observations in category}} \).

SAMPLE EXPECT is the expected value of the average measure for this sample. These values always advance with category. This is a description of the sample, not a Rasch parameter.

INFIT MNSQ is the average of the INFIT mean-squares associated with the responses in each category. The expected values for all categories are 1.0.

OUTFIT MNSQ is the average of the OUTFIT mean-squares associated with the responses in each category. The expected values for all categories are 1.0. This statistic is sensitive to grossly unexpected responses.

Note: Winsteps always reports the MNSQ values in Table 3.2. An approximation to their standardized values can be obtained by using the number of observations in the category as the degrees of freedom, and then looking at the plot below.

STRUCTURE CALIBRATN, the calibrated measure of the transition from the category below to this category. This is an estimate of the Rasch model parameter, Fj. Use this for anchoring in Winsteps. (This corresponds to Fj in the Di+Fj parameterization of the "Rating Scale" model, and is similarly applied as the Fij of the Dij=Di+Fij of the "Partial Credit" model.) The bottom category no prior transition, and so that the measure is shown as NONE. The Rasch-Andrich threshold is expected to increase with category value. This parameter, sometimes called the Step Difficulty, Step Calibration, Rasch-Andrich threshold, Tau or Delta, indicates how difficult it is to observe a category, not how difficult it is to perform it. Disordering of these estimates (so that they do not ascend in value up the rating scale), sometimes called "disordered deltas", indicates that the category is relatively rarely observed, i.e., occupies a narrow interval on the latent variable, and so may indicate substantive problems with the rating (or partial credit) scale category definitions. These Rasch-Andrich thresholds are relative pair-wise measures of the transitions between categories. They are the points at which adjacent category probability curves intersect. They are not the measures of the categories. See plot below.

CATEGORY MEASURE, the sample-free measure corresponding to this category. ( ) is printed where the matching calibration is infinite. The value shown corresponds to the measure .25 score points (or LOWADJ= and HIADJ=) away from the extreme. This is the best basis for the inference: "ratings averaging x imply measures of y" or "measures of y imply ratings averaging x". This is implied by the Rasch model parameters.

STRUCTURE MEASURE, item measure add to the calibrated measure of this transition from the category below to this category. For structures with only a single item, this is an estimate of the Rasch model parameter, Dij = Di + Fij. (This corresponds to the Dij parameterization of the "Partial Credit" model.) The bottom category has no prior transition, and so that the measure is shown as NONE. The Rasch-Andrich threshold is expected to increase with category value, but these can be disordered. "Dgi + Fgj" locations are plotted in Table 2.4, where "g" refers to the ISGROUPS= assignment.

STRUCTURE S.E. is an approximate standard error of the Rasch-Andrich threshold measure.

SCORE-TO-MEASURE
These values are plotted in Table 21, "Expected Score" ogives. They are useful for quantifying category
measures. *This is implied by the Rasch model parameters.*

AT CAT is the measure (on an item of 0 logit measure) corresponding to an expected score equal to the category label, which, for the rating (or partial credit) scale model, is where this category has the highest probability. See plot below.

( ) is printed where the matching calibration is infinite. The value shown corresponds to the measure .25 score points (or LOWADJ= and HIADJ=) away from the extreme.

--ZONE-- is the range of measures from an expected score from 1/2 score-point below to the category to 1/2 score-point above it, the *Rasch-half-point thresholds.* Measures in this range (on an item of 0 measure) are expected to be observed, on average, with the category value. See plot below.

50% CUMULATIVE PROBABILITY gives the location of median probabilities, i.e. these are *Rasch-Thurstone thresholds,* similar to those estimated in the "Graded Response" or "Proportional odds" models. At these calibrations, the probability of observing the categories below equals the probability of observing the categories equal or above. The .5 or 50% cumulative probability is the point on the variable at which the category interval begins. *This is implied by the Rasch model parameters.*

**COHERENCE**

M->C shows what percentage of the measures that were expected to produce observations in this category actually did. Do the measures imply the category?

**Guttman’s Coefficient of Reproducibility** is the count-weighted average of the M->C, i.e., Reproducibility = sum (COUNT * M->C) / sum(COUNT * 100)

C->M shows what percentage of the observations in this category were produced by measures corresponding to the category. Does the category imply the measures?

**ESTIM DISCR** (when DISCRIM=Y) is an estimate of the local discrimination when the model is parameterized in the form: log-odds = aj (Bn - Di - Fj)

**OBSERVED - EXPECTED RESIDUAL DIFFERENCE** (when shown) is the residual difference between the observed and expected counts of observations in the category.

residual difference % = (observed count - expected count) * 100 / (expected count)  
residual difference value = observed count - expected count  
These are shown if at least one residual percent >=1%. This indicates that the Rasch estimates have not converged to their maximum-likelihood values, due to lack of convergence, anchoring, or a large data set. For example,

(a) iteration was stopped early using Ctrl+F or the pull-down menu option.

(b) iteration was stopped when the maximum number of iterations was reached  MJMLE=  

(c) the convergence criteria LCONV= and RCONV= are not small enough for this data set.

(d) anchor values (PAFILE=, IAFILE= and/or SAFILE=) are in force which do not allow maximum likelihood estimates to be obtained.

**ITEM MEASURE ADDED TO MEASURES**, is shown when the rating (or partial credit) scale applies to only one item, e.g., when ISGROUPS=0. Then all measures in these tables are adjusted by the estimated item measure.

<table>
<thead>
<tr>
<th>CATEGORY PROBABILITIES: MODES - Structure measures at intersections</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
</tr>
<tr>
<td>R</td>
</tr>
<tr>
<td>O</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>L</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>T</td>
</tr>
</tbody>
</table>

197
Curves showing how probable is the observation of each category for measures relative to the item measure. Ordinarily, 0 logits on the plot corresponds to the item measure, and is the point at which the highest and lowest categories are equally likely to be observed. The plot should look like a range of hills. Categories which never emerge as peaks correspond to disordered Rasch-Andrich thresholds. These contradict the usual interpretation of categories as a being sequence of most likely outcomes.

Null, Zero, Unobserved Categories

STKEEP=YES and Category 2 has no observations:

```
<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>OBSERVED</th>
<th>OBSVD SAMPLE</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>STRUCTURE</th>
<th>CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABEL</td>
<td>SCORE</td>
<td>COUNT</td>
<td>%</td>
<td>AVRGE EXPECT</td>
<td>MNSQ</td>
<td>MNSQ</td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td>-------</td>
<td>---</td>
<td>-------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>378</td>
<td>20</td>
<td>-.67</td>
<td>-.73</td>
<td>.96</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>620</td>
<td>34</td>
<td>-.11</td>
<td>-.06</td>
<td>.81</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>852</td>
<td>46</td>
<td>1.34</td>
<td>1.33</td>
<td>1.00</td>
</tr>
</tbody>
</table>
```

Category 2 is an incidental (sampling) zero. The category is maintained in the response structure.

STKEEP=NO and Category 2 has no observations:

```
<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>OBSERVED</th>
<th>OBSVD SAMPLE</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>STRUCTURE</th>
<th>CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABEL</td>
<td>SCORE</td>
<td>COUNT</td>
<td>%</td>
<td>AVRGE EXPECT</td>
<td>MNSQ</td>
<td>MNSQ</td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td>-------</td>
<td>---</td>
<td>-------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>378</td>
<td>20</td>
<td>-.87</td>
<td>-1.03</td>
<td>1.08</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>620</td>
<td>34</td>
<td>.13</td>
<td>.33</td>
<td>.85</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>852</td>
<td>46</td>
<td>2.24</td>
<td>2.16</td>
<td>1.00</td>
</tr>
</tbody>
</table>
```

Category 2 is a structural (unobservable) zero. The category is eliminated from the response structure.

218. Table 4.1, 5.1, 8.1, 9.1 Fit plots

(controlled by FRANGE=, LOCAL=, MNSQ=, OUTFIT=)
These tables are plots of the t standardized fit statistics, INFIT or OUTFIT, against the parameter estimates. INFIT is a t standardized information-weighted mean square statistic, which is more sensitive to unexpected behavior affecting responses to items near the person’s measure level. OUTFIT is a t standardized outlier-sensitive mean square fit statistic, more sensitive to unexpected behavior by persons on items far from the person’s measure level. The standardization is approximate. Its success depends on the distribution of persons and items. Consequently, the vertical axis is only a guide and should not be interpreted too rigidly. The NORMAL= variable controls the standardization method used.

Letters on the plot indicate the misfitting person or items. Numbers indicate non-extreme fit or multiple references. The letters appear on Tables 6, 17, 18, 19 for persons, and Tables 10, 13, 14, 15 for items.

Use MNSQ= to change the y-axis to mean-squares.

**219. Table 5.2, 9.2 Fit plots**

If both infit and outfit plots are requested, then a plot of outfit against infit is also produced to assist with the identification of the different patterns of observations they diagnose. For interpretation of fit statistics, see dichotomous and polytomous fit statistics.

**220. Table 6.1, 10.1 Person and item statistics**

(controlled by FITI=, FITP=, USCALE=, UIMEAN=, UPMEAN=, UDECIM=, LOCAL=, ISORT=, OUTFIT=, PSORT=, TOTAL=)
<table>
<thead>
<tr>
<th>Entry Number</th>
<th>Raw Score</th>
<th>Count</th>
<th>Measure</th>
<th>S.E.</th>
<th>INFIT</th>
<th>ZSTD</th>
<th>OUTFIT</th>
<th>OBS%</th>
<th>EXP%</th>
<th>ACT</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>40</td>
<td>74</td>
<td>2.18</td>
<td>.21</td>
<td>2.41</td>
<td>6.3</td>
<td>11.4</td>
<td>9.0</td>
<td>A</td>
<td>40.5</td>
</tr>
<tr>
<td>5</td>
<td>35</td>
<td>74</td>
<td>2.42</td>
<td>.22</td>
<td>2.30</td>
<td>5.6</td>
<td>3.62</td>
<td>7.3</td>
<td>B</td>
<td>52.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>93.0</td>
<td>74.0</td>
<td>0.0</td>
<td>0.23</td>
<td>1.08</td>
<td>0.0</td>
<td>68.8</td>
<td>65.2</td>
<td></td>
</tr>
<tr>
<td>S.D.</td>
<td></td>
<td>30.9</td>
<td>0.0</td>
<td>1.41</td>
<td>0.6</td>
<td>2.3</td>
<td>0.87</td>
<td>13.1</td>
<td>10.2</td>
<td></td>
</tr>
</tbody>
</table>

**MISFIT ORDER** for Tables 6 and 10 is controlled by **OUTFIT=**. The reported **MEAN** and **S.D.** include extreme scores, which are not shown in these Tables.

**ENTRY NUMBER** is the sequence number of the person, or item, in your data, and is the reference number used for deletion or anchoring.

"PERSONS" or "ITEMS", etc. is the item name or person-identifying label.

**RAW SCORE** is the raw score corresponding to the parameter, i.e., the raw score by a person on the test, or the sum of the scored responses to an item by the persons.

**COUNT** is the number of data points used to construct measures.

**MEASURE** is the estimate (or calibration) for the parameter, i.e., person ability (theta, B, beta, etc.), or the item difficulty (b, D, delta, etc.). Values are reported in logits with two decimal places, unless rescaled by **USCALE=**, **UIMEAN=**, **UPMEAN=**, **UDECIM=**. If the score is extreme, a value is estimated, but as MAXIMUM (perfect score) or MINIMUM (zero score). No measure is reported if the element is DROPPED (no valid observations remaining) or DELETED (you deleted the person or item). The difficulty of an item is defined to be the point on the latent variable at which its high and low categories are equally probable. **SAFILE=** can be used to alter this definition.

If unexpected results are reported, check whether **TARGET=** or **CUTLO=** or **CUTHI=** are specified. **INESTIMABLE** is reported if all observations are eliminated as forming part of extreme response strings.

**ERROR** is the standard error of the estimate. For anchored values, an "A" is shown on the listing and the error reported is that which would have been obtained if the value had been estimated. Values are reported in logits with two decimal places, unless rescaled by **USCALE=**, **UDECIM=**

**INFIT** is a t standardized information-weighted mean square statistic, which is more sensitive to unexpected behavior affecting responses to items near the person's measure level.

**MNSQ** is the mean-square infit statistic with expectation 1. Values substantially less than 1 indicate dependency in your data; values substantially greater than 1 indicate noise. See dichotomous and polytomous fit statistics.

<table>
<thead>
<tr>
<th>Value</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;2.0</td>
<td>Off-variable noise is greater than useful information. Degrades measurement.</td>
</tr>
<tr>
<td>&gt;1.5</td>
<td>Noticeable off-variable noise. Neither constructs nor degrades measurement</td>
</tr>
<tr>
<td>0.5 - 1.5</td>
<td>Productive of measurement</td>
</tr>
<tr>
<td>&lt;0.5</td>
<td>Overly predictable. Misleads us into thinking we are measuring better than we really are. (Attenuation paradox.)</td>
</tr>
</tbody>
</table>

Always remedy the large misfits first. Misfits <1.0 are only of concern when shortening a test.

**ZSTD** is the infit mean-square fit statistic t standardized to approximate a theoretical "unit normal", mean 0 and variance 1, distribution. ZSTD (standardized as a z-score) is used of a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's t-distribution value has been adjusted to a unit normal value. The standardization is shown on RSA, p.100-101. When LOCAL=Y, then EMP is shown, indicating a local {0,1} standardization. When LOCAL=LOG, then LOG
is shown, and the natural logarithms of the mean-squares are reported. More exact values are shown in the Output Files.

Ben Wright advises: "ZSTD is only useful to salvage non-significant MNSQ>1.5, when sample size is small or test length is short."

OUTFIT is a t standardized outlier-sensitive mean square fit statistic, more sensitive to unexpected behavior by persons on items far from the person's measure level.

<table>
<thead>
<tr>
<th>Value</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;2.0</td>
<td>Off-variable noise is greater than useful information. Degrades measurement.</td>
</tr>
<tr>
<td>&gt;1.5</td>
<td>Noticeable off-variable noise. Neither constructs nor degrades measurement</td>
</tr>
<tr>
<td>0.5 - 1.5</td>
<td>Productive of measurement</td>
</tr>
<tr>
<td>&lt;0.5</td>
<td>Overly predictable. Misleads us into thinking we are measuring better than we really are. (Attenuation paradox.)</td>
</tr>
</tbody>
</table>

Always remedy the large misfits first. Misfits <1.0 are usually only of concern when shortening a test.

MNSQ is the mean-square outfit statistic with expectation 1. Values substantially less than 1 indicate dependency in your data; values substantially greater than 1 indicate the presence of unexpected outliers. See dichotomous and polytomous fit statistics.

ZSTD is the outfit mean-square fit statistic t standardized similarly to the INFIT ZSTD. ZSTD (standardized as a z-score) is used of a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's t-distribution value has been adjusted to a unit normal value.

Ben Wright advises: "ZSTD is only useful to salvage non-significant MNSQ>1.5, when sample size is small or test length is short."

PTBIS CORR (reported when PTBIS=Yes) is the point-biserial correlation, rpbis, between the individual item (or person) response "scores" and the total person (or item) test score (less the individual response "scores"). Negative values for items often indicate mis-scoring, or rating (or partial credit) scale items with reversed direction. Letters indicating the identity of persons or items appearing on the fit plots appear under PTBIS. For adaptive tests, an rpbis near zero is expected.

The formula for this product-moment correlation coefficient is:

\[ r_{pbis} = \frac{\sum (x-xbar)(y-ybar)}{\sqrt{\sum (x-xbar)^2 \sum (y-ybar)^2}} \]

where x = observation for this item (or person), y = total score (including extreme scores) for person omitting this item (or for item omitting this person).

PTMEA CORR. (reported when PTBIS=N) is the point-measure correlation, rpm or RPM, between the observations on an item (as fractions with the range 0,1) and the corresponding person measures, or vice-versa. Since the point-biserial loses its meaning in the presence of missing data, specify PTBIS=N when there are missing data or when CUTLO= or CUTHI= are specified. The point-measure correlation has a range of -1 to +1.

EXACT MATCH
OBS% Observed% is the percent of data points which are within 0.5 score points of their expected values, i.e., that match predictions.
EXP% Expected% is the percent of data points that are predicted to be within 0.5 score points of their expected values.

ESTIM DISCRIM is an estimate of the item discrimination, see DISCRIM=

ASYMPTOTE LOWER and UPPER are estimates of the upper and lower asymptotes for dichotomous items, see ASYMPTOTE=
WEIGH is the weight assigned by IWEIGHT= or PWEIGHT=

DISPLACE is the displacement of the reported MEASURE from its data-derived value. This should only be shown with anchored measures.

G is the grouping code assigned with ISGROUPS=

M is the model code assigned with MODELS=

"BETTER FITTING OMITTED" appears in fit-ordered Tables, where items better than FITI=, or persons better than FITP=, are excluded.

Above the Table are shown the "real" separation and reliability coefficients from Table 3.

Selection of persons and items for misfit tables

Report measure in Tables 6 and 10, if any of:

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Less than</th>
<th>Greater than</th>
</tr>
</thead>
<tbody>
<tr>
<td>t standardized INFIT</td>
<td>- (FITP or FITI)</td>
<td>FITP or FITI</td>
</tr>
<tr>
<td>t standardized OUTFIT</td>
<td>- (FITP or FITI)</td>
<td>FITP or FITI</td>
</tr>
<tr>
<td>mean-square INFIT</td>
<td>1 - (FITP or FITI)/10</td>
<td>1 + (FITP or FITI)/10</td>
</tr>
<tr>
<td>mean-square OUTFIT</td>
<td>1 - (FITP or FITI)/10</td>
<td>1 + (FITP or FITI)/10</td>
</tr>
<tr>
<td>point-biserial correlation</td>
<td>negative</td>
<td></td>
</tr>
</tbody>
</table>

To include every person, specify FITP=0. For every item, FITI=0.

Usually OUTFIT=N, and Tables 6 and 10 are sorted by by maximum (Infit mean-square, Outfit mean-square)
When OUTFIT=Y, Tables 6 and 10 are sorted by Infit Mean-square.


Controlled by USCALE=, UMEAN=, UDECIM=, LOCAL=, FITLOW=, FITHIGH=, MNSQ=.

Specify CHART=YES to produce Tables like this.

With MNSQ=Yes:

```
<table>
<thead>
<tr>
<th>ENTRY</th>
<th>MEASURE</th>
<th>INFIT MEAN-SQUARE</th>
<th>OUTFIT MEAN-SQUARE</th>
<th>PUPIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
<td>*</td>
<td>: . : *</td>
<td>A : . : *</td>
<td>JACKSON, SOLOMON</td>
</tr>
<tr>
<td>53</td>
<td>*</td>
<td>: * . :</td>
<td>K : . : *</td>
<td>SABOL, ANDREW</td>
</tr>
<tr>
<td>32</td>
<td>*</td>
<td>* . : *</td>
<td>W * . :</td>
<td>ROSSNER, JACK</td>
</tr>
<tr>
<td>21</td>
<td>*</td>
<td>* : a * :</td>
<td>EISEN, NORM L.</td>
<td></td>
</tr>
</tbody>
</table>
```

The fit information is shown in graphical format to aid the eye in identifying patterns and outliers. The fit bars are positioned by FITLOW= and FITHIGH=.

With MNSQ=No:

```
<table>
<thead>
<tr>
<th>ENTRY</th>
<th>MEASURE</th>
<th>INFIT t standardized</th>
<th>OUTFIT t standardized</th>
<th>TAPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>E</td>
<td>: . : .</td>
<td>: . : .</td>
<td>4-1-3-4-2-1-4</td>
</tr>
<tr>
<td>15</td>
<td>*</td>
<td>: . : *</td>
<td>: * :</td>
<td>1-3-2-4-1-3</td>
</tr>
<tr>
<td>14</td>
<td>*</td>
<td>: . : *</td>
<td>: * :</td>
<td>1-4-2-3-4-1</td>
</tr>
<tr>
<td>12</td>
<td>*</td>
<td>: . : *</td>
<td>: * :</td>
<td>1-3-2-4-3</td>
</tr>
</tbody>
</table>
```

202
222. **Table 6.5, 10.5 Most unexpected observations**

These tables display the unexpected responses in the Guttman scalogram format. The Guttman Scalogram of unexpected responses shows the persons and items with the most unexpected data points (those with the largest standardized residuals) arranged by measure, such that the high value observations are expected in the top-left of the data matrix, near the "high", and the low values are expected in the bottom of the matrix, near the "low". The category values of unexpected observations are shown. Expected values (with standardized residuals less than |2|) are shown by ".". Missing values, if any, are left blank.

**UCOUNT=** sets the maximum number of persons and items to report in the anti-Guttman matrices in Tables 6.4, 6.5, 10.4, 10.5.

<table>
<thead>
<tr>
<th>PUPIL</th>
<th>MEASURE</th>
<th>ACT</th>
</tr>
</thead>
<tbody>
<tr>
<td>FATALE, NATASHA</td>
<td>4.77</td>
<td>..................1.</td>
</tr>
<tr>
<td>SCHATTNER, GAIL</td>
<td>3.55</td>
<td>...................0</td>
</tr>
<tr>
<td>STOLLER, DAVE</td>
<td>.96</td>
<td>0.10...0......222</td>
</tr>
<tr>
<td>SABOL, ANDREW</td>
<td>-1.59</td>
<td>..............1....1</td>
</tr>
</tbody>
</table>

223. **Table 6.4, 10.4 Most misfitting response strings**

These tables display the unexpected responses in the most misfitting response strings in a Guttman scalogram format. **UCOUNT=** sets the maximum number of persons and items to report in the anti-Guttman matrices in Tables 6.4, 6.5, 10.4, 10.5.

**ACT** | **OUTMNSQ** | **PUPIL**
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>WATCH A RAT</td>
<td>5.68 A</td>
<td>1110.0...2...2...2...1112.221112.2</td>
</tr>
<tr>
<td>WATCH BUGS</td>
<td>1.99 B</td>
<td>......0......2...2...2......11.21.</td>
</tr>
<tr>
<td>LEARN WEED NAMES</td>
<td>1.29 D</td>
<td>....0...0........................................1</td>
</tr>
</tbody>
</table>

The items (or persons) are ordered by descending mean-square misfit. Each column corresponds to a person. The entry numbers are printed vertically. The responses are ordered so that the highest expected responses are to the left (high), the lowest to the right (low). The category values of unexpected observations are shown. Expected values (with standardized residuals less than |2|) are shown by ".". Missing values, if any, are left blank.

224. **Table 6.6, 10.6 Most unexpected response list**

This shows the most unexpected responses sorted by unexpectedness (standardized residual). Large standardized residuals contribute to large outfit mean-square fit statistics. **UCOUNT=** sets the maximum number of "most unexpected responses" to report in Tables 6.6, 10.6.

**TABLE 6.6 LIKING FOR SCIENCE** (Wright & Masters

**INPUT:** 75 KIDS, 25 ACTS **MEASURED:** 75 KIDS, 25 ACTS, 3 CATS **WINSTEPS** 3.58.0

203
DATA is the response code in the data file
OBSERVED is the code's value after rescoring
EXPECTED is the predicted observation based on the person and item estimated measures
RESIDUAL is (OBSERVED - EXPECTED), the difference between the observed and expected values
ST. RES. is the standardized residual, the unexpectedness of the residual expressed as a unit normal deviate
MEASDIFF is the difference between the ability and difficulty estimates. This produces the EXPECTED value.
ACT is the item entry number
KID is the person entry number
ACT is the item label
KID is the person label

225. **Table 7.1, 11.1 Misfitting responses**

(controlled by FITI=, FITP=, MNSQ=, OUTFIT=)

These tables show the persons or items for which the t standardized outfit (or infit, if OUTFIT=N) statistic is greater than the misfit criterion (FITP= or FITI=). Persons or items are listed in descending order of misfit. The response codes are listed in their sequence order in your data file. The residuals are standardized response score residuals, which have a modelled expectation of 0, and a variance of 1. Negative residuals indicate that the observed response was less correct (or, for rating (or partial credit) scales, lower down the rating scale) than expected. The printed standardized residual is truncated, not rounded, so that its actual value is at least as extreme as that shown. Standardized residuals between -1 and 1 are not printed. For exact details, see XFILE=.

"X" indicates that the item (or person) obtained an extreme score. "M" indicates a missing response.

For Table 7, the diagnosis of misfitting persons, persons with a t standardized fit greater than FITP= are reported. Selection is based on the OUTFIT statistic, unless you set OUTFIT=N in which case the INFIT statistic is used.

For Table 11, the diagnosis of misfitting items, items with a t standardized fit greater than FITI= are reported. Selection is based on the OUTFIT statistic, unless you set OUTFIT=N in which case the INFIT statistic is used.

**TABLE OF POORLY FITTING ITEMS**

<table>
<thead>
<tr>
<th>NUMBER</th>
<th>NAME</th>
<th>POSITION</th>
<th>MEASURE</th>
<th>INFIT (ZSTD)</th>
<th>OUTFIT</th>
<th>MISFIT OVER 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>WATCH A RAT</td>
<td></td>
<td>2.00</td>
<td>5.8</td>
<td>A</td>
<td>8.1</td>
</tr>
<tr>
<td></td>
<td>RESPONSE:</td>
<td>1:</td>
<td>0 2 1 1</td>
<td>1 2 0 2 0 0 1 0 1 1 0 0 0 0 1 1 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Z-RESIDUAL:</td>
<td>X</td>
<td>2 3 3</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RESPONSE:</td>
<td>26:</td>
<td>1 2 0 2</td>
<td>1 0 0 1 1 0 1 0 0 0 0 0 0 2 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Z-RESIDUAL:</td>
<td>3 6 2</td>
<td>-2</td>
<td>-2</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

/ This letter on fit plots

5 FIND BOTTLES AND CANS | 2.21 | 5.2 | B | 6.5 |
| RESPONSE: 1: 1 2 0 0 1 2 0 0 1 0 0 0 0 0 1 1 0 1 |
| Z-RESIDUAL: X 2 4 6 -2 |

226. **Table 7.2, 17.3, 18.3, 19.3 Diagnostic KeyForms for Persons**

This table displays a version of Table 2.2 with the responses filled in for each person. These can be useful for diagnosis and individual reporting. The responses are the original values in the data file, as far as possible.

**TABLE 7.2 LIKING FOR SCIENCE (Wright & Masters p.18) sf.out Sep 26 8:59 2000**

**INPUT: 76 PUPILS, 25 ACTS ANALYZED: 74 PUPILS, 25 ACTS, 36 CATS WINSTEPS v3.08**

**KEY: .1.=OBSERVED, 1=EXPECTED, (1)=OBSERVED, BUT VERY UNEXPECTED.**
<table>
<thead>
<tr>
<th>ENTRY NUMBER</th>
<th>DATA CODE</th>
<th>VALUE</th>
<th>COUNT</th>
<th>%</th>
<th>MEASURE</th>
<th>MEAN</th>
<th>MNSQ CORR.</th>
<th>ACT</th>
<th>ACT</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>-0.87</td>
<td>0.44</td>
<td>.5</td>
<td>-0.28</td>
<td>WATCH BIRDS</td>
<td>0 Dislike</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>35</td>
<td>47</td>
<td>0.23</td>
<td>0.11</td>
<td>.4</td>
<td>-0.51</td>
<td>''</td>
<td>1 Neutral</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>36</td>
<td>49</td>
<td>1.83</td>
<td>0.22</td>
<td>.7</td>
<td>.62</td>
<td>''</td>
<td>2 Like</td>
</tr>
</tbody>
</table>

MISSING *** | 1 1* | 1.04 | .01 | |

The vertical line of numbers corresponds to the person measure, and indicates the expected (average) responses. Responses marked .0. or .1. or .2. are observed. Responses (1) and (2) are observed and statistically significantly unexpected, |t| ≥ 2, p<5%. Responses shown merely as 0 or 1 are expected, but not observed.

**Table 10.3, 13.3, 14.3, 15.3, 25.3** Item option/distracter frequencies

Item OPTION FREQUENCIES are output if DistractorS=Y. These show occurrences of each of the valid data codes in CODES=, and also of MISSCORE= in the input data file. Counts of responses forming part of extreme scores are included. Only items included in the corresponding main table are listed.

OSORT= controls the ordering of options within items. The standard is the order of data codes in CODES=.
SCORE VALUE is the value assigned to the data code by means of NEWSCORE=, KEY1=, IVALUEA=, etc. '*' means the data code is ignored, i.e., regarded as not administered. MISSCORE=1 scores missing data as "1".

DATA COUNT is the frequency of the data code in the data file - this includes observations for both non-extreme and extreme persons and items. For counts only for non-extreme persons and items, see the DISFILE= DATA % is the percent of scored data codes. For dichotomies, the % are the p-values for the options.

For data with score value "****", the percent is of all data codes, indicated by "***".

AVERAGE MEASURE is the observed, sample-dependent, average measure of persons in this analysis who responded in this category (adjusted by PWEIGHT=). This is a quality-control statistic for this analysis. (It is not the sample-independent value of the category, which is obtained by adding the item measure to the "score at category", in Table 3.2 or higher, for the rating (or partial credit) scale corresponding to this item.) For each observation in category k, there is a person of measure Bn and an item of measure Di. Then: average measure = \sum(Bn - Di) / count of observations in category.

An "****" indicates that the average measure for a higher score value is lower than for a lower score value. This contradicts the hypothesis that "higher score value implies higher measure, and vice-versa".

S.E. MEAN is the standard error of the mean (average) measure of the sample of persons who responded in this category (adjusted by PWEIGHT=).

OUTFIT MEAN-SQUARE is the ratio of observed variance to expected variance for observed responses in this category. Values greater than 1.0 indicate unmodeled noise. Values less than 1.0 indicate loss of information.

PTMEA CORR is the correlation between the occurrence, scored 1, or non-occurrence, scored 0, of this category or distractor and the person measures. With PTBIS=YES, the correlation is between the occurrence and the person raw score, indicated by PTBIS CORR. When this correlation is high positive for a correct MCQ option, then the item exhibits convergent validity. When this correlation is low or negative for incorrect MCQ options, then the item exhibits discriminant validity. Krus, D. J. & Ney, R. G. (1978) Convergent and discriminant validity in item analysis. Educational and Psychological Measurement, 38, 135-137.

ITEM (here, ACT) is the name or label of the item.

Data codes and Category labels are shown to the right of the box, if CLFILE= or CFILE= is specified.

228. Table 12.2, 12.12 Item distribution maps

(controlled by MRANGE=, MAXPAG=, NAMLMP=, ISORT=)

In Table 12.2, the full item names are shown located at their calibrations, along with the person distribution. You can use NAMLMP= to control the number of characters of each name reported. "QSMSQ" summarize the distributions. An "M" marker represents the location of the mean measure. "S" markers are placed one sample standard deviation away from the mean. "T" markers are placed two sample standard deviations away.

MAXPAG= controls the length of the Table. MRANGE= controls the displayed range of measures. ISORT= controls the sort order within rows. You can adjust these values from the "Specification" pull-down menu and then select Table 12 from the Output Tables menu. Thus you can experiment without needing to rerun the analysis.

Where there are more items than can be shown on one line, the extra items are printed on subsequent lines, but the latent variable "|") does not advance and is left blank.

In subtable 12.12, the items are arranged by easiness. The item hierarchy is reversed.

Items arranged by measure: Look for the hierarchy of item names to spell out a meaningful construct from easiest (highest p-value or highest average rating) at the bottom to hardest (lowest p-value or lowest average rating) at the top.
Subtable 12.12:
Items arranged by easiness: Look for the hierarchy of item names to spell out a meaningful construct from easiest at the top to hardest at the bottom.
The double line || indicates the two sides have opposite orientations.

KIDS MAP OF ACTS

Subtable 12.5 Item map with expected score zones

This Table shows the items positioned at the lower edge of each expected score zone. The expected score zone above label.2 extends from 1.5 to 2. Above label.1, the zone extends from 0.5 to 1.5. Lower than label.1 is the
zone from 0 to 0.5. If you put the item number at the start of the item labels after &END, you can show only the item numbers on this plot by using NAMLMP= or IMAP=. Column headings are the category labels that match the (rescored) category numbers in CFILE=.

Where there are more items than can be shown on one line, the extra items are printed on subsequent lines, but the latent variable "|" does not advance and is left blank.

<table>
<thead>
<tr>
<th>KIDS MAP OF ACTS – Expected score zones</th>
<th>Like</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;more&gt;</td>
<td>Neutral</td>
</tr>
<tr>
<td>5</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>XXX</td>
</tr>
<tr>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>XXX</td>
</tr>
<tr>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>XXX</td>
</tr>
<tr>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 12.6 Item map with 50% cumulative probabilities

This Table shows the items positioned at median 50% cumulative probability (the Rasch-Thurstone thresholds) at the lower edge of each rating probability zone. Above label 2, the most probable category is 2. Below label 1, the most probable category is 0. Between label 1 and label 2 is the zone which can be thought of as corresponding to a rating of 1. If you put the item number at the start of the item labels after &END, you can show only the item...
numbers on this plot by using NAMLMP= or IMAP=. Columns are headed by the (rescored) categories in CFILE=.

Where there are more items than can be shown on one line, the extra items are printed on subsequent lines, but the latent variable "|" does not advance and is left blank.

231. Table 13.1, 14.1, 15.1, 25.1, 26.1 Item statistics

(controlled by USCALE=, UMEAN=, UDECIM=, LOCAL=, ISORT=, TOTAL=, DISCRIMINATION=, ASYMPTOTE=, PVALUE=)
Above the Table are shown the "real" separation and reliability coefficients from Table 3.

**ENTRY NUMBER** is the sequence number of the item, in your data, and is the reference number used for deletion or anchoring.

**RAW SCORE** is the raw score corresponding to the parameter, i.e., the sum of the scored responses to an item by the persons. If TOTALSCORE=Yes, then RAW SCORE includes extreme (zero, perfect) scores. If TOTALSCORE=No, then RAW SCORE excludes extreme (zero, perfect) scores.

**COUNT** is the number of data points summed to make the RAW SCORE.

**MEASURE** is the estimate (or calibration) for the parameter, i.e., the item difficulty (b, D, delta, etc.). If the score is extreme, a value is estimated, but as MAXIMUM (zero score) or MINIMUM (perfect score). No measure is reported if the element is DROPPED (no valid observations remaining) or DELETED (you deleted the item).

The difficulty of an item is defined to be the point on the latent variable at which its high and low categories are equally probable. SAFILE= can be used to alter this definition.

If unexpected results are reported, check whether TARGET= or CUTLO= or CUTHI= are specified. INESTIMABLE is reported if all observations are eliminated as forming part of extreme response strings. To make such measures estimable, further data (real or artificial) is required including both extreme and non-extreme observations.

**MINIMUM ESTIMATE MEASURE** - the sample obtained the extreme maximum (perfect) score on this item, so it has been estimated with an extreme maximum measure, see Extreme scores. Fit statistics are not computable, but they would correspond to perfect fit to the Rasch model.

**MAXIMUM ESTIMATE MEASURE** - the sample obtained the extreme minimum measure, see Extreme scores. Fit statistics are not computable, but they would correspond to perfect fit to the Rasch model.

**ERROR** is the standard error of the estimate. For anchored values, an "A" is shown on the listing and the error reported is that which would have been obtained if the value had been estimated.

**INFIT** is a t standardization mean square statistic, which is more sensitive to unexpected behavior affecting responses to items near the person's measure level.

**MNSQ** is the mean-square infit statistic with expectation 1. Values substantially less than 1 indicate dependency in your data; values substantially greater than 1 indicate noise. See dichotomous and polytomous fit statistics.

<table>
<thead>
<tr>
<th>Value</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;2.0</td>
<td>Off-variable noise is greater than useful information. Degrades measurement.</td>
</tr>
<tr>
<td>&gt;1.5</td>
<td>Noticeable off-variable noise. Neither constructs nor degrades measurement</td>
</tr>
<tr>
<td>0.5 - 1.5</td>
<td>Productive of measurement</td>
</tr>
<tr>
<td>&lt;0.5</td>
<td>Overly predictable. Misleads us into thinking we are measuring better than we really are. (Attenuation paradox.)</td>
</tr>
</tbody>
</table>

Always remedy the large misfits first. Misfits <1.0 are only of concern when shortening a test.

**ZSTD** is the infit mean-square fit statistic t standardized to approximate a theoretical "unit normal", mean 0 and
variance 1, distribution. ZSTD (standardized as a z-score) is used if a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's t-distribution value has been adjusted to a unit normal value. The standardization is shown on RSA, p.100-101. When LOCAL=Y, then EMP is shown, indicating a local [0,1] standardization. When LOCAL=LOG, then LOG is shown, and the natural logarithms of the mean-squares are reported. More exact values are shown in the Output Files.

Ben Wright advises: “ZSTD is only useful to salvage non-significant MNSQ>1.5, when sample size is small or test length is short.”

OUTFIT is a t standardized outlier-sensitive mean square fit statistic, more sensitive to unexpected behavior by persons on items far from the person's measure level.

<table>
<thead>
<tr>
<th>Value</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;2.0</td>
<td>Off-variable noise is greater than useful information. Degrades measurement.</td>
</tr>
<tr>
<td>&gt;1.5</td>
<td>Noticeable off-variable noise. Neither constructs nor degrades measurement</td>
</tr>
<tr>
<td>0.5 - 1.5</td>
<td>Productive of measurement</td>
</tr>
<tr>
<td>&lt;0.5</td>
<td>Overly predictable. Misleads us into thinking we are measuring better than we really are. (Attenuation paradox.)</td>
</tr>
</tbody>
</table>

Always remedy the large misfits first. Misfits <1.0 are usually only of concern when shortening a test.

MNSQ is the mean-square outfit statistic with expectation 1. Values substantially less than 1 indicate dependency in your data; values substantially greater than 1 indicate the presence of unexpected outliers. See dichotomous and polytomous fit statistics.

ZSTD is the outfit mean-square fit statistic t standardized similarly to the INFIT ZSTD. ZSTD (standardized as a z-score) is used if a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's t-distribution value has been adjusted to a unit normal value.

Ben Wright advises: “ZSTD is only useful to salvage non-significant MNSQ>1.5, when sample size is small or test length is short.”

PTMEA CORR. (reported when PTBIS=N) is the point-measure correlation, rpm or RPM, between the observations on an item (as fractions with the range 0,1) and the corresponding person measures, or vice-versa. Since the point-biserial loses its meaning in the presence of missing data, specify PTBIS=N when data are missing or CUTLO= or CUTHI= are specified.

PTBIS CORR (reported when PTBIS=Yes) is the point-biserial correlation, rpbis, between the individual item response "scores" and the total person test score (less the individual item response "scores"). Negative values for items often indicate mis-scoring, or rating (or partial credit) scale items with reversed direction. Letters indicating the identity of items appearing on the fit plots appear under PTBIS. For adaptive tests, an rpbis near zero is expected.

The formula for this product-moment correlation coefficient is:

\[ r_{pbis} = \frac{\text{sum} \{ (x-x \text{bar})(y-y \text{bar}) \} }{ \sqrt{ \text{sum} \{ (x-x \text{bar})^2 \} } \sqrt{ \text{sum} \{ (y-y \text{bar})^2 \} } } \]

where \( x \) = observation for this item (or person), \( y \) = total score (including extreme scores) for person omitting this item

EXACT MATCH
OBS% Observed% is the percent of data points which are within 0.5 score points of their expected values, i.e., that match predictions.
EXP% Expected% is the percent of data points that are predicted to be within 0.5 score points of their expected values.

ESTIM DISCRIM is an estimate of the item discrimination, see DISCRIM=

ASYMPTOTE LOWER and UPPER are estimates of the upper and lower asymptotes for dichotomous items, see ASYMPTOTE=
P-VALUE is the observed proportion correct or average rating, see PVALUE=

WEIGH is the weight assigned by IWEIGHT= or PWEIGHT=

DISPLACE is the displacement of the reported MEASURE from its data-derived value. This should only be shown with anchored measures.

G is the grouping code assigned with ISGROUPS=

M is the model code assigned with MODELS=

232. Table 16.3 Person distribution map

(controlled by MRANGE=, MAXPAG=, NAMLMP=, PSORT=)

In Table 16.3, the full person names are shown with an item distribution. You can use NAMLMP= to control the number of characters of each name reported. "QSMSQ" summarize the distributions. An "M" marker represents the location of the mean measure. "S" markers are placed one sample standard deviation away from the mean. "T" markers are placed two sample standard deviations away. MAXPAG= controls the length of the Table. MRANGE= controls the displayed range of measures. PSORT= controls the sort order within rows.

Where there are more persons than can be shown on one line, the persons are printed on subsequent lines, but the latent variable "|" does not advance and is left blank.

Persons arranged by measure: Look for the hierarchy of person names to spell out a meaningful distribution from highest scoring at the top to lowest scoring at the bottom.
### Table 17.1, 18.1, 19.1 Person statistics

(controlled by USCALE=, UMEAN=, UDECIM=, LOCAL=, PSORT=, TOTAL=)

<table>
<thead>
<tr>
<th>ENTRY NUMBER</th>
<th>TOTAL SCORE</th>
<th>COUNT</th>
<th>MEASURE</th>
<th>S.E.</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>PTMEA</th>
<th>KID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>18</td>
<td>-2.94</td>
<td>.82</td>
<td>.61</td>
<td>-1.2</td>
<td>.29</td>
<td>2.0</td>
</tr>
<tr>
<td>...</td>
<td>34</td>
<td>12</td>
<td>1.94</td>
<td>.98</td>
<td>1.74</td>
<td>1.4</td>
<td>.71</td>
<td>.9</td>
</tr>
<tr>
<td>35</td>
<td>3</td>
<td>18</td>
<td>-6.62</td>
<td>1.85</td>
<td>MINIMUM ESTIMATED MEASURE</td>
<td>Helen F</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.7</td>
<td>14.0</td>
<td>-.37</td>
<td>1.03</td>
<td>MINIMUM ESTIMATED MEASURE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.4</td>
<td>0</td>
<td>2.22</td>
<td>.17</td>
<td>.94</td>
<td>1.2</td>
<td>1.29</td>
<td>1.1</td>
</tr>
</tbody>
</table>

Above the Table are shown the "real" separation and reliability coefficients from Table 3.

- **ENTRY NUMBER** is the sequence number of the person in your data, and is the reference number used for deletion or anchoring.
- **RAW SCORE** is the raw score corresponding to the parameter, i.e., the raw score by a person on the test. If TOTALSCORE=Yes, then RAW SCORE includes extreme (zero, perfect) scores. If TOTALSCORE=No, then RAW SCORE excludes extreme (zero, perfect) scores.
- **COUNT** is the number of data points summed to make the RAW SCORE.
- **MEASURE** is the estimate for the parameter, i.e., person ability (theta, B, beta, etc.). If the score is extreme, a value is estimated, but as MAXIMUM (perfect score) or MINIMUM (zero score). No measure is reported if the element is DROPPED (no valid observations remaining) or DELETED (you deleted the person or item). The difficulty of an item is defined to be the point on the latent variable at which its high and low categories are equally probable. SAFILE= can be used to alter this definition.
- If unexpected results are reported, check whether TARGET= or CUTLO= or CUTHI= are specified.
- INESTIMABLE is reported if all observations are eliminated as forming part of extreme response strings. To make such measures estimable, further data (real or artificial) is required including both extreme and non-extreme observations.
- **ERROR** is the standard error of the estimate. For anchored values, an "A" is shown on the listing and the error reported is that which would have been obtained if the value had been estimated.
- **INFIT** is a t standardized information-weighted mean square statistic, which is more sensitive to unexpected behavior affecting responses to items near the person's measure level.
MNSQ is the mean-square infit statistic with expectation 1. Values substantially less than 1 indicate dependency in your data; values substantially greater than 1 indicate noise. See dichotomous and polytomous fit statistics.

<table>
<thead>
<tr>
<th>Value</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;2.0</td>
<td>Off-variable noise is greater than useful information. Degrades measurement.</td>
</tr>
<tr>
<td>&gt;1.5</td>
<td>Noticeable off-variable noise. Neither constructs nor degrades measurement</td>
</tr>
<tr>
<td>0.5 - 1.5</td>
<td>Productive of measurement</td>
</tr>
<tr>
<td>&lt;0.5</td>
<td>Overly predictable. Misleads us into thinking we are measuring better than we really are. (Attenuation paradox.)</td>
</tr>
</tbody>
</table>

Always remedy the large misfits first. Misfits <1.0 are only of concern when shortening a test.

ZSTD is the infit mean-square fit statistic that standardized to approximate a theoretical "unit normal", mean 0 and variance 1, distribution. ZSTD (standardized as a z-score) is used of a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's t-distribution value has been adjusted to a unit normal value. The standardization is shown on RSA, p.100-101. When LOCAL=Y, then EMP is shown, indicating a local (0,1) standardization. When LOCAL=LOG, then LOG is shown, and the natural logarithms of the mean-squares are reported. More exact values are shown in the Output Files.

Ben Wright advises: "ZSTD is only useful to salvage non-significant MNSQ>1.5, when sample size is small or test length is short."

OUTFIT is a t standardized outlier-sensitive mean square fit statistic, more sensitive to unexpected behavior by persons on items far from the person's measure level.

<table>
<thead>
<tr>
<th>Value</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;2.0</td>
<td>Off-variable noise is greater than useful information. Degrades measurement.</td>
</tr>
<tr>
<td>&gt;1.5</td>
<td>Noticeable off-variable noise. Neither constructs nor degrades measurement</td>
</tr>
<tr>
<td>0.5 - 1.5</td>
<td>Productive of measurement</td>
</tr>
<tr>
<td>&lt;0.5</td>
<td>Overly predictable. Misleads us into thinking we are measuring better than we really are. (Attenuation paradox.)</td>
</tr>
</tbody>
</table>

Always remedy the large misfits first. Misfits <1.0 are usually only of concern when shortening a test.

MNSQ is the mean-square outfit statistic with expectation 1. Values substantially less than 1 indicate dependency in your data; values substantially greater than 1 indicate the presence of unexpected outliers. See dichotomous and polytomous fit statistics.

ZSTD is the outfit mean-square fit statistic standardized similarly to the INFIT ZSTD. ZSTD (standardized as a z-score) is used of a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's t-distribution value has been adjusted to a unit normal value.

Ben Wright advises: "ZSTD is only useful to salvage non-significant MNSQ>1.5, when sample size is small or test length is short."

PTMEA CORR. (reported when PTBIS=N) is the point-measure correlation, rpm or RPM, between the observations on an item (as fractions with the range 0,1) and the corresponding item measures. Since the point-biserial loses its meaning in the presence of missing data, specify PTBIS=N when data are missing or CUTLO= or CUTHI= are specified.

PTBIS CORR (reported when PTBIS=Yes) is the point-biserial correlation, rpbis, between the individual person "scores" and the total item test score (less the individual person response "scores"). Letters indicating the identity of persons appearing on the fit plots appear under PTBIS. For adaptive tests, an rpbis near zero is expected.

The formula for this product-moment correlation coefficient is:

\[ r_{pbis} = \frac{\text{sum} \{(x-x\ bar)(y-y\ bar)\}}{\sqrt{\text{sum} \{(x-x\ bar)^2\} \text{sum} \{(y-y\ bar)^2\}}} \]

where \( x \) = observation for this person, \( y \) = total score (including extreme scores) for item omitting this
WEIGH is the weight assigned by IWEIGHT= or PWEIGHT=.

DISPLACE is the displacement of the reported MEASURE from its data-derived value. This should only be shown with anchored measures.

234. Table 20 Complete score-to-measure table on test of all items

A measure and standard error is estimated for every possible score on a test composed of all non-extreme items included in the analysis. This can also be written with SCOREFILE=. The measures corresponding to extreme scores (all items right, or all items wrong) are marked by "E" and estimated using the EXTRSC= criterion. A graph of the score to measure conversion is also reported. "*" indicates the conversion. Since the 'S' and 'F' models specify that not all item levels are encountered, measures complete tests are only approximated here. In the Table of Measures on Complete Test:

<table>
<thead>
<tr>
<th>SCORE</th>
<th>MEASURE</th>
<th>S.E.</th>
<th>SCORE</th>
<th>MEASURE</th>
<th>S.E.</th>
<th>SCORE</th>
<th>MEASURE</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-6.17E</td>
<td>1.83</td>
<td>5</td>
<td>-1.97</td>
<td>.83</td>
<td>10</td>
<td>2.55</td>
<td>.89</td>
</tr>
<tr>
<td>1</td>
<td>-4.86</td>
<td>1.08</td>
<td>6</td>
<td>-1.19</td>
<td>.92</td>
<td>11</td>
<td>3.37</td>
<td>.89</td>
</tr>
<tr>
<td>2</td>
<td>-3.94</td>
<td>.85</td>
<td>7</td>
<td>-.23</td>
<td>1.00</td>
<td>12</td>
<td>4.21</td>
<td>.93</td>
</tr>
<tr>
<td>3</td>
<td>-3.27</td>
<td>.79</td>
<td>8</td>
<td>.80</td>
<td>.97</td>
<td>13</td>
<td>5.23</td>
<td>1.12</td>
</tr>
<tr>
<td>4</td>
<td>-2.64</td>
<td>.78</td>
<td>9</td>
<td>1.72</td>
<td>.92</td>
<td>14</td>
<td>6.60E</td>
<td>1.84</td>
</tr>
</tbody>
</table>

CURRENT VALUES, UMEAN=0.00 USCALE=1.00
TO SET MEASURE RANGE AS 0-100, UMEAN=48.307 USCALE=7.827
TO SET MEASURE RANGE TO MATCH RAW SCORE RANGE, UMEAN=6.763 USCALE=1.096
TEST SLOPE=.40 INTERCEPT=-.02

TEST SLOPE and INTERCEPT

These are estimated from the relationship:

\[
\log \left( \frac{\text{SCORE}}{\text{MAXIMUM} - \text{SCORE}} \right) = \text{TEST SLOPE} \times \left( \text{MEASURE} - \text{INTERCEPT} \right) / \text{USCALE}
\]

with extreme scores (zero and maximum) omitted.

TEST SLOPE is the slope of the Test ICC relative to a standard logistic ogive.

INTERCEPT usually approximates the mean item difficulty.

<table>
<thead>
<tr>
<th>SCORE</th>
<th>MEASURE</th>
<th>S.E.</th>
<th>SCORE</th>
<th>MEASURE</th>
<th>S.E.</th>
<th>SCORE</th>
<th>MEASURE</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-6.17E</td>
<td>1.83</td>
<td>5</td>
<td>-1.97</td>
<td>.83</td>
<td>10</td>
<td>2.55</td>
<td>.89</td>
</tr>
<tr>
<td>1</td>
<td>-4.86</td>
<td>1.08</td>
<td>6</td>
<td>-1.19</td>
<td>.92</td>
<td>11</td>
<td>3.37</td>
<td>.89</td>
</tr>
<tr>
<td>2</td>
<td>-3.94</td>
<td>.85</td>
<td>7</td>
<td>-.23</td>
<td>1.00</td>
<td>12</td>
<td>4.21</td>
<td>.93</td>
</tr>
<tr>
<td>3</td>
<td>-3.27</td>
<td>.79</td>
<td>8</td>
<td>.80</td>
<td>.97</td>
<td>13</td>
<td>5.23</td>
<td>1.12</td>
</tr>
<tr>
<td>4</td>
<td>-2.64</td>
<td>.78</td>
<td>9</td>
<td>1.72</td>
<td>.92</td>
<td>14</td>
<td>6.60E</td>
<td>1.84</td>
</tr>
</tbody>
</table>

CURRENT VALUES, UMEAN=0.00 USCALE=1.00
TO SET MEASURE RANGE AS 0-100, UMEAN=48.307 USCALE=7.827
TO SET MEASURE RANGE TO MATCH RAW SCORE RANGE, UMEAN=6.763 USCALE=1.096
TEST SLOPE=.40 INTERCEPT=-.02

RAW SCORE-MEASURE OGIVE FOR COMPLETE TEST
The statistical information is \((USCALE/S.E.)^2\). For the **test information function**, plot \((USCALE/S.E.)^2\) against the "MEASURE" or the "SCORE". The plot can be drawn with EXCEL. It is easier to obtain the information directly from the SCOREFILE=.

### Table 20.2

<table>
<thead>
<tr>
<th>SCORE</th>
<th>MEASURE</th>
<th>S.E.</th>
<th>NORMED S.E.</th>
<th>FREQUENCY %</th>
<th>CUM.FREQ. %</th>
<th>PERCENTILE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-6.17E</td>
<td>1.83</td>
<td>147</td>
<td>107</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>-4.86</td>
<td>1.08</td>
<td>225</td>
<td>63</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>-3.94</td>
<td>0.85</td>
<td>278</td>
<td>50</td>
<td>1 2.9</td>
<td>1 2.9</td>
</tr>
<tr>
<td>3</td>
<td>-3.27</td>
<td>0.79</td>
<td>318</td>
<td>46</td>
<td>2 5.9</td>
<td>3 8.8</td>
</tr>
<tr>
<td>4</td>
<td>-2.64</td>
<td>0.78</td>
<td>355</td>
<td>46</td>
<td>2 5.9</td>
<td>5 14.7</td>
</tr>
<tr>
<td>5</td>
<td>-1.97</td>
<td>0.83</td>
<td>394</td>
<td>49</td>
<td>2 5.9</td>
<td>7 20.6</td>
</tr>
<tr>
<td>6</td>
<td>-1.19</td>
<td>0.92</td>
<td>440</td>
<td>54</td>
<td>3 8.8</td>
<td>10 29.4</td>
</tr>
<tr>
<td>7</td>
<td>-0.23</td>
<td>1.00</td>
<td>496</td>
<td>59</td>
<td>12 35.3</td>
<td>22 64.7</td>
</tr>
<tr>
<td>8</td>
<td>0.80</td>
<td>0.97</td>
<td>557</td>
<td>57</td>
<td>5 14.7</td>
<td>27 79.4</td>
</tr>
<tr>
<td>9</td>
<td>1.72</td>
<td>0.92</td>
<td>610</td>
<td>54</td>
<td>4 11.8</td>
<td>31 91.2</td>
</tr>
<tr>
<td>10</td>
<td>2.55</td>
<td>0.89</td>
<td>660</td>
<td>52</td>
<td>1 2.9</td>
<td>32 94.1</td>
</tr>
<tr>
<td>11</td>
<td>3.37</td>
<td>0.89</td>
<td>707</td>
<td>52</td>
<td>2 5.9</td>
<td>34 100.0</td>
</tr>
<tr>
<td>12</td>
<td>4.21</td>
<td>0.93</td>
<td>756</td>
<td>54</td>
<td>0</td>
<td>34 100.0</td>
</tr>
<tr>
<td>13</td>
<td>5.23</td>
<td>1.12</td>
<td>817</td>
<td>66</td>
<td>0</td>
<td>34 100.0</td>
</tr>
<tr>
<td>14</td>
<td>6.60E</td>
<td>1.84</td>
<td>897</td>
<td>108</td>
<td>0</td>
<td>34 100.0</td>
</tr>
</tbody>
</table>
The columns in the Table of Sample Norms and Frequencies are:

**Measures on the Complete Test:**

- **SCORE**: raw score on a complete test containing all calibrated items.
- **MEASURE**: measure corresponding to score.

  *If a person did not take all items or items are weighted, then that person is stratified with the measure on the complete test nearest the person's estimated measure (as reported in Table 18), regardless of that person's observed score.*

- **S.E.**: standard error of the measure (model).

  The statistical information is \((\text{USCALE}/\text{S.E.})^2\)

**Statistics for this sample:**

- **NORMED**: measures linearly locally-rescaled so that the mean person measure for this sample is 500 and the sample measure sample standard deviation is 100. Equivalent to \(\text{UPMEAN}=500, \text{USCALE}=100/(\text{Person S.D.})\)
- **S.E.**: standard error of the normed measure.
- **FREQUENCY**: count of sample with measures at or near (for missing data) the complete test measure
- **%**: percentage of sample included in FREQUENCY.
- **CUM.FREQ.**: count of sample with measures near or below the test measure, the cumulative frequency.
- **%**: percentage of sample include in CUM. FREQ.
- **PERCENTILE**: mid-range percentage of sample below the test measure, constrained to the range 1-99.

Logit measures support direct probabilistic inferences about relative performances between persons and absolute performances relative to items. Normed measures support descriptions about the location of subjects within a sample (and maybe a population). Report the measures which are most relevant to your audience.

This Table is easy to paste into Excel. use Excel's "data", "text to columns" feature to put the scores and measures into columns.

Table 20 shows integer raw scores, unless there are decimal weights for \(I\text{WEIGHT}=\). In which case, scores to 1 decimal place are shown.

To obtain other decimal raw scores for short tests, go to the Graphs pull-down menu. Select "Test Characteristic Curve". This displays the score-to-measure ogive. Click on "Copy data to clipboard". Open Excel. Paste. There will be to three columns. The second column is the measure, the third column is the raw score.

Score-to-measure Table 20 can be produced from known item measures \((I\text{FILE}=)\) and, for polytomies, known rating scale structure difficulties \((S\text{FILE}=)\). 

- **IAFILE=** (IFILE=): the item anchor file
- **SAFILE=** (SFILE=): the structure/step anchor file (if not dichotomies)
- **CONVERGE=L**: only logit change is used for convergence
- **LCONV=0.005**: logit change too small to appear on any report.
- **STBIAS=NO**: no estimation bias correction with anchor values

The data file comprises two dummy data records, so that every item has a non extreme score, e.g.,

For dichotomies:
- Record 1: 10101010101
- Record 2: 01010101010

For a rating scale from 1 to 5:
- Record 1: 15151515151
- Record 2: 51515151515

**235. Table 20.3 Complete score-to-calibration table for tests based on whole sample**

This Table, which must be selected explicitly with \(T\text{FILE}=\), or as a subtable with "Request Subtable" on the "Output Tables" menu, shows an estimated item calibration for all possible rates of success by persons, i.e., the item measure corresponding to every observable \(p\)-value for the entire sample. To select this Table, use "Request subtables" in the Output Tables menu,
TABLE OF ITEM MEASURES ON COMPLETE SAMPLE
FOR GROUPING "0", MODEL "R", ACT NUMBER: 1  WATCH BIRDS
+--------------------------------------------------------------------------+
| SCORE  MEASURE    S.E. | SCORE  MEASURE    S.E. | SCORE  MEASURE    S.E. |
|------------------------+------------------------+------------------------|
|     0     8.71E   1.81 |    50     1.91     .22 |   100     -.51     .22 |
|     1     7.47    1.02 |    51     1.86     .22 |   101     -.56     .22 |
|     2     6.72     .74 |    52     1.82     .22 |   102     -.61     .22 |
|     3     6.26     .62 |    53     1.77     .22 |   103     -.66     .23 |
+--------------------------------------------------------------------------+

SCOREraw score on this item of a complete sample containing all calibrated persons.
MEASUREmeasure corresponding to score.
S.E.standard error of the measure.
The statistical information is (USCALE/S.E.)²

If an item is not taken by all persons or persons are weighted, then that item is stratified with the measure on the complete sample nearest the estimated measure, regardless of the observed score.

236. Table 21 Probability curves
(controlled by MRANGE=, CURVES=)
To produce these using other software, see GRFILE=

The probability of each response is shown across the measurement continuum. The measure to be used for determining the probability of any particular response is the difference between the measure of the person and the calibration of the item. For dichotomies, only one curve is shown plotting the probability of scoring a "1" (correct), and also of scoring a "0" (incorrect) for any measure relative to item measure. For 'S' and 'F' models these curves are approximations.

When there are more than two categories, the probability of each category is shown.

218
For response structures with three or more categories, two further graphs can be drawn. The second graph depicts the expected score ogive. The vertical "*" characters correspond to integer expected scores, and the "|" characters correspond to half-point expected scores, the Rasch-half-point thresholds. The intervals between the Rasch-half-point thresholds can be thought of as the intervals corresponding to the observed categories. For the purposes of inference, measures in the zone on the x-axis between "|" and "|" correspond, on average, to the rating given on the 'y' axis, '1'. Similarly ratings on the y-axis can be thought of as corresponding to measures in the matching zone on the x-axis. The degree to which the data support this is given by the COHERENCE statistics in Table 3. Empirical item characteristic curves are shown in Table 29 and from the Graphs menu.

The third graph is of the zone curves which indicate the probability of an item score at or below the stated category for any particular difference between person measure and item calibration. The area to the left of the "0" ogive corresponds to "0". The right-most area corresponds to the highest category. The P=0.5 intercepts are the median cumulative probabilities. "|" indicate the Rasch-Thurstone thresholds.

237. Table 22.1 Sorted observed data matrix (scalogram)

The observations are printed in order of person and item measures, with most able persons listed first, the easiest items printed on the left. This scalogram shows the extent to which a Guttman pattern is approximated. The zoned responses are in Table 22.2 and the original responses in Table 22.3.
238. **Table 22.2 Guttman scalogram of zoned responses.**

The scalogram is that of Table 22.1, but with each observation marked as to whether it conforms with its expectation or not. Observations within 0.5 rating points of their expectation are deemed to be in their expected categories, and are reported with their category values, e.g., ‘1’, ‘2’, etc. These ratings support the overall inferential relationship between observations and measures. Observations more than 0.5 rating points away from their expectations, i.e., in a "wrong" category, are marked with a letter equivalent: ‘A’ = ‘0’, ‘B’ = ‘1’, ‘C’ = ‘2’, etc. These contradict observation-to-measure inferences. The proportion of in- and out-of-category observations are reported by the COHERENCE statistics in Table 3.

<table>
<thead>
<tr>
<th>PERSON</th>
<th>ITEM</th>
<th>ACT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1111112</td>
<td>1 221</td>
<td>21 22</td>
</tr>
<tr>
<td>8920311251427643569784035</td>
<td></td>
<td>890123125427634569780435</td>
</tr>
</tbody>
</table>

239. **Table 22.3 Guttman scalogram of original codes.**

The observations are printed in order of person and item measures, with most able persons listed first, the easiest items printed on the left. This scalogram shows the original codes in the data file. Here is the Scalogram for Example 5, a computer-adaptive, multiple-choice test.
240. Table 23.0, 24.0 Variance components scree plot for items or persons

This Table shows the variance decomposition in the observations.

<table>
<thead>
<tr>
<th>Total variance in observations</th>
<th>Variance explained by measures</th>
<th>Unexplained variance (total)</th>
<th>Unexpl var explained by 1st contrast</th>
<th>Unexpl var explained by 2nd contrast</th>
<th>Unexpl var explained by 3rd contrast</th>
<th>Unexpl var explained by 4th contrast</th>
<th>Unexpl var explained by 5th contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical</td>
<td>Modeled</td>
<td>Empirical</td>
<td>Modeled</td>
<td>Modeled</td>
<td>Modeled</td>
<td>Modeled</td>
<td>Modeled</td>
</tr>
<tr>
<td>Total variance in observations</td>
<td>142.7</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Variance explained by measures</td>
<td>117.7</td>
<td>82.5%</td>
<td>84.5%</td>
<td>82.5%</td>
<td>84.5%</td>
<td>84.5%</td>
<td>84.5%</td>
</tr>
<tr>
<td>Unexplained variance (total)</td>
<td>25.0</td>
<td>17.5%</td>
<td>15.5%</td>
<td>17.5%</td>
<td>15.5%</td>
<td>17.5%</td>
<td>15.5%</td>
</tr>
<tr>
<td>Unexpl var explained by 1st contrast</td>
<td>4.5</td>
<td>3.1%</td>
<td>17.8%</td>
<td>4.5</td>
<td>3.1%</td>
<td>17.8%</td>
<td>4.5</td>
</tr>
<tr>
<td>Unexpl var explained by 2nd contrast</td>
<td>3.0</td>
<td>2.1%</td>
<td>11.9%</td>
<td>3.0</td>
<td>2.1%</td>
<td>11.9%</td>
<td>3.0</td>
</tr>
<tr>
<td>Unexpl var explained by 3rd contrast</td>
<td>2.4</td>
<td>1.7%</td>
<td>9.6%</td>
<td>2.4</td>
<td>1.7%</td>
<td>9.6%</td>
<td>2.4</td>
</tr>
<tr>
<td>Unexpl var explained by 4th contrast</td>
<td>1.7</td>
<td>1.2%</td>
<td>6.9%</td>
<td>1.7</td>
<td>1.2%</td>
<td>6.9%</td>
<td>1.7</td>
</tr>
<tr>
<td>Unexpl var explained by 5th contrast</td>
<td>1.6</td>
<td>1.1%</td>
<td>6.4%</td>
<td>1.6</td>
<td>1.1%</td>
<td>6.4%</td>
<td>1.6</td>
</tr>
</tbody>
</table>

Table of STANDARDIZED RESIDUAL variance: the standardized residuals form the basis of this computation, set by PRCOMP= (in Eigenvalue units): variance components are rescaled so that the total unexplained variance has its expected summed eigenvalue.

Empirical: variance components for the observed data
Model: variance components expected for the data if exactly fit the Rasch model
Total variance in observations: total variance in the observations around their Rasch expected values in standardized residual units
Variance explained by measures: variance explained by the item difficulties, person abilities and rating scale structures.
Unexplained variance (total): variance not explained by the Rasch measures
Unexpl var explained by 1st, 2nd, ... contrast: size of the first, second, ... contrast (component) in the principal component decomposition of residuals

Variance Explained
In this example, 82.5% is explained by the measures based on the standardized residuals from the observations. If the data fit the model perfectly, 84.5% would be explained. The unexplained variance in the data is 17.5%. This includes the Rasch-predicted randomness and any departures in the data from Rasch criteria, e.g., multidimensionality.

The variance terms are computed in the following way. The average of the responses made by all persons to all items is computed. Extreme scores and missing data are omitted. Adjustment is made of number of categories in polytomies. This average "standard" response is what every response would be if all items were equally difficult and all persons were equally able. The variance to be explained by the Rasch measures is caused by the spread of the item difficulties, person abilities, and disparate polytomous structures.

i) The empirical total observed response variance is the sum of the (observed - standard responses)**2.

ii) The modeled variance explained by the measures is the sum of (Rasch expectations - standard responses)**2.

iii) The modeled unexplained variance is sum of (Rasch-model response variances).

iv) The modeled total observed response variance is the sum of (i) and (iii).

(v) The empirical unexplained variance is the larger of:
   (a) The sum of the (observed - Rasch expectation)**2.
   (b) The absolute value of (i) - (ii).

(vi) The empirical explained variance is (i) - (v).

All values are locally-rescaled so that the total unexplained variance matches the sum of the eigenvalues, i.e., the variance to be explained, by the PCA of residuals. In general, the empirical variances are more conservative, misfit- or dimensionality-attenuated, values. The modeled values correspond to unidimensional data fitting the...
Rasch model. This is analogous to “real” and “model” standard errors.

Effect of misfit or gross dimensionality
Here is part of Table 23 for an analysis which contains 3 items negatively-correlated with the others - symptomatic of miskeying or gross multi-dimensionality. The empirical explained variance is conspicuously less than the model explained variance.

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

<table>
<thead>
<tr>
<th></th>
<th>Empirical</th>
<th>Modeled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total variance in observations</td>
<td>25.7 100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Variance explained by measures</td>
<td>5.7 22.1%</td>
<td>47.8%</td>
</tr>
<tr>
<td>Unexplained variance (total)</td>
<td>20.0 77.9%</td>
<td>52.2%</td>
</tr>
<tr>
<td>Unexpl var explained by 1st contrast</td>
<td>5.3 20.6%</td>
<td></td>
</tr>
</tbody>
</table>

VARIANCE COMPONENT SCREE PLOT

Scree plot of the variance component percentage sizes, logarithmically scaled.
T, TV: total variance in the observations, always 100%
M, MV: variance in the observations explained by the Rasch measures
U, UV: unexplained variance
1, U1: first contrast (component) in the residuals
2, U2: second contrast (component) in the residuals, etc.

241. Table 23.2, 24.2 Principal components plots of item or person loadings

Please do not interpret this as a usual factor analysis. These plots show contrasts between opposing factors, not loadings on one factor. For more discussion, see dimensionality and contrasts.

Quick summary:
(a) the X-axis is the measurement axis. So we are not concerned about quadrants, we are concerned about vertical differences. The Table 23 plots show contrasts between types of items: those at the top vs. those at the bottom. The Table 24 plots show contrasts between types of persons: those at the top vs. those at the bottom.

(b) “How much” is important. See the Variance Table explained in Table 23.0. Important differences have eigenvalues greater than 2.0.

(c) If the difference is important, it suggests that we divide the test into two pieces: the items in the top half of the plot and the items in the bottom half. Perform two separate analyses and cross-plot and correlate the person measures. We will then see for whom the differences are important. Usually, for a carefully designed instrument,
it is such a small segment that we decide it is not worth thinking of the test as measuring two dimensions. Tables 23.4 and 24.4 also help us think about this.

These plots show the contrasts by plotting the loading on each component against the item calibration (or person measure). The contrast shows items (or persons) with different residual patterns. A random pattern with few high loadings is expected.

The horizontal axis is the Rasch dimension. This has been extracted from the data prior to the analysis of residuals.

Letters “A” and “a” identify items (persons) with the most opposed loadings on the first contrast in the residuals. On subsequent contrasts, the items retain their first contrast identifying letters.

In the residuals, each item (person) is modeled to contribute one unit of randomness. Thus, there are as many residual variance units as there are items (or persons). For comparison, the amount of person (item) variance explained by the item (person) measures is approximated as units of that same size.

In this example, based on the FIM™ data, Example 10 using exam10hi.txt data, the first contrast in the standardized residuals separates the items into 3 clusters. To identify the items, see Tables 23.3, 24.3. You will see that items A and B have a psycho-social aspect.

In this example, the dimension is noticeable, with strength of around 3 out of 13 items. This is in the residual variance, i.e., in the part of the observations unexplained by the measurement model. But, hopefully, most of the variance in the observations has been explained by the model. The part of that explained variance attributable to the Persons is shown in variance units locally-rescaled to accord with the residual variances. In this example, the variance explained by the measures is equivalent to 16 items. Consequently, though the secondary dimension in the items is noticeable, we do not expect it to have much practical impact on person measurement.

For items:

STANDARDIZED RESIDUAL CONTRAST 1 PLOT
Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

<table>
<thead>
<tr>
<th>Empirical</th>
<th>Modeled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total variance in observations</td>
<td>29.5 100.0% 100.0%</td>
</tr>
<tr>
<td>Variance explained by measures</td>
<td>16.5 56.0% 56.3%</td>
</tr>
<tr>
<td>Unexplained variance (total)</td>
<td>13.0 44.0% 100.0% 43.7%</td>
</tr>
<tr>
<td>Unexplained variance in 1st contrast</td>
<td>3.3 11.1% 25.2%</td>
</tr>
</tbody>
</table>

-2 -1 0 1 2

.9 + A
.8 + B
.7 +
.6 + C
.5 + O
.4 + T
.3 + R
.2 + A
.1 + S
+.0 +
+.1 + L
+.2 + D
+.3 + I
+.4 + N
+.5 + G
+.6 + f
+.7 + e
+.8 + d
+.9 + c
+1.0 + b
+1.1 + a

223
The plot shows a contrast in the residuals for PERSONS. Each letter is a person up to a maximum of 52 persons, A-Z and a-z. For persons 53-181, "1" means that there is one person at that location on the plot. "2" means that there are two persons, etc.

**Table 23.3, 24.3 Principal components analysis/contrast of residuals**

Please do not interpret this as a usual factor analysis. These plots show **contrasts** between opposing factors, **not loadings on one factor**. For more discussion, see dimensionality and contrasts.

This Table decomposes the matrix of item (Table 23) or person (Table 24) correlations based on residuals to identify possible other contrasts (dimensions) that may be affecting response patterns. Specify PRCOMP=S or =R or =L to obtain this Table.

Prior to this first contrast, the Rasch dimension has been extracted from the data. Residuals are those parts of the observations not explained by the Rasch dimension. According to Rasch specifications, these should be random and show no structure. The contrasts show conflicting local patterns in inter-item (or inter-person) correlations based on residuals or their transformations. Letters "E", "b", etc. relate items (or persons) to their loadings on the first contrast. In this Table, Bladder and Bowel contrast with Dressing. Since Bladder and Bowel misfit, they load on a second dimension in the data.
Table 23.4, 24.4 Person or item contrast

Please do not interpret this as a usual factor analysis. These plots show contrasts between opposing factors, not loadings on one factor. For more discussion, see dimensionality and contrasts.

The effect of the contrast between the oppositely loading items (Table 23) or persons (Table 24) at the top and bottom of the contrast plot is shown here.

Responses by persons (or on items) to the items (or by the persons) with extreme positive loadings and the items (or persons) with extreme negative loadings are identified. These responses are seen to be higher, as expected, or lower than the model predicts. Counts of these are obtained for each person (or item). The persons (or items) showing the biggest impact of this contrast are listed first. Items and persons showing the most contrast are chosen for this Table based on the "Liking for Science" data.

Table 23.4, showing the impact of the first contrast in the item residuals on persons:

<table>
<thead>
<tr>
<th>ACT contrast 1 CONTRASTING RESPONSES BY PUPILS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Person scoring higher than expected on top 7 items</td>
</tr>
<tr>
<td>5 2 0</td>
</tr>
<tr>
<td>5 2 0</td>
</tr>
</tbody>
</table>

Table 24.4, showing the impact of the first contrast in the person residuals on items:

<table>
<thead>
<tr>
<th>PUPIL contrast 1 CONTRASTING RESPONSES BY ACTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Items biased towards top 9 persons</td>
</tr>
<tr>
<td>9 0 0</td>
</tr>
<tr>
<td>8 1 0</td>
</tr>
<tr>
<td>7 2 0</td>
</tr>
</tbody>
</table>

Items biased towards bottom 9 persons |
| 0 4 5 | 0 9 0 | 13 GROW GARDEN |
| 0 5 4 | 0 9 0 | 2 READ BOOKS ON ANIMALS |
| 0 5 4 | 0 9 0 | 15 READ ANIMAL STORIES |
Table 23.99, 24.99 Largest residual correlations for items or persons

These Tables show items (Table 23.99, formerly Table 23.1) and persons (Table 24.99, formerly 24.1) that may be locally dependent. Specify PRCOMP=R (for score residuals) or PRCOMP=S or Y (for standardized residuals) or PRCOMP=L (for logit residuals) to obtain this Table. Residuals are those parts of the data not explained by the Rasch model. High correlation of residuals for two items (or persons) indicates that they may not be locally independent, either because they duplicate some feature of each other or because they both incorporate some other shared dimension.

Missing data are deleted pairwise if both of a pair are missing or PRCOMP=O (for observations), otherwise missing data are replaced by their Rasch expected residuals of 0.

<table>
<thead>
<tr>
<th>RESID</th>
<th>ENTRY</th>
<th>ENTRY</th>
</tr>
</thead>
<tbody>
<tr>
<td>.71</td>
<td>17 Q. PROBLEM SOLVING</td>
<td>18 R. MEMORY</td>
</tr>
<tr>
<td>.69</td>
<td>14 N. COMPREHENSION</td>
<td>15 O. EXPRESSION</td>
</tr>
<tr>
<td>.61</td>
<td>9 I. BED TRANSFER</td>
<td>10 J. TOILET TRANSFER</td>
</tr>
<tr>
<td>.54</td>
<td>4 D. UPPER BODY DRESSING</td>
<td>8 H. BOWEL</td>
</tr>
<tr>
<td>.46</td>
<td>16 P. SOCIAL INTERACTION</td>
<td>17 Q. PROBLEM SOLVING</td>
</tr>
<tr>
<td>.45</td>
<td>16 P. SOCIAL INTERACTION</td>
<td>18 R. MEMORY</td>
</tr>
<tr>
<td>.40</td>
<td>10 J. TOILET TRANSFER</td>
<td>11 K. TUB, SHOWER</td>
</tr>
<tr>
<td>-.37</td>
<td>10 J. TOILET TRANSFER</td>
<td>18 R. MEMORY</td>
</tr>
<tr>
<td>-.36</td>
<td>10 J. TOILET TRANSFER</td>
<td>17 Q. PROBLEM SOLVING</td>
</tr>
</tbody>
</table>

Note: Redundant correlations of 1.0 are not printed. If A has a correlation of 1.0 with B, and also with C, assume that B and C also have a correlation of 1.0.

This Table is used to detect dependency between pairs of items or persons. When raw score residual correlations are computed, it corresponds to Wendy Yen's Q3 statistic. is used to detect dependency between pairs of items or persons. Yen suggests a small positive adjustment to the correlation of size 1/(L-1) where L is the test length. Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. Applied Psychological Measurement, 8, 125-145. Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. Journal of Educational Measurement, 30, 187-213. Yen suggest a small adjustment for bias

Table 27.1, 28.1 Subtotal summaries on one line

These summarize the measures from the main analysis for all items or persons selected by ISUBTOT= (Table 27) or PSUBTOT= (Table 28), including extreme scores. Histograms are shown in Tables 27.2 and 28.2. PSUBTOTAL= is useful for quantifying the impact of a test on different types of test-takers.

<table>
<thead>
<tr>
<th>KID</th>
<th>MEAN</th>
<th>S.E.</th>
<th>OBSERVED</th>
<th>MEDIAN</th>
<th>REAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>-.36</td>
<td>.37</td>
<td>2.15</td>
<td>-.25</td>
<td>1.55</td>
</tr>
<tr>
<td>18</td>
<td>-.36</td>
<td>.52</td>
<td>2.14</td>
<td>-.25</td>
<td>1.47</td>
</tr>
<tr>
<td>17</td>
<td>-.37</td>
<td>.54</td>
<td>2.17</td>
<td>-.25</td>
<td>1.63</td>
</tr>
</tbody>
</table>

The first line, "***", is the total for all persons (or items)
The remaining codes are those in the person (or item) columns specified by $S9W1$ or whatever, using the column selection rules. In this example, "F" is the code for "Female" in the data file. "M" for "Male". It is seen that the two distributions are almost identical.

The statistical significance of the difference between the two subtotal means is:
\[
t = (\text{mean measure of } "F" - \text{mean measure of } "M") / \sqrt{(\text{S.E. Mean } "F")^2 + (\text{S.E. Mean } "M")^2}
\]

REAL SEPARATION is the separation coefficient defined as "true" sample standard deviation / measurement error based on misfit-inflated error variance.

### 246. Table 27.2, 28.2 Subtotal measure histograms

These show the distributions of the measures from the main analysis for each sub-sample. Summaries are shown in Tables 27.1 (Items, $ISUBTOT=$) and 28.1 (Persons, $PSUBTOT=$).

Here is the measure distribution of the total sample:

```
| T | S | M | S | T |
---|---|---|---|---|
-5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 |
```

Read the counts vertically so that the center count is 12 observed near -0.4.

**M** = Mean, **S** = one sample standard deviation from mean, **T** = two sample standard deviations from mean.

Here is the measure distribution of the total sample standardized to a sample size of 1000:

```
| T | S | M | S | T |
---|---|---|---|---|
2 | 2 | 5 | 5 | 5 | 8 | 4 | 4 | 1 | 2 | 5 |
9 | 9 | 8 | 8 | 8 | 6 | 3 | 3 | 5 | 9 | 8 |
```

Here is the measure distribution of one sub-sample, specified as $S9W1=\"F\", using the column selection rules.

```
| T | S | M | S | T |
---|---|---|---|---|
1 | 1 | 1 | 2 | 7 | 3 | 2 | 1 |
```

Here is the measure distribution of the sub-sample standardized to a sample size of 1000:

```
| T | S | M | S | T |
---|---|---|---|---|
5 | 5 | 5 | 1 | 8 | 6 | 1 | 5 |
6 | 6 | 6 | 2 | 9 | 7 | 2 | 6 |
```

Here is the measure distribution of the sub-sample standardized so that the total at any measure is 1,000, but in proportion to the observed counts in each sub-sample:

```
| T | S | M | S | T |
---|---|---|---|---|
-5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 |
```

Here is the measure distribution of the sub-sample standardized so that the total any measure is 1,000, but based on the proportion observed when the sizes of all sub-samples are adjusted to be equal:

```
| T | S | M | S | T |
---|---|---|---|---|
1 |
```

Read the counts vertically so that the center count is 12 observed near -0.4.

**M** = Mean, **S** = one sample standard deviation from mean, **T** = two sample standard deviations from mean.
247. **Table 29 Empirical ICCs and option frequencies**

This Table display both the model expected score ogive, the predicted Item Characteristic Curve (ICC), and also the empirical observed average score per measure, the empirical ICC. See also Graphs Menu.

It also shows the relative frequency of each item response code for each measure level.

The model ICC is shown by `.`, this is the expected average score on the item for persons at each measure relative to the item.

Observed average scores on the items are shown by `x`. When `. ` and `x` coincide, `*` is shown.

You can use this plot to compute the empirical item discrimination.

For a multiple-choice item, with C as an incorrect distractor, this could like the plot below. The dots are the Rasch ogive for an incorrect answer.

The model ICC is shown by `. `. Observed average scores on the items are shown by `x`. When `. ` and `x` coincide, `*` is shown.
For a polytomous item, the percentage frequencies of responses, and the model ordinal probability, of each category are shown at each measure level. The dots are the model curve. "4" is the empirical frequency for category 4. "**" means a "4" and a "."/

Table 30 Differential item functioning DIF bias analysis

Table 30 supports the investigation of item bias, Differential Item Functioning (DIF), i.e., interactions between individual items and types of persons. Specify DIF= for person classifying indicators in person labels. Item bias and DIF are the same thing. The widespread use of "item bias" dates to the 1960's, "DIF" to the 1980's. The reported DIF is corrected to test impact, i.e., differential average performance on the whole test. Use ability stratification to look for non-uniform DIF using the selection rules. Tables 30.1 and 30.2 present the same information from different perspectives.

Table 31 supports person bias, Differential Person Functioning (DPF), i.e., interactions between individual persons and classifications of items.

Table 33 reports bias or interactions between classifications of items and classifications of persons.

In these analyses, persons and items with extreme scores are excluded, because they do not exhibit differential ability across items. For background discussion, see DIF and DPF considerations.

Example output:
You want to examine item bias (DIF) between Females and Males. You need a column in your Winsteps person label that has two (or more) demographic codes, say "F" for female and "M" for male (or "0" and "1" if you like dummy variables) in column 9.

**Table 30.1** is best for pairwise comparisons, e.g., Females vs. Males.

DIF specification is: DIF=S9W1

<table>
<thead>
<tr>
<th>KID</th>
<th>DIF</th>
<th>DIF</th>
<th>KID</th>
<th>DIF</th>
<th>DIF</th>
<th>DIF</th>
<th>JOINT</th>
<th>MantelHanzl</th>
<th>TAP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CLASS MEASURE S.E.</td>
<td>CLASS MEASURE S.E.</td>
<td>CONTRAST S.E.</td>
<td>t</td>
<td>d.f.</td>
<td>Prob.</td>
<td>Prob.</td>
<td>Size</td>
<td>Number</td>
</tr>
<tr>
<td>-----</td>
<td>---------------------</td>
<td>---------------------</td>
<td>----------------</td>
<td>------</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
<td>------</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td>2.85</td>
<td>.89</td>
<td>1.24</td>
<td>.70</td>
<td>1.61</td>
<td>1.42</td>
<td>.2049</td>
<td>1.95</td>
<td>13</td>
</tr>
<tr>
<td>F</td>
<td>5.25</td>
<td>1.90</td>
<td>-3.89</td>
<td>.90</td>
<td>-1.37</td>
<td>2.10</td>
<td>.5188</td>
<td>.2528</td>
<td>4</td>
</tr>
<tr>
<td>F</td>
<td>-2.78</td>
<td>.89</td>
<td>-3.89</td>
<td>.90</td>
<td>1.10</td>
<td>1.26</td>
<td>.87</td>
<td>31</td>
<td>.3884</td>
</tr>
</tbody>
</table>

Size of Mantel-Haenszel slice = .100

DIF Specification defines the columns used to identify DIF classifications, using the selection rules.

Reading across the Table 30.1 columns:

- KID CLASS identifies the CLASS of persons. KID is specified with PERSON=, e.g., the first CLASS is "F"
- DIF MEASURE is the difficulty of this item for this class, with all else held constant, e.g., -5.24 is the local difficulty for Class F of Item 4.
- 5.31< reports that this measure corresponds to an extreme minimum score. EXTRSCORE= controls extreme score estimate.
- DIF S.E. is the standard error of the DIF MEASURE
- DIF CONTRAST is the difference between the DIF MEASUREs, i.e., size of the DIF across the two classifications of persons, e.g., 2.85 - 1.24 = 1.61. A positive DIF contrast indicates that the item is more difficult for the left-hand-listed CLASS.
- JOINT S.E. is the standard error of the DIF CONTRAST = sqrt(first DIF S.E.^2 + second DIF S.E.^2), e.g., 1.13 = sqrt(.89^2 + .70^2)
- t gives the DIF significance as a unit normal deviate = DIF CONTRAST / JOINT S.E. The t-test is a two-sided test for the difference between two means (i.e., the estimates) based on the standard error of the means (i.e., the standard error of the estimates). The null hypothesis is that the two estimates are the same, except for measurement error.
- d.f. is the joint degrees of freedom. This is shown as the sum of the sizes of two classifications - 2 for the two measure estimates, but this estimate of d.f. is somewhat high, so interpret the t-test conservatively, e.g., d.f. = (17 F + 17 M - 2) = 32.
- Prob. is the probability of the reported t with the reported d.f., but interpret this conservatively. If you wish to make a Bonferroni multiple-comparison correction, compare this Prob. with your chosen significance level, e.g., .05, divided by the number of entries in this Table.
- Mantel-Hanzel reports Mantel-Haenszel (1959) DIF test for dichotomies or Mantel (1963) for polytomies using MHSLICE=

Prob. is the probability of observing these data (or worse) when there is no DIF. Reported when computable.

Size is an estimate of the DIF (scaled by USCALE=). Reported when computable. Otherwise +. and -. indicate direction.

TAP Number is the item entry number. TAP is specified by ITEM=

Name is the item label.

<table>
<thead>
<tr>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
</table>
Table 30.2 is best for multiple comparisons, e.g., regions against the national average.

Table of the two-sided t distribution:

<table>
<thead>
<tr>
<th>d.f.</th>
<th>p=.05</th>
<th>p=.01</th>
<th>d.f.</th>
<th>p=.05</th>
<th>p=.01</th>
<th>d.f.</th>
<th>p=.05</th>
<th>p=.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.71</td>
<td>63.66</td>
<td>1</td>
<td>2.08</td>
<td>2.83</td>
<td>1</td>
<td>2.08</td>
<td>2.83</td>
</tr>
<tr>
<td>2</td>
<td>4.30</td>
<td>9.93</td>
<td>2</td>
<td>2.07</td>
<td>2.82</td>
<td>2</td>
<td>2.07</td>
<td>2.82</td>
</tr>
<tr>
<td>3</td>
<td>3.18</td>
<td>5.84</td>
<td>3</td>
<td>2.06</td>
<td>2.80</td>
<td>3</td>
<td>2.06</td>
<td>2.80</td>
</tr>
<tr>
<td>4</td>
<td>2.78</td>
<td>4.60</td>
<td>4</td>
<td>2.05</td>
<td>2.79</td>
<td>4</td>
<td>2.05</td>
<td>2.79</td>
</tr>
<tr>
<td>5</td>
<td>2.57</td>
<td>4.03</td>
<td>5</td>
<td>2.05</td>
<td>2.78</td>
<td>5</td>
<td>2.05</td>
<td>2.78</td>
</tr>
<tr>
<td>6</td>
<td>2.45</td>
<td>3.71</td>
<td>6</td>
<td>2.04</td>
<td>2.76</td>
<td>6</td>
<td>2.04</td>
<td>2.76</td>
</tr>
<tr>
<td>7</td>
<td>2.37</td>
<td>3.50</td>
<td>7</td>
<td>2.04</td>
<td>2.75</td>
<td>7</td>
<td>2.04</td>
<td>2.75</td>
</tr>
<tr>
<td>8</td>
<td>2.31</td>
<td>3.36</td>
<td>8</td>
<td>2.03</td>
<td>2.74</td>
<td>8</td>
<td>2.03</td>
<td>2.74</td>
</tr>
<tr>
<td>9</td>
<td>2.26</td>
<td>3.25</td>
<td>9</td>
<td>2.02</td>
<td>2.73</td>
<td>9</td>
<td>2.02</td>
<td>2.73</td>
</tr>
<tr>
<td>10</td>
<td>2.23</td>
<td>3.17</td>
<td>10</td>
<td>2.02</td>
<td>2.72</td>
<td>10</td>
<td>2.02</td>
<td>2.72</td>
</tr>
<tr>
<td>Inf.</td>
<td>1.96</td>
<td>2.58</td>
<td>Inf.</td>
<td>1.96</td>
<td>2.58</td>
<td>Inf.</td>
<td>1.96</td>
<td>2.58</td>
</tr>
</tbody>
</table>

249. Table 31 Differential person functioning DPF bias/interaction analysis

Table 31 supports person bias, Differential Person Functioning (DPF), i.e., interactions between individual persons and classifications of items. Specify DPF= for classifying indicators in item labels. Use difficulty stratification to look for non-uniform DPF using the selection rules.

Table 30 supports the investigation of item bias, Differential Item Functioning (DIF), i.e., interactions between individual items and types of persons.

Table 33 reports bias or interactions between classifications of items and classifications of persons.

In these analyses, persons and items with extreme scores are excluded, because they do not exhibit
differential ability across items. For background discussion, see DIF and DPF considerations.

**Example output:**

**Table 31.1**

<table>
<thead>
<tr>
<th>TAP CLASS</th>
<th>DPF CLASS</th>
<th>DPF S.E.</th>
<th>TAP CLASS</th>
<th>DPF CLASS</th>
<th>DPF S.E.</th>
<th>DPF CONTRAST</th>
<th>JOINT S.E.</th>
<th>t</th>
<th>d.f.</th>
<th>Prob.</th>
<th>Number</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DPF Specification defines the columns used to identify Differential Person Function classifications, using the selection rules.

- **TAP CLASS** is the item class.
- **DPF MEASURE** is the ability of the person for this item class, with all else held constant.
- **DPF S.E.** is the standard error of the measure.
- **DPF CONTRAST** is the difference in the person ability measures, i.e., size of the DPF, for the two classifications of items.
- **JOINT S.E.** is the standard error of the DPF CONTRAST.
- **t** gives the DPF significance as a Student's t-test. The t-test is a two-sided test for the difference between two means (i.e., the estimates) based on the standard error of the means (i.e., the standard error of the estimates). The null hypothesis is that the two estimates are the same, except for measurement error.
- **d.f.** is the joint degrees of freedom. This is shown as the sum of the sizes of two classifications - 2 for the two measure estimates, but this estimate of d.f. is somewhat high, so interpret the t-test conservatively.
- **Prob.** is the probability of the reported t with the reported d.f., but interpret this conservatively. If you wish to make a Bonferroni multiple-comparison correction, compare this Prob. with your chosen significance level, e.g., .05, divided by the number of entries in this Table.

-5.24 > reports that this measure corresponds to an extreme maximum score. **EXTRSCORE** controls extreme score estimate.

5.30 < reports that this measure corresponds to an extreme minimum score. **EXTRSCORE** controls extreme score estimate.

**Table 31.2**

<table>
<thead>
<tr>
<th>TAP CLASS</th>
<th>OBSERVATIONS</th>
<th>BASELINE</th>
<th>DPF MEASURE</th>
<th>DPF S.E.</th>
<th>DPF CONTRAST</th>
<th>JOINT S.E.</th>
<th>t</th>
<th>d.f.</th>
<th>Prob.</th>
<th>Number</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This displays a list of the local difficulty/ability estimates underlying the paired DPF analysis. These can be plotted directly from the Plots menu.

- **TAP CLASS:** is the item class - these are the subtests (item classifications) for which differential person functioning is to be investigated.

- **OBSERVATIONS:**
  - **COUNT** is the number of observations of the classification.
  - **AVERAGE** is the average observation on the classification.
  - **EXPECT** is the expected average value of the observations based on all items.

- **BASELINE:**
  - **SCORE MEASURE** is the overall ability measure based on all items.

Average of DPF measures across CLASS for a person is not the BASELINE MEASURE because score-
to-measure conversion is non-linear.

DPF: Differential Person Function
SCORE is the difference between average of the observations observed and expected for this class
MEASURE is the local ability of the person for this class
SIZE is the difference between the local and overall ability of this person. >> indicate extreme scores.
S.E. is the approximate standard error of the difference.
t is an approximate Student's t-test would be the MEASURE divided by the S.E. with a little less than (COUNT-2) degrees of freedom.
Probability is not reported, because the t-test is too inexact.

250. Table 32 Control specifications

This gives the setting of every Winsteps control variable. It can be accessed as Table 32 from a control file, or from the Output Files pull-down menu. It is written to a temporary text file, but can be "saved as" to a permanent file.

; Values of Control Specifications
; CONTROL FILE = C:\e\Ab6.0\bsteps\mrwe\data\kct.txt
; OUTPUT REPORT FILE = C:\e\Ab6.0\bsteps\mrwe\data\ZOU713ws.txt
; DATE AND TIME = Feb 11 23:29 2004
ALPHANUM = ASCII = Y
BATCH = N
CATREF = 0
CFILE = CHART = N
CLFILE = CODES = 01
CONVERGE = E

251. Table 33 DIF/DPF/DGF interactions by class

This Table identifies interactions between classifications of persons (identified by DIF=) and classifications of items (identified by DPF=) using the column selection rules. Differential average classification-group performance (DGF) is powerful when looking for latent classes among the persons. For more details, see Table 30 (DIF) and Table 31 (DPF).

Table 33.1

<table>
<thead>
<tr>
<th>KID</th>
<th>DIF</th>
<th>DIF</th>
<th>KID</th>
<th>DIF</th>
<th>DIF</th>
<th>DIF</th>
<th>JOINT</th>
<th>TAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASS</td>
<td>SIZE</td>
<td>S.E.</td>
<td>CLASS</td>
<td>SIZE</td>
<td>S.E.</td>
<td>CONTRAST</td>
<td>S.E.</td>
<td>t</td>
</tr>
<tr>
<td>F</td>
<td>-.03</td>
<td>.28</td>
<td>M</td>
<td>.05</td>
<td>.27</td>
<td>-.08</td>
<td>.39</td>
<td>.20</td>
</tr>
<tr>
<td>F</td>
<td>-.07</td>
<td>.58</td>
<td>M</td>
<td>.04</td>
<td>.55</td>
<td>-.11</td>
<td>.80</td>
<td>.14</td>
</tr>
<tr>
<td>F</td>
<td>.55</td>
<td>.88</td>
<td>M</td>
<td>-.49</td>
<td>.90</td>
<td>1.04</td>
<td>1.25</td>
<td>-.83</td>
</tr>
</tbody>
</table>

This Table contrasts, for each item class, the size and significance of the Differential Item Functioning for pairs of person classifications.

Table 33.2 - These can be plotted directly from the Plots menu.

<table>
<thead>
<tr>
<th>KID</th>
<th>OBSERVATIONS</th>
<th>BASELINE</th>
<th>DIF</th>
<th>DIF</th>
<th>DIF</th>
<th>DIF</th>
<th>DIF</th>
<th>TAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASS</td>
<td>COUNT</td>
<td>AVERAGE</td>
<td>EXPECT</td>
<td>SCORE</td>
<td>SIZE</td>
<td>S.E.</td>
<td>t</td>
<td>CLASS</td>
</tr>
<tr>
<td>F</td>
<td>187</td>
<td>.42</td>
<td>.41</td>
<td>.00</td>
<td>-.03</td>
<td>.28</td>
<td>-.11</td>
<td>1</td>
</tr>
<tr>
<td>M</td>
<td>187</td>
<td>.39</td>
<td>.39</td>
<td>.00</td>
<td>.05</td>
<td>.27</td>
<td>.17</td>
<td>1</td>
</tr>
<tr>
<td>F</td>
<td>34</td>
<td>.85</td>
<td>.85</td>
<td>.01</td>
<td>-.07</td>
<td>.58</td>
<td>-.12</td>
<td>2</td>
</tr>
<tr>
<td>M</td>
<td>34</td>
<td>.76</td>
<td>.77</td>
<td>.00</td>
<td>.04</td>
<td>.55</td>
<td>.08</td>
<td>2</td>
</tr>
<tr>
<td>F</td>
<td>17</td>
<td>.88</td>
<td>.92</td>
<td>-.04</td>
<td>.55</td>
<td>.88</td>
<td>.63</td>
<td>3</td>
</tr>
<tr>
<td>M</td>
<td>17</td>
<td>.88</td>
<td>.84</td>
<td>.04</td>
<td>-.49</td>
<td>.90</td>
<td>-.55</td>
<td>3</td>
</tr>
</tbody>
</table>
This shows the local size of the DIF for each person class on each item class. The reported t value is approximate, and should be interpreted conservatively.

**Table 33.3**

<table>
<thead>
<tr>
<th>TAP</th>
<th>DPF</th>
<th>DPF</th>
<th>DPF</th>
<th>DPF</th>
<th>JOINT</th>
<th>KID</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASS</td>
<td>SIZE</td>
<td>S.E.</td>
<td>CLASS</td>
<td>SIZE</td>
<td>S.E.</td>
<td>CONTRAST</td>
</tr>
<tr>
<td>1</td>
<td>.03</td>
<td>.28</td>
<td>2</td>
<td>.07</td>
<td>.58</td>
<td>-.04</td>
</tr>
<tr>
<td>1</td>
<td>.03</td>
<td>.28</td>
<td>3</td>
<td>-.55</td>
<td>.88</td>
<td>.58</td>
</tr>
<tr>
<td>2</td>
<td>.07</td>
<td>.58</td>
<td>3</td>
<td>-.55</td>
<td>.88</td>
<td>.62</td>
</tr>
<tr>
<td>1</td>
<td>-.05</td>
<td>.27</td>
<td>2</td>
<td>-.04</td>
<td>.55</td>
<td>.00</td>
</tr>
<tr>
<td>1</td>
<td>-.05</td>
<td>.27</td>
<td>3</td>
<td>.49</td>
<td>.90</td>
<td>-.54</td>
</tr>
<tr>
<td>2</td>
<td>-.04</td>
<td>.55</td>
<td>3</td>
<td>.49</td>
<td>.90</td>
<td>-.54</td>
</tr>
</tbody>
</table>

This Table contrasts, for each person class, the size and significance of the Differential Person Functioning for pairs of item classifications.

**Table 33.4** - These can be plotted directly from the Plots menu.

This shows the local size of the DP for each item class on each person class. The reported t value is approximate, and should be interpreted conservatively.

**252. Table 34 Columnar statistical comparison and scatterplot**

To automatically produce this Excel scatterplot of two sets of measures or fits statistics:
Select **Compare Statistics** on the Plots pull-down menu. If this is too big for your screen see **Display too big**.

---

234
Measures, standard errors, fit statistics indicate which statistic (or which field of the IFILE= or PFILE=) is to be the basis of the comparison.

Display with columns generates a the line-printer graphical-columns plot. It is displayed as Table 34. The first column is the Outfit Mean-Square of this analysis. The third column is the Outfit Mean-Square of the Right File (exam12lopf.txt in this case). The second column is the difference. The fourth column is the identification, according to the current analysis. Persons or items are matched and listed by Entry number.

+-----------------------------------------------------------------------------+
| PERSON | Outfit MnSq Difference | exam12lopf.txt | File Compa |
|-------------------+-------------------------+-------------------+-----------|
|      .   *        |      *     .            |   *  .            |  1  21101 |
|      .       *    |         *  .            |      .     *      |  2  21170 |
|  *   .            |            .*           |   *  .            |  3  21174 |
|     *.            |            *            |     *.            | 35  22693 |
+-----------------------------------------------------------------------------+

Display with Excel scatterplot initiates a graphical scatterplot plot. If the statistics being compared are both measures, then a 95% confidence interval is shown. This plot can be edited with all Excel tools.

253. Table 0 The title page

This page contains the authorship and version information. It appears in the Report Output File, but not on the Output Table menu.

WINSTEPS is updated frequently. Please refer to the web page for current version numbers and recent enhancements at www.winsteps.com
254. **Table 0.1 Analysis identification**

This shows the Title, Control file, Output report file, date and time of this analysis.

If the output file name has the form ZOU???ws.txt, then it is a temporary file which will be erased when Winsteps terminates. You can open this file from the “Edit” menu and save it as a permanent file.

```
TABLE 0.1 LIKING FOR SCIENCE (Wright & Masters p. ZOU042ws.txt Oct 9 9:00 2002

TITLE= LIKING FOR SCIENCE (Wright & Masters p.18)
CONTROL FILE: C:\WINSTEPS\sf.txt
OUTPUT FILE: C:\WINSTEPS\ZOU042ws.txt
DATE: Oct 9 9:00 2002
ACT DELETIONS: 2-3 10-15 17-19 21 24
76 PUPIL Records Input.
```

255. **Table 0.2 Convergence report**

(controlled by LCONV=, RCONV=, CONVERGE=, MPROX=, MJMLE=, CUTLO=, CUTHI=)

```
TABLE 0.2 LIKING FOR SCIENCE (Wright & Masters p. ZOU042ws.txt Oct 9 9:00 2002
INPUT: 76 PUPILS, 25 ACTS  WINSTEPS 3.36

CONVERGENCE TABLE
+----------------------------------------------------------------------------+
<p>| PROX   ACTIVE COUNT | EXTREME 5 RANGE | MAX LOGIT CHANGE |</p>
<table>
<thead>
<tr>
<th>ITERATION PUPILS ACTS CATS PUPILS ACTS MEASURES STRUCTURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1       76      25     3  3.59    1.62        3.1355    -.1229</td>
</tr>
<tr>
<td>2       74      12     3  4.03    1.90         .3862    -.5328</td>
</tr>
<tr>
<td>3       74      12     3  4.19    1.96         .1356    -.0783</td>
</tr>
</tbody>
</table>
WARNING: DATA MAY BE AMBIGUOUSLY CONNECTED INTO 6 SUBSETS, see Connection Ambiguities
+----------------------------------------------------------------------------+
<p>| JMLE     MAX SCORE | MAX LOGIT | LEAST CONVERGED | CATEGORY STRUCTURE |</p>
<table>
<thead>
<tr>
<th>ITERATION RESIDUAL* CHANGE PUPIL ACT CAT RESIDUAL CHANGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1        -2.04       .2562       7  5*      2       -.72    .0003</td>
</tr>
</tbody>
</table>
Standardized Residuals N(0,1)  Mean: .03 S.D.: 1.24
Look for scores and residuals in last line to be close to 0, and standardized residuals to be close to mean 0.0, S.D. 1.0.

The meanings of the columns are:

PROX       normal approximation algorithm - for quick initial estimates
ITERATION  number of times through your data to calculate estimates
ACTIVE COUNT number of elements participating in the estimation process after elimination of deletions and perfect/zero scores
    PERSONS person parameters
    ITEMS item parameters
    CATS rating scale categories - shows 2 for dichotomies
EXTREME 5 RANGE
    PERSONS The current estimate of the spread between the average measure of the top 5 persons and the average measure of the bottom 5 persons.
    ITEMS The current estimate of the spread between the average measure of the top 5 items and the average measure of the bottom 5 items.
MAX LOGIT CHANGE
    MEASURES maximum logit change in any person or item estimate. This is expected to decrease gradually until convergence, i.e., less than LCONV=.
    STRUCTURE maximum logit change in any structure measure estimate - for your information - need not be
as small as MEASURES.

**JMILE**  JMLE joint maximum likelihood estimation - for precise estimates
**ITERATION** number of times through your data to calculate estimates
It is unusual for more than 100 iterations to be required
**MAX SCORE RESIDUAL** maximum score residual (difference between integral observed score and decimal expected score) for any person or item estimate - used to compare with RCONV=. This number is expected to decrease gradually until convergence acceptable.
* indicates to which person or item the residual applies.

**MAX LOGIT CHANGE** maximum logit change in any person or item estimate - used to compare with LCONV=. This number is expected to decrease gradually until convergence is acceptable.

**LEAST CONVERGED** element numbers are reported for the person, item and category farthest from meeting the convergence criteria.
* indicates whether the person or the item is farthest from convergence.
the CAT (category) may not be related to the ITEM to its left. See Table 3.2 for details of unconverged categories.

**CATEGORY RESIDUAL** maximum count residual (difference between integral observed count and decimal expected count) for any response structure category - for your information. This number is expected to decrease gradually. Values less than 0.5 have no substantive meaning.

**STRUCTURE CHANGE** maximum logit change in any structure calibration. Not used to decide convergence, but only for your information. This number is expected to decrease gradually.

Standardized Residuals These are modeled to have a unit normal distribution. Gross departures from mean of 0.0 and standard deviation of 1.0 indicate that the data do not conform to the basic Rasch model specification that randomness in the data be normally distributed.

### Table 0.3 Control file

This Table shows the control file used in the analysis. It includes the extra specifications and expands SPFILE= commands. This is also appended to the LOGFILE=, when specified.

A complete listing of control variables is obtained using "Control variable file=" from the Output Files pull-down menu, which is also Table 32.

"Extra Specifications" are listed after &END.

```
TABLE 0.3 LIKING FOR SCIENCE (Wright & Masters p. ZOU042ws.txt Oct 9 9:00 2002
INPUT: 76 PUPILS, 25 ACTS MEASURED: 75 PUPILS, 12 ACTS, 3 CATS WINSTEPS 3.36
--------------------------------------------------------------------------------

&INST
TITLE='LIKING FOR SCIENCE +
   +(Wright & Masters p.18)' ;demonstrates continuation line
ITEMS=ACT
PERSONS=PUPIL
ITEM1=1
NI=25
NAMLMP=20
XWIDE=2
NAME1=51
isubtot=$s1W1
psubtot=$s2W1
pweight=$s2w1
iweight=$s3w1
dif=$s3W1
dpf=$s4W1
CODES=000102
```
Select by clicking on "Probability Cat. Curves" or from the Graphs menu. If you don't see all this on your screen, you may have your screen resolution set to 800 x 600 pixels. Try setting it to 1024 x 768. Windows "Start", "Settings", "Control Panel", "Display", "Settings", Move "Screen resolution" slider to the right.

**Traceline:** You can identify an individual traceline by single-left-clicking on it. Its description will then appear below the plot. Click elsewhere on the plot to remove the selection indicators. You can remove a traceline by double-left-clicking on it. Click on the command button, e.g., "Probability Curves", to return the plot to its initial appearance.

**Adjust** or fine-tune minimum or maximum enables you to change the x- and y-axis ranges. **Adjust number** enables you to change the number of x- and y- tick marks and labels.
**Copy Plot to Clipboard** places this screen on the clipboard. Open a graphics program, such as *Paint*, and paste in the image for editing. This only copies the part of the plot that is visible on your screen. Maximize the chart window and increase your screen resolution if the entire plot is not visible. To increase screen resolution: Windows "Start", "Settings", "Control Panel", "Display", "Settings" and move the "Screen resolution" slider to the right.

**Copy Plot Data to Clipboard** places the plotted data on the clipboard. Use paste special to paste as a picture meta-file, bitmap or as a text listing of the data points.

**Next Curve** takes you to the curves for the next grouping.

**Previous Curve** takes you to the curves for the previous grouping.

**Click for Relative (Absolute) x-axis** plots relative to difficulty of the current item or relative to the latent trait.

**Select Curves** enables you to jump to the set of curves you want to see by clicking on the list that is displayed.

**Background color** changes the background color behind the plot, or the color of the selected line on the plot.

Click on **Close Box** in upper right corner to close.

This shows the probability of observing each ordered category according to the Rasch model. To identify a category, click on it:

![Diagram](image)

The caption can be clicked on and moved. "2" is the category score. "25% independent" is the category description from CFILE= or CLFILE=.

To delete the line corresponding to a category, double-click on it:

![Diagram](image)

For individual items, the horizontal scaling can be changed from relative to item difficulty to relative to the latent variable by clicking on "Click for Absolute x-axis":

![Diagram](image)
258. Empirical category curves

These are the empirical (data-describing) category curves. They are obtained by clicking on "Empirical Cat. Curves" or from the Graphs menu. The width of each empirical interval can be adjusted by the "Empirical Interval" control. The smoothing of the empirical curves by means of cubic splines is adjusted by the smoothing control.

259. Expected score ICC

Select by clicking on "Expected Score ICC" or from the Graphs menu. Expected Score ICC plots the model-expected item characteristic curve. This shows the Rasch-model prediction for each measure relative to item difficulty. Its shape is always ascending monotonic. The dashed lines indicate the Rasch-half-point thresholds correspond to expected values of .5 score points. The intervals on the x-axis demarcated by dashed lines are the zones within which the expected score rounds to each observed category. To remove the dashed lines, double-click on them.

260. Empirical ICC

Select by clicking on "Empirical ICC" or from the Graphs menu. This shows the empirical (data-descriptive) item characteristic curve. Each black "x" represents observations in an interval on the latent variable. The "x" is positioned at the average rating (y-axis) at the average measure (x-axis) for observations close by. "Close by" is set by the empirical slider beneath the plot. The blue lines are merely to aid the eye discern the trend. The curve can be smoothed with the "smoothing" slider.
This shows the joint display of the expected and empirical ICCs. The boundary lines indicate the upper and lower 95% two-sided confidence intervals (interpreted vertically). When an empirical point lies outside of the boundaries, then some unmodeled source of variance maybe present in the observations. **Double-click** on a line on this plot to remove it from the display. The solid red "model" line is generated by the relevant Rasch model. For a test of dichotomous items, these red curves will be the same for every item. The empirical blue line is constructed from the observed frequencies along the variable, marked by x. The empirical ("x") x- and y-coordinates are the means of the measures and ratings for observations in the interval. The upper green line (and the lower grey line) are at 1.96 model standard errors above (and below) the model "red line", i.e., form a two-sided 95% confidence band. The distance of these lines from the red line is determined by the number of observations in the interval, not by their fit.

261. **Empirical randomness**

Select by clicking on "Empirical Randomness" or from the **Graphs** menu. Empirical intervals are set with the "Empirical Interval" slider. This displays the local value of the mean-square statistics. The Outfit mean-square statistic (standardized residual chi-square divided by its degrees of freedom) is the red line. The Infit mean-square statistic (ratio of observed to expected residual variance) is the blue line.

262. **Cumulative probabilities**
Select by clicking on "Cumulative Probabilities" or from the **Graphs** menu. This shows the modeled category probability curves accumulated so that the left-hand curve (red) is the probability of being observed in the lowest category. The next curve (blue) is the probability of being observed in the lowest or next lowest category. And so on to the right. The points of intersection between these curves and the 0.5 probability line are the Rasch-Thurstone thresholds. The points at which being observed in this category (or below) and the category abover (or higher) are equal. These curves are always in the order of the category scores.

**263. Item information function**

Select by clicking on "Item Information" or from the **Graphs** menu. This shows the (Ronald A.) Fisher information in responses made to items. It is the same as the binomial variance (dichotomies) or the multinomial variance (polytomies).

**264. Category information function**

Select by clicking on "Category Information" or from the **Graphs** menu. This shows the item information partitioned according to the probability of observing each category.

**265. Conditional probability curves**

Select by clicking on "Conditional Probabilities" or from the **Graphs** menu. Conditional probabilities of observing
adjacent categories. These are a series of Rasch dichotomous ogives. The intercepts with 0.5 probability are the Rasch-Andrich thresholds. They can be disordered relative to the latent variable.

266. Test characteristic curve

Select by clicking on "Test CC" or from the Graphs menu. This is the test characteristic curve, the score-to-measure ogive for this set of items. It is always monotonic ascending. See also Table 20.

267. Test information function

Select by clicking on "Test Information" or from the Graphs menu. This shows the Fisher information for the test (set of items) on each point along the latent variable. The information is the inverse-square of the person measure at that location of the variable. See also Table 20.

268. Test randomness

Select by clicking on "Test Randomness" or from the Graphs menu. Empirical intervals are set with the "Empirical Interval" slider. This displays the local value of the mean-square statistics. The Outfit mean-square statistic (standardized residual chi-square divided by its degrees of freedom) is the red line. The Infit mean-square statistic (ratio of observed to expected residual variance) is the blue line.
Select by clicking on "Multiple Item ICCs" or from the **Graphs** menu. This enables the display of multiple model and empirical ICCs on the same graph. Click on the "Model" and "Empirical" curves you wish to display. Click again on your selection to clear it.

Displayed are the selected model and empirical ICCs.

**270. Compare statistics**

From the **Plots** menu, this enables the simple graphical or tabular comparison of equivalent statistics from two runs. External files must be in **IFILE=** or **PFILE=** format.
There are several decisions to make:
1. Do you want to plot person (row) or item (column) statistics?
2. Which statistic for the x-axis?
3. Which statistic for the y-axis?
4. Do you want to use the statistic from this analysis or from the PFILE= or IFILE= of another analysis?
5. Do you want to display in the statistics as Columns in a Table or as an Excel scatterplot or both?

If you are using the statistic from a PFILE= or IFILE= and Winsteps selects the wrong column, then identify the correct column using the "Statistic field number" area.

When two measures are compared, then their standard errors are used to construct confidence bands:

Here the item calibrations in the current analysis are being compared with the item calibrations in file IFILE=SFIF.txt from another analysis. This is the columnar output:

```
<table>
<thead>
<tr>
<th>Measures</th>
<th>Differences</th>
<th>Measures</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-2</td>
<td></td>
<td>SFIF.txt</td>
</tr>
<tr>
<td></td>
<td>3-4</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>n101</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Month</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>n102</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sign</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>n103</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Phone number</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>n104</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ticket</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>n105</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>building</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>nm01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>student ticket</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>nm02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>menu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>nm03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>sweater</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>nm04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Forbidden City</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>nm05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>public place</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>nm06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>post office</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>nm07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>sign on wall</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>nh01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>supermarket</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>nh02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>advertisement</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>nh03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>vending machine</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>nh04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>outside store</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>nh05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>stairway</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>nh06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>gas station</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>nh07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Taipei</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>nh08</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>window at post office</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>nh09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>weather forecast</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>nh10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>section of newspaper</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>nh11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>exchange rate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>li01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>open</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>li02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>vending machine</td>
</tr>
</tbody>
</table>
```

and the plotted output:

```
Plot data-point label...

How are the plotted datapoints to be labeled?

<table>
<thead>
<tr>
<th>Marker</th>
<th>Entry number</th>
<th>Help</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>Entry+Label</td>
<td>Cancel</td>
</tr>
</tbody>
</table>

Only part of the label?

$s1w4$
```

We are selecting only the first 4 characters of the item label, e.g., "nl01" and plotting only the Label:
The points are plotted by their labels. The squiggly lines are the confidence bands. They are not straight because the standard errors of the points differ greatly. You can use the Excel functions to draw in straight lines by eye and remove the plotted curved lines. In this plot, the dotted line is the empirical equivalence (best-fit) line. The empirical identity line is shown in another plot by selecting the tab on the bottom of the Excel screen (green arrow). You can edit the data points by selecting the tab labeled “Z...” (red arrow).

271. **Bubble charts**

From the **Plots** menu, Bubble charts show measures and fit values graphically. They are featured in Bond & Fox. For successful operation, **Excel must be available on your computer**.

To produce these charts, Winsteps writes out the requisite values into a temporary file. Excel is then launched automatically. It reads in the temporary file and follows instructions to create a bubble plot. The plot is displayed, and the Excel worksheet becomes available for any alterations that are desired. The Excel worksheet may be made permanent using “Save As”.

Selecting "Bubble Chart" from the **Plots** menu:

Here is the vertical-measure bubble-plot for the KCT data as it appears on your computer screen. Bubbles are sized by their **standard errors**.
Or when it is formatted for printing portrait-orientation on legal-size paper, using Excel printer set-up:

And the horizontal-measure bubble-plot for the KCT data as it appears on your computer screen:

**Bubble Resizing:** The relative bubble sizes are set by the standard errors of the measures. But their overall standard size is set by Excel. To change their overall sizes, right click on the edge of a circle. Then use Excel's "Format Data Series", "Options", "Scale bubble size to:"
272. **Keyform plot**

From the **Plots menu**, Keyforms are self-measuring and diagnosing forms, such as the KeyMath diagnostic profile, and the KeyFIM. These are a powerful application of Rasch measurement for instantaneous use. These are plotted in horizontal or vertical format using Excel - but be patient, Excel is somewhat slow to display them. The keyform can be plotted with either horizontal or vertical orientation. In earlier version of Winsteps, these were specified by **KEYFORM**=

The 7 columns in the Excel Worksheet are:

For points in the KeyForm:

<table>
<thead>
<tr>
<th>COLUMN</th>
<th>The horizontal location (x-axis) in the vertical layout or vertical location (y-axis) in the horizontal layout.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEASURE</td>
<td>The measure (y-axis) in the vertical layout or (x-axis) in the horizontal layout.</td>
</tr>
<tr>
<td>POINT-LABEL</td>
<td>The value with which to label the point. Use the Excel addin at <a href="http://www.winsteps.com/ministep.htm">www.winsteps.com/ministep.htm</a></td>
</tr>
</tbody>
</table>

For column (row headings)

<table>
<thead>
<tr>
<th>COLUMN</th>
<th>The horizontal location (x-axis) in the vertical layout or vertical location (y-axis) in the horizontal layout.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEAD-MEASURE</td>
<td>The top-of-column measure (y-axis) in the vertical layout or end-of-row (x-axis) in the horizontal layout.</td>
</tr>
<tr>
<td>ITEM-ID</td>
<td>The item number</td>
</tr>
<tr>
<td>ITEM-LABEL</td>
<td>The item identifying label</td>
</tr>
</tbody>
</table>

Example: For the first 3 items of the "Liking For Science" Data

<table>
<thead>
<tr>
<th>COLUMN</th>
<th>MEASURE</th>
<th>POINT-LABEL</th>
<th>ITEM-ID</th>
<th>ITEM-LABEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-3.87</td>
<td>0</td>
<td>1</td>
<td>WATCH BIRDS</td>
</tr>
<tr>
<td>1</td>
<td>-2.79</td>
<td>-</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-1.42</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.94</td>
<td>-</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3.02</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-4.52</td>
<td>0</td>
<td>2</td>
<td>READ BOOKS ON ANIMALS</td>
</tr>
<tr>
<td>2</td>
<td>-3.44</td>
<td>-</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-1.08</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.29</td>
<td>-</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2.37</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-1.94</td>
<td>0</td>
<td>3</td>
<td>READ BOOKS ON PLANTS</td>
</tr>
<tr>
<td>3</td>
<td>-0.86</td>
<td>-</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.50</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3.87</td>
<td>-</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4.95</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-4.93</td>
<td>0</td>
<td>5</td>
<td>Raw Score</td>
</tr>
<tr>
<td>5</td>
<td>-3.28</td>
<td>1</td>
<td>5</td>
<td>Raw Score</td>
</tr>
<tr>
<td>5</td>
<td>-1.64</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.04</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.59</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.23</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5.03</td>
<td>6</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>6.00</td>
<td>Measure</td>
<td>7</td>
<td>Measure</td>
</tr>
<tr>
<td>9</td>
<td>6.00</td>
<td>S.E.</td>
<td>9</td>
<td>S.E.</td>
</tr>
</tbody>
</table>

| 7 | -5 | -5 |
| 9 | -5 | 1.97 |
| 7 | -4 | -4 |
| 9 | -4 | 1.32 |
| 7 | -3 | -3 |
| 9 | -3 | 1.32 |
| 7 | -2 | -2 |
| 9 | -2 | 1.29 |
| 7 | -1 | -1 |
| 9 | -1 | 1.29 |
with Excel, produces vertical plots like: (These can be plotted directly from the Plots menu.)

<table>
<thead>
<tr>
<th>WATCH BIRDS</th>
<th>READ BOOKS ON ANIMALS</th>
<th>READ BOOKS ON PLANTS</th>
<th>Raw Score</th>
<th>Measure</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6</td>
<td>5</td>
<td>2.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>3</td>
<td>1.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>1</td>
<td>4</td>
<td>1.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>3</td>
<td>1.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>0</td>
<td>-1</td>
<td>1.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>0</td>
<td>0</td>
<td>1.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>-</td>
<td>-6</td>
<td>1.97</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

with Excel, produces horizontal plots like: (These can be plotted directly from the Plots menu.)

273. **DIF Plot**

From the Plots menu, DIF Plot 30 produces the plotted output corresponding to Table 30. It shows differential item functioning (DIF).
In the DIF= box specify the column in the person label that identifies the DIF classification for each person.

Select what item identification is to be displayed on the Excel plot. In this case, the item entry number. The person classifications will be identified by their column codes.

There are three standard plots:
- **t-value** reports simple t-tests of the item DIF against the overall item difficulty
- **Relative Measure** reports the size of the item DIF relative to the overall item difficulty
- **Local Measure** reports the difficulty of the item for each person classification

### 274. DPF Plot

From the Plots menu, DPF Plot 31 produces the plotted output corresponding to Table 31. It shows differential person functioning (DPF).
This brings up

In the DPF= box specify the column in the item label that identifies the DPF classification for each person.

Select what person identification is to be displayed on the Excel plot. In this case, the person label. The item classifications will be identified by their column codes.

There are three standard plots:
t-value reports simple t-tests of the item DIF against the overall item difficulty
Relative Measure reports the size of the item DIF relative to the overall item difficulty
Local Measure reports the difficulty of the item for each person classification

275. DIF-DPF Plot

From the Plots menu, DIF-DPF Plot 33 produces the plotted output corresponding to Table 33. It shows differential functioning between classes of items and classes of persons, DIF & DPF, also called differential group functioning (DGF).

This brings up

In the upper DIF= box specify the column in the person label that identifies the DIF classification for each person.
In the lower DPF= box specify the column in the item label that identifies the DPF classification for each person.

There are four standard plots:
Relative Measure-ip reports the relative measures of the item classes as points and the person classes as columns
Relative Measure-pi reports the relative measure of the person classes as points and the item classes as columns
t-value-ip reports simple t-tests of the item class measures against the overall measures for each person class
t-value-pi reports simple t-tests of the person class measures against the overall measures for each item
276. Advice to novice analysts

In test construction, the rule is "all items must be about the same thing, but then be as different as possible"! The central idea is that there is a latent variable which we are attempting to measure people on. The empirical definition of the latent variable is the content of the items. Essentially, we should be able to summarize on the items into a sentence which matches our intended definition of the latent variable. Latent variables can be very broad, e.g., "psychological state" or "educational achievement", or very narrow, e.g., "degree of paranoia" or "ability to do long division".

In other words, all items share something in common, but each item also brings in something that the others don't have.

Of course, this never happens perfectly. So what we need is:

(a) all items to point in the same direction, so that a higher rating (or "correct" answer) on the item indicates more of the latent variable. The first entry on the Diagnosis menu displays correlations. Items with negative correlations probably need their scoring reversed with IVVALUE=.

(b) what the items share overwhelms what they don't share

(c) what the items don't all share, i.e., what is different about each of the items, is unique to each item or shared by only a few items.

What they all (or almost all) share, is usually thought of as the "test construct", the "major dimension", or the "Rasch dimension", or the "first factor in the data". This is what test validation studies focus on. Evaluating or confirming the nature of this construct.

What is unique to each item, or to clusters of only a few items, are "subdimensions", "secondary dimensions", "secondary contrasts", "misfit to the Rasch dimension", etc. We are concerned to evaluate: (i) are they a threat to scores/measures on the major dimension? (ii) do they manifest any useful information?

There are always as many contrasts in a test as there are items (less one). So how do we proceed?

(a) We want the first dimension to be much larger than all other dimensions, and for all items to have a large positive loading on it. This is essentially what the point-biserial correlation tells us in a rough way, and Rasch analysis tells us in a more exact way.

(b) We want so few items to load on each subdimension that we would not consider making that subdimension into a separate instrument. In practice, we would need at least 5 items to load heavily on a contrast, maybe more, before we would consider those items as a separate instrument. Then we crossplot and correlate scores or measures on the subdimension against scores on the rest of the test to see its impact.

(c) When a contrast has 2 items or less heavily loading on it - it may be interesting, but it is only a wrinkle in this test. For instance, when we look at a two item contrast, we may say, "That is interesting, we could make a test of items like these!" But to make that test, we would need to write new items and collect more data. Its impact on this test is obviously minimal.

In reporting your results, you would want to:

(a) Describe, and statistically support, what most items share: the test construct.

(b) Identify, describe and statistically support, sub-dimensions big enough to be split off as separate tests. Then contrast scores/measures on those subdimensions with scores/measures on the rest of the test.

(c) Identify smaller sub-dimensions and speculate as to whether they could form the basis of useful new tests.

In all this, statistical techniques, like Rasch analysis and factor analysis, support your thinking, but do not do your
thinking for you!

In what you are doing, I suggest that you choose the simplest analytical technique that enables you to tell your story, and certainly choose one you understand!

277. Anchored estimation

Anchoring or fixing parameter estimates (measures) is done with IAFILE= for items, PAFILE= for persons, and SAFILE= for response structures.

From the estimation perspective under JMLE, anchored and unanchored items appear exactly alike. The only difference is that anchored values are not changed at the end of each estimation iteration, but unanchored estimates are. JMLE converges when "observed raw score = expected raw score based on the estimates". For anchored values, this convergence criterion is never met, but the fit statistics etc. are computed and reported by Winsteps as though the anchor value is the "true" parameter value. Convergence of the overall analysis is based on the unanchored estimates.

Using pre-set "anchor" values to fix the measures of items (or persons) in order to equate the results of the current analysis to those of other analyses is a form of "common item" (or "common person") equating. Unlike common-item equating methods in which all datasets contribute to determining the difficulties of the linking items, the current anchored dataset has no influence on those values. Typically, the use of anchored items (or persons) does not require the computation of equating or linking constants. During an anchored analysis, the person measures are computed from the anchored item values. Those person measures are used to compute item difficulties for all non-anchored items. Then all non-anchored item and person measures are fine-tuned until the best possible overall set of measures is obtained. Discrepancies between the anchor values and the values that would have been estimated from the current data can be reported as displacements. The standard errors associated with the displacements can be used to compute approximate t-statistics to test the hypothesis that the displacements are merely due to measurement error.

278. Average measures, distractors and rating scales

The "average measure" for a category is the average ability of the people who respond in that category or to that distractor (or distracter. The term "distractor" has been in use since at least 1934, and was perhaps originated by Paul Horst in 1933). This is an empirical value. It is not a Rasch-model parameter.

The "step difficulty" (Rasch-Andrich threshold, step calibration, etc.) is an expression of the log-odds of being observed in one or other of the adjacent categories. This is a model-based value. It is a Rasch-model parameter.

Our theory is that people who respond in higher categories (or to the correct MCQ option) should have higher average measures. This is verified by "average measure".

Often there is also a theory about the rating scale, such as "each category in turn should be the most probable one to be observed as one advances along the latent variable." If this is your theory, then the "step difficulties" should also advance. But alternative theories can be employed. For instance, in order to increase item discrimination one may deliberately over-categorize a rating scale - visual-analog scales are an example of this. A typical visual analog-scale has 101 categories. If these functioned operationally according to the "most probable" theory, it would take something like 100 logits to get from one end of the scale to the other.

The relationship between "average measure" and "step difficulties" or "item difficulties" is complex. It is something like:

\[
\text{step difficulty} = \log \left( \frac{\text{count in lower category}}{\text{count in higher category}} \right) + \left( \text{average of the measures across both categories} \right) - \text{normalizer}
\]

normalized so that: \(\text{sum(step calibrations)} = 0\)

So that, the higher the frequency of the higher category relative to the lower category, the lower (more negative) the step calibration (and/or item difficulty)
and the higher the average of the person measures across both categories, the higher (more positive) the step calibration (and/or item difficulty)

but the step calibrations are estimated as a set, so that the numerical relationship between a pair of categories is influenced by their relationships with every other category. This has the useful consequence that even if a category is not observed, it is still possible to construct a set of step calibrations for the rating scale as a whole.

Rules of Thumb:
In general, this is what we like to see:
1. More than 10 observations per category (or the findings may be unstable, i.e., non-replicable)
2. A smooth distribution of category frequencies. The frequency distribution is not jagged. Jaggedness can indicate categories which are very narrow, perhaps category transitions have been defined to be categories. But this is sample-distribution-dependent.
3. Clearly advancing average measures.
4. Average measures near their expected values.
5. Observations fit of the observations with their categories: Outfit mean-squares near 1.0. Values much above 1.0 are much more problematic than values much below 1.0.

279. Automating file selection

Use the Winsteps "Batch" pull-down menu to do this.

Assigning similar names to similar disk files can be automated using Batch commands.

For example, suppose you want to analyze your data file, and always have your output file have suffix "O.TXT", the PFILE have suffix ".PF" and the IFILE have suffix ".IF". Key in your control file and data, say "ANAL1", omitting PFILE= and IFILE= control variables. Then key in the following Batch script file, called, say, MYBATCH.BAT (for Windows-95 etc.) or MYBATCH.CMD (for Windows-2000 etc.), using your word processor, saving the file as an ASCII or DOS text file:

REM the MYBATCH.BAT batch file to automate WINSTEPS
START /w ..\WINSTEPS BATCH=YES %1 %1O.TXT PFILE=%1.PF IFILE=%1.IF

For WINSTEPS, specify BATCH=YES to close all windows and terminate the program when analysis is complete.

To execute this, type at the DOS prompt (or using the Winsteps "Batch" pull-down menu:
C:> MYBATCH ANAL1 (Press Enter Key)
This outputs the tables in ANAL1O.TXT, PFILE= in ANAL1.PF and IFILE= in ANAL1.IF.

You can also edit the files WINBATCH.BAT or WINBATCH.CMD. These can be executed from the DOS prompt or from the Winsteps Batch pull-down menu. See Running Batch Files.

280. Batch mode example: Score Tables

Winsteps produces a Score Table in Table 20 which is for responses to every active item. You can use the Batch File feature to automate the production of Score Tables for subtests of items.

(1) Do the calibration run:
   title="Item calibration"
   ni=3          ; these match your data file
   item1=1
   name1=1
   codes=123
   ISGROUPS=0
   ifile= if.txt ; item calibrations
   sfile= sf.txt ; rating (or partial credit) scale strucure
   data=data.txt
   &end
(2) Set up the control file for the batch run:

```plaintext
; This is scon.txt
title="Produce score table"
ni=3               ; match your data file
item1=1
name1=1
codes=123
ISGROUPS=0
iafile = if.txt    ; item anchor file
safilen = sf.txt   ; rating (or partial credit) scale structure anchor file
pafilen = *        ; dummy person measures
1 0               ; only logit change is used for convergence
2 0

CONVERGE=L         ; only logit change is used for convergence
LCONV=0.005        ; logit change too small to appear on any report.
&end
```

(3) Set up the Batch (.bat or .cmd) file and run it. Use IDFILE=* to select the items you want (or ISELECT=)

```plaintext
rem this is score.cmd
del sc*.txt
start /w ..\winsteps sc.txt dummy SCOREFILE=sc001.txt batch=yes idfile=* +3 * title=item3
start /w ..\winsteps sc.txt dummy SCOREFILE=sc010.txt batch=yes idfile=* +2 * title=item2
start /w ..\winsteps sc.txt dummy SCOREFILE=sc100.txt batch=yes idfile=* +1 * title=item1
start /w ..\winsteps sc.txt dummy SCOREFILE=sc011.txt batch=yes idfile=* +2 +3 * title=items23
start /w ..\winsteps sc.txt dummy SCOREFILE=sc101.txt batch=yes idfile=* +1 +3 * title=item13
start /w ..\winsteps sc.txt dummy SCOREFILE=sc110.txt batch=yes idfile=* +1 +2 * title=items12
start /w ..\winsteps sc.txt dummy SCOREFILE=sc111.txt batch=yes idfile=* +1 +2 +3 * title=items123
.....
copy sc*.txt scores.txt
```

(4) The Score Tables are in file scores.txt

281. **Biserial correlation**

If the sample is normally distributed (i.e., conditions for the computation of the biserial exist), then to obtain the biserial correlation from the point-biserial:

Biserial = Point-biserial * f(P-value)

Example: Specify PTBISERIAL=Yes and PVALUE=Yes. Display Table 14.

<table>
<thead>
<tr>
<th>ENTRY</th>
<th>RAW</th>
<th>MODEL</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>PTBIS</th>
<th>P-</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NUMBER</td>
<td>SCORE</td>
<td>COUNT</td>
<td>MEASURE</td>
<td>S.E.</td>
<td>MNSQ</td>
<td>ZSTD</td>
<td>MNSQ</td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>-------</td>
<td>-------</td>
<td>------</td>
<td>------</td>
<td>-----</td>
<td>------</td>
</tr>
<tr>
<td>8</td>
<td>27</td>
<td>34</td>
<td>-2.35</td>
<td>.54</td>
<td>.59</td>
<td>-1.3</td>
<td>.43</td>
</tr>
</tbody>
</table>

Point-biserial = .65. P-value = .77. Then, from the Table below, f(P-value) = 1.3861, so Biserial correlation = .65 * 1.39 = .90

Here is the Table of p-value and f(p-value).

<table>
<thead>
<tr>
<th>p-va</th>
<th>f(p-val)</th>
<th>p-va</th>
<th>f(p-val)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.99</td>
<td>3.8335</td>
<td>0.01</td>
<td>3.7335</td>
</tr>
<tr>
<td>0.98</td>
<td>2.8914</td>
<td>0.02</td>
<td>2.8914</td>
</tr>
</tbody>
</table>

256
282. Category boundaries and thresholds

Conceptualizing rating scales and partial-credit response structures for communication can be challenging. Rasch measurement provides several approaches. Choose the one that is most meaningful for you.

Look at this excerpt of Table 3.2:

<table>
<thead>
<tr>
<th>DATA</th>
<th>QUALITY CONTROL</th>
<th>STEP</th>
<th>EXPECTATION</th>
<th>.5 Cumul.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category Counts</td>
<td>Cum.</td>
<td>Avge</td>
<td>Exp.</td>
<td>OUTFIT</td>
</tr>
<tr>
<td>Score</td>
<td>Used</td>
<td>%</td>
<td>%</td>
<td>Meas</td>
</tr>
<tr>
<td>0 891 6% 6%</td>
<td>-.04</td>
<td>-.07</td>
<td>1.3</td>
<td>( -1.12)</td>
</tr>
<tr>
<td>1 383 7% 9%</td>
<td>-.47</td>
<td>-.33</td>
<td>-.74</td>
<td>-.25</td>
</tr>
<tr>
<td>2 1017 7% 15%</td>
<td>1.07</td>
<td>1.17</td>
<td>.8</td>
<td>-.15</td>
</tr>
<tr>
<td>3 12638 85% 100%</td>
<td>2.15</td>
<td>2.14</td>
<td>1.0</td>
<td>-.91</td>
</tr>
</tbody>
</table>

Here are three ways of conceptualizing and communicating the transition, threshold, boundary between category 1 and category 2:

(1) Rasch-half-point thresholds. Someone at the boundary between "1" and "2" would have an expected rating of
1.5, or 1000 persons at the boundary between "1" and "2" would have an average rating of 1.5. This boundary is the "EXPECTATION Measure at 2 -0.5" which is -0.01 logits, the Rasch-half-point threshold. To illustrate this, use the model item characteristic curve. The expected score ogive / model ICC (Table 21.2 - second on list in Graphs menu). The CAT+.25, CAT-0.5, AT CAT, and CAT-25 columns in the ISFILE= plot points on this ogive. The expected score ogive relates most directly to the estimation of the Rasch parameters. Since it is only one line, it is also convenient for summarizing performance at any point on the latent variable by one number. Crucial points are the points on the variable corresponding to the lower category value + 0.5, i.e., more than the higher adjacent category value - 0.5. These Rasch-half-point thresholds are "average score thresholds" or "Rasch-ICC thresholds".

(2) Rasch-Thurstone thresholds. Someone at the boundary between "1" and "2" would have a 50% chance of being rated 1 or below, and a 50% chance of being rated 2 or above. This is the Rasch-Thurstone threshold of -0.02 logits. To illustrate this, use the cumulative probability curves. The cumulative probability curves (Table 21.3 - and third on list in Graphs menu). The 50%PRB columns in the ISFILE= are the crucial points on these curves, and are the Rasch-Thurstone thresholds, useful for identifying whether a person is most likely to respond below, at or above a certain category.

(3) Rasch-Andrich thresholds. Someone at the boundary between "1" and "2" would have an equal chance of being rated 1 or 2. This is the Rasch-Step Calibration (Rasch-Andrich Threshold) of -0.15 logits. To illustrate this, use the category probability curves. The probability curves (Table 21.1 - and top of list in Graphs menu). The Structure MEASURE in the ISFILE= gives the point of equal probability between adjacent categories. The points of highest probability of intermediate categories are given by the AT CAT values. These probability curves relate most directly to the Rasch parameter values, also called Rasch-Andrich thresholds. They are at the intersection of adjacent probability curves, and indicate when the probability of being observed in the higher category starts to exceed that of being observed in the adjacent lower one. This considers the categories two at a time, but can lead to misinference if there is Rasch-Andrich threshold disordered.

d) Empirical average measures. For any particular sample, there is the average ability of the people who scored in any particular category of any particular item. This is the "Average Measure" reported in Table 3.2. This is entirely sample-dependent. It is not reported in ISFILE=

### 283. Column and classification selection and transformation

**Selection:** Several specifications require or allow the choice of columns in the item label, person label or data record. They include DIF=, DPF=, IAFILE=, IMAP=, IPMATRIX=, ISORT=, ISUBTOT=, IWEIGHT=, PAFILE=, PMAP=, PSORT=, PSUBTOT=, PWEIGHT=. There are several formats:

- **Specification = C number or $C number (can be followed by - number, E number, W number)**
  - selects one or a block of columns in the person data record (or item label)
  - **PWEIGHT = $C203W10** the person weighting is in column 203 of the data record, with a width of 10 columns.
  - This always works if the columns are within the person or item label. If the columns referenced are in the data record, but outside the person label, the information may not be available.
  - **Specification = number or $S number or S number**
  - selects one column or a block of columns in the person or item label
  - **DIF=3** the column containing the DIF classification identified is the third column in the person label
  - **PMAP=SS5** print on the person map, Table 16, the character in column 5 of the person label

- **Specification = number - number or number E number (also commencing with $S or S)**
selects a block of columns from the person or item label
ISORT = 3 - 7  sort the items in Table 15 by the contents of columns 3 to 7 of the item labels.

Specification = number W number (also commencing with $S or S)
selects a block of columns from the person or item label
PSUBTOTAL = 12W3  subtotal the persons using classifiers starting column 12 with a width of 3 columns of the person label.

Specification = @Fieldname
selects one or a block of columns as specified by a prior @Fieldname= instruction
@AGEGROUPED = 2-3; the age group-classification is in columns 2-3 of the person label
PSUBTOTAL = @AGEGROUPED  ; subtotal the persons by age classifier

Specification = measure or fit stratification, e.g., MA3
M = Measure
A = Ascending or D = Descending
1 or higher integer: number of strata
e.g., MA3 = Measures Ascending in 3 ability strata

F = Fit
I = Infit, O = Outfit
M = Mean-square, L = Log-mean-square, S = t standardized
A = Ascending or D = Descending
1 or higher integer: number of strata
e.g., FILD2 = Fit - Infit - Log-scaled - Descending -2 fit strata

Stratum for this value = Max(1+ Number of strata * (Value - Min)/(Max - Min), Number of strata)

Specification = selection + selection
+ signs can be used to concatenate selections
IMAP = 3-4 + 8W3  show in Table 12 columns 3, 4, 8, 9, 10 of the item label.

Specification = "..." or '...
constants in quotes can be included with selected columns.
PMAP = 3 + "/" + 6  show in Table 15 column 3, a/ and column 6 of the person label, e.g., F/3

Some specifications have several options, e.g.,
IAFILE = value can be one of the following:
IAFILE = 3-7  the item anchor values are in columns 3-7 of the item label
IAFILE = *  the item anchor values follow this specification in a list
IAFILE = file name  the item anchor values are in the file called "file name"

IAFILE=, PAFILE=, etc., value is first checked for "**". If not this, it is parsed as a selection. If not a valid selection, it is parsed as a file name. Consequently, a mis-specified value may produce a "file not found" error.

Transformation: Selected codes can be transformed into other codes (of the same or fewer characters) for Tables 27, 28, 30, 31, 33:

First produce the Table with the transformation box blank. Inspect the reported codes. Transform and combine them using the transformation box in a second run. In this example, codes 1,A,B,C,D,E,F,S will all be converted to 1. 2,G,H,I,J,K,L,M,N,O,P,Q,R will all be converted to 2. T,U,V,W will all be converted to T. Codes X,Y,Z and any others will be unchanged.
In each line in the transformation box, the code at the beginning (extreme left) of each line is the code into which it and all other blank-separated codes on the line are transformed. Ranges of codes are indicated by -. To specify a blank or hyphen as a code, place them in quotes: " " and "-". Codes are matched to the transformation box starting at the top left and line by line until a match is found, or there is no match.

284. Comparing estimates with other Rasch software

There are many Rasch-specific software packages and IRT packages which can be configured for Rasch models. Each implements particular estimation approaches and other assumptions or specifications about the estimates. Comparing or combining measures across packages can be awkward. There are three main considerations:

(a) choice of origin or zero-point
(b) choice of user-scaling multiplier.
(c) handling of extreme (zero and perfect) scores.

Here is one approach:

Produce person measures from Winsteps and the other computer program on the same data set. For Winsteps set USCALE=1 and UIMEAN=0.

Cross-plot the person measures with the Winsteps estimates on the x-axis. (This is preferential to comparing on item estimates, because these are more parameterization-dependent.)

Draw a best-fit line through the measures, ignoring the measures for extreme scores.

The slope is the user-scaling multiplier to apply. You can do this with USCALE= slope.

The intercept is the correction for origin to apply when comparing measures. You can do this with UIMEAN= y-axis intercept.

The departure of extreme scores from the best-fit line requires adjustment. You can do this with EXTRSCORE=.

This may take multiple runs of Winsteps. If the measures for perfect scores are above the best-fit line, and those for zero scores are below, then decrease EXTRSCORE= in 0.1 increments or less. If vice-versa, then increase EXTRSCORE= in 0.1 increments or less.

With suitable choices of UIMEAN=, USCALE= and EXTRSCORE=, the crossplotted person measures should approximate the identity line.

The item estimates are now as equivalent as they can be even if, due to different choice of parameterization or estimation procedure, they appear very different.

You may notice scatter of the person measures around the identity line or obvious curvature. These could reflect differential weighting of the items in a response string, the imposition of prior distributions, the choice of approximation to the logistic function, the choice of parameterization of the Rasch model or other reasons. These are generally specific to each software program and become an additional source of error when comparing measures.

285. Connection ambiguities

Winsteps attempts to estimate an individual measure for each person and item within one frame of reference. Usually this happens. But there are exceptions. The data may not be "well-conditioned" (Fischer G.H., Molenaar, I.W. (eds.) (1995) Rasch models: foundations, recent developments, and applications. New York: Springer-Verlag. p. 41-43).

Extreme scores (zero and perfect scores) imply measures that our beyond the current frame of reference. Winsteps uses Bayesian logic to provide measures corresponding to those scores.

More awkward situations are shown in this data set. It is Examsubs.txt

Title = "Example of subset reporting"
Name1 = 1
Item1 = 9
NI = 10
&End
Item 1 dropped as extreme perfect score
Item 2 in subset 2
There are 10 items. The first item is answered correctly by all who responded to it. So it is estimated as extreme and dropped from further analysis. Then the first person Alf responds incorrectly to all non-extreme items and is dropped.

After eliminating Item 1 and Alf, Subset 1: Item 6 has a Guttman pattern. It distinguishes between those who succeeded on it from those who failed, with no contradiction to that distinction in the data. So there is an unknown logit distance between those who succeeded on Item 6 and those who failed on it. Consequently the difficulty of Item 6 is uncertain.

The remaining subsets have measures that can be estimated within the subset, but have unknown distance from the persons and items in the other subsets.

Subset 2: Items 2, 3 and Ben, Carl.
Subset 3: Items 4, 5 and David, Edward.
Subset 4: Frank
Subset 5: Items 7, 8 and George, Henry
Subset 6: Items 9, 10 and Ivan, Jack

Under these circumstance, Winsteps reports one of an infinite number of possible solutions. Fit statistics and standard errors are usually correct. Reliability coefficients are accidental. Measure comparisons within subsets are correct. Across-subset measure comparisons are accidental.

A solution would be to anchor two equivalent items (or two equivalent persons) in the different subsets to the same values - or juggle the anchor values to make the mean of each subset the same (or whatever). Or else do separate analyses. Or construct a real or dummy data records which include 0 & 1 responses to all items.

This data set causes Winsteps to report, near the top of the iteration screen, and in the output file:

WARNING: DATA MAY BE AMBIGUOUSLY CONNECTED INTO 6 SUBSETS
SUBSET 1 OF 1 ITEMS includes ITEM 6: Item 6
SUBSET 2 OF 2 PERSONS includes ITEM 2 and PERSON 2: Ben
SUBSET 3 OF 2 PERSONS includes ITEM 4 and PERSON 4: David
SUBSET 4 OF 1 PERSONS includes PERSON 6: Frank
SUBSET 5 OF 2 PERSONS includes ITEM 7 and PERSON 7: George
SUBSET 6 OF 2 PERSONS includes ITEM 9 and PERSON 9: Ivan

Winsteps reports an example entry number and person name from each subset, so that you can compare their response strings.

286. Convergence considerations

For early runs, set the convergence criteria loosely, or use Ctrl+F to stop the iterations.

If in doubt, set the convergence criteria very tightly for your final report, e.g.,

CONVERGE=B ; both LCONV= and RCONV= apply
LCONV=.0001 ; largest logit change .0001 logits
RCONV=.01 ; largest score residual .01 score points
MJMLE= 0 ; unlimited JMLE iterations
and be prepared to stop the iterations manually, if necessary, using "Finish Iterating", Ctrl+F, on the File pull-down menu.

Remember that convergence criteria tighter than the reported standard error of a measure are somewhat of a numerical fiction.

The Rasch model is non-linear. This means that estimates cannot be obtained immediately and exactly, as can be done with the solution of simultaneous linear equations. Instead, estimates are obtained by means of a series of guesses at the correct answer.

The initial guess made by Winsteps is that all items are equally difficult and all persons equally able. The expected responses based on this guess are compared with the data. Some persons have performed better than expected, some worse. Some items are harder than expected, some easier. New guesses, i.e., estimates, are made of the person abilities and item difficulties, and also the rating (or partial credit) scale structures where relevant.

The data are again examined. This is an "iteration" through the data. Expectations are compared with observations, and revised estimates are computed.

This process continues until the change of the estimates is smaller than specified in LCONV=, or the biggest difference between the marginal "total" scores and the expected scores is smaller than specified in RCONV=. The precise stopping rule is controlled by CONVERGE=, When the estimates are good enough, the iterative process has "converged". Then iteration stops. Fit statistics are computed and the reporting process begins.

There are standard convergence criteria which are suitable for most small and medium-sized complete data sets. LCONV= is harder to satisfy for small complete data sets and many sparse data designs, RCONV= for large complete data sets.

Anchored analyses
Anchor values always misalign somewhat with the current dataset unless they are estimated from it. Thus, the maximum residuals can never reduce to zero. Convergence occurs when the maximum logit change is too small to be meaningful. Accordingly, RCONV= is unproductive and only LCONV= is useful. Suggested specifications are:

CONVERGE = L ; only LCONV is operative
LCONV = .005 ; smaller than visible in standard, two decimal, output.

Missing data
For some data designs much tighter criteria are required, particularly linked designs with large amounts of missing data. For instance, in a vertical equating design of 5 linked tests, standard convergence occurred after 85 iterations. Tightening the convergence criteria, i.e., using smaller values of LCONV= and RCONV=, convergence occurred after 400 iterations. Further iterations made no difference as the limit of mathematical precision in the computer for this data set had been reached. The plot shows a comparison of the item difficulties for the 5 linked tests estimated according to the standard and strict convergence criteria.

CONVERGE = B
LCONV = .001 ; 10 time stricter than usual
RCONV = .01 ; 10 times stricter than usual

Note that convergence may take many iterations, and may require manual intervention to occur: Ctrl+F.
287. Decimal and percentage data

Winsteps analyzes ordinal variables expressed as integers, cardinal numbers, in the range 0-254, i.e., 255 ordered categories.

**Percentage observations:**
Observations may be presented for Rasch analysis in the form of percentages in the range 0-100. These are straightforward computationally but are often awkward in other respects.

A typical specification is

```
XWIDE = 3
CODES = "  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19+
+ 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39+
+ 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59+
+ 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79+
+ 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99100"
STKEEP = Yes ; to keep intermediate unobserved categories
```

Since it is unlikely that all percentages will be observed, the rating (or partial credit) scale structure will be difficult to estimate. Since it is even more unlikely that there will be at least 10 observations of each percentage value, the structure will be unstable across similar datasets.

It is usually better from a measurement perspective (increased person "test" reliability, increased stability) to collapse percentages into shorter rating (or partial credit) scales, e.g., 0-10, using IREFER= and IVALUE= or NEWSCORE=.

**Decimal observations:**
When observations are reported in fractional or decimal form, e.g., 2.5 or 3.9, multiply them by suitable multipliers, e.g., 2 or 10, to bring them into exact integer form.

Specify STKEEP=NO, if the range of observed integer categories includes integers that cannot be observed.

**Continuous and percentage observations:**
These are of two forms:

(a) Very rarely, observations are already in the linear, continuous form of a Rasch variable. Since these are in the form of the measures produced by Winsteps, they can be compared and combined with Rasch measures using standard statistical techniques, in the same way that weight and height are analyzed.

(b) Observations are continuous or percentages, but they are not (or may not be) linear in the local Rasch context. Examples are "time to perform a task", "weight lifted with the left hand". Though time and weight are reported in linear units, e.g., seconds and grams, their implications in the specific context is unlikely to be linear. "Continuous" data are an illusion. All data are discrete at some level. A major difficulty with continuous data is determining the precision of the data for this application. This indicates how big a change in the observed data constitutes a meaningful difference. For instance, time measured to .001 seconds is statistically meaningless in the Le Mans 24-hour car race - even though it may decide the winner!

To analyze these forms of data, segment them into ranges of observably different values. Identify each segment with a category number, and analyze these categories as rating scales. It is best to start with a few, very wide segments. If these produce good fit, then narrow the segments until no more statistical improvement is evident. The general rule is: if the data analysis is successful when the data are stratified into a few levels, then it may be successful if the data are stratified into more levels. If the analysis is not successful at a few levels, then more levels will merely be more chaotic. Signs of increasing chaos are increasing misfit, categories "average measures" no longer advancing, and a reduction in the sample "test" reliability.

**May I suggest that you start by stratifying your data into 2 levels?** (You can use Excel to do this.) Then analyze the resulting 2 category data. Is a meaningful variable constructed? If the analysis is successful (e.g., average measures per category advance with reasonable fit and sample reliability), you could try stratifying into more levels.

Example: *My dataset contains negative numbers such as "-1.60", as well as positive numbers such as "2.43". The*
range of potential responses is -100.00 to +100.00.

Winsteps expects integer data, where each advancing integer indicates one qualitatively higher level of performance (or whatever) on the latent variable. The maximum number of levels is 0-254. There are numerous ways in which data can be recoded. On is to use Excel. Read your data file into Excel. Its "Text to columns" feature in the "Data" menu may be useful. Then apply a transformation to the responses, for instance, recoded response = integer ( (observed response - minimum response)*100 / (maximum response - minimum response) )

This yields integer data in the range 0-100, i.e., 101 levels. Set the Excel column width, and "Save As" the Excel file in ".prn" (formatted text) format. Or you can do the same thing in SAS or SPSS and then use the Winsteps SAS/SPSS menu.

288. Dependency and unidimensionality

Question: To calibrate item difficulty, I am using data from 75 subjects. Most of the subjects have been tested repeatedly, between two and 9 times each. The reason for this was that I assumed that by training and time (with natural development) the subjects ability was different between different testing situations. Now the referee has asked me to verify that "the requirement of local independence is not breached". How can I check this?

Unidimensionality can be violated in many different ways. If you run all known statistical tests to check for violations (even with your subjects tested only once), your data would undoubtedly fail some of them - (for technical details of some of these tests see Fischer & Molenaar, "Rasch Models", chapter 5.) Consequently, the question is not "are my data perfectly unidimensional" - because they aren't. The question becomes "Is the lack of unidimensionality in my data sufficiently large to threaten the validity of my results?"

Imagine that you accidentally entered all your data twice. Then you know there is a lack of local independence. What would happen? Here is what happened when I did this with the dataset exam12lo.txt:

Data in once:

```
SUMMARY OF 35 MEASURED PERSONS

+-----------------------------------------------------------------------------+
| RAW          RAW                          MODEL         INFIT        OUTFIT    |
| SCORE COUNT MEASURE ERROR MNSQ ZSTD MNSQ ZSTD |                                         |
|-----------------------------------------------------------------------------|
| MEAN 38.2 13.0 -.18 .32 1.01 -.1 1.02 .0 |                                         |
| S.D. 10.1 13.0 .99 .06 .56 1.4 .57 1.3 |                                         |
| MAX. 54.0 13.0 1.44 .59 2.36 2.9 2.28 2.5 |                                         |
| MIN. 16.0 13.0 -.92 .29 .23 -2.9 .24 -2.3 |                                         |
|-----------------------------------------------------------------------------|
| REAL RMSE .36 ADJ.SD .92 SEPARATION 2.55 PERSON RELIABILITY .87 |                                         |
| MODEL RMSE .33 ADJ.SD .94 SEPARATION 2.85 PERSON RELIABILITY .89 |                                         |
| S.E. OF PERSON MEAN = .17 |                                         |
|-----------------------------------------------------------------------------|
| PERSON RAW SCORE-TO-MEASURE CORRELATION = .99 CRONBACH ALPHA (KR-20) PERSON RAW SCORE RELIABILITY = .89 |                                         |
|-----------------------------------------------------------------------------|
```

Data in twice:

```
SUMMARY OF 13 MEASURED ITEMS

+-----------------------------------------------------------------------------+
| RAW          RAW                          MODEL         INFIT        OUTFIT    |
| SCORE COUNT MEASURE ERROR MNSQ ZSTD MNSQ ZSTD |                                         |
|-----------------------------------------------------------------------------|
| MEAN 102.9 35.0 0.0 .20 1.08 -.2 1.02 -.2 |                                         |
| S.D. 23.6 35.0 .93 .03 .58 2.3 .53 2.0 |                                         |
| MAX. 145.0 35.0 2.45 .31 2.16 3.9 2.42 4.3 |                                         |
| MIN. 46.0 35.0 -1.65 .18 .31 -4.2 .39 -3.3 |                                         |
|-----------------------------------------------------------------------------|
| REAL RMSE .24 ADJ.SD .90 SEPARATION 3.81 ITEM RELIABILITY .94 |                                         |
| MODEL RMSE .20 ADJ.SD .91 SEPARATION 4.53 ITEM RELIABILITY .95 |                                         |
| S.E. OF ITEM MEAN = .27 |                                         |
|-----------------------------------------------------------------------------|
```

Imagine that you accidentally entered all your data twice. Then you know there is a lack of local independence.
<table>
<thead>
<tr>
<th>RAW</th>
<th>MODEL</th>
<th>INFIT</th>
<th>OUTFIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAW</td>
<td>SCORE</td>
<td>COUNT</td>
<td>MEASURE</td>
</tr>
<tr>
<td>-----</td>
<td>-------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>MEAN</td>
<td>38.2</td>
<td>13.0</td>
<td>-1.38</td>
</tr>
<tr>
<td>S.D.</td>
<td>10.1</td>
<td>.0</td>
<td>.99</td>
</tr>
<tr>
<td>MAX.</td>
<td>54.0</td>
<td>13.0</td>
<td>1.44</td>
</tr>
<tr>
<td>MIN.</td>
<td>16.0</td>
<td>13.0</td>
<td>-2.92</td>
</tr>
</tbody>
</table>

PERSON RAW SCORE-TO-MEASURE CORRELATION = .99
CRONBACH ALPHA (KR-20) PERSON RAW SCORE RELIABILITY = .89

REAL RMSE .36 ADJ.SD .92 SEPARATION 2.55 PERSON RELIABILITY .87
MODEL RMSE .33 ADJ.SD .94 SEPARATION 2.85 PERSON RELIABILITY .89
S.E. OF PERSON MEAN = .12

SERIAL RMSE .17 ADJ.SD .92 SEPARATION 5.48 ITEM RELIABILITY .97
MODEL RMSE .14 ADJ.SD .92 SEPARATION 6.48 ITEM RELIABILITY .98
S.E. OF ITEM MEAN = .27

There is almost no difference in the person report. The biggest impact the lack of local independence has in this situation is to make the item standard errors too small. Consequently you might report item results as statistically significant that aren't.

So, with your current data, you could adjust the size of the standard errors to their biggest "worst case" size:

Compute k = number of observations in your data / number of observations if each person had only been tested once

Adjusted standard error = reported standard error * sqrt (k).

This would also affect Reliability computations:
Adjusted separation = reported separation / sqrt(k)
Adjusted Reliability = Rel. / ( k + Rel. - Rel.*k) = Adj.Sep**2 / (1+Adj.Sep.*2)

The size of the mean-square fit statistics does not change, but you would also need to adjust the size of the t standardized fit statistics (if you use them). This is more complicated. It is probably easiest to read them off the plot from Rasch Measurement Transactions 17:1 shown below.

Look at your current mean-square and significance. Find the point on the plot. Go down to the x-axis. Divide the value there by k. Go to the same mean-square value contour. The "worst case" lower statistical significance value is on the y-axis.

Another noticeable aspect of your current data could be that there are misfitting subjects who were tested 9 times, while fitting persons are tested only twice. This would introduce a small distortion into the measurement system. So, arrange all the Tables in fit order, and look at each end, do some subjects appear numerous times near the end of a Table? If so, drop out those subjects and compare item calibrations with and without those subjects. If there is no meaningful difference, then those subjects are merely at the ends of the probabilistic range predicted by the Rasch model.
289. Dichotomous mean-square fit statistics

For a general introduction, see Diagnosing Misfit

<table>
<thead>
<tr>
<th>Responses:</th>
<th>Diagnosis</th>
<th>INFIT</th>
<th>OUTFIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy--Items--Hard Pattern</td>
<td>MnSq</td>
<td>MnSq</td>
<td></td>
</tr>
<tr>
<td>111</td>
<td>0110110100</td>
<td>000  Modelled/Ideal</td>
<td>1.1</td>
</tr>
<tr>
<td>000</td>
<td>0000111111</td>
<td>111  Miscode</td>
<td>4.3</td>
</tr>
<tr>
<td>011</td>
<td>1111100000</td>
<td>000  Carelessness/Sleeping</td>
<td>1.0</td>
</tr>
<tr>
<td>111</td>
<td>1111000000</td>
<td>001  Lucky Guessing</td>
<td>1.0</td>
</tr>
<tr>
<td>101</td>
<td>0101010101</td>
<td>010  Response set/Miskey</td>
<td>2.3</td>
</tr>
<tr>
<td>111</td>
<td>1000111110</td>
<td>000  Special knowledge</td>
<td>1.3</td>
</tr>
<tr>
<td>111</td>
<td>1111000000</td>
<td>000  Guttman/Deterministic</td>
<td>0.5</td>
</tr>
<tr>
<td>111</td>
<td>1010110010</td>
<td>000  Imputed outliers *</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Right|Transition|Wrong Expectation: 1.0

Overall pattern:
high - low - high OUTFIT sensitive to outlying observations
>>1.0 unexpected outliers
<<1.0 overly predictable outliers

low - high - low INFIT sensitive to pattern of inlying observations
>>1.0 disturbed pattern
<<1.0 Guttman pattern

* as when a tailored test is filled out by imputing all "right" response to easier items and all "wrong" to harder items.

The exact details of these computations have been lost, but the items appear to be uniformly distributed about 0.4 logits apart, extracted from Linacre, Wright (1994) Rasch Measurement Transactions 8:2 p. 360

The Z-score standardized Student's-t statistics report, as unit normal deviates, how likely it is to observe the reported mean-square values, when the data fit the model. The term Z-score is used of a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's t-distribution value has been adjusted to a unit normal value.

290. DIF - DPF - bias - interactions concepts

Computation
The DIF (differential item functioning) or DPF (differential person functioning) analysis proceeds with all items and persons, except the item or person currently targeted, anchored at the measures from the main analysis.
(estimated from all persons and items, including the currently targeted ones). The item or person measure for the current classification is then computed, along with its S.E. Mathematically, it is unlikely that no bias effects will be observed, or that bias sizes will cancel out exactly. The DIF contrast in Table 30 and 31 is the difference between the DIF sizes, and is a log-odds estimate, equivalent to a Mantel-Haenszel DIF size. The $t$ is the DIF contrast divided by the joint S.E. of the two DIF measures. It is equivalent to the Mantel-Haenszel significance test, but has the advantage of allowing missing data. This analysis is the same for all item types supported by Winsteps (dichotomies, rating or partial credit scales, etc.).

To replicate this with Winsteps yourself:

From a run of all the data, produce a PFILE=pf.txt and a SFILE=sf.txt

Then for each person classification of interest:

PAFILE=pf.txt
SAFILE=sf.txt
PSELECT=?????X ; to select only the person classification of interest
IFILE = X.txt ; item difficulties for person classification of interest
CONVERGE=L ; only logit change is used for convergence
LCONV=0.005 ; logit change too small to appear on any report.

Do this for each class.
The IFILE= values should match the values shown in Table 30.2
To graph the ICCs for different DIF classes on the same plot, see DIF item characteristic curves.

Classification sizes
There is no minimum size, but the smaller the classification size (also called reference groups and focal groups), the less sensitive the DIF test is statistically. Generally, results produced by classifications sizes of less than 30 are too much influenced by idiosyncratic behavior to be considered dependable.

Effect of imprecision in person or item estimates
This computation treats the person measures (for DIF) or the item measures (for DPF) as point estimates (i.e., exact). You can inflate the reported standard errors to allow for the imprecision in those measures. Formula 29 of Wright and Panchapakesan (1969), www.rasch.org/memo46.htm, applies. You will see there that, for dichotomies, the most by which imprecision in the baseline measures can inflate the variance is 25%. So, if you multiply the DIF/DPF point estimate S.E. by sqrt(1.25) = 1.12 (and divide the $t$ by 1.12), then you will be as conservative as possible in computing the DIF/DPF significance.

Impact on Person/Item Measurement
Unless DIF/DPF is large and mostly in one direction, the impact of DIF/DPF on person/item measurement is generally small. Wright & Douglas (1976) Rasch Item Analysis by Hand. "In other work we have found that when [test length] is greater than 20, random values of [item calibration misestimation] as high as 0.50 have negligible effects on measurement."

Wright & Douglas (1975) Best Test Design and Self-Tailored Testing. "They allow the test designer to incur item discrepancies, that is item calibration errors, as large as 1.0. This may appear unnecessarily generous, since it permits use of an item of difficulty 2.0, say, when the design calls for 1.0, but it is offered as an upper limit because we found a large area of the test design domain to be exceptionally robust with respect to independent item discrepancies."

DIF/DPF statistical significance
Table 30.1 shows pair-wise test of the statistical significance of DIF across classes. Table 30.2 shows statistical significance of DIF for a class against the average difficulty. A statistical test for DIF for multiple classes on one item is a "fixed effects" chi-square of homogeneity. For L measures, $D_i$, with standard errors $SE_i$, a test of the hypothesis that all L measures are statistically equivalent to one common "fixed effect" apart from measurement error is a chi-square statistics with L-1 d.f. where $p>.05$ (or $.01$) indicates statistically equivalent estimates.

$$ \chi^2 = \sum_{i=1}^{L} \frac{D_i^2}{SE_i^2} - \left( \sum_{i=1}^{L} \frac{D_i}{SE_i^2} \right)^2 / \left( \sum_{i=1}^{L} \frac{1}{SE_i^2} \right) $$
Non-Uniform DIF or DPF
To investigate this with the Winsteps, include in the item or person label a stratification variable, indicating, low, middle or high performers (or item difficulties). Use this as the classification variable for DIF or DPF. A graphical approach is to PSELECT= the people you want, and then draw their empirical ICCs for items. Several of these can be combined into one plot by copying the plotted data and pasting to Excel, then using the Excel graphing function.

The Mathematics of Winsteps DIF and DPF Estimation

Algebraically, the general model is in two stages:

**Stage 1:** \[ \log \left( \frac{P_{nij}}{P_{ni(j-1)}} \right) = B_n - D_{gi} - F_{gj} \]

Where \( B_n \) is the ability of person \( n \), \( D_{gi} \) is the difficulty of person \( i \) in classification \( g \), \( F_{gj} \) is the Rasch-Andrich threshold measure of category \( j \) relative to category \( j-1 \) for items in item-grouping \( g \).

For the Rasch dichotomous model, all items are in the same item-grouping (so that \( g \) is omitted), and there are only two categories, with \( F_1 = 0 \).

For the Andrich rating-scale model, all items are in the same item-grouping (so that \( g \) is omitted), and there are more than two categories, with \( \sum(F_j)=0 \).

For the Masters' partial-credit model, each item is in its own item-grouping (\( g=i \)), and there are more than two categories, with \( \sum(F_{ij})=0 \). To reparameterize into the conventional partial-credit model formulation, \( D_i + F_{ij} = D_{ij} \).

Estimates of \( b_n, d_{gi} \) and \( f_{gj} \) are obtained.

**Stage 2:**

**Table 30:** For person-subsample (ps) DIF: \[ \log \left( \frac{P_{nij}}{P_{ni(j-1)}} \right) = b_n - d_{gi} - f_{gj} - DIF(ps) \]

**Table 31:** For item-subsample (is) DPF: \[ \log \left( \frac{P_{nij}}{P_{ni(j-1)}} \right) = b_n - d_{gi} - f_{gj} + DPF(is) \]

**Table 33:** For person-subsample item-subsample (ps)(is) DIPF: \[ \log \left( \frac{P_{nij}}{P_{ni(j-1)}} \right) = b_n - d_{gi} - f_{gj} + DIPF(ps)(is) \]

Estimates of \( b_n, d_{gi} \) and \( f_{gj} \) anchored (fixed) from stage 1. The estimates of DIF, DPF or DIPF are the maximum likelihood estimates for which the marginal residuals for the subsamples from the stage 1 analysis are the sufficient statistics. All these computations are as sample-distribution-free as is statistically possible, except when the subsampling is based on the sample-distribution (e.g., when persons are stratified into subsamples according to their ability estimates.)

Different forms of DIF detection
A cross-plot of item difficulties derived from independent runs of the focal and reference classifying-groups, is basically reporting "Is the instrument working differently for the two sample classifications?", and, if so, "Where are the most conspicuous differences?" In the old days, when much analysis was done by hand, this would identify which items to choose for more explicitly constructed DIF tests, such as Mantel-Haenszel. From these plots we can get approximate DIF t-tests. This approach is obviously useful - maybe more useful than the item-by-item DIF tests. But it allows DIF in an item to change the person measures, and to alter the difficulties of other items and to change the rating (or partial credit) scale structure. To apply this "differential test functioning" approach to DIF detection, perform independent analyses of each sample class, produce IFILE= and cross-plot the measures using the Compare Statistics plot.

But, it is the item-by-item DIF tests that have traditional support. So, for these, we need to hold everything else
constant while we examine the DIF of each item. This is what Mantel-Haenszel does (using person raw scores), or the Winsteps DIF Table does (using person measures).

The Winsteps DIF table is equivalent to doing:
(a) The joint run of all person classifications, producing anchor values for person abilities and rating (or partial credit) scale structure.
(b) The classification A run with person abilities and rating (or partial credit) scale structure anchored to produce classification A item difficulties.
(c) The classification B run with person abilities and rating (or partial credit) scale structure anchored to produce classification B item difficulties.
(d) Pairwise item difficulty difference t-tests between the two sets of item difficulties (for classifications A and B).

Lord's Chi-square DIF method takes a different approach, automatically looking for a core of stable items, but it is accident-prone and appears to overdetect DIF. In particular, if items were slightly biased, 50% against boys and 50% against girls, it would be accidental which set of items would be reported as "unbiased" and which as "biased".

Mantel-Haenszel method. See MHSLICE=.

ANOVA method. This can be facilitated by Winsteps.
(1) Identify the relevant demographic variable in the person label, and set ITEM1= at the variable, and NAMLEN=1.
(2) Perform a standard Winsteps analysis
(3) Use USCALE=, UMEAN= and UDECIM= to transform the person measures into convenient "class intervals": integers with lowest value 1, and highest value 10 for 10 class intervals.
(4) Write out an XFILE= selecting only:
   person measure (class interval)
   standardized residual
   person label (demographic variable)
(5) Read this file into your statistics software.
(6) Transform the demographic variable into 1 and 2.
(7) Perform the "fully randomized" ANOVA with standardized residual as the dependent variable, and person measure and person label as the independent variables.

291. DIF item characteristic curves

Here is a way to show the item characteristic curves (ICCs) for different DIF classifications on the same plot:

```
TITLE='DATA SET UP FOR MULTIPLE PERSON CLASSIFICATIONS (DIF GROUPS)'
NAME=1
$DIF = $SI1W ; DIF classification in column 1 of Person Label. Codes are 1 and 2.
ITEM=3
NI = 32 ; 10 FOR MAIN ANALYSIS: 10 FOR EACH GROUP + 2 BLANKS
CODES=01
IWEIGHT=*
11-32 0 ; WEIGHTED ZERO: FOR DIF ICC GRAPHS ONLY
* ; ISELECT=0 ; enter this in Specification pull-down menu to report 1st 10 items only
&END
0-01 ; 0-02 ; 0-03 ; 0-04 ; 0-05 ; 0-06 ; 0-07 ; 0-08 ; 0-09 ; 0-10
BLANK
1-01 ; 10 items (12-21), weighted 0 for classification 1
1-02
1-03
1-04
1-05
```
Specifying the items on the Multiple ICCs screen produces:

![Item Characteristic Curves](image_url)

The same data layout can be obtained without a rectangular copy as follows:
TITLE='DATA SET UP FOR MULTIPLE PERSON CLASSIFICATIONS (DIF GROUPS)'
NAME1=1
$DIF = $SIW1 ; DIF classification in column 1 of Person Label. Codes are 1 and 2.
ITEM1=3
NI = 32 ; 10 FOR MAIN ANALYSIS: 10 FOR EACH GROUP + 2 BLANKS
CODES=01
IWEIGHT=* 
11-32 0 ; WEIGHTED ZERO: FOR DIF ICC GRAPHS ONLY 
ISELECT=0 ; enter this in Specification pull-down menu to report 1st 10 items only
FORMAT = (13A,T3,11A,T3,10A) ; this copies the items 3 times.
EDFILE = *
1-15 23-32 X ; blank out top right: persons 1-15, items 23-32
16-25 12-21 X ; blank out bottom center: persons 16-25, item 12-21 
RFILE = reformat.txt ; look at to check that the data looks like the example above
&END
0-01 ; original 10 items
0-02
0-03
0-04
0-05
0-06
0-07
0-08
0-09
0-10
BLANK
1-01 ; 10 items (12-21), weighted 0 for classification 1
1-02
1-03
1-04
1-05
1-06
1-07
1-08
1-09
1-10
BLANK
2-01 ; 10 items (23-32), weighted 0 to classification 2
2-02
2-03
2-04
2-05
2-06
2-07
2-08
2-09
2-10
END LABELS
1 0111001100  ; original data
1 1111111000
1 1111111001
1 1101101100
1 0110001000
1 1011010100
1 1111101110
1 1111110100
1 0111110000
1 1111110000
1 1111111000
1 1111100000
1 0000000000
1 0000000000
1 0000000000
1 0111110000
2 1111111100
2 1111111100
2 1111110100
2 1111000000
2 0000000000
2 0100000000
2 1000000000
2 0111110010
2 1111100110

271
292. Dimensionality: contrasts & variances

Please do not interpret Rasch-residual-based Principal Components Analysis (PCAR) as a usual factor analysis. These components show contrasts between opposing factors, not loadings on one factor. Criteria have yet to be established for when a deviation becomes a dimension. So PCA is indicative, but not definitive, about secondary dimensions.

Example from Table 23.0 from Example0.txt:

<table>
<thead>
<tr>
<th></th>
<th>Empirical</th>
<th>Modeled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total variance in observations</td>
<td>127.9 100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Variance explained by measures</td>
<td>102.9 80.5%</td>
<td>82.0%</td>
</tr>
<tr>
<td>Unexplained variance (total)</td>
<td>25.0 19.5%</td>
<td>100.0% 18.0%</td>
</tr>
<tr>
<td>Unexplained variance in 1st contrast</td>
<td>4.6 3.6%</td>
<td>18.5%</td>
</tr>
<tr>
<td>Unexplained variance in 2nd contrast</td>
<td>2.9 2.3%</td>
<td>11.8%</td>
</tr>
<tr>
<td>Unexplained variance in 3rd contrast</td>
<td>2.3 1.8%</td>
<td>9.2%</td>
</tr>
<tr>
<td>Unexplained variance in 4th contrast</td>
<td>1.7 1.4%</td>
<td>6.9%</td>
</tr>
<tr>
<td>Unexplained variance in 5th contrast</td>
<td>1.6 1.3%</td>
<td>6.5%</td>
</tr>
</tbody>
</table>

The Rasch dimension explains 80.5% of the variance in the data: good! The largest secondary dimension, "the first contrast in the residuals" explains 3.6% of the variance - somewhat greater than around 1% that would be observed in data like these simulated to fit the Rasch model. Check this by using the SIMUL= option in Winsteps to simulate a Rasch-fitting dataset with same characteristics as this dataset. Then produce this Table for it. Also see: www.rasch.org/rmt/rmt191h.htm

The eigenvalue of the biggest residual contrast is 4.6 - this indicates that it has the strength of about 5 items, somewhat bigger than the strength of two items, the smallest amount that could be considered a "dimension". Contrast the items at the top and bottom of the first PCAR plot to identify what this secondary dimension reflects.

Rules of Thumb:
- Variance explained by measures > 60% is good.
- Unexplained variance explained by 1st contrast (size) < 3.0 is good.
- Unexplained variance explained by 1st contrast < 5% is good.

But there are plenty of exceptions ....

Analytical Note:
Winsteps performs an unrotated "principal components" factor analysis. (using Hotelling's terminology). If you would like to rotate axes, have oblique axes, or perform a "common factor" factor analysis of the residuals, Winsteps can write out the matrices of residual item (or person) correlations, see the "Output Files" pull down menu or ICORFIL= and PCORFIL=. You can import these into any statistics software package.

The purpose of PCA of residuals is not to construct variables (as it is with "common factor" analysis), but to explain variance. First off, we are looking for the contrast in the residuals that explains the most variance. If this contrast is at the "noise" level, then we have no shared second dimension. If it does, then this contrast is the "second" dimension in the data. (The Rasch dimension is hypothesized to be the first). Similarly we look for a third dimension, etc. Rotation, oblique axes, the "common factor" approach, all reapportion variance, usually in an attempt to make the factor structure more clearly align with the items, but, in so doing, the actual variance structure and dimensionality of the data is masked.

In Rasch analysis, we are trying to do the opposite of what is usually happening in factor analysis. In Rasch analysis of residuals, we want **not** to find contrasts, and, if we do, we want to find the least number of contrasts above the noise level, each, in turn, explaining as much variance as possible. This is exactly what unrotated PCA does.

In conventional factor analysis of observations, we are hoping desperately to find shared factors, and to assign
the items to them as clearly and meaningfully as possible. In this endeavor, we use a whole toolbox of rotations, obliquenesses and choices of diagonal self-correlations (i.e., the "common factor" approach).

But, different analysts have different aims, and so Winsteps provides the matrix of residual correlations to enable the analyst to perform whatever factor analysis is desired!

**The Rasch Model: Expected values, Model Variances, and Standardized Residuals**

The Rasch model constructs linear measures from ordinal observations. It uses dis ordering of the observations across persons and items to construct the linear frame of reference. Perfectly ordered observations would accord with the ideal model of Louis Guttman, but lack information as to the distances involved.

Since the Rasch model uses dis ordering in the data to construct distances, it predicts that this dis ordering will have a particular ideal form. Of course, empirical data never exactly accord with this ideal, so a major focus of Rasch fit analysis is to discover where and in what ways the dis ordering departs from the ideal. If the departures have substantive implications, then they may indicate that the quality of the measures is compromised.

A typical Rasch model is:

\[ \log \left( \frac{P_{niki}}{P_{niki(k-1)}} \right) = B_n - D_i - F_k \]

where

- \( P_{niki} \) = the probability that person \( n \) on item \( i \) is observed in category \( k \), where \( k=0,m \)
- \( P_{niki(k-1)} \) = the probability that person \( n \) on item \( i \) is observed in category \( k-1 \)
- \( B_n \) = the ability measure of person \( n \)
- \( D_i \) = the difficulty measure of item \( i \)
- \( F_k \) = the structure calibration from category \( k-1 \) to category \( k \)

This predicts the observation \( X_{ni} \). Then

\[ X_{ni} = E_{ni} \pm \sqrt{V_{ni}} \]

where

- \( E_{ni} = \sum (kP_{niki}) \) for \( k=0,m \).
  This is the expected value of the observation.
- \( V_{ni} = \sum (k^2P_{niki}) - (E_{ni})^2 \) for \( k=0,m \).
  This is the model variance of the observation about its expectation, i.e., the predicted randomness in the data.

The Rasch model is based on the specification of "local independence". This asserts that, after the contribution of the measures to the data has been removed, all that will be left is random, normally distributed noise. This implies that when a residual, \( (X_{ni} - E_{ni}) \), is divided by its model standard deviation, it will have the characteristics of being sampled from a unit normal distribution. That is:

\[ \frac{(X_{ni} - E_{ni})}{\sqrt{V_{ni}}} \]

the standardized residual of an observation, is specified to be \( N(0,1) \)

The bias in a measure estimate due to the misfit in an observation approximates \( (X_{ni} - E_{ni}) * \text{ S.E.}^2(\text{measure}) \)

**Principal Components Analysis of Residuals**

"Principal Component Analysis (PCA) is a powerful technique for extracting structure from possibly high-dimensional data sets. It is readily performed by solving an eigenvalue problem, or by using iterative algorithms which estimate principal components [as in Winsteps]. ... some of the classical papers are due to Pearson (1901); Hotelling (1933); ... PCA is an orthogonal transformation of the coordinate system in which we describe our data. The new coordinate values by which we represent the data are called principal components. It is often the case that a small number of principal components is sufficient to account for most of the structure in the data. These are sometimes called factors or latent variables of the data." (Schölkopf, D., Smola A.J., Müller K.-R., 1999, Kernel Principal Component Analysis, in Schölkopf at al. "Advances in Kernel Methods", London: MIT Press).


The standardized residuals are modeled to have unit normal distributions which are independent and so
uncorrelated. A PCA of Rasch standardized residuals should look like a PCA of random normal deviates. Simulation studies indicate that the largest component would have an eigenvalue of about 1.4 and they get smaller from there. But there is usually something else going on in the data, so, since we are looking at residuals, each component contrasts deviations in one direction ("positive loading") against deviation in the other direction ("negative loading"). As always with factor analysis, positive and negative loading directions are arbitrary. Each component in the residuals only has substantive meaning when its two ends are contrasted. This is a little different from PCA of raw observations where the component is thought of as capturing the "thing".

Loadings are plotted against Rasch measures because deviation in the data from the Rasch model is often not uniform along the variable (which is actually the "first" dimension). It can be localized in easy or hard items, high or low ability people. The Wright and Masters "Liking for Science" data is an excellent example of this.

Total, Explained and Unexplained Variances

The decomposition of the total variance in the data set proceeds as follows for the standardized residual, PRCOMP=S and raw score residual PRCOMP=R, option.

(i) The average person ability measure, b, and the average item difficulty measure, d, are computed.

(ii) The expected response, Ebd, by a person of average ability measure to an item of average difficulty measure is computed. (If there are multiple rating or partial credit scales, then this is done for each rating or partial credit scale.)

(iii) Each observed interaction of person n, of estimated measure Bn, with item i, of estimated measure Di, produces an observation Xni, with an expected value, Eni, and model variance, Vni.

   The raw-score residual, Zni, of each Xni is Zni = Xni-Eni.
   The standardized residual, Zni, of each Xni is Zni = (Xni-Eni)/sqrt(Vni).

Empirically:

(iv) The piece of the observation available for explanation by Bn and Di is approximately Xni - Ebd.

   In raw-score residual units, this is Cni = Xni-Ebd
   In standardized residual units, this is Cni = (Xni-Ebd)/sqrt(Vni)

   The total variance sum-of-squares in the data set available for explanation by the measures is: VAvailable = \( \sum(Cni^2) \)

(v) The total variance sum of squares predicted to be unexplained by the measures is: VUnexplained = \( \sum(Zni^2) \)

(vi) The total variance sum of squares explained by the measures is: VExplain = VAvailable - VUnexplained

   If VExplain is negative, see below.

Under model conditions:

(viii) The total variance sum of squares explained by the measures is:

   Raw-score residuals: VMexplained = \( \sum((Eni-Ebd)^2) \)
   Standardized residuals: VMexplained = \( \sum((Eni-Ebd)^2/Vni) \)

(ix) The total variance sum of squares predicted to be unexplained by the measures is:

   Raw score residuals: VMunexplained = \( \sum(Vni) \)
   Standardized residuals: VMunexplained = \( \sum(Vni/Vni) = \sum(1) \)

x) total variance sum-of-squares in the data set predicted to be available for explanation by the measures is:

VMAvailable = VMexplained + VMUnexplained

Negative Variance Explained

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)
Total variance in observations = 20.3 100.0%
Variance explained by measures = -23.7 -116.2%

According to this Table, the variance explained by the measures is less than the theoretical minimum of 0.00. This "negative variance" arises when there is unmodeled covariance in the data. In Rasch situations this happens when the randomness in the data, though normally distributed when considered overall, is skewed when partitioned by measure difference. A likely explanation is that some items are reverse-coded. Check that all
correlations are positive by viewing the Diagnosis Menu, Table A. If necessary, use IREFER= to recode items. If there is no obvious explanation, please email your control and data file to www.winsteps.com

**Principal Components Analysis of Standardized Residuals**

(i) The standardized residuals for all observations are computed. Missing observations are imputed to have a standardized residual of 0, i.e., to fit the model.

(ii) Correlation matrices of standardized residuals across items and across persons are computed. The correlations furthest from 0 (uncorrelated) are reported in Tables 23.99 and 24.99.

(iii) In order to test the specification that the standardized residuals are uncorrelated, it is asserted that all randomness in the data is shared across the items and persons. This is done by placing 1’s in the main diagonal of the correlation matrix. This accords with the “Principal Components” approach to Factor Analysis. ("General" Factor Analysis attempts to estimate what proportion of the variance is shared across items and persons, and reduces the diagonal values from 1’s accordingly. This approach contradicts our purpose here.)

(iv) The correlation matrices are decomposed. In principal, if there are L items (or N persons), and they are locally independent, then there are L item components (or N person components) each of size (i.e., eigenvalue) 1, the value in the main diagonal. But there are expected to be random fluctuations in the structure of the randomness. However, eigenvalues of less than 2 indicate that the implied substructure (dimension) in these data has less than the strength of 2 items (or 2 persons), and so, however powerful it may be diagnostically, it has little strength in these data.

(v) If items (or persons) do have commonalities beyond those predicted by the Rasch model, then these may appear as shared fluctuations in their residuals. These will inflate the correlations between those items (or persons) and result in components with eigenvalues greater than 1. The largest of these components is shown in Table 23.2 and 24.3, and sequentially smaller ones in later subtables.

(vi) In the Principal Components Analysis, the total variance is expressed as the sum of cells along the main diagonal, which is the number of items, L, (or number of persons, N). This corresponds to the total unexplained variance in the dataset, VUnexplained.

(vii) The variance explained by the current contrast is its eigenvalue.

**Example: Item Decomposition**

From Table 23.2: The Principal Components decomposition of the standardized residuals for the items, correlated across persons. Winsteps reports:

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

<table>
<thead>
<tr>
<th></th>
<th>Empirical</th>
<th>Modeled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total variance in observations</td>
<td>1452.0</td>
<td>100.0% 100.0%</td>
</tr>
<tr>
<td>Variance explained by measures</td>
<td>1438.0</td>
<td>99.0% 98.6%</td>
</tr>
<tr>
<td>Unexplained variance (total)</td>
<td>14.0</td>
<td>1.0% 1.4%</td>
</tr>
<tr>
<td>Unexpl var explained by 1st contrast</td>
<td>2.7</td>
<td>.2%</td>
</tr>
</tbody>
</table>

The first contrast has an eigenvalue size of 2.7. This corresponds to 2.7 items. There are 14 active items, so that the total unexplained variance in the correlation matrix is 14 units.

The "Modeled" column shows what this would have looked like if these data fit the model exactly.

Conclusion: Though this contrast has the strength of 3 items, and so might be independently constructed from these data, its strength is so small that it is barely a ripple on the total measurement structure.

Caution: The 1st contrast may be an extra dimension, or it may be a local change in the intensity of this dimension:

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)

<table>
<thead>
<tr>
<th></th>
<th>Empirical</th>
<th>Modeled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total variance in observations</td>
<td>97.1</td>
<td>100.0% 100.0%</td>
</tr>
<tr>
<td>Variance explained by measures</td>
<td>58.1</td>
<td>59.8% 59.0%</td>
</tr>
<tr>
<td>Unexplained variance (total)</td>
<td>39.0</td>
<td>40.2% 100.0% 41.0%</td>
</tr>
</tbody>
</table>
The first contrast comprises items A-E. But their mean-squares are all less than 1.0, indicating they do not contradict the Rasch variable, but are rather too predictable. They appear to represent a local intensification of the Rasch dimension, rather than a contradictory dimension.

**Comparison with Rasch-fitting data**

Winsteps makes it easy to compare empirical PCA results with the results for an equivalent Rasch-fitting data set. From the Output Files menu, make a “Simulated Data” file, call it, say, test.txt

From the Files menu, Restart Winsteps. Under “Extra specifications”, type in "data=test.txt".

Exactly the same analysis is performed, but with Rasch-fitting data. Look at the Dimensionality table:

```
<table>
<thead>
<tr>
<th>item</th>
<th>MEASURE</th>
<th>ENTRY</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>-2</td>
<td>-1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>COUNT</td>
<td></td>
</tr>
</tbody>
</table>
```

```
<table>
<thead>
<tr>
<th>Item MEASURE</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>ENTRY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

```
<table>
<thead>
<tr>
<th>contrast</th>
<th>LOADING</th>
<th>MEASURE</th>
<th>MNSQ</th>
<th>MNSQ</th>
<th>NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.74</td>
<td>.71</td>
<td>.78</td>
<td>.70</td>
<td>A</td>
</tr>
<tr>
<td>1</td>
<td>.65</td>
<td>.26</td>
<td>.79</td>
<td>.68</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>.64</td>
<td>1.34</td>
<td>.87</td>
<td>.84</td>
<td>C</td>
</tr>
<tr>
<td>1</td>
<td>.56</td>
<td>.64</td>
<td>.85</td>
<td>.80</td>
<td>D</td>
</tr>
<tr>
<td>1</td>
<td>.51</td>
<td>-.85</td>
<td>.84</td>
<td>.60</td>
<td>E</td>
</tr>
</tbody>
</table>
```

Repeat this process several times, simulating a new dataset each time. If they all look like this, we can conclude that the value of 2.7 for the 1st contrast in the residuals is negligibly bigger than the 2.2 expected by chance.

**General Advice**

A question here is “how big is big”? Much depends on what you are looking for. If you expect your instrument to have a wide spread of items and a wide spread of persons, then your measures should explain most of the variance. But if your items are of almost equal difficulty (as recommended, for instance, by G-Theory) and your persons are of similar ability (e.g., hospital nurses at the end of their training) then the measures will only explain a small amount of the variance.

Ben Wright recommends that the analyst split the test into two halves, assigning the items, top vs. bottom of the first component in the residuals. Measure the persons on both halves of the test. Cross-plot the person measures. If the plot would lead you to different conclusions about the persons depending on test half, then there is a multidimensionality. If the plot is just a fuzzy straight line, then there is one, perhaps somewhat vague, dimension.
**Rules of Thumb**

"Reliability" (= Reproducibility) is "True" variance divided by Observed variance. If an acceptable, "test reliability" (i.e., reproducibility of this sample of person measures on these items) is 0.8, then an acceptable Rasch "data reliability" is also 0.8, i.e., "variance explained by measures" is 4 times "total unexplained variance".

In the unexplained variance, a "secondary dimension" must have the strength of at least 3 items, so if the first first contrast has "units" (i.e., eigenvalue) less than 3 (for a reasonable length test) then the test is probably unidimensional. (Of course, individual items can still misfit).

Negative variance can occur when the unexpectedness in the data is not random. An example is people who flatline an attitude survey. Their unexpected responses are always biased towards one category of the rating (or partial credit) scale.


**293. Dimensionality: when is a test multidimensional?**

For more discussion see dimensionality and contrasts.

"I can not understand the residual contrast analysis you explained. For example, in Winsteps, it gave me the five contrasts’ eigenvalues: 3.1, 2.4, 1.9, 1.6, 1.4. (I have 26 items in this data). The result is the same as when I put the residuals into SPSS."

Reply:

Unidimensionality is never perfect. It is always approximate. The Rasch model constructs from the data parameter estimates along the unidimensional latent variable that best concurs with the data. But, though the Rasch measures are always unidimensional and linear, their concurrence with the data is never perfect. Imperfection results from multi-dimensionality in the data and other causes of misfit.

Multidimensionality always exists to a lesser or greater extent. The vital question is: "Is the multi-dimensionality in the data big enough to merit dividing the items into separate tests, or constructing new tests, one for each dimension?"

In your example, the first contrast has eigenvalue of 3.1. This means that the contrast between the strongly positively loading items and the strongly negatively loading items on the first contrast in the residuals has the strength of about 3 items. Since positive and negative loading is arbitrary, you must look at the items at the top and the bottom of the contrast plot. Are those items substantively different? Are they so different that they merit the construction of two separate tests?

It may be that two or three off-dimension items have been included in your 26 item instrument and should be dropped. But this is unusual for a carefully developed instrument. It is more likely that you have a "fuzzy" or "broad" dimension, like mathematics. Mathematics includes arithmetic, algebra, geometry and word problems. Sometimes we want a “geometry test”. But, for most purposes, we want a “math test”.

If in doubt, split your 26 items into two subtests, based on +ve and -ve loadings on the first residual contrast. Measure everyone on the two subtests. Cross-plot the measures. What is their correlation? Do you see two versions of the same story about the persons, or are they different stories? Which people are off-diagonal? Is that important? If only a few people are noticeably off-diagonal, or off-diagonal deviance would not lead to any action, then you have a substantively unidimensional test. A straightforward way to obtain the correlation is to write out a PFILE= output file for each subtest. Read the measures into EXCEL and have it produce their Pearson product-moment correlation. If R1 and R2 are the reliabilities of the two subtests, and C is the correlation of their ability estimates reported by Excel, then their latent (error-disattenuated) correlation approximates C / sqrt (R1*R2). If this approaches 1.0, then the two subtests are statistically telling the same story. But you may have a "Fahrenheit-Celsius" equating situation if the best fit line on the plot departs from a unit slope.
You can do a similar investigation for the second contrast of size 2.4, and third of size 1.9, but each time the motivation for doing more than dropping an off-dimension item or two becomes weaker. Since random data can have eigenvalues of size 1.4, there is little motivation to look at your 5th contrast.

294. Disjoint strings of responses

When the responses are not arranged in one continuous string in the record, instruct WinStepS to skip over or ignore the gaps.

Example: The 18 item string is in columns 40 to 49 and then 53 to 60 of your data file. The person-id is in columns 11-30. Data look like:

```
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
this results in disordering of the “average measures” or “observed averages”, the average abilities of the people observed in each category, and also large mean-square fit statistics. The “scale structure measures”, also called “step calibrations”, “step difficulties”, “step measures”, “Rasch-Andrich thresholds”, “deltas”, “taus”, etc., remain ordered.

(ii) Disorder in the “step calibrations” or “disordered Rasch-Andrich thresholds” implies less frequently observed intermediate categories, i.e., that they correspond to narrow intervals on the latent variable.

In this example, the FIM categories are correctly ordered, but the frequency of level 2 has been reduced by removing some observations from the data. Average measures and fit statistics remain well behaved. The disordering in the “step calibrations” now reflects the relative infrequency of category 2. This infrequency is pictured in plot of probability curves which shows that category 2 is never a modal category in these data. The step calibration values do not indicate whether measurement would be improved by collapsing levels 1 and 2, or collapsing levels 2 and 3, relative to leaving the categories as they stand.

<table>
<thead>
<tr>
<th>FIM LEVEL</th>
<th>COUNT</th>
<th>AVERAGE MEASURE</th>
<th>INFIT MNSQ</th>
<th>OUTFIT MNSQ</th>
<th>STEP CALIBRATN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>96</td>
<td>-2.81</td>
<td>.90</td>
<td>.96</td>
<td>NONE</td>
</tr>
<tr>
<td>2</td>
<td>44</td>
<td>-1.96</td>
<td>.88</td>
<td>.92</td>
<td>-1.49</td>
</tr>
<tr>
<td>3</td>
<td>101</td>
<td>-1.03</td>
<td>1.02</td>
<td>.98</td>
<td><strong>2.33</strong></td>
</tr>
<tr>
<td>4</td>
<td>168</td>
<td>-.30</td>
<td>1.07</td>
<td>1.22</td>
<td>-1.29</td>
</tr>
<tr>
<td>5</td>
<td>210</td>
<td>.82</td>
<td>.96</td>
<td>.88</td>
<td>.05</td>
</tr>
<tr>
<td>6</td>
<td>146</td>
<td>2.30</td>
<td>.75</td>
<td>.82</td>
<td>1.97</td>
</tr>
<tr>
<td>7</td>
<td>101</td>
<td>3.27</td>
<td>.87</td>
<td>.89</td>
<td>3.09</td>
</tr>
</tbody>
</table>

296. Displacement measures

DISPLACE column should only appear with anchored or TARGET= runs. Otherwise its appearance indicates lack of convergence.

The DISPLACE value is the size of the change in the parameter estimate that would be observed in the next estimation iteration if this parameter was free (unanchored) and all other parameter estimates were anchored at their current values.

For a parameter (item or person) that is anchored in the main estimation, DISPLACE indicates the size of disagreement between an estimate based on the current data and the anchor value.

For an unanchored item, if the DISPLACE value is large enough to be of concern, then the convergence criteria are not tight enough LCONV=, RCONV=, CONVERGE=, MUCON=
It is calculated using Newton-Raphson estimation.

Person: \[ \text{DISPLACE} = \frac{(\text{observed marginal score} - \text{expected marginal score})}{\text{(model variance of the marginal score)}} \]

Item: \[ \text{DISPLACE} = -\frac{(\text{observed marginal score} - \text{expected marginal score})}{\text{(model variance of the marginal score)}} \]

DISPLACE approximates the displacement of the estimate away from the statistically better value which would result from the best fit of your data to the model. Each DISPLACE value is computed as though all other parameter estimates are exact. Only meaningfully large values are displayed. They indicate lack of convergence, or the presence of anchored or targeted values. The best fit value can be approximated by adding the displacement to the reported measure or calibration. It is computed as:

\[ \text{DISPLACE} = \frac{(\text{observed score} - \text{expected score based on reported measure})}{\text{(Rasch-model-derived score variance)}} \]

This value is the Newton-Raphson adjustment to the reported measure to obtain the measure estimated from the current data. In BTD, p. 64, equation 3.7.11: \( d_i(j) \) is the anchor value, \( d_i(j+1) \) is the value estimated from the current data, and \( d_i(j+1) - d_i(j) \) is the displacement, given by the right-hand term of the estimation equation, also in step 6 of www.rasch.org/rmt/rmt102t.htm. In RSA, p. 77, equation 4.4.6, \( d_i(t) \) is the anchor value, \( d_i(t+1) \) is the value estimated from the current data, and \( d_i(t+1) - d_i(t) \) is the displacement, given by the right-hand term of the estimation equation, also in step 6 of www.rasch.org/rmt/rmt122q.htm

**Standard Error of the Displacement Measure**

<table>
<thead>
<tr>
<th>ENTRY</th>
<th>RAW</th>
<th>NUMBER</th>
<th>COUNT</th>
<th>SCORE</th>
<th>MEASURE</th>
<th>S.E.</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>PTMEA</th>
<th>MNSQ</th>
<th>ZSTD</th>
<th>DISPLACE</th>
<th>TAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>35</td>
<td>35</td>
<td>2.00A</td>
<td>.74</td>
<td>.69</td>
<td>-.61</td>
<td>.22</td>
<td>.5</td>
<td>.00</td>
<td>-3.90</td>
<td>1-2-4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Since the reported "measure" is treated as a constant when "displacement" is computed, the S.E. of the reported "measure" actually is the same as the S.E. of the displacement. The DISPLACE column shows the displacement in the same units as the MEASURE. This is logits when USCALE=1, the default. If the anchored measure value is considered to be exact, i.e., a point-estimate, then the S.E. standard error column indicates the standard error of the displacement. The statistical significance of the Displacement is given by:

\[ t = \frac{\text{Displacement}}{\text{S.E.}(\text{Displacement})} \]

with approximately COUNT degrees of freedom.

This evaluates how likely the reported size of the displacement is, if its "true" size is zero. But both the displacements and their standard errors are estimates, so the t-value may be slightly mis-estimated. Consequently allow for a margin of error when interpreting the t-values.

If the anchored measure value has a standard error obtained from a different data set, then the standard error of the displacement is:

\[ \text{S.E.}(\text{Displacement}) = \sqrt{\text{S.E.}^2 + \text{S.E.}(\text{anchor value from original data})^2} \]

**When does large displacement indicate that an item or person should be unanchored or omitted?**

This depends on your purpose. If you are anchoring items in order to measure three additional people to add to your measured database of thousands, then item displacement doesn't matter.

Anchor values should be validated before they are used. Do two analyses:
(a) with no items anchored (i.e., all items floating), produce person and item measures.
(b) with anchored items anchored, produce person and item measures.

Then cross-plot the item difficulties for the two runs, and also the person measures. The person measures will usually form an almost straight line.

For the item difficulties, unanchored items will form a straight-line. Some anchored items may be noticeably off the line. These are candidates for dropping as anchors. The effect of dropping or unanchoring a "displaced" anchor item is to realign the person measures by roughly \( \frac{\text{displacement}}{(\text{number of remaining anchored items})} \).

Random displacements of less than 0.5 logits are unlikely to have much impact in a test instrument.
In other work we have found that when [test length] is greater than 20, random values of [discrepancies in item calibration] as high as 0.50 [logits] have negligible effects on measurement." (Wright & Douglas, 1976, "Rasch Item Analysis by Hand")

"They allow the test designer to incur item discrepancies, that is item calibration errors, as large as 1.0 [logit]. This may appear unnecessarily generous, since it permits use of an item of difficulty 2.0, say, when the design calls for 1.0, but it is offered as an upper limit because we found a large area of the test design domain to be exceptionally robust with respect to independent item discrepancies." (Wright & Douglas, 1975, "Best Test Design and Self-Tailored Testing.")

Most DIF work seems to be done by statisticians with little interest in, and often no access to, the substantive material. So they have no qualitative criteria on which to base their DIF acceptance/rejection decisions. The result is that the number of items with DIF is grossly over-reported (Hills J.R. (1989) Screening for potentially biased items in testing programs. Educational Measurement: Issues and practice. 8(4) pp. 5-11).

297. Edit Initial Settings (Winsteps.ini) file

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

To change the Winsteps standard starting directory from a short-cut icon:

Right-click the short-cut to Winsteps.exe (this may be in your \windows\start menu directory)
Click on "Properties"
Select "Shortcut"
Type the desired directory into "Start in"
Click on "Apply"

Altering the Initial Settings

1) Pull-down the "Edit" menu
2) Click on "Edit initial settings"

3) The Settings display. If this is too big for your screen see Display too big.

Winsteps.ini is the name of a file in your Winsteps folder which maintains various settings for Winsteps. It is a Text file. Here are the typical contents:

Editor="C:\Program Files\Windows NT\Accessories\wordpad.exe"
Temporary directory="C:\DOCUME~1\Mike\LOCALS~1\Temp"
Filter=All Files (*.*)
Excel=C:\Program Files\Microsoft Office\Office\EXCEL.EXE
SPSS=C:\WINDOWS\system32\NOTEPAD.EXE
Reportprompt=Y
Extraask=Y
Closeonexit=Ask
Welcome=Yes
OFOPTIONS=NTPYYYN
XFIELDS="XXXXXXXXXXXXXXXXXXXXXXXXXXXX"
SIGNS=;,,,..

(1) Editor path: Editor="C:\Program Files\Windows NT\Accessories\wordpad.exe"
This specifies the word processor or text editor that is to be used by Winsteps - WordPad or your own text editor. If Output Tables do not display, this may not be a working version of WordPad, see Changing Wordpad.

(2) Excel path: Excel=C:\Program Files\Microsoft Office\Office\EXCEL.EXE
This provides a fast link to Excel from the File pull-down menu. The path to any program can be placed here. If Winsteps does not find Excel automatically, please go to the Windows "Start" menu, do a "Find" or "Search" for "Files and Folders" to locate it, and enter its path here. It is not part of Winsteps, and may not be present on your computer.

(3) SPSS path: SPSS=C:\SPSS.EXE
This provides a fast link to SPSS from the File pull-down menu. The path to any program can be placed here. If Winsteps does not find SPSS automatically, please go to the Windows "Start" menu, do a "Find" or "Search" for "Files and Folders" to locate it, and enter its path here. It is not part of Winsteps, and may not be installed on your computer.

(4) Filter for file selection: Filter=All Files (*.*)
This is the selection for the Files dialog box, used when setting up or saving files. If everything is in Text files, then specify: Filter=Text Files (*.txt)

(5) Temporary directory for work files: Temporary directory="C:\Temp"
Temporary Output and Table files are placed:
(a) In the same directory as your Input file (if possible)
or (b) in the "Temporary directory", which is ordinarily "C:\TEMP"
Other temporary files are placed in the "Temporary directory."
Files ending "....ws.txt" can be deleted when Winsteps is not running.

(6) Prompt for output file name: Reportprompt=No specifies that your standard report output file will be a temporary file.
Reportprompt=Yes produces a prompt for a file name on the screen. You can always view the Report output file from the Edit menu, and save it as another file.

(7) Prompt for Extra Specifications: Extraask=No specifies that there will be no extra specifications to be entered.

(8) Show "Welcome" Help? Welcome=Yes displays the "Easy Start" help message. Welcome=No does not.

(9) Close output windows on exit? Closeonexit=Ask
This choice can be overridden when Winsteps stops.

(10) Character that starts .... SIGNS=;,,,,
This enables the processing of files in accordance with international usages. For instance, it may be more
convenient to output files with decimal commas, use semicolons as separators, and indicate comments with
exclamation marks:

| Character that starts a comment: | ! |
| Field separator or delimiter: | ; |
| Decimal point or sign: | , |

(11) OFOPTIONS=, XFIELDS= and other settings are internal switches. Deleting them does no harm. They will
be automatically reset.

OK: Click on this to accept the settings. Some settings go into effect the next time your start Winsteps.

Cancel: Do not change the settings.

Help: Displays this page of the Help file

To reset all settings back to their standard values, find Winsteps.ini in your Winsteps folder and delete it. Standard
values will be instituted next time you start Winsteps.

298. Equating and linking tests

Test Equating and linking are usually straightforward with *Winsteps*, but do require clerical care. The more
thought is put into test construction and data collection, the easier the equating will be.

Imagine that Test A (the more definitive test, if there is one) has been given to one sample of persons, and Test B
to another. It is now desired to put all the items together into one item hierarchy, and to produce one set of
measures encompassing all the persons.

Initially, **analyze each test separately**. Go down the "Diagnosis" pull-down menu. If the tests don't make sense
separately, they won't make sense together.

There are several equating methods which work well with *Winsteps*. Test equating is discussed in Bond & Fox
"Applying the Rasch model", and earlier in Wright & Stone, "Best Test Design", George Ingebo "Probability in the
Measure of Achievement" - all available from www.rasch.org/books.htm

Concurrent or One-step Equating
All the data are entered into one big array. This is convenient but has its hazards. Off-target items can introduce
noise and skew the equating process. **CUTLO**= and **CUTHI**= may remedy targeting deficiencies. Linking designs
forming long chains require much tighter than usual convergence criteria. Always cross-check results with those
obtained by one of the other equating methods.

Common Item Equating
This is the best and easiest equating method. The two tests share items in common, preferably at least 5 spread
out across the difficulty continuum.

Step 1. From the separate analyses, **crossplot** the difficulties of the common items, with Test B on the y-axis and
Test A on the x-axis. The slope of the best-fit line i.e., the line though the point at the means of the common items
and through the (mean + 1 S.D.) point should have slope near 1.0. If it does, then the intercept of the line with the
x-axis is the equating constant.

First approximation: Test B measures in the Test A frame of reference = Test B measure + x-axis intercept.

Step 2. Examine the scatterplot. Points far away from the best fit line indicate items that have behaved differently
on the two occasions. You may wish to consider these to be no longer common items. Drop the items from the
plot and redraw the best fit line. Items may be off the diagonal, or exhibiting large misfit because they are off-
target to the current sample. This is a hazard of vertical equating. **CUTLO**= and **CUTHI**= may remedy targeting
deficiencies.
Step 3a. If the best-fit slope remains far from 1.0, then there is something systematically different about Test A and Test B. You must do "Celsius - Fahrenheit" equating. Test A remains as it stands.
The slope of the best fit is: slope = (S.D. of Test B common items) / (S.D. of Test A common items)
Include in the Test B control file:
USCALE = the value of 1/slope
UMEAN = the value of the x-intercept
and reanalyze Test B. Test B is now in the Test A frame of reference, and the person measures from Test A and Test B can be reported together.

Step 3b. The best-fit slope is near to 1.0. Suppose that Test A is the "benchmark" test. Then we do not want responses to Test B to change the results of Test A.
From a Test A analysis produce IFILE= and SFILE= (if there are rating or partial credit scales).
Edit the IFILE= and SFILE= to match Test B item numbers and rating (or partial credit) scale.
Use them as an IAFILE= and SAFILE= in a Test B analysis.
Test B is now in the same frame of reference as Test A, so the person measures and item difficulties can be reported together.

Step 3c. The best-fit slope is near to 1.0. Test A and Test B have equal status, and you want to use both to define the common items.
Use the MFORMS= command to combine the data files for Test A and Test B into one analysis. The results of that analysis will have Test A and Test B items and persons reported together.

Partial Credit items
"Partial credit" values are much less stable than dichotomies. Rather than trying to equate across the whole partial credit structure, one usually needs to assert that, for each item, a particular "threshold" or "step" is the critical one for equating purposes. Then use the difficulties of those thresholds for equating. This relevant threshold for an item is usually the transition point between the two most frequently observed categories - the Rasch-Andrich threshold - and so the most stable point in the partial credit structure.

Stocking and Lord iterative procedure
The Stocking and Lord (1983) present an iterative common-item procedure in which items exhibiting DIF across tests are dropped from the link until no items exhibiting inter-test DIF remain. A known hazard is that if the DIF distribution is skewed, the procedure trims the longer tail and the equating will be biased. To implement the Stocking and Lord procedure in Winsteps, code each person (in the person id label) according to which test form was taken. Then request a DIF analysis of item x person-test-code (Table 30). Drop items exhibiting DIF from the link, by coding them as different items in different tests.


Common Person Equating
Some persons have taken both tests, preferably at least 5 spread out across the ability continuum.

Step 1. From the separate analyses, crossplot the abilities of the common persons, with Test B on the y-axis and Test A on the x-axis. The slope of the best-fit line i.e., the line though the point at the means of the common
persons and through the \((\text{mean} + 1 \text{ S.D.})\) point should have slope near 1.0. If it does, then the intercept of the line with the x-axis is the equating constant.

First approximation: Test B measures in the Test A frame of reference = Test B measure + x-axis intercept.

Step 2. Examine the scatterplot. Points far away from the best fit line indicate persons that have behaved differently on the two occasions. You may wish to consider these to be no longer common persons. Drop the persons from the plot and redraw the best fit line.

Step 3a. If the best-fit slope remains far from 1.0, then there is something systematically different about Test A and Test B. You must do "Celsius - Fahrenheit" equating. Test A remains as it stands. The slope of the best fit is: \(\text{slope} = (\text{S.D. of Test B common persons}) / (\text{S.D. of Test A common persons})\)
Include in the Test B control file:
USCALE = the value of \(1/\text{slope}\)
UMEAN = the value of the x-intercept
and reanalyze Test B. Test B is now in the Test A frame of reference, and the person measures from Test A and Test B can be reported together.

Step 3b. The best-fit slope is near to 1.0. Suppose that Test A is the "benchmark" test. Then we do not want responses to Test B to change the results of Test A.
From a Test A analysis produce PFILE=
Edit the PFILE= to match Test B person numbers
Use it as a PFILE= in a Test B analysis.
Test B is now in the same frame of reference as Test A, so the person measures and person difficulties can be reported together.

Step 3c. The best-fit slope is near to 1.0. Test A and Test B have equal status, and you want to use both to define the common persons.
Use your text editor or word processor to append the common persons' Test B responses after their Test A ones, as in the desing below. Then put the rest of the Test B responses after the Test A responses, but aligned in columns with the common persons's Test B responses. Perform an analysis of the combined data set. The results of that analysis will have Test A and Test B persons and persons reported together.

<table>
<thead>
<tr>
<th>Test A items</th>
<th>Test B items</th>
</tr>
</thead>
<tbody>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
</tbody>
</table>
| --------------|--------------| Common Person
| --------------|--------------|
| --------------|--------------| Common Person
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>--------------</td>
<td>--------------</td>
</tr>
</tbody>
</table>
| --------------|--------------| Common Person
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------</td>
</tr>
</tbody>
</table>
| --------------|--------------| Common Person
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------</td>
</tr>
</tbody>
</table>

Virtual Equating of Test Forms
The two tests share no items or persons in common, but the items cover similar material.

Step 1. Identify pairs of items of similar content and difficulty in the two tests. Be generous about interpreting "similar" at this stage.
Steps 2-4: simple: The two item hierarchies (Table 1 using short clear item labels) are printed and compared, equivalent items are identified. The sheets of paper are moved vertically relative to each other until the overall hierarchy makes the most sense. The value on Test A corresponding to the zero on Test B is the UMEAN= value to use for Test B. If the item spacing on one test appear expanded or compressed relative to the other test, use USCALE= to compensate.

Or:

Step 2. From the separate analyses, crossplot the difficulties of the pairs of items, with Test B on the y-axis and Test A on the x-axis. The slope of the best-fit line i.e., the line though the point at the means of the common items and through the (mean + 1 S.D.) point should have slope near 1.0. If it does, then the intercept of the line with the x-axis is the equating constant.

First approximation: Test B measures in the Test A frame of reference = Test B measure + x-axis intercept.

Step 3. Examine the scatterplot. Points far away from the best fit line indicate items that are not good pairs. You may wish to consider these to be no longer paired. Drop the items from the plot and redraw the best fit line.

Step 4. The slope of the best fit is: slope = (S.D. of Test B common items) / (S.D. of Test A common items)
Include in the Test B control file:
USCALE = the value of 1/slope
UMEAN = the value of the x-intercept
and reanalyze Test B. Test B is now in the Test A frame of reference, and the person and item measures from Test A and Test B can be reported together.

Random Equivalence Equating
The samples of persons who took both tests are believed to be randomly equivalent. Or, less commonly, the samples of items in the tests are believed to be randomly equivalent.

Step 1. From the separate analyses of Test A and Test B, obtain the means and sample standard deviations of the two person samples (including extreme scores).

Step 2. To bring Test B into the frame of reference of Test A, adjust by the difference between the means of the person samples and user-rescale by the ratio of their sample standard deviations.

Include in the Test B control file:
USCALE = value of (S.D. person sample for Test A) / (S.D. person sample for Test B)
UMEAN = value of (mean for Test A) - (mean for Test B * USCALE)
and reanalyze Test B.

Check: Test B should now report the same sample mean and sample standard deviation as Test A.

Test B is now in the Test A frame of reference, and the person measures from Test A and Test B can be reported together.

Linking Tests with Common Items
Here is an example:
A. The first test (50 items, 1,000 students)
B. The second test (60 items, 1,000 students)
C. A linking test (20 items from the first test, 25 from the second test, 250 students)

Here is a typical Rasch approach. It is equivalent to applying the "common item" linking method twice.

(a) Rasch analyze each test separately to verify that all is correct.

(b) Cross-plot the item difficulties for the 20 common items between the first test and the linking test. Verify that the link items are on a statistical trend line parallel to the identity line. Omit from the list of linking items, any items that have clearly changed relative difficulty. If the slope of the trend line is not parallel to the identity line (45 degrees), then the test discrimination has changed. The test linking will use a "best fit to trend line" conversion:
Corrected measure on test 2 in test 1 frame-of-reference =
\[\frac{(\text{observed measure on test 2} - \text{mean measure of test 2 link items}) \times (\text{SD of test 1 link items})}{\text{SD of test 1 link items}} + \text{mean measure of test 1 link items}\]

(c) Cross-plot the item difficulties for the 25 common items between the second test and the linking test. Repeat (b).

(d1) If both trend lines are approximately parallel to the identity line, than all three tests are equally discriminating, and the simplest equating is "concurrent". Put all 3 tests in one analysis. You can use the MFORMS= command to put all items into one analysis. You can also selectively delete items using the Specification pull-down menu in order to construct measure-to-raw score conversion tables for each test, if necessary. Or you can use a direct arithmetical adjustment to the measures based on the mean differences of the common items: www.rasch.org/memo42.htm "Linking tests".

(d2) If best-fit trend lines are not parallel to the identity line, then tests have different discriminations. Equate the first test to the linking test, and then the linking test to the second test, using the "best fit to trend line" conversion, shown in (b) above. You can also apply the "best fit to trend" conversion to Table 20 to convert every possible raw score.

299. Estimation bias correction - warnings

At least two sources of estimation error are reported in the literature.

An "estimation bias" - this is usually negligibly small after the administration of 10 dichotomous items (and fewer rating scale items). Its size depends on the probability of observing extreme score vectors. For a two item test, the item measure differences are twice their theoretical values, reducing as test length increases. This can be corrected. STBIAS= does this approximately, but is only required if exact probability inferences are to be made from logit measure differences. The experimental XMILE= option in Winsteps does this more exactly. But these corrections have their drawbacks.

A "statistical inflation". Since error variance always adds to observed variance, individual measures are always reported to be further apart (on average) than they really are. This cannot be corrected, in general, at an individual- measure level, because, for any particular measurement it cannot be known to what extent that measurement is biased by measurement error. However, if it is hypothesized that the persons, for instance, follow a normal distribution of known mean and standard deviation, this can be imposed on the estimates (as in MMLE) and the global effects of the estimate dispersion inflation removed. This is done in some other Rasch estimation software.

Estimation Bias

All Rasch estimation methods have some amount of estimation bias (which has no relationship with demographic bias). The estimation algorithm used by Winsteps, JMLE, has a slight bias in measures estimated from most datasets. The effect of the bias is to spread out the measures more widely than the data indicate. In practice, a test of more than 20 dichotomous items administered to a reasonably large sample will produce measures with inconsequential estimation bias. Estimation bias is only of concern when exact probabilistic inferences are to be made from short tests or small samples. Ben Wright opted for JMLE in the late 1960's because users were rarely concerned about such exact inferences, but they were concerned to obtain speedy, robust, verifiable results from messy data sets with unknown latent parameter distributions. Both of the identifiable sources of error are reduced by giving longer tests to bigger samples. With short tests, or small samples, other threats to validity tend to be of greater concern than the inflationary ones.

If estimation bias would be observed even with an infinitely large sample (which it would be with JMLE), then the estimation method is labeled "statistically inconsistent" (even though the estimates are predictable and logical). This sounds alarming but the inconsistency is usually inconsequential, or can be easily corrected in the unlikely event that it does have substantive consequences.

The JMLE joint likelihood estimation algorithm produces estimates that have a usually small statistical bias. This bias increases the spread of measures and calibrations, but usually less than the standard error of measurement.
The bias quickly becomes insignificantly small as the number of persons and items increases. The reason that JMLE is statistically inconsistent under some conditions, and noticeably biased for short tests or small samples, is that it includes the possibility of extreme scores in the estimation space, but cannot actually estimate them. Inconsistency doesn't really matter, because it asks "if we have infinite data, will the estimation method produce the correct answer?" Estimation bias, also called statistical bias, is more important because it asks "How near to correct are the estimates with finite data?" In practice, JMLE bias is smaller than the other sources of noise in the data. See Ben Wright's comments at [www.rasch.org/memo45.htm](http://www.rasch.org/memo45.htm)

For paired comparisons and very short tests, estimation can double the apparent spread of the measures, artificially inflating test reliability. This can be eliminated by specifying PAIRED=YES.

Correcting for bias may be helpful when it is desired to draw exact probabilistic inferences for small, complete datasets without anchoring.

Correcting for bias may be misleading, or may be supressed by Winsteps, in the presence of missing data or anchored persons or items.

Bias correction can produce apparently inconsistent measures if bias-corrected measures, estimated from an unanchored analysis, are then used to anchor that same dataset.

Estimation correction methods:

STBIAS=YES implements a variant of the simple bias correction proposed in Wright, B.D. and Douglas, G.A. (1977). Best procedures for sample-free item analysis. Applied Psychological Measurement, 1, 281-294. With large samples, a useful correction for bias is to multiply the estimated measures by (L-1)/L, where L is the smaller of the average person or item response count, so, for paired comparisons, multiply by 0.5. This is done automatically when PAIRED=YES.

XMLE=YES implements an experimental more sophisticated bias correction.

Other Rasch programs may or may not attempt to correct for estimation bias. When comparing results from other programs, try both STBIAS=Y and STBIAS=N to find the closest match. See also XMLE=

Estimation methods with less bias under sum circumstances include CMLE and MMLE, but these have other limitations or restrictions which are deemed to outweigh their benefits for most uses.

Technical information:

Statistical estimation bias correction with JMLE is relevant when you wish to make exact probabilistic statements about differences between measures for short tests or small samples. The (L-1)/L correction applies to items on short dichotomous tests with large samples, where L is the number of non-extreme items on a test. For long dichotomous tests with small samples, the correction to person measures would be (N-1)/N. Consequently Winsteps uses a bias correction on dichotomous tests for items of (L-1)/L and for persons of (N-1)/N.

The reason for this correction is because the sample space does not match the estimation space. The difference is extreme score vectors. Estimation bias manifests itself as estimated measures which are more dispersed than the unbiased measures. The less likely an extreme score vector, the smaller the correction to eliminate bias. Extreme score vectors are less likely with polytomies than with dichotomies so the bias correction is smaller. For example, if an instrument uses a rating scale with m categories, then Winsteps corrects the item measures by (m-1)(L-1)/((m-1)(L-1)+1) and person measures by (m-1)(N-1)/((m-1)(N-1)+1) - but these are rough approximations.

Since the mathematics of bias correction is complex and involves person and item parameter distributions, dichotomies or polytomies, anchored parameter values, person and item weighting, and missing data patterns, I have devised an experimental algorithm, XMLE, which attempts to do an essentially exact adjustment for bias under all conditions. You can activate this in Winsteps with XMLE=Yes. You can compare results of an XMLE=YES run with those of a standard run with STBIAS=NO, to see the size of the bias with your data. This is more exact than STBIAS=YES. It is experimental because the algorithm has yet to be checked for its behavior under many conditions, and so may malfunction - (it definitely does with extreme scores, for which reasonable bias-corrected estimates have yet to be devised).
With most Rasch software using CMLE, PMLE or MMLE bias correction of item measures is not done because the estimation bias in the item difficulties is generally very small. Bias correction of person abilities is not done though estimation bias exists.

Interaction terms are computed in an artificial situation in which the abilities and difficulties estimates are treated as known. Estimation bias is a minor effect in the interaction estimates. It would tend to increase very slightly the probability that differences between interaction estimates are reported as significant. So this is another reason to interpret DIF tests conservatively. If the number of relevant observations for an interaction term is big enough for the DIF effect to be regarded as real, and not a sampling accident, then the estimation bias will be very small. In the worst case, the multiplier would be of the order of \((C-1)/C\) where \(C\) is the number of relevant observations.

**Comparing Estimates**

Bigsteps and Winsteps should produce the same estimates when

(a) they are run with very tight convergence criteria, e.g.,
    RCONV=.00001
    LCONV=.00001
    MUCON=0

(b) they have the same statistical bias adjustment
    STBIAS=YES ; estimates will be wider spread
    or
    STBIAS=NO ; estimates will be narrower

(c) they have the same extreme score adjustment
    EXTRSC=0.5

The item estimates in BTD were produced with statistical bias adjustment, but with convergence criteria that would be considered loose today. Tighter convergence produces a wider logit spread. So the BTD item estimates are slightly more central than Winsteps or Bigsteps.

Winsteps and Bigsteps are designed to be symmetric. Transpose persons and items, and the only change is the sign of the estimates and an adjustment for local origin. The output reported in BTD (and by most modern Rasch programs) is not symmetric. So the person measure estimates in BTD are somewhat different.

**300. Estimation methods: JMLE, PROX, XMLE**

Winsteps implements three methods of estimating Rasch parameters from ordered qualitative observations: JMLE, PROX and XMLE. Estimates of the Rasch measures are obtained by iterating through the data. Initially all unanchored parameter estimates (measures) are set to zero. Then the PROX method is employed to obtain rough estimates. Each iteration through the data improves the PROX estimates until they are usefully good. Then those PROX estimates are the initial estimates for JMLE which fine-tunes them, again by iterating through the data, in order to obtain the final JMLE estimates. The iterative process ceases when the convergence criteria are met. These are set by MJMLE=, CONVERGE=, LCONV= and RCONV=. Depending on the data design, this process can take hundreds of iterations (Convergence: Statistics or Substance?). When only rough estimates are needed, force convergence by pressing Ctrl+F or by selecting “Finish iterating” on the File pull-down menu.

**Extreme scores:** (perfect, maximum possible scores, and zero, minimum possible scores) are dropped from the main estimation procedure. Their measures are estimated separately using EXTRSC=.

**Missing data:** most Rasch estimation methods do not require that missing data be imputed, or that there be case-wise or list-wise omission of data records with missing data. For datasets that accord with the Rasch model, missing data lower the precision of the measures and lessen the sensitivity of the fit statistics, but do not bias the measure estimates.

**Likelihood:** Using the current parameter estimates (Rasch measures), the probability of observing each data point is computed, assuming the data fit the model. The probabilities of all the data points are multiplied together
to obtain the likelihood of the entire data set. The parameter estimates are then improved (in accordance with the estimation method) and a new likelihood for the data is obtained. The values of the parameters for which the likelihood of the data has its maximum are the "maximum likelihood estimates" (Ronald A. Fisher, 1922).

**JMLE** "Joint Maximum Likelihood Estimation" is also called UCON, "Unconditional maximum likelihood estimation". It was devised by Wright & Panchapakesan. In this formulation, the estimate of the Rasch parameter (for which the observed data are most likely, assuming those data fit the Rasch model) occurs when the observed raw score for the parameter matches the expected raw score. "Joint" means that the estimates for the persons (rows) and items (columns) and rating scale structures (if any) of the data matrix are obtained simultaneously. The iterative estimation process is described at [Iteration](#).

**Advantages** - these are implementation dependent, and are implemented in Winsteps:
1. independence from specific person and item distributional forms.
2. flexibility with missing data
3. the ability to analyze test lengths and sample sizes of any size
4. symmetrical analysis of person and item parameters so that transposing rows and columns does not change the estimates
5. flexibility with person, item and rating scale structure anchor values
6. flexibility to include different variants of the Rasch model in the same analysis (dichotomous, rating scale, partial credit, etc.)
7. unobserved intermediate categories of rating scales can be maintained in the estimation with exact probabilities.
8. all non-extreme score estimable (after elimination of extreme scores and rarely-observed Guttman subsets)
9. all persons with the same total raw scores on the same items have the same measures; all items with the same raw scores across the same persons have the same measures.

**Disadvantages:**
11. measures for extreme (zero, perfect) scores for persons or items require post-hoc estimation.
12. estimates are statistically inconsistent
13. estimation bias, particularly with small samples or short tests, inflates the logit distance between estimates.
14. chi-squares reported for fit tests (particularly global fit tests) may be somewhat inflated, exaggerating misfit to the Rasch model.

Comment on (8): An on-going debate is whether measures should be adjusted up or down based on the misfit in response patterns. With conventional test scoring and Rasch JMLE, a lucky guess counts as a correct answer exactly like any other correct answer. Unexpected responses can be identified by fit statistics. With the three-parameter-logistic item-response-theory (3-PL IRT) model, the score value of an unexpected correct answer is diminished whether it is a lucky guess or due to special knowledge. In Winsteps, responses to off-target items (the locations of lucky guesses and careless mistakes) can be trimmed with [CUTLO=](#) and [CUTHI=](#), or be diminished using [TARGET=](#). III

Comment on (13): JMLE exhibits some estimation bias in small data sets (for reasons, see XMLE below), but this rarely exceeds the precision (model standard error of measurement, SEM) of the measures. Estimation bias is only of concern when exact probabilistic inferences are to be made from short tests or small samples. It can be exactly corrected for paired-comparison data with [PAIRED=](#). For other data, it can be approximately corrected with [STBIAS=](#) Yes, but, in practice, this is not necessary (and sometimes not advisable).

**PROX** is the Normal Approximation Algorithm devised of Cohen (1979). This algorithm capitalizes on the similar shapes of the logistic and normal ogives. It models both the persons and the items to be normally distributed. The variant of **PROX** implemented in Winsteps allows missing data. The form of the estimation equations is:

\[
\text{Ability of person} = \text{Mean difficulty of items encountered} + \\
\log \left( \frac{(\text{observed score} - \text{minimum possible score on items encountered})}{(\text{maximum possible score on items encountered} - \text{observed score})} \right) \times \sqrt{1 + \frac{\text{variance of difficulty of items encountered}}{2.9}}
\]

In Winsteps, PROX iterations cease when the variance of the items encountered does not increase substantially from one iteration to the next.

**Advantages** - these are implementation dependent, and are implemented in Winsteps:
Computationally the fastest estimation method.

Disadvantages
(1) Person and item measures assumed to be normally distributed.
(11)-(14) of JMLE

Other estimation methods in common use (but not implemented in Winsteps):

- **Gaussian least-squares** finds the Rasch parameter values which minimize the overall difference between the observations and their expectations, \( \text{Sum}(X_{ni} - Eni)^2 \) where the sum is overall all observations, \( X_{ni} \) is the observation when person encounters item \( i \), and \( Eni \) is the expected value of the observation according to the current Rasch parameter estimates. For Effectively, off-target observations are down-weighted, similar to \( \text{TARGET}=\text{Yes} \) in Winsteps.

- **Minimum chi-square** finds the Rasch parameter values which minimize the overall statistical misfit of the data to the model, \( \text{Sum}((X_{ni} - Eni)^2 / V_{ni}) \) where \( V_{ni} \) is the modeled binomial or multinomial variance of the observation around its expectation. Effectively off-target observations are up-weighted to make them less improbable.

Gaussian least-squares and Minimum chi-square:
Advantages - *these are implementation dependent*:
(1)-(8) All those of JMLE.

Disadvantages:
(9) persons with the same total raw scores on the same items generally have different measures; items with the same raw scores across the same persons generally have different measures.
(11)-(13) of JMLE
(14) global fit tests uncertain.

- **CMLE.** Conditional maximum likelihood estimation. Item difficulties are structural parameters. Person abilities are incidental parameters, conditioned out for item difficulty estimation by means of their raw scores. The item difficulty estimates are those that maximize the likelihood of the data given the person raw scores and assuming the data fit the model. The item difficulties are then used for person ability estimation using a JMLE approach.

Advantages - *these are implementation dependent*:
(1), (6)-(9) of JMLE
(3) the ability to analyze person sample sizes of any size
(5) flexibility with item and rating scale structure anchor values
(12) statistically-consistent item estimates
(13) minimally estimation-biased item estimates
(14) exact global fit statistics

Disadvantages:
(2) limited flexibility with missing data
(3) test length severely limited by mathematical precision of the computer
(4) asymmetric analysis of person and item parameters so that transposing rows and columns changes the estimates
(5) no person anchor values
(11) of JMLE
(13) estimation-biased of person estimates small but uncertain

- **MMLE.** Marginal maximum likelihood estimation. Item difficulties are structural parameters. Person abilities are incidental parameters, integrated out for item difficulty estimation by imputing a person measure distribution. The item difficulties are then used for person ability estimation using a JMLE approach.

Advantages - *these are implementation dependent*:
(3), (6)-(9) of JMLE
(1) independence from specific item distributional forms.
(2) flexibility with missing data extends to minimal length person response strings
(5) flexibility with item and rating scale structure anchor values
(11) extreme (zero, perfect) scores for persons are used for item estimation.
(12) statistically-consistent item estimates
(13) minimally estimation-biased item estimates
(14) exact global fit statistics

Disadvantages:
(1) specific person distribution required
(4) asymmetric analysis of person and item parameters so that transposing rows and columns changes the estimates
(5) no person anchor values
(11) measures for extreme (zero, perfect) scores for specific persons or items require post-hoc estimation.
(13) estimation-biased of person estimates small but uncertain

**PMLE.** Pairwise maximum likelihood estimation. Person abilities are incidental parameters, conditioned out for item difficulty estimation by means of pairing equivalent person observations. The item difficulties are then used for person ability estimation using a JMLE approach.

**Advantages** - these are implementation dependent:
(1), (3), (6), (7) of JMLE
(5) flexibility with item and rating scale structure anchor values
(8) all persons with the same total raw scores on the same items have the same measure
(12) statistically-consistent item estimates

Disadvantages:
(11) of JMLE
(2) reduced flexibility with missing data
(4) asymmetric analysis of person and item parameters so that transposing rows and columns changes the estimates
(5) no person anchor values
(8) items with the same total raw scores across the same persons generally have different measures.
(13) estimation-biased or item and person estimates small but uncertain
(14) global fit tests uncertain.
(15) uneven use of data in estimation renders standard errors and estimates less secure

**WMLE.** Warm's (1989) Weighted Maximum Likelihood Estimation. Standard MLE estimates are the maximum values of the likelihood function and so statistical modes. Warm shows that the likelihood function is skewed, leading to an additional source of estimation bias. The mean likelihood estimate is less biased. Warm suggests an unbiasing correction that can be applied, in principle, to any MLE method, but there are computational constraints. Even when feasible, this fine tuning appears to be less than the relevant standard errors and have no practical benefit. It is not currently implemented in Winsteps.

**XMLE.** "Exclusory Maximum Likelihood Estimation", implements Linacre's (1989) XCON algorithm in Winsteps. Statistical "consistency" is the property that an estimation method will yield the "true" value of a parameter when there is infinite data. Statistical "estimation bias" is the degree to which an estimate differs from its "true" value with a finite amount of data. JMLE is statistically inconsistent under some conditions, and noticeably estimation-biased for short tests or small samples, because it includes the possibility of extreme scores in the estimation space, but cannot estimate them. The XMLE algorithm removes the possibility of extreme response vectors from the estimation space, to a first approximation. This makes XMLE consistent, and much less estimation-biased than JMLE. In fact XMLE is even less biased than CMLE for small samples, this is because CMLE only eliminates the possibility of extreme person response vectors, not the possibility of extreme item response vectors.

XMLE and JMLE use the same estimation methods. The difference is in the probability terms used in the estimation equations.
For JMLE, for the dichotomous case,
\[
\ln (\frac{P_{ni1}}{P_{ni0}}) = B_n - D_i
\]
where \(P_{ni1}\) is the probability that person \(n\) succeeds on item \(i\). For XMLE,
\[
R_{ni1} = P_{ni1} - \text{Product}(P_{mi1}) - \text{Product}(P_{nj1}) + \text{Product}(P_{mi1}) \times \text{Product}(P_{nj1})
\]
where \( m = 1, N \) and \( j = 1, L \), so that \( \text{Product}(P_{mi1}) \) is the likelihood of a perfect-score for person \( n \), and \( \text{Product}(P_{nj1}) \) is the likelihood of the sample all succeeding on item \( i \). Similarly, 
\[
R_{ni0} = P_{ni0} - \text{Product}(P_{mi0}) - \text{Product}(P_{nj0}) + \text{Product}(P_{mi0}) \times \text{Product}(P_{nj0})
\]
So the JMLE estimation equation for person \( n \) or item \( i \) is based on 
\[
\text{Expected raw score} = \sum \left( \frac{R_{ni1}}{R_{ni1} + R_{ni0}} \right)
\]

**Example:** Consider a two-item dichotomous test. Possible person scores are 0, 1, 2. Person scores of 0 and 2 are dropped from estimation as extreme. The remaining very large sample of \( N \) persons all score 1 success so all have the same measure, 0 logits for convenience. Twice as many successes are observed on item 1 as item 2. Under these conditions, in the estimation sample, a success on item 1 requires a failure on item 2 and vice-versa. So, according to the Rasch model, the logit distance between item 1 and item 2 = \( \log(\text{frequency of success on item 1} / \text{frequency of success on item 2}) = \log(2) \). And the expected score on item 1 is 2/3 and on item 2 is 1/3.

JMLE considers observations of item 1 and item 2 to be independent of the total raw score, and computes the distance between item 1 and item 2 = \( \log(\text{frequency of success on item 1} / \text{frequency of success on item 2}) - \log(\text{frequency of failure on item 1} / \text{frequency of failure on item 2}) \). This is the worst case of JMLE estimation bias and occurs with pairwise comparison data. For such data, this estimation-bias of 2 can be automatically corrected with PAIRED=Yes. As test length increases, the bias reduces and is considered to be non-consequential for test lengths of 10 items or more, Wright’s Memo 45.

Considerations with XMLE=YES include:
1. Anchoring values changes the XMLE probabilities. Consequently, measures from a Table 20 score table do not match measures from the estimation run. Consequently, it may be necessary to estimate item calibrations with XMLE=YES. Then anchor the items and perform XMLE=NO.
2. Items and persons with extreme (zero and perfect) scores are deleted from the analysis.
3. For particular data structures, measures for finite scores may not be calculable.

Advantages - these are implementation dependent, and are implemented in Winsteps:
1. (1)-(8) of JMLE
2. (12) estimates are statistically consistent
3. (13) estimation bias is small

Disadvantages:
1. (11) measures for extreme (zero, perfect) scores for persons or items require post-hoc estimation, and even then may not be estimable
2. (14) global fit tests uncertain


**301. Exact Match: OBS% and EXP%**
Suppose your dataset consists of observations, \( \{X_{ni}\} \), of person \( n \) on item \( i \). Based on the Rasch parameters (measures), there is an expected value \( E_{ni} \) corresponding to each observation \( X_{ni} \). \( E_{ni} \) is obtained by a calculation from the Rasch model.

When the absolute value of \( (X_{ni}-E_{ni}) \) is less than 0.5 then the observed data point is within 0.5 score points of its expected value, so the match is the closest possible. Thus, across all observations of item \( i \),

- \( \text{Count} \ ( (X_{ni}-E_{ni}) < 0.5 ) = A \) - these observations are of the closest categories to their expectations
- \( \text{Count} \ ( (X_{ni}-E_{ni}) = 0.5 ) = B \) - these observations are on the borderline of matching their expectations
- \( \text{Count} \ ( (X_{ni}-E_{ni}) > 0.5 ) = C \) - these observations are at least one category away from their expectations

So that \( A+B+C = \text{Count} \ (X_{ni}) \)

\[ \text{OBS\%} = \frac{100 \times (A + B/2)}{A + B + C} \]

Each possible value of \( X_{ni} \) has a probability according to the Rasch model. Based on these, the expected value of \( \text{OBS\%} \) can be computed, this is the \( \text{EXP\%} \). So, if the possible values of \( X_{ni} \) are \( j=0,1,2,...,m \), with probabilities \( P_{nij} \), then

- \( A = \sum ( (j-E_{ni})<0.5 ) \times P_{nij} \)
- \( B = \sum ( (j-E_{ni})=0.5 ) \times P_{nij} \)
- \( C = \sum ( (j-E_{ni})>0.5 ) \times P_{nij} \)

So that \( A+B+C = \text{Count} \ (X_{ni}) \)

\[ \text{EXP\%} = \frac{100 \times (A + B/2)}{A + B + C} \]

If \( \text{OBS\%} < \text{EXP\%} \) then the local data are more random than the model predicts.
If \( \text{OBS\%} > \text{EXP\%} \) then the local data are more predictable than the model predicts.

302. **Exporting Tables to EXCEL**

Exporting Winsteps Tables to Excel is easy.

Produce the Winsteps Table using the "Output Tables" pull-down menu.

"Select" with the mouse the part you want (Usually column heading lines to the bottom of the Table) then right-click "Copy"

Start Excel (e.g., from the Winsteps Files menu)

Paste into Excel - top left cell. Everything will go into the first column.

Under "Data" go "Text to columns"

Excel usually gets everything exactly right, so that each Winsteps table column is in a separate Excel column.

Done!

303. **Extra Specifications prompt**

WINSTEPS expects to find the control variables in your control file. You may, however, specify one or more control variables on the "Extra specifications" line. These variables supersede instructions in the control file. This is useful for making temporary changes to the control variables. **There are special rules for blanks** (see Example 3). You can turn off the Extra Specifications prompt from the Edit Initial Settings menu.

Example 1: You want to verify that your data is correctly formatted, so you only want to do one \( \text{JMLE} \) iteration this time, i.e., you want to set \( \text{JMLE}=1 \) for this run only:

Please enter name of WINSTEPS control file: \( \text{SF.TXT} \)(Enter)

Please enter name of report output file: \( \text{SFO.TXT} \)(Enter)
Extra specifications? (e.g., MJMLE = 1), or press Enter:
MJMLE = 1 (Enter)

Note:
Extra specifications? (e.g., MJMLE = 1), or press Enter:
MJMLE = 1 (Enter)
is invalid because there are blanks in MJMLE = 1.

Example 2: You want to produce the fit plot in Table 4 with specially chosen ranges on the axes:
Please enter name of WINSTEPS control file: SF.TXT (Enter)
Please enter name of report output file: SFO.TXT (Enter)
Extra specifications? (e.g., MJMLE = 1), or press Enter:
TABLES = 0001 MRANGE = 3 FRANGE = 4 (Enter)

Example 3: To put blanks in an Extra Specification. Put the whole specification within " " (double quotes).
Put the argument within ' ' (single quotes). E.g., You want the title TITLE = to be: Analysis B, and UMEAN = 50.
Please enter name of WINSTEPS control file: SF.TXT (Enter)
Please enter name of report output file: SFO.TXT (Enter)
Extra specifications? (e.g., MJMLE = 1), or press Enter:
"Title = 'Analysis B' " UMEAN = 50

304. Extreme scores: what happens

Extreme scores are the lowest and highest possible scores for persons on items, or for items by persons. They include zero and perfect scores. They are shown in the Tables as MINIMUM ESTIMATE MEASURE and MAXIMUM ESTIMATE MEASURE.

Mathematically, they correspond to infinite or indefinite measures on the latent variable and so are not directly estimable. Accordingly persons or items with extreme scores are dropped for the duration of the measurement estimation process. The extreme persons are dropped casewise. The extreme items are dropped listwise.

Sometimes the effect of dropping extreme items and persons is to make other items and persons extreme. If so, these are also dropped. If the data have a Guttman pattern, ultimately all items and persons are dropped and the measures for that data set are reported as inestimable.

After the measures of all non-extreme items and persons have been estimated, then the extreme scores are reinstated. Reasonable extreme measures are imputed for them (using a Bayesian approach), so that all persons and items have measures.

See Extremescore =

305. Global fit statistics

Winsteps reports global fit statistics and approximate global log-likelihood chi-square statistic in Table 3.1. The variance tables report the relative sizes of explained and unexplained variances.

The chi-square value is approximate. It is based on the current reported estimates which may depart noticeably from the "true" maximum likelihood estimates for these data. The degrees of freedom are the number of datapoints used in the free estimation (i.e., excluding missing data, data in extreme scores, etc.) less the number of free parameters. The number of free parameters is the least number of parameters from which all the Rasch expected observations could be constructed. For complete data, this is the lower of the number of different observed person marginal scores or the number of different observed item marginal scores, less one for identifying the local origin.

If you wish to compute your own global (or any other) fit test, the response-level probabilities, residuals etc. are reported in the XFILE=. For instance, for a global fit test, you could add up all the log-probabilities. Then chi-square estimate = - 2 * log-probability. A different chi-square estimate is the sum of squared-standardized residuals. You can count up the number of free parameters. For complete dichotomous data, it is usually the minimum of (number of different person marginal raw scores, number of different item marginal scores) - 1.
Deviance statistics are more trustworthy. They are the difference between the chi-squares of two analyses, with d.f. of the difference between the number of free parameters estimated.

The Rasch model is an idealization, never achieved by real data. Accordingly, given enough data, we expect to see statistically significant misfit the model. If the current data do not misfit, we merely have to collect more data, and they will! In essence, the null hypothesis of this significance test is the wrong one! We learn nothing from testing the hypothesis, "Do the data fit the model (perfectly)?" Or, as usually expressed in social science, "Does the model fit the data (perfectly)?" Perfection is never obtained in empirical data. What we really want to test is the hypothesis "Do the data fit the model usefully?" And, if not, where is the misfit, and what is it? Is it big enough in size (not "statistical significance") to cause trouble? This is the approach used in much of industrial quality-control, and also in Winsteps.

306. GradeMap interface

To produce this plot from GradeMap, an analysis and graphical program distributed in association with the book "Constructing Measures: An Item Response Modeling Approach" by Mark Wilson (2004) Mahwah NJ: Lawrence Erlbaum Associates. bearcenter.berkeley.edu/GradeMap/

The GradeMap option on the Output Files menu displays this box which enables a simple conversion of Winsteps control and data files into GradeMap format. These files can then be imported by GradeMap - see the GradeMap User Guide. The files are displayed after they are created by Winsteps, so that you can edit them if you wish. Excel is used if it is available. If changes are made, save the file in tab-separated .txt format - which is expected by GradeMap. Here is a Model Specification File from Example0.txt.
Example GradeMap Dialog:
Winsteps: Create Model Specification file
Winsteps: Create Student Data file
Winsteps: Launch GradeMap
GradeMap screen displays:
User name: admin
Password: bear
Menu: File
Menu: New Project
Remove: Yes
Menu: System
Menu: Import Model Specification
Import data: (your Model Specification file, e.g., lfsitems.txt)
Menu: File
Menu: Import Student Data
Import Student Data: (your Student Data file, e.g., lfschildren.txt)
Menu: View
Menu: Select Item Set
Click on: T
Click on: OK
Menu: Estimation Tasks
Menu: Compute
Accept: Yes
EAP: No
Menu: Reports & Maps etc.

Be patient! GradeMap operates slowly with large files.

307. Guttman patterns

Psychometrician Louis Guttman (1916-1987) perceived the ideal test to be one in which a person succeeds on all the items up to a certain difficulty, and then fails on all the items above that difficulty. When persons and items are ordered by raw score, this produces a data set with a “Guttman pattern”. This is data is not analyzable in the usual way by Rasch analysis, because each person or item in turn becomes an extreme score. Here is a Guttman pattern with dichotomous data:

Easy->Hard items (columns)
1111111 Most able person (rows)
1111110
1111100
It is sometimes useful to make this type of data estimable by adding a dummy reversed-Guttman record:

Least able person

<table>
<thead>
<tr>
<th>1111100</th>
<th>Least able person</th>
</tr>
</thead>
<tbody>
<tr>
<td>1100000</td>
<td></td>
</tr>
<tr>
<td>1100000</td>
<td></td>
</tr>
<tr>
<td>1000000</td>
<td></td>
</tr>
</tbody>
</table>

0000000 Least able person

00000001 < Dummy person record

^ Dummy item record

or by anchoring the most extreme items (or persons) a conveniently long distance apart, e.g., 10 logits:

PAFILE=*  
1 10 ; anchor the first (highest score) person at 10 logits  
8 0 ; anchor the last (lowest score) person at 0 logits  
*

&END
END LABELS

308. Half-rounding

Rounding occurs when a number, such as 5.032 is to be displayed with only two decimal places, e.g., as 5.03.

The general rule followed by Winsteps is to round to the nearest displayable value. Examples:

5.034 rounds to 5.03  
5.036 rounds to 5.04  
-5.034 rounds to -5.03  
-5.036 rounds to -5.04

Rounding errors may arise with 5.035 and -5.035. Winsteps intends to round these away from zero to 5.04 and -5.04.

In practice, the computer arithmetic sometimes loses precision due to hardware limitations, so that 5.035 becomes an internal number like 5.034999997 - a number which is the same computationally as 5.035. But this value half-rounds to 5.03. This behavior is impossible to predict, because most displayed numbers are the result of a long chain of internal computations, .

Increasing the value of UDECIM= may display more decimal places, and so better display the numerical results.

In recent versions, Winsteps adds .0000005 to positive numbers, and subtracts that amount from negative numbers, in order that .005 will almost always round to .01.
309. How big an analysis can I do?

WINSTEPS will operate with a minimum of two observations per item or person, or even 1 observation with anchoring. This is useful for producing score tables with anchored items and dummy response strings. For statistically stable measures to be estimated, 30 observations per element are needed:

www.rasch.org/rmt/rmt74m.htm

The upper limit of WINSTEPS is 10,000,000+ persons. Winsteps can analyze 30,000+ items. There can be up to 255 ordinally numbered categories per item (polytomies, rating scales, partial credit etc.).

The widest logit range of measures that maintains computational precision is 90 logits, but useful results can be reported with a measure range up to 700 logits wide.

310. How long will an analysis take?

A PC with a math co-processor processes about 1,000,000 observations per minute. Most analyses have reached convergence within 20 iterations, so a rule of thumb is:

length of analysis in minutes = (number of persons)*(length of test)*2/100,000

311. Information - item and test

Fisher information is the amount of information data provide about a parameter. For item information, this is the amount of information the response to an item provides about a person parameter. For test information, this is the amount of information all of the items encountered by a person provide about a person parameter.

All dichotomous items have the same Rasch-derive item information. For a response of probability p, the item information is the binomial variance of the probability, p(1-p). This function is:

For polytomous items, the information function depends on the rating scale structure. The better targeted the item is on the person the more Fisher information the item provides about the person parameter. This motivates some item selection techniques for computer-adaptive testing.

The test information function is the sum of the item information functions. The standard error of a measure is the inverse square-root of the test information at the location of the maximum-likelihood parameter estimate (or higher). Here is the test information function for the Knox Cube Test in Example 1.
Test (sample) reliability is a summary of the interaction between the sample distribution and the test information function.

312. Item difficulty: definition

As modeled in Winsteps, the difficulty (challenge, easiness, etc.) of an item (task, prompt, etc.) is the point on the latent variable (unidimensional continuum) at which the highest and lowest category have equal probability of being observed.

For a dichotomous item, this is the point at which each category has a 50% probability of being observed.

For a Rasch-Andrich rating-scale item, this definition implies that the sum of the rating-scale-structure measures sum to zero relative to the item difficulty, i.e., the sum of the Rasch-Andrich thresholds is zero, i.e., \( \sum(F_j) = 0 \).

For a Masters partial-credit item, this definition implies that the item difficulty is the average of the difficulties of the Rasch-Masters thresholds for the item, i.e., \( D_i = \text{average}(D_{ij}) \), so that reparameterizing, \( D_{ij} = D_i + F_j \), then \( \sum(F_{ij}) = 0 \) for each item i.

313. Item discrimination or slope estimation

The Rasch model specifies that item discrimination, also called the item slope, be uniform across items. This supports additivity and construct stability. Winsteps estimates what the item discrimination parameter would have been if it had been parameterized. The Rasch slope is the average discrimination of all the items. It is not the mean of the individual slopes because discrimination parameters are non-linear. Mathematically, the average slope is set at 1.0 when the Rasch model is formulated in logits, or 1.70 when it is formulated in probits (as 2-PL and 3-PL usually are). 0.59 is the conversion from logits to probits.

The empirical discrimination is computed after first computing and anchoring the Rasch measures. In a post-hoc analysis, a discrimination parameter, \( a_i \), is estimated for each item. The estimation model is of the form:

\[
\log \left( \frac{P_{nij}}{P_{ni}(j-1)} \right) = a_i \left( B_{n} - \hat{D}_{i} - \hat{F}_{j} \right)
\]

This has the appearance of a 2-PL IRT or “Generalized Partial Credit” model, but differs because the discrimination or slope parameter is not used in the estimation of the other parameters. The reported values of item discrimination, \( \text{DISCR} \), are a first approximation to the precise value of \( a_i \) obtained from the Newton-Raphson estimation equation:

\[
\hat{a}_i = 1 + \frac{\sum_{n} (X_{ni} - P_{ni})(\theta_\alpha - b_i)}{\sum_{n} P_{ni}(1 - P_{ni})(\theta_\alpha - b_i)^2}
\]

The possible range of \( a_i \) is \(-\infty \) to \( +\infty \), where \( +\infty \) corresponds to a Guttman data pattern (perfect discrimination) and \( -\infty \) to a reversed Guttman pattern. Rasch estimation usually forces the average item discrimination to be near 1.0. Consequently an estimated discrimination of 1.0 accords with Rasch model expectations. Values greater than 1.0 indicate over-discrimination, and values less than 1.0 indicate under-discrimination. Over-discrimination is thought to be beneficial in many raw-score and IRT item analyses. High discrimination usually corresponds to
low MNSQ values, and low discrimination with high MNSQ values.

From an informal simulation study, Edward Wolfe reports Winsteps discrimination to have a .88 correlation with the generating slope parameters for a 2-PL dataset. BILOG has a .95 correlation.

Table 29.1 allows you to estimate the empirical item discrimination, at least as well as a 2-PL IRT computer program. This is because 2-PL discrimination estimation is degraded by the imputation of a person distribution and constraints on discrimination values. It is also skewed by accidental outliers which your eye can disregard. When Discrimination=Yes, exact computation is done in the measure tables.

In Table 29.1 draw in the line that, to your eye, matches the central slope of the empirical item characteristic curve (ICC).

Estimate the logit distance from where the line intercepts the .0 score value to where it intercepts the 1.0 score value (for dichotomies). The logit distance here is about 4.0 logits.

Use the central logit measure to logit discrimination line in this nomogram to estimate discrimination. In this nomogram, a logit distance of 4.0 logits, corresponds to a logit discrimination of 1.0, in accordance with model prediction. Steeper slopes, i.e., higher discriminations, correspond to shorter distances.
314. Iterations - PROX & JMLE

The Rasch model formulates a non-linear relationship between non-linear raw scores and linear measures. So, estimating measures from scores requires a non-linear process. This is performed by means of iteration. Two estimation methods are used, PROX and JMLE.

In Winsteps, initially every person is estimated to have the same ability measure at the origin of the measurement scale. Each item is estimated to have the same difficulty measure, also at the origin of the measurement scale. Each rating scale structure parameter, Rasch-Andrich threshold, is also estimated to be 0.

In Winsteps, the first phase of estimation uses the PROX (normal approximation) estimation algorithm. This takes the initial set of estimates and produces revised estimates:

\[ B_n = \mu_n + \sqrt{1 + \sigma_n^2 / 2.9} \log \frac{R_n}{(N_n - R_n)} \]

where \( B_n \) is the revised ability estimate for person \( n \), \( \mu_n \) is the mean difficulty of the items encountered by person \( n \), and \( \sigma_n \) is the standard deviation of those item difficulties. \( R_n \) is the observed raw score for person \( n \) and \( N_n \) is a perfect score on those same items. Similarly, for the items,

\[ D_i = \mu_i - \sqrt{1 + \sigma_i^2 / 2.9} \log \frac{R_i}{(N_i - R_i)} \]

where \( D_i \) is the revised difficulty estimate for item \( i \), \( \mu_i \) is the mean ability of the persons encountering by item \( i \), and \( \sigma_i \) is the standard deviation of those person abilities. \( R_i \) is the observed raw score on item \( i \) and \( N_i \) is a perfect score by those same persons.

To update these PROX estimates, Winsteps traverses the data computing the values of all the terms on the right-side of the estimation equations. This traversal is called an "iteration". When the increase in the range of the person or item measures is smaller than 0.5 logits, or when \( M_{PROX} = \) is reached, iteration ceases.

Initial estimates of the Rasch-Andrich threshold between category \( k \) and category \( k-1 \) are obtained from \( \log (\text{observed frequency of category } k-1 / \text{observed frequency of category } k) \) normalized to sum to zero across the thresholds of a rating scale.

The PROX estimates become the starting values for JMLE (Joint Maximum Likelihood Estimation). Using these person, item and rating scale structure estimates, Winsteps computes the expected value, according to the Rasch model, corresponding to each observation in term. After iterating through the entire data set, the marginal sums of these expected values, the person expected raw scores and the item expected raw scores, are compared with their observed (empirical) values. If a person's expected raw score is less than that person's observed raw score, then the ability estimate raised. If the person's expected raw score is greater than the observed score, then the ability estimate is lowered. If the item's expected raw score is less than the observed score, then the difficulty estimate is lowered. If the item's expected raw score is greater than the observed score, then the difficulty estimate is raised.

The estimation equations for JMLE are derived in RSA, where Newton-Raphson iteration is employed.

\[ y' = y + \frac{\text{observed score} - \text{Rasch expected score based on current estimates}}{\text{modeled variance}} \]

where \( y \) is a current estimated person measure and \( y' \) is the improved estimate.

Newton-Raphson estimation has proved unstable with sparse data sets and also with rating scales which have alternating very high and very low frequency categories. Accordingly, Winsteps implements a more robust proportional-curve-fitting algorithm to produce JMLE estimates. The relationship between raw scores and measures is always monotonic, so the characteristic curve for each person or item parameter is modeled to have the local form of a logistic ogive:

\[ y = a \cdot \log( (x-l)/(h-x) ) + c \]

where \( y \) is an estimated measure, \( a = \) slope of the ogive, \( x = \) a raw score, \( l = \) the known minimum possible raw score for the parameter, \( h = \) the known maximum possible raw score for the parameter, \( c = \) location of ogive relative to local origin.

Values of \( x \) are obtained form the current estimated measure \( y \) and a nearby measure \( (y + d) \). From these, \( a \) and \( c \) are estimated. The revised measure \( y' \) is obtained by evaluating the equation using the observed raw score as the value of \( x \). In the plot below for Example 1, the current estimate, \( y \), is -3 logits, a nearby estimate, \( y+d \), is -2 logits. These both estimate raw scores on the currently-estimated test characteristic curve (TCC, the remainder of
which is not yet known). The violet line is the logistic ogive going through these two known points. It is close to the putative TCC. The observed score of "5" is then found on the logistic ogive and an improved estimate is obtained. After all the person and item estimates are improved, the estimated TCC changes and this estimation process is repeated by performing another iteration through the data.

For the rating scale structure, the estimate, $y_k$, for Rasch-Andrich threshold $k$ is improved by

$$y_k' = y_k - \log \left( \frac{\text{observed count category } k}{\text{observed count category } k-1} \right) + \log \left( \frac{\text{estimated count category } k}{\text{estimated count category } k-1} \right)$$

When the various convergence criteria are satisfied, iteration ceases and the final estimates are obtained. These are used in computing fit statistics.

Example: Here is the iteration Table for example0.txt:

<table>
<thead>
<tr>
<th>CONVERGENCE TABLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROX</td>
</tr>
<tr>
<td>ITERATION</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

In the top section of the Convergence Table are reported the number of active persons, items and categories. The range of item and person measures at the end of each PROX iteration is shown, also the biggest change in any person or item, and in any Rasch-Andrich threshold. PROX iteration ceases with iteration 3 because the "KIDS" (persons) and "ACTS" (items) range has increased by less than 0.5 logits.

In the lower section, for each JMLE iteration, the maximum score residual, the biggest difference between any
observed and expected marginal score is shown. Also the biggest change in any measure. Iteration ceases when the values, in iteration 10, are less than the convergence criteria.

315. Local Dependence

In some data designs, data are collected from the same persons more than once, or observations are collected on equivalent items. Consequently there is reason to suspect that local dependence exists in the data. What is its impact on the Rasch measures?

Local dependence usually squeezes or stretches the logit measures, but does not usually change cut-points much when they are expressed in raw-score terms.

Here is an experiment to determine whether local dependence is a problem.

Assuming that data from the same persons may be a problem, select from your cases one of each different response string. This will make the data as heterogeneous as possible. Perform an analysis of this data set and see if that changes your conclusions markedly. If it does, then local dependence may be a concern. If it doesn't then local dependence is having no substantive impact.

Using Excel, a method of obtaining only one of each different response string:

0. Import the data into excel as a "character" column
1. from the Excel data pull down menu choose -> filter -> advanced filter
2. under "action" choose "copy to another location"
3. click "list range" and highlight the range of element numbers - if you want the whole column click on the letter at the top of the column
4. click "copy to" and choose an empty column, e.g., column J.
5. click "unique records only"
6. click "OK"
7. look at column J. The data are unique.

316. Logit and probit

When USCALE=1 (or USCALE= is omitted), measures are reported in logits. When USCALE=0.59, measures are reported in approximated probits.

**Logit:** A logit (log-odds unit) is a unit of interval measurement which is well-defined within the context of a single homogeneous test. When logit measures are compared between tests, their probabilistic meaning is maintained but their substantive meanings may differ. This is often the case when two tests of the same construct contain items of different types. Consequently, logit measures underlying different tests must be *equated* before the measures can be meaningfully compared. This situation is parallel to that in Physics when some temperatures are measured in degrees Fahrenheit, some in Celsius, and others in Kelvin.

As a first step in the *equating* process, plot the pairs of measures obtained for the same elements (e.g., persons) from the two tests. You can use this plot to make a quick estimate of the nature of the relationship between the two logit measurement frameworks. If the relationship is not close to linear, the two tests may not be measuring the same thing.

**Logarithms:** In Rasch measurement all logarithms, "log", are "natural" or "Napierian", sometime abbreviated elsewhere as "ln". "Logarithms to the base 10" are written log10. Logits to the base 10 are called "lods".

**Logit-to-Probability Conversion Table**

<table>
<thead>
<tr>
<th>Logit</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.0</td>
<td>99%</td>
</tr>
<tr>
<td>4.6</td>
<td>99%</td>
</tr>
<tr>
<td>4.0</td>
<td>98%</td>
</tr>
<tr>
<td>3.0</td>
<td>95%</td>
</tr>
<tr>
<td>2.2</td>
<td>90%</td>
</tr>
<tr>
<td>2.0</td>
<td>88%</td>
</tr>
</tbody>
</table>

Logit difference between ability measure and item calibration & Probability of success on a dichotomous item
Example with dichotomous data:
In Table 1, it is the distance between each person and each item which determines the probability.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4</td>
<td>80%</td>
</tr>
<tr>
<td>1.1</td>
<td>75%</td>
</tr>
<tr>
<td>1.0</td>
<td>73%</td>
</tr>
<tr>
<td>0.8</td>
<td>70%</td>
</tr>
<tr>
<td>0.5</td>
<td>62%</td>
</tr>
<tr>
<td>0.4</td>
<td>60%</td>
</tr>
<tr>
<td>0.2</td>
<td>55%</td>
</tr>
<tr>
<td>0.1</td>
<td>52%</td>
</tr>
<tr>
<td>0</td>
<td>50%</td>
</tr>
</tbody>
</table>

Example with dichotomous data:
In Table 1, it is the distance between each person and each item which determines the probability.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4</td>
<td>80%</td>
</tr>
<tr>
<td>1.1</td>
<td>75%</td>
</tr>
<tr>
<td>1.0</td>
<td>73%</td>
</tr>
<tr>
<td>0.8</td>
<td>70%</td>
</tr>
<tr>
<td>0.5</td>
<td>62%</td>
</tr>
<tr>
<td>0.4</td>
<td>60%</td>
</tr>
<tr>
<td>0.2</td>
<td>55%</td>
</tr>
<tr>
<td>0.1</td>
<td>52%</td>
</tr>
<tr>
<td>0</td>
<td>50%</td>
</tr>
</tbody>
</table>

The two persons are at 3.7 logits. The three items are at 4.5 logits. The difference is 3.7 - 4.5 = -0.8 logits. From the logit table above, this is predicted as 30% probability of success for persons like these on items like these.

**Inference with Logits**
Logit distances such as 1.2 logits are exactly correct for individual dichotomous items. 1.2 logits is the distance between 50% success and 80% success.

Logit distances are also exactly correct for describing the relative performance on adjacent categories of a rating scale, e.g., if in a Likert Scale, “Agree” and “Strongly Agree” are equally likely to be observed at a point on the latent variable, then 1.2 logits higher, “Strongly Agree” is likely to be observed 8 times, for every 2 times that “Agree” is observed.

For sets of dichotomous items, or performance on a rating scale item considered as a whole, the direct interpretation of logits no longer applies. The mathematics of a probabilistic interpretation under these circumstances is complex and rarely worth the effort to perform. Under these conditions, logits are usually only of
mathematical value for the computation of fit statistics - if you wish to compute your own.

Different tests usually have different probabilistic structures, so that interpretation of logits across tests are not the same as interpretation of logits within tests. This is why test equating is necessary.

**Logits and Probits**

Logits are the "natural" unit for the logistic ogive. Probits are the "natural" units for the unit normal cumulative distribution function, the "normal" ogive. Many statisticians are more familiar with the normal ogive, and prefer to work in probits. The normal ogive and the logistic ogive are similar, and a conversion of 1.7 approximately aligns them.

When the measurement units are probits, the dichotomous Rasch model is written:

\[
\log \left( \frac{P}{1-P} \right) = 1.7 \times (B - D)
\]

To have the measures reported in probits, set USCALE = 0.59 = 1/1.7

**Some History**

Around 1940, researchers focussed on the "normal ogive model". This was an IRT model, computed on the basis that the person sample has a unit normal distribution N(0,1).

The "normal ogive" model is: Probit (P) = theta - Di

where theta is a distribution, not an individual person.

But the normal ogive is difficult to compute. So they approximated the normal ogive (in probit units) with the much simpler-to-compute logistic ogive (in logit units). The approximate relationship is: logit = 1.7 probit.

IRT philosophy is still based on the N(0,1) sample distribution, and so a 1-PL IRT model is: \( \log(P/(1-P)) = 1.7 \times (\theta - D) \)

where theta represents a sample distribution. Di is the "one parameter".

The Rasch model takes a different approach. It does not assume any particular sample or item distribution. It uses the logistic ogive because of its mathematical properties, not because of its similarity to the cumulative normal ogive.

The Rasch model parameterizes each person individually, Bn. As a reference point it does not use the person mean (norm referencing). Instead it conventionally uses the item mean (criterion referencing). In the Rasch model there is no imputation of a normal distribution to the sample, so probits are not considered.

The Rasch model is: \( \log(P/(1-P)) = Bn - Di \)

Much IRT literature asserts that 1-PL = Rasch model. This is misleading. The mathematical equations can look similar, but their motivation is entirely different.

If you want to approximate the "normal ogive IRT model" with Rasch software, then
(a) adjust the person measures so the person mean = 0: UPMEAN=0
(b) adjust the user-scaling: probits = logits/1.7: USCALE=0.59

After this, the sample may come close to having an N(0,1) sample distribution - but not usually! So you can force S.D. = 1 unit, by setting USCALE = 1 / person S.D.

**317. Misfit diagnosis: infit outfit mean-square standardized**

What do Infit Mean-square, Outfit Mean-square, Infit Zstd (z-standardized), Outfit Zstd (z-standardized) mean?

Outfit: outlier-sensitive fit statistic. This is based on the conventional chi-square statistic. This is more sensitive to unexpected observations made persons on items that are relatively very easy or very hard for them (and vice-versa).
**Infit:** inlier-pattern-sensitive fit statistic. This is based on the chi-square statistic with each observation weighted by its statistical information (model variance). This is more sensitive to unexpected patterns of observations made persons on items that are roughly targeted on them (and *vice-versa*).

**Mean-square:** this is the chi-square statistic divided by its degrees of freedom. Consequently its expected value is close to 1.0. Values greater than 1.0 (underfit) indicate unmodelled noise or other source of variance in the data - these degrade measurement. Values less than 1.0 (overfit) indicate that the model predicts the data too well - causing summary statistics, such as reliability statistics, to report inflated statistics. See further dichotomous and polytomous mean-square statistics.

**Example of computation:**
The outfit mean-square is the accumulation of squared-standardized-residuals divided by their count, which is their expectation. The infit mean-square is the accumulation of information-weighted residuals divided by their expectation. The information an observation is its model variance. For dichotomies, this is the binominal variance = \( p(1-p) \)

Outfit mean-square = \( \frac{\text{sum} (\text{observed-residual}^2 / \text{model variance})}{\text{count}} \)
Infit mean-square = \( \frac{\text{sum} (\text{observed information-weighted residual variance})}{\text{sum} (\text{modeled information-weighted residual variance})} \)

Outlying observations have smaller information and so have less information than on-target observations. If all observations have the same amount of information, then the information cancels out. Then Infit mean-square = Outfit mean-square.

For dichotomous data. Two observations: Model \( p=0.5 \), observed=1. Model \( p=0.25 \), observed =1.
Outfit mean-square = \( \frac{\text{sum} ((1-0.5)^2/0.25 + (1-0.25)^2/0.0625))}{2} = (1 + 3)/2 = 2 \)
Infit mean-square = \( \frac{\text{sum} ((1-0.5)^2 + (1-0.25)^2)}{(0.25 + 0.0625)} = (0.25 + 0.56)/0.25 = 1.84. \ The off-target observation has less influence.

**Z-Standardized:** these report the statistical significance (probability) of the chi-square (mean-square) statistics occurring by chance when the data fit the Rasch model. The values reported are unit-normal deviates, in which .05% 2-sided significance corresponds to 1.96. Overfit is reported with negative values. These are also called "t-statistics" reported with infinite degrees of freedom.

**General rules:**
First, investigate negative point-measure or point-biserial correlations. Look at the Distractor Tables, 10.3. Remedy miskeys, data entry errors, etc.

Then, the general rule is Investigate outfit before infit, mean-square before t standardized, high values before low values.

There is an asymmetry in the implications of out-of-range high and low mean-squares (or positive and negative t-statistics). High mean-squares (or positive t-statistics) are a much greater threat to validity than low mean-squares (or negative fit statistics).

Poor fit does not mean that the Rasch measures (parameter estimates) aren't linear. The Rasch model forces its estimates to approximate linearity. Misfit means that the reported estimates, though effectively linear, provide a distorted picture of the data.

High outfit mean-squares may be the result of a few random responses by low performers. If so, drop with PDFILE=. these performers when doing item analysis, or use EDFILE= to change those response to missing.

High infit mean-squares indicate that the items are mis-performing for the people on whom the items are targeted. This is a bigger threat to validity, but more difficult to diagnose than high outfit.

Mean-squares show the size of the randomness, i.e., the amount of distortion of the measurement system. 1.0 are their expected values. Values less than 1.0 indicate observations are too predictable (redundancy, model overfit). Values greater than 1.0 indicate unpredictability (unmodeled noise, model underfit). Mean-squares
usually vaerage to 1.0, so if there are high values, there must also be low ones. Examine the high ones first, and temporarily remove them from the analysis if necessary, before investigating the low ones.

Zstd are t-tests of the hypotheses "do the data fit the model (perfectly)?" ZSTD (standardized as a z-score) is used of a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student's t-distribution value has been adjusted to a unit normal value. They show the improbability (significance). 0.0 are their expected values. Less than 0.0 indicate too predictable. More than 0.0 indicates lack of predictability. If mean-squares are acceptable, then Zstd can be ignored. They are truncated towards 0, so that 1.00 to 1.99 is reported as 1. So a value of 2 means 2.00 to 2.99, i.e., at least 2. For exact values, see Output Files. If the test involves less than 30 observations, it is probably too insensitive, i.e., "everything fits". If there are more than 300 observations, it is probably too sensitive, i.e., "everything misfits".

The Wilson-Hilferty cube root transformation converts the mean-square statistics to the normally-distributed z-standardized ones. For more information, please see Patel's "Handbook of the Normal Distribution" or www.rasch.org/rmt/rmt162g.htm.

Anchored runs:
Anchor values may not exactly accord with the current data. To the extent that they don't, they fit statistics may be misleading. Anchor values that are too central for the current data tend to make the data appear to fit too well. Anchor values that are too extreme for the current data tend to make the data appear noisy.

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

In general, mean-squares near 1.0 indicate little distortion of the measurement system, regardless of the Zstd value.

Evaluate high mean-squares before low ones, because the average mean-square is usually forced to be near 1.0.

Outfit mean-squares: influenced by outliers. Usually easy to diagnose and remedy. Less threat to measurement.
Infit mean-squares: influenced by response patterns. Usually hard to diagnose and remedy. Greater threat to measurement.

Extreme scores always fit the Rasch model exactly, so they are omitted from the computation of fit statistics. If an extreme score has an anchored measure, then that measure is included in the fit statistic computations.

Question: Does it mean that these mean-square values, >2 etc, are not sample size dependent?
Answer: Correct as a general rule-of-thumb. The mean-squares are already corrected for sample size: they are the chi-squares divided by their degrees of freedom, i.e., sample size. The mean-squares answer "how big is the impact of the misfit". The t-statistics answer "how unlikely to be observed when the data fit the model." We eagerly await the theoretician who devises a statistical test for the hypothesis "the data fit the Rasch model usefully." (as opposed to the current tests for perfectly).

Question: Is this contradicting the usual statistical advice about model-data fit?
Statisticians are usually concerned with "how likely are these data to be observed, assuming they accord with the model?" If it is too unlikely (i.e., significant misfit), then the verdict is "these data don't accord with the model."

The practical concern is: "In the imperfect empirical world, data never exactly accord with the Rasch model, but do these data deviate seriously enough for the Rasch measures to be problematic?" The builder of my house followed the same rule (regarding Pythagoras theorem) when building my bathroom. It looked like the walls were square enough for practical purposes. Some years later, I installed a full-length rectangular mirror - then I discovered that the walls were not quite square enough for my purposes (so I had to do some cosmetic adjustments) - so there is always a judgment call. The table of mean-squares is my judgment call as a "builder of Rasch measures".

308
Ben Wright's Infit and Outfit statistics (e.g., RSA p. 100) are initially computed as mean-square statistics (i.e., chi-square statistics divided by their degrees of freedom). Their likelihood (significance) is then computed. This could be done directly from chi-square tables, but the convention is to report them as unit normal deviates (i.e., t-statistics corrected for their degrees for freedom). I prefer to call them z-statistics, but the Rasch literature has come to call them t-statistics, so now I do to. It is confusing because they are not strictly Student t-statistics (for which one needs to know the degrees of freedom).

<table>
<thead>
<tr>
<th>Classification</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>Explanation</th>
<th>Investigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noisy</td>
<td>Noisy</td>
<td>Lack of convergence</td>
<td>Final values in Table 0 large?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noisy</td>
<td>Loss of precision</td>
<td>Many categoies? Large limits?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noisy</td>
<td>Anchoring</td>
<td>Displacements reported?</td>
<td></td>
</tr>
<tr>
<td>Hard Item</td>
<td>Noisy</td>
<td>Bad item</td>
<td>Ambiguous or negative wording?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Muted</td>
<td></td>
<td>Debatable or misleading options?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Muted</td>
<td></td>
<td>At end of test?</td>
<td></td>
</tr>
<tr>
<td>Item</td>
<td>Noisy</td>
<td>Qualitatively different item</td>
<td>Anchor value incorrectly applied?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Muted</td>
<td>Incompatible anchor or value</td>
<td>Similar items?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Muted</td>
<td></td>
<td>Overlapping items?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Muted</td>
<td></td>
<td>Item correlated with other variable?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Muted</td>
<td></td>
<td>Are there alternative anchora?</td>
<td></td>
</tr>
<tr>
<td>Rating scale</td>
<td>Noisy</td>
<td>Extreme category overlap</td>
<td>Poor category wording?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Muted</td>
<td></td>
<td>Combine or omit categories?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Muted</td>
<td></td>
<td>Weigh mod for scale?</td>
<td></td>
</tr>
<tr>
<td>Person</td>
<td>Noisy</td>
<td></td>
<td>Scammer failure?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noisy</td>
<td>Processing error</td>
<td>Unexpected wrong answer?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noisy</td>
<td>Clinical error</td>
<td>Unexpected errors at start?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noisy</td>
<td>Idiosyncratic person</td>
<td>Unexpected errors at end?</td>
<td></td>
</tr>
<tr>
<td>High Person</td>
<td>?</td>
<td></td>
<td>Caseness?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noisy</td>
<td>Sleeping</td>
<td>Unexpected wrong answer?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noisy</td>
<td>Resting</td>
<td>Unexpected errors at start?</td>
<td></td>
</tr>
<tr>
<td>Low Person</td>
<td>?</td>
<td></td>
<td>Guessing?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noisy</td>
<td>Guesing</td>
<td>Unexpected right answer?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noisy</td>
<td>Response set</td>
<td>Systematic response pattern?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noisy</td>
<td>&quot;Special&quot; knowledge</td>
<td>Content of unexpected answers?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Muted</td>
<td></td>
<td>Did not reach end of test?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Muted</td>
<td></td>
<td>Only answered easy items?</td>
<td></td>
</tr>
<tr>
<td>Person/Judge Rating</td>
<td>Noisy</td>
<td>Extreme category overlap</td>
<td>Euchanism? Defiance?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Muted</td>
<td></td>
<td>Conservation? Resistance?</td>
<td></td>
</tr>
<tr>
<td>Judge Rating</td>
<td>Muted</td>
<td></td>
<td>Apparent untrustworthiness</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Muted</td>
<td></td>
<td>Collusion?</td>
<td></td>
</tr>
</tbody>
</table>

The relationship between mean-square and z-standardized t-statistics is shown in this plot. Basically, the standardized statistics are insensitive to misfit with less than 30 observations and overly sensitive to misfit when there are more than 300 observations.
318.  **Missing data**

One of Ben Wright's requirements for valid measurement, derived from the work of L.L. Thurstone, is that "Missing data must not matter." Of course, missing data always matters in the sense that it lessens the amount of statistical information available for the construction and quality-control of measures. Further, if the missing data, intentionally or unintentionally, skew the measures (e.g., incorrect answers are coded as "missing responses"), then missing data definitely do matter. But generally, missing data are missing essentially at random (by design or accident) or in some way that will have minimal impact on the estimated measures (e.g., adaptive tests).

Winsteps does not require complete data in order to make estimates. One reason that Winsteps uses JMLE is that it is very flexible as regards estimable data structures. For each parameter (person, item or Rasch-Andrich threshold) there are sufficient statistics: the marginal raw scores and counts of the non-missing observations. During Winsteps estimation, the observed marginal counts and the observed and expected marginal scores are computed from the same set of non-missing observations. Missing data are skipped over in these additions. When required, Winsteps can compute an expected value for every observation (present or missing) for which the item and person estimates are known.

The basic estimation algorithm used by Winsteps is:

Improved parameter estimate = current parameter estimate + \( \frac{(\text{observed marginal score} - \text{expected marginal score})}{(\text{modeled variance of the expected marginal score})} \)

The observed and expected marginal scores are obtained by summing across the non-missing data. The expected score and its variance are obtained by Rasch estimation using the current set of parameter estimates, see *RSA*.

If data are missing, or observations are made, in such a way that measures cannot be constructed unambiguously in one frame of reference, then the message

**WARNING: DATA MAY BE AMBIGUOUSLY CONNECTED INTO nnn SUBSETS**

is displayed on the Iteration screen to warn of ambiguous connection.

319.  **Mixed and Saltus models**

Rasch models are grounded in the concept of the unidimensional latent variable, i.e., the items defining the latent variable operate in the same way for all members of the target population. Of course, this is a fiction. But reality can often be made to cooperate.

But there are occasions when a population is comprised of different classes of persons with the items comprising a different latent variable for each class. The classes are called "Latent Classes".

Standard Rasch "latent trait" models can be extended to allow for latent classes. These are called "Mixture Models" (Rost, 1990). The Saltus model (Mark Wilson, 1989) is a mixed model in which segments of items are modeled to shift their difficulties together, and by the same amount, for different latent classes. In these models,
the different latent variables are defined by item difficulties, but individual respondents are not assigned to a particular class, but rather the probability that each respondent belongs to each class is reported.

Winsteps does not do a mixture or Saltus analysis directly, but it can provide much of the same information, and also can indicate whether a more rigorous latent class analysis is likely to be productive.

Here is an approach:

Step 1. Identify meaningful potential respondent classes, e.g., male/female, high/low performers. The Winsteps PCA analysis (e.g., Table 24.4) may help identify potential classes.

Step 2. Mark in the person label the class codes. The Microsoft Word “rectangle copy” function may be useful. High and low performers do not need to be flagged, instead the MA2 function can be used.

Step 3. Perform DIF analysis based on the class codes. Items displaying strong DIF may be exhibiting class-related behavior.

Step 4. Flag the items by class in the item identification.

Step 5. Look for item-classification by person-classification interactions (differential classification-grouped functioning, DGF, Table 33). These would approximate the Saltus findings.


320. Multiple t-tests

Question: Winsteps Tables report many t-tests. Should Bonferroni adjustments for multiple comparisons be made?

Reply: It depends on how you are conducting the t-tests. For instance, in Table 30.1. If your hypothesis (before examining any data) is “there is no DIF for this CLASS in comparison to that CLASS on this item”, then the reported probabilities are correct.

If you have 20 items, then one is expected to fail the $p \leq .05$ rule. So if your hypothesis (before examining any data) is “there is no DIF in this set of items for any CLASS”, then adjust individual t-test probabilities accordingly.

In general, we do not consider the rejection of a hypothesis test to be "substantively significant", unless it is both very unlikely (i.e., statistically significant) and reflects a discrepancy large enough to matter (i.e., to change some decision). If so, even if there is only one such result in a large data set, we may want to take action. This is much like sitting on the proverbial needle in a haystack. We take action to remove the needle from the haystack, even though statistical theory says, “given a big enough haystack, there will probably always be a needle in it somewhere.”

A strict Bonferroni correction for $n$ multiple significance tests at joint level $\alpha$ is $\alpha / n$ for each single test. This accepts or rejects the entire set of multiple tests. In an example of a 100 item test with 20 bad items (.005 < $p$ < .01), the threshold values for cut-off with $p \leq .05$ would be: 0.0005, so that the entire set of items is accepted.

Benjamini and Hochberg (1995) suggest that an incremental application of Bonferroni correction overcomes some of its drawbacks. Here is their procedure:

i) Perform the $n$ single significance tests.
ii) Number them in ascending order by probability $P(i)$ where $i=1,n$ in order.
iii) Identify $k$, the largest value if $i$ for which $P(i) = \alpha * i/n$
iv) Reject the null hypothesis for $i = 1, k$

In an example of a 100 item test with 20 bad items (.005 < $p$ < .01), the threshold values for cut-off with $p \leq .05$...
would be: 0.0005 for the 1st item, .005 for the 10th item, .01 for the 20th item, .015 for the 30th item. So that \( k \) would be at least 20 and perhaps more. All 20 bad items have been flagged for rejection.


321. **Null or unobserved categories**

There are two types of unobserved or null categories: structural zeroes and incidental/sampling zeroes.

Structural null categories occur when rating scale categories are number 10, 20, 30,… instead of 1,2,3. To force Winsteps to eliminate non-existent categories 11, 12, 13, either rescore the data \texttt{IVALUE=} or specify \texttt{STKEEP=NO}.

For intermediate incidental null zeroes, imagine this scenario: The Wright & Masters “Liking for Science” data are rescored from 0,1,2 to 0,1,3 with a null category at 2. the categories now mean “disagree, neutral, agree-ish, agree”. We can imagine that no child in this sample selected the half-smile of agree-ish. The category frequencies of categories 0,1,2,3 are 378, 620, 0, 852

The three Rasch-Andrich threshold parameters are -.89, +infinity, -infinity. The -infinity is because the second parameter is of the order log(620/0). The –infinity is because the third parameter is of the order log(0/852).

Mark Wilson's 1991 insight was that the leap from the 2nd to the 4th category is of the order log(620/852). This is all that is needed for immediate item and person estimation. But it is not satisfactory for anchoring rating scales. In practice however, a large value substitutes satisfactorily for infinity. So, a large value such as 40 logits is used for anchoring purposes. Thus the approximated parameters become -.89, 40.89, -40.00 for \texttt{SAFILE=}.

With these anchored threshold values, the expected category frequencies become: 378.8, 619.4, .0, 851.8. None of these are more than 1 score point away from their observed values, and each represents a discrepancy of .2% or less of its category count.

Extreme incidental null categories (unobserved top or bottom categories) are essentially out of range of the sample and so the sample provides no direct information about their estimates. To estimate those estimates requires us to make an assertion about the form of the rating scale structure. The Rasch “Poisson” scale is a good example. All its infinitude of thresholds are estimable because they are asserted to have a specific form. But see Example 12 for a different approach to this situation.

Our recommendation is that structural zeroes be rescored out of the data. If categories may be observed next time, then it is better to **include a dummy data record** in your data file which includes an observation of the missing category and reasonable values for all the other item responses that accord with that missing category. This one data record will have minimal impact on the rest of the analysis.

322. **One observation per respondent**

**Question:** I'm trying to analyze a dataset where there are four test forms, and on each test form there is only one 4-point polytomous item. That is, each student took one and only one test question. Can this type of dataset be calibrated using Winsteps?

**Reply:** If there is only one response per person, there is not enough information to construct measures, but only enough to order the people by the raw score of that one response. But …

If the people taking each of the 4 forms are supposed to be randomly equivalent, then we can **equate** the forms, and discover how a “3” on one form relates to a “3” on another form. To do this:

Enter the 4 forms as 4 items in Winsteps.
For each "item" enter the column of responses.
Anchor the rows at 0.
Set \texttt{ISGROUPS=}0
Run the analysis.

The measure corresponding to each score on each item is given in Table 3.2, "Score at Cat", and shown in Table
2.2. Use the measures in the "At Cat." column to correspond to the polytomous observations in summary analyses.

**Example:** The responses to the 4 forms, A, B, C, D, were:

A 1 3 2 4  
B 2 4 3 1 1 3  
C 3 2 2 3 1 4 1  
D 4 4 3 2 1  

Note that the order of the persons within form doesn't matter, and the number of respondents per form doesn't matter. Here is the *Winsteps* control file:

```plaintext
Title = "Measurement with 4 forms"
NI=4
Item=1
Name=1 ; there aren't any row names.
Codes=1234
ISGROUPS=0 ; allow each form its own rating (or partial credit) scale
Item=Form ; rename to remind ourselves
Person=Row ; Rows are anchored at zero, and so are all equivalent.
Pafile="
1-7 0 ; anchor all rows at "0". 7 is the largest number of students who took any form. 
*
CONVERGE=L ; only logit change is used for convergence
LCONV=0.005 ; logit change too small to appear on any report.
&end
A ; the 4 items are the 4 forms
B
C
D
END LABELS
1234 ; responses per form entered as columns with students in any order.
3424
2323
.111
.34.
.1\.

Resulting Table 2.2:

**Table 2.2 Measurement with 4 forms**

<table>
<thead>
<tr>
<th>EXPECTED SCORE: MEAN (&quot;:&quot; INDICATES HALF-POINT THRESHOLD)</th>
<th>NUM</th>
<th>FORM</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 2 3 4 5 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 2 3 4 5 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 2 3 4 5 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 2 3 4 5 6</td>
</tr>
</tbody>
</table>

**Table 3.2:**

**Summary of Category Structure. Model="R"**

<table>
<thead>
<tr>
<th>FORM ITEM DIFFICULTY MEASURE OF .00 ADDED TO MEASURES</th>
</tr>
</thead>
<tbody>
<tr>
<td>CATEGORY OBSERVED</td>
</tr>
<tr>
<td>LABEL SCORE COUNT</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>
### 323. Order of elements in Control file

<table>
<thead>
<tr>
<th>Element</th>
<th>Condition for Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>&amp;INST (optional, for backwards compatibility only)</td>
<td></td>
</tr>
<tr>
<td>TITLE= title of analysis</td>
<td>recommended</td>
</tr>
<tr>
<td>ITEM1= starting column of items</td>
<td>Required</td>
</tr>
<tr>
<td>Ni= number of items</td>
<td>Required</td>
</tr>
<tr>
<td>ISGROUPS= grouping information</td>
<td>optional, with GRPFRM=N (the standard)</td>
</tr>
<tr>
<td>MODELS= model information</td>
<td>optional, with MODFRM=N (the standard)</td>
</tr>
<tr>
<td>RESCORE= rescore information</td>
<td>optional, with RESFRM=N (the standard)</td>
</tr>
<tr>
<td>KEY1= key information</td>
<td>optional, if KEYFRM= omitted (the standard)</td>
</tr>
<tr>
<td>KEY2= ..</td>
<td>optional, if KEYFRM= omitted (the standard)</td>
</tr>
<tr>
<td>KEYn= ..</td>
<td>optional, if KEYFRM= omitted (the standard)</td>
</tr>
<tr>
<td>(n=1 to 99, number of largest key)</td>
<td></td>
</tr>
<tr>
<td>other control variables</td>
<td>optional</td>
</tr>
<tr>
<td>: comments</td>
<td>optional</td>
</tr>
</tbody>
</table>

#### Control file format:

- \&INST (optional, for backwards compatibility only)
- TITLE= title of analysis (recommended)
- ITEM1= starting column of items (Required)
- Ni= number of items (Required)
- ISGROUPS= grouping information (optional, with GRPFRM=N (the standard))
- MODELS= model information (optional, with MODFRM=N (the standard))
- RESCORE= rescore information (optional, with RESFRM=N (the standard))
- KEY1= key information (optional, if KEYFRM= omitted (the standard))
- KEY2= .. (optional, if KEYFRM= omitted (the standard))
- KEYn= .. (optional, if KEYFRM= omitted (the standard)) (n=1 to 99, number of largest key)
- other control variables (optional)
- : comments (optional)

### 324. Partial Credit model

The "Partial Credit" Rasch model was devised for multiple-choice questions in which credit is given for almost-correct distractors. But there is no reason to believe that the almost-correctness of distractors to different questions is the same. Consequently, each item is modeled to have its own response structure.

This model was extended to any questionnaire using ordered polytomies in which the response structure is

---

```
<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>STRUCTURE</th>
<th>SCORE-TO-MEASURE</th>
<th>50% CUM.</th>
<th>COHERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABEL</td>
<td>MEASURE</td>
<td>S.E.</td>
<td>AT CAT.</td>
<td>ZONE---</td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>1</td>
<td>NONE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.00   1.15</td>
<td>-.42</td>
<td>1.01</td>
<td>.00</td>
</tr>
<tr>
<td>3</td>
<td>.00   1.00</td>
<td>.42</td>
<td>.00</td>
<td>1.01</td>
</tr>
<tr>
<td>4</td>
<td>.00   1.15</td>
<td>1.59</td>
<td>1.01</td>
<td>+INF</td>
</tr>
</tbody>
</table>
```
modeled to be unique to each item.

Winsteps estimates response structures by item groupings, using ISGROUPS=. From this perspective, the Andrich Rating Scale Model includes all items in one grouping. The Masters Partial Credit Model allocates each item to its own grouping.

The conventional representation of the Partial Credit model is

\[
\log \left( \frac{P_{nij}}{P_{n(i-1)}} \right) = B_n - D_{ij}
\]

Winsteps parameterizes \( D_{ij} \) as \( D_i + F_{ij} \) where \( \text{sum}(F_{ij}) = 0 \). And \( D_i \) is the average \( (D_{ij}) \).

\[
\log \left( \frac{P_{nij}}{P_{n(i-1)}} \right) = B_n - D_i - F_{ij}
\]

**Algebraically these two representations are identical.**

Thus every item has a mean difficulty, \( D_i \). This simplifies communication, because the results of a Partial Credit analysis now have the same form as any other polytomous analysis supported by Winsteps.

**325. Plausible values**

Plausible values are estimates intended to represent the distribution of measures that could produce the observed scores. They were developed for large-scale educational assessments from which group-level measures are to be obtained, but with data too thin to support individual-level measurement.

Winsteps is designed for individual measurement. When this is possible, then group-level reporting can be done, e.g., with PSUBTOT=. The Winsteps estimate approximates the mean of the plausible-value distribution. For Rasch software that produces plausible values, see www.winsteps.com/rasch.htm.

**326. Plotting with EXCEL**

*First try the "Plots" pull-down menu.*

**Plotting:** This is conveniently and flexibly done with EXCEL:

(A) Check that your copy of EXCEL works.

(B) Download the free Excel chart-labeler add-in from www.appsfro.com/Utilities/ChartLabeler.htm

(C) Run XYChartLabeler.exe
The Excel add-in "XY Chart Labels" is added to the Excel Tools pull-down menu.

(D) To plot
Write PFILE= or IFILE= from two analyses, or copy from Output Tables.

Copy and paste each of these into EXCEL:
Use "Data" "Text to Columns" to put values in columns

Put the columns to be plotted next to each other:
x-axis values to left of y-axis values.
Highlight numbers to be cross-plotted.

**To make a plot:**
Click on "Chart Wizard"
Click on "XY Scatter"
"Next"
"Next"
Fill in "Chart Title" and "Value" names
"Next"
Click "A new sheet"
"Finish"

**On the plot:**
Click "Series 1"
"Clear"
Right click a data point
Click "Format data"
Click "Marker none" (the data points will soon disappear!! - Don't worry!)
Click "OK"
Right click a gridline
"Clear"

**Add point labels:**
Click on "Chart" tab
Click "Tools"
Click "XY Chart labels"
Click "XY Add"
Click "Centered"
Click "Select label _"
Click "Sheet 1" tab
Highlight point labels
Click red marker
Click "OK"
Point-labeled XY plot appears.

**Use the plot to communicate:**
Click on plot
Use handles to make the plot square
If drawing toolbar is not active:
   Right click a toolbar
      Click on Drawing
Click on "line" tool
Draw in a useful line.

### 327. Poisson counts

The Winsteps program can analyze Poisson count data, with a little work.

Poisson counts are a rating (or partial credit) scale with pre-set structure. The structure measures are $\log_e(n)$, $n=1$ upwards.

You can define a structure anchor file in this way:

```
XWIDE=2
STKEEP=YES
CODES=000102030405060708091011121314.......979899
SAFILEe=*
   0 0 ; the value corresponding to $\log(1)$ - the pivot point for the item measure
   2 .693 ; the value corresponding to $\log(2)$
   3 1.099 ; the value corresponding to $\log(3)$
   ....
   99 4.595 ; the value corresponding to $\log(99)$
```

Arrange that the observations have an upper limit much less than 99.

You may find that you need to multiply all structure measures by a constant to adjust the "natural" form of the Poisson counts to the actual discrimination of the empirical Poisson process. (The *Facets* program does this
You need to adjust the constant so that the average mean-square is about 1.0. (See RMT 14:4 about using mean-squares to adjust logit user-scaling.)

But my experience with running Poisson counts in the Facets program (which supports them directly) is that most "Poisson count" data do not match the Poisson process well, and are more usefully parameterized as a rating (or partial credit) scale. There is nearly always some other aspect of the situation that perturbs the pure Poisson process.

328. Point-measure correlation

PTMEA, the point-measure correlation is reported instead of PTBIS when PTBIS=N or PTBIS=RPM is specified. PTMEA or RPM is the point-measure correlation, \( r_{pm} \). It is computed in the same way as the point bi-serial, except that Rasch measures replace total scores. Since the point-biserial loses its meaning in the presence of missing data, specify PTBIS=N when data are missing or CUTLO= or CUTHI= are specified.

The formula for this product-moment correlation coefficient is:

\[
\begin{align*}
\text{r}_{\text{pbis}} &= \frac{\sum (x-x_{\text{bar}})(y-y_{\text{bar}})}{\sqrt{\sum (x-x_{\text{bar}})^2 \sum (y-y_{\text{bar}})^2}} \\
\end{align*}
\]

where \( x \) = observation for this item (or person), \( y \) = measure for this item (or person). The range is -1 to +1. Its maximum values approximate the point-biserial correlation, reported in Rasch Measurement Transactions 5:4 to be:

329. Polytomous mean-square fit statistics

<table>
<thead>
<tr>
<th>Response String</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>RPM (PTMEA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy..................</td>
<td>Hard</td>
<td>MnSq</td>
<td>MnSq</td>
</tr>
</tbody>
</table>

I. modelled:

<table>
<thead>
<tr>
<th>String</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>RPM (PTMEA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>33333132210000001011</td>
<td>.98</td>
<td>.99</td>
<td>.78 Stochastically</td>
</tr>
<tr>
<td>3133233231220000000</td>
<td>.98</td>
<td>1.04</td>
<td>.81 monotonic in form,</td>
</tr>
<tr>
<td>3333333112230000000</td>
<td>1.06</td>
<td>.97</td>
<td>.87 strictly monotonic</td>
</tr>
<tr>
<td>33333331110010200001</td>
<td>1.03</td>
<td>1.00</td>
<td>.81 in meaning</td>
</tr>
</tbody>
</table>

II. overfitting (muted):

<table>
<thead>
<tr>
<th>String</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>RPM (PTMEA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3322222222111111100</td>
<td>.18</td>
<td>.22</td>
<td>.92 Guttman pattern</td>
</tr>
<tr>
<td>33333222222111110000</td>
<td>.31</td>
<td>.35</td>
<td>.97 high discrimination</td>
</tr>
<tr>
<td>32222222222111111110</td>
<td>.21</td>
<td>.26</td>
<td>.89 low discrimination</td>
</tr>
<tr>
<td>323232321212101010</td>
<td>.52</td>
<td>.54</td>
<td>.82 tight progression</td>
</tr>
</tbody>
</table>
III. limited categories:
33333333322222222222 .24 .24 .87 high (low) categories
22222222211111111111 .24 .34 .87 central categories
33333322222222111111 .16 .20 .93 only 3 categories

IV. informative-noisy:
32222222201111111130 .94 1.22 .55 noisy outliers
33233332212333000000 1.25 1.09 .77 erratic transitions
331333302300101000 1.49 1.40 .72 noisy progression
33333333300000000000 1.37 1.20 .87 extreme categories

V. non-informative:
22222222222222222222 .85 1.21 .00 one category
12121212121212121212 1.50 1.96 -.09 central flip-flop
01230123012301230123 3.62 4.61 -.19 rotate categories
03030303030303030303 5.14 6.07 -.09 extreme flip-flop
032002010113311002 2.99 3.59 -.01 random responses

VI. contradictory:
111111122322111111 1.75 2.02 .00 folded pattern †
111111111111112222222222 2.56 3.20 -.87 central reversal
22222222223333333333 2.11 4.13 -.87 high reversal
00111111111112222222223 4.00 5.58 -.92 Guttman reversal
00000000003333333333 8.30 9.79 -.87 extreme reversal


The z-score standardized statistics report, as unit normal deviates, how likely it is to observe the reported mean-square values, when the data fit the model. The term z-score is used of a t-test result when either the t-test value has effectively infinite degrees of freedom (i.e., approximates a unit normal value) or the Student’s t-distribution value has been adjusted to a unit normal value.

†"folded data" can often be rescued by imposing a theory of "not reached" and "already passed" on to the observations. For instance, in archaeological analysis, the absence of bronze implements can mean a "stone age" or an "iron age" society. A useful recoding would be "1" = "stone age", "2" = "early bronze", "3" = "bronze", "2=>4" = "late bronze", "1=>5" = "iron age". This can be done iteratively to obtain the most self-consistent set of 4's and 5's. (Folding is discussed in Clive Coombes’ "A Theory of Data".)

330. Quality-control misfit selection criteria

Rasch measurement does not make any presumptions about the underlying distribution of the parameters. Maximum likelihood estimation expects "errors" in the observations to be more or less normally distributed around their expected values. Since all observations are integral values, this expectation can be met only asymptotically as the number of persons and items becomes infinite. The information-weighted fit statistic, "infit", and the outlier-sensitive fit statistic, "outfit", are described in BTD and RSA. Possible values, and hence interpretation, of these statistics is influenced by the observed distribution the person and item statistics. This is particularly true of their t standardized values which are designed to follow standard normal (0,1) distributions. The local significance of these statistics is best interpreted in terms of their means and sample standard deviations reported in Table 3. Start investigating the misfit causing the most extreme values of these statistics, and stop your investigation when the observed responses become coherent with your intentions.

The fit statistics reported will not exactly match those printed in BTD or RSA, or those produced by another program. This is because the reported values of these statistics are the result of a continuing process of development in statistical theory and practice. Neither "correct" fit statistics nor "correct" values exist, but see the Appendices for guidance.

Report measure in Tables 6 and 10 if any of:
Statistic Less than Greater than
\[ t \text{ standardized INFIT} \quad -(\text{FITP or FITI}) \quad \text{FITP or FITI} \]
\[ t \text{ standardized OUTFIT} \quad -(\text{FITP or FITI}) \quad \text{FITP or FITI} \]
\[ \text{mean-square INFIT} \quad 1-(\text{FITP or FITI})/10 \quad 1+(\text{FITP or FITI})/10 \]
\[ \text{mean-square OUTFIT} \quad 1-(\text{FITP or FITI})/10 \quad 1+(\text{FITP or FITI})/10 \]

or point-biserial correlation negative

To include every person, specify \text{FITP}=0. For every item, \text{FITI}=0.

For Table 7, the diagnosis of misfitting persons, persons with a \text{t standardized fit} greater than \text{FITP}= are reported. Selection is based on the OUTFIT statistic, unless you set \text{OUTFIT}=N in which case the INFIT statistic is used.

For Table 11, the diagnosis of misfitting items, items with a \text{t standardized fit} greater than \text{FITI}= are reported. Selection is based on the OUTFIT statistic, unless you set \text{OUTFIT}=N in which case the INFIT statistic is used.

331. **Rank order data**

Rankings and partial rankings, with or without ties, can be conveniently analyzed using ISGROUPS=0.

Each row is an element to be ranked.

Each column is a set of rankings. In the item label, place any interesting demographics about the person doing the ranking.

Note: if every ranking set includes every element, and ties are not allowed, then elements can be columns and ranking sets as rows. ISGROUPS=0 is not required.

In general, we allow each ranking (column) to define its own "ranking scale". This is equivalent to the Partial Credit model.

Measures for the elements are obtained. Measures for the ranking sets are meaningless and can be ignored.

Fit statistics, DIF and DPF analysis, and contrast analysis of residuals are all highly informative.

332. **Rectangular copying**

To copy a rectangle of numbers:

1. Select the lines of text that include the rectangle of numbers.
2. Copy the lines to the clipboard
3. Paste the lines into a word-processing document or an Excel spreadsheet cell.
4. Set the font of the lines to Courier.

5A. In Word, select rectangles with Alt+Mouse (see below)
5B. In TextPad, select rectangles with Alt+Mouse
5C. In WordPerfect, select "Edit > Select > Rectangle"
5D. In Excel, use "Data > Text to Columns" to select the column of numbers into a column.

You could also display the column of numbers on your computer screen and do a graphical copy. PrintScreen saves the screen to the clipboard, then paste into Paint and do a rectangle selection of what you want. Paste the selection into your document as a Figure.

**Rectangular copy-and-paste with Microsoft Word**

In Word, ctrl-A the whole document.
Select a "Courier" font. Now everything lines up neatly in columns.
333. Reliability and separation of measures

The Winsteps "person reliability" is equivalent to the traditional "test" reliability. Low values indicate a narrow range of person measures, or a small number of items. To increase person reliability, test persons with more extreme abilities (high and low), lengthen the test. Improving the test targeting may help slightly.

The Winsteps "item reliability" has no traditional equivalent. Low values indicate a narrow range of item measures, or a small sample. To increase "item reliability", test more people. In general, low item reliability means that your sample size is too small for stable item estimates based on the current data. If you have anchored values, then it is the item reliability of the source from which the anchor values emanate which is crucial, not the current sample.

The traditional "test reliability", as defined by Charles Spearman in 1904 is the "true person variance / observed person variance" for this sample on these test items. So it is really a "person sample reliability" rather than a "test reliability", where reliability = reproducibility of person ordering. The "true person variance" cannot be known, but it can be approximated. KR-20 approximates it by summarizing item point-biserials. Cronbach Alpha approximates it with an analysis of variance. Winsteps approximates it using the measure standard errors.

The "model" person reliability (including measures for extreme scores) is an upper bound to this value, when persons are ordered by measures.

The "real" person reliability (including measures for extreme scores) is a lower bound to this value, when persons are ordered by measures.

KR-20 value is an estimate of the value when persons are ordered by raw scores. CRONBACH ALPHA (KR-20) KID RAW SCORE RELIABILITY is the conventional "test" reliability index. It reports an approximate test reliability based on the raw scores of this sample. It is only reported for complete data. An apparent paradox is that extreme scores have perfect precision, but extreme measures have perfect imprecision.

Winsteps computes upper and lower boundary values for the True Reliability. The lower boundary is the Real Reliability. The upper boundary is the Model Reliability. The unknowable True Reliability lies somewhere between these two. As contradictory sources of noise are remove from the data, the True Reliability approaches the Model Reliability.

Conventionally, only a Person ("Test") Reliability is reported. The relationship between raw-score-based reliability (i.e., KR-20, Cronbach Alpha) and measure-based reliability is complex, see www.rasch.org/rmt/rmt113i.htm - in general, Cronbach Alpha overestimates reliability, Rasch underestimates it. So, when it is likely that the Rasch reliability will be compared with conventional KR-20 or Cronbach Alpha reliabilities (which are always computed assuming the data match their assumptions), then include extreme persons and report the higher Rasch reliability, the "Model" reliability, computed on the assumption that all unexpectedness in the data is in accord with Rasch model predictions. The big differences between Score and Measure reliabilities occur when

(a) there are extreme scores. These increase score reliability, but decrease measure reliability.
(b) missing data. Missing data always decreases measure reliability. If the missing data are imputed at their expected values (in order to make conventional reliability formulas computable), they increase score reliability.

Winsteps attempts to adjust the raw-score reliability for this inflation in the raw-score reliability, but can only do the adjustment in an approximate way.

Winsteps also reports an item reliability, "true item variance / observed item variance". When this value is low, it
indicates that the sample size may be too small for stable comparisons between items.

Anchored values are treated as though they are the “true values” of the MLE estimates. Their local standard errors are estimated using the current data in the same way as unanchored MLE standard error estimates. It is the measures (anchored or unanchored) and local standard errors that are used in the reliability computations. If you wish to compute reliabilities using different standard error estimates (e.g., the ones when the anchor values were generated), then please perform a separate reliability computation (using Excel).

You can easily check the Winsteps reliability estimate computation yourself.

Read the Winsteps PFILE into an Excel spreadsheet.

Compute the STDEVP standard deviation of the person measures. Square it. This is the “Observed variance”.

“Model” Reliability: Take the standard ERROR column. Square each entry. Sum the squared entries. Divide that sum by the count of entries. This is the “Model Error variance” estimate. Then,

\[
\text{Model Reliability} = \frac{\text{True Variance}}{\text{Observed Variance}} = \frac{(\text{Observed Variance} - \text{Model Error Variance})}{\text{Observed Variance}}.
\]

“Real” Reliability: Take the standard ERROR column. Square each entry, SE^2. In another column, put SE^2*Maximum [1.0, INFIT mean-square). Divide that sum by the count of entries. This is the “Real Error variance” estimate. Then,

\[
\text{Real Reliability} = \frac{\text{True Variance}}{\text{Observed Variance}} = \frac{(\text{Observed Variance} - \text{Real Error Variance})}{\text{Observed Variance}}.
\]

Separation and Reliability
The crucial elements in the computation of reliability are the “True” variance and the Error variance. These are squared distances and so difficulty to conceptualize directly. It is easier to think of their square-roots, the “True” standard deviation (TSD) and the root-mean-square standard error (RMSE).

SEPARATION is the ratio of the PERSON (or ITEM) ADJ.S.D., the “true” standard deviation, to RMSE, the error standard deviation. It provides a ratio measure of separation in RMSE units, which is easier to interpret than the reliability correlation. This is analogous to the Fisher Discriminant Ratio. SEPARATION^2 is the signal-to-noise ratio, the ratio of “true” variance to error variance.

RELIABILITY is a separation reliability. The PERSON (or ITEM) reliability is equivalent to KR-20, Cronbach Alpha, and the Generalizability Coefficient. See much more at Reliability. The relationship between separation SEPARATION and RELIABILITY is

\[
\text{RELIABILITY} = \frac{\text{SEPARATION}^2}{1 + \text{SEPARATION}^2}
\]

or

\[
\text{SEPARATION} = \left(\frac{\text{RELIABILITY}}{1 - \text{RELIABILITY}}\right)^{0.5}.
\]

<table>
<thead>
<tr>
<th>Error</th>
<th>True</th>
<th>Obs</th>
<th>Signal-Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>S.D.</td>
<td>Var Var.</td>
<td>Separation= True SD/Error RMSE</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>16</td>
<td>17</td>
</tr>
</tbody>
</table>


334. Right-click functions

Mouse Right-Click on the Winsteps screen
Right-click anywhere on the Winsteps screen, and a menu will be displayed.
During estimation: It is the File menu.
After estimation: It is the Table menu.
Mouse Right-Click on the Task Bar
Right-click on the task bar, to obtain the window-control menu.

Mouse Right-Click on File Dialog Box:
Right-click a file name, and get the Send To menu.

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

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

335. Rules for assigning values to control variables (key-words)

Do not worry about these unless WINSTEPS does not respond to your control file the way you expected. If possible, compare your control file with what is shown in Table 0 of your report output file in order to isolate the problem.

1. Values are assigned to control variables by typing the name of the control variable (or enough of it to disambiguate it), an equal sign, "=" , and the value, e.g.
   TABLES=11011011100
   or
2. You must use one line for each assignment, but **continuation lines** are permitted. To continue a line, put a + at the end of the line. Then put a + at the start of the text in the next line. The two lines will be joined together so that the + signs are squeezed out, e.g.,

```
TITLE = "Analysis of medical+ research data"
```

is interpreted as

```
TITLE = "Analysis of medical research data"
```

*Continuation lines"+" are helpful to make control files fit on your screen.*

```
CODES = 01020304+
  +05060708
```

is interpreted as

```
CODES = 0102030405060708
```

To **comment out** a continuation line:

```
; CODES = 01020304+
  +05060708
```

or

```
; CODES = 01020304+
  +05060708
```

3. The control variables may be listed *in any order*.

4. Character strings must be enclosed in 'single quotes' or "double quotes" when they contain blanks, e.g.,

```
TITLE="Rasch Analysis of Test Responses"
```

or

```
TITLE='Rasch Analysis of Test Responses'
```

Quotes are not required for single words containing no blanks, e.g. PFILE=kctpf.txt

5. The control variables may be in upper or lower case or mixed, e.g., Pfile = Person.txt

6. Blanks before or after control variables, and before or after equal signs are ignored, e.g.  
```
TITLE="Test Report"
```

and  
```
TITLE = "Test Report"
```

are equally correct.

7. Commas at the end of lines are ignored, so equally correct are:  
```
NAME1 = 33,
```

and  
```
NAME1 = 33
```

8. Control variables can be made into comments, and so be ignored, by entering a semi-colon in column 1, e.g.  
```
; FITP=3 is ignored
```

9. When all control variables (required or optional) have been assigned values, type **&END** (in upper or lower case) on the next line, e.g.,

```
Title ="A 30 Item test"
NI = 30
  ; this is a comment: person names in columns 1-20.
ITEM1 = 21
&END
```
336. **Shortcut Keys**

Some very frequent operations are quicker to perform using shortcut keystrokes than by menu selection. Here are the shortcut keys implemented in Winsteps:

Alt+ hold down the Alt key and press the letter key.
Ctrl+ hold down the Ctrl key and press the letter key.

Alt+A start Another copy of Winsteps
Alt+E Edit the control file
Alt+H display Help file
Alt+R Restart Winsteps with this control file
Alt+S Specification entry
Alt+X Exit this Winsteps, and restart with this control file

Ctrl+F Finish iterating
Ctrl+O Open a control or output file
Ctrl+Q Quit Winsteps
Ctrl+S Save on-screen activity log
Ctrl+P Print on-screen activity log

Esc Escape from action

337. **Specifying how data are to be recoded**

You will need to choose how this is done.

First, use CODES= to specify the response codes in your data file.

If there is only one type of recoding to be done, use

NEWSCORE=

If this one type of rescoring only applies to some of the items, also use

RESCORE=

If the rescoring is more complex, use

IREFER= and IVVALUE=

If the items are multiple-choice, use

KEYn=

If missing values are not to be ignored, i.e., treated as not-administered, you will need

MISSCORE=

If alphabetical codes are used to express two-digit numbers in one column, use

ALPHANUM=

338. **Standard errors: model and real**

A standard error quantifies the precision of a measure or an estimate. It is the standard deviation of an imagined error distribution representing the possible distribution of observed values around their "true" theoretical value. This precision is based on information within the data. The quality-control fit statistics report on accuracy, i.e., how closely the measures or estimates correspond to a reference standard outside the data, in this case, the Rasch model.

**Model "Ideal" Standard Error**

The highest possible precision for any measure is that obtained when every other measure is known, and the data fit the Rasch model. For well-constructed tests with clean data (as confirmed by the fit statistics), the model
standard error is usefully close to, but slightly smaller than, the actual standard error. The "model" standard error is the "best case" error. It is the asymptotic value for JMLE. For dichotomous data this is, summed over items i=1,L for person n, or over person n=1,N for item i:

$$S.E. = \sqrt{\sum (P_{ni}(1-P_{ni}))}$$

For polytomies (rating scales, partial credit, etc.), with categories j=0,m:

$$S.E. = \sqrt{\sum_{i} \left(\sum_{j} \left(\sum_{j} P_{nij} \cdot \left(m - \sum_{j=0}^{m} P_{nij}\right)^2\right)\right)}$$

Misfit-Inflated "Real" Standard Error

Wright and Panchapakesan (1969) discovered an important result for tests in which each examinee takes more than a handful of items, and each item is taken by more than a handful of examinees: the imprecision introduced into the target measure by using estimated measures for the non-target items and examinees is negligibly small. Consequently, in almost all data sets except those based on very short tests, it is only misfit of the data to the model that increases the standard errors noticeably above their model "ideal" errors. Misfit to the model is quantified by fit statistics. But, according to the model, these fit statistics also have a stochastic component, i.e., some amount of misfit is expected in the data. Discovering "perfect" data immediately raises suspicions! Consequently, to consider that every departure of a fit statistic from its ideal value indicates failure of the data to fit the model is to take a pessimistic position. What it is useful, however, is to estimate "real" standard errors by enlarging the model "ideal" standard errors by the model misfit encountered in the data.

Recent work by Jack Stenner shows that the most useful misfit inflation formula is

$$\text{Real S.E. of an estimated measure} = \text{Model S.E.} \times \text{Maximum} \left[1.0, \sqrt{\text{INFIT mean-square}}\right]$$

In practice, this "Real" S.E. sets an upper bound on measure imprecision. It is the "worst case" error. The actual S.E. lies between the "model" and "real" values. But since we generally try to minimize or eliminate the most aberrant features of a measurement system, we will probably begin by focusing attention on the "Real" S.E. as we establish that measurement system. Once we become convinced that the departures in the data from the model are primarily due to modelled stochasticity, then we may base our decision-making on the usually only slightly smaller "Model" S.E. values.

What about Infit mean-squares less than 1.0? These indicate overfit of the data to the Rasch model, but do not reduce the standard errors. Instead they flag data that is lacking in randomness, i.e., is too deterministic. Guttman data are like this. Their effect is to push the measures further apart. With perfect Guttman data, the mean-squares are zero, and the measures are infinitely far apart. It would seem that inflating the S.E.s would adjust for this measure expansion, but Jack Stenner's work indicates that this is not so. In practice, some items overfit and some underfit the model, so that the overall impact of low infit on the measurement system is diluted.

339. Starting Winsteps from the DOS prompt

Winsteps can also be invoked from the DOS prompt in a DOS window. At the prompt enter

C:/>WINSTEPS(Enter)

Winsteps proceeds with its standard operations.

You can enter control and output files directly on the prompt line.

C:/>WINSTEPS SF.txt SF.OUT(Enter)

Winsteps starts analysis immediately. You will not be prompted for "Extra Specifications"

You can also enter extra specifications here:

C:/>WINSTEPS SF.txt SF.OUT chart=yes distractors=no(Enter)

Leave no spaces within specifications, or place them in quotes, e.g.,

C:/>WINSTEPS SF.txt SF.OUT "chart = yes" "distractors = no"(Enter)

To perform the previous analysis again, with a temporary report output file:

C:/>WINSTEPS @(Enter)

@ is replaced by the top control file on the Files= menu. If no output file is specified, then a temporary one is
used.

For Batch file operation, see Batch=

**340. Subtest scoring**

A test or protocol may consist of a series of subtests. Code each item in its item label with what subtest it belongs to.

1. Analysis of subtests.
   Items and Persons: Use ISELECT= in your Winsteps control file to select the relevant subtest. This performs an independent analysis of the items and persons for the subtest.

2. Reporting of subtests.
   Items: In an overall analysis, the items of individual subtests can be reported after applying ISELECT= from the Specification pull-down box.
   Persons: a measure for each person on each subtest can be obtained by specifying the subtest character in DPF= and producing Table 31 or the DPF plot.

**341. Transposing the data matrix**

To transpose the rows and columns of the data matrix, select Transpose on the Output Files menu.

```
Control variable file =
ITEM File IFILE =
PERSON File PFILE =
Structure File SFILE =

Category/Option,Distracter File DISFILE =
ITEM-Structure File ISFILE =
Response File RFILE =
Score File SCFILE =
Observation File XFILE =

Matrix File IPMATRIX =
Correlation File ICORFILE =
Correlation File PCORFILE =
Graphics File GRFILE =
Guttmanized File GUTTMAN =
Simulated Data File SIMUL =

Transposed Data File TRANSPOSE =

GradeMap Item and Student files
```

then
Original data: data codes are those in the data file.
Scored data: data after applying NEWSCORE=, IVALUE=, KEY=, etc.
Recounted data: data after applying STKEEP=No, etc.
Permanent file: file name for the transposed file. Enter in the white box, or use the "Browse" function.
The transposed file is created and displayed.
Temporary file: use a temporary file: this can be "saved as" a permanent file.
The transposed file is created and displayed.
Launch Winsteps: launch Winsteps with the permanent transposed file as its control file.
Display original: show the original, untransposed, control file.
Done: transposing actions completed
Cancel: exit from this routine
Help: show Help file.

producing, for Exam1.txt:

; Transposed from: C:\WINSTEPS\examples\exam1.txt
&INST
TITLE = "TRANSPOSED: KNOX CUBE TEST"
ITEM = KID
PERSON = TAP
NI = 35           ; ACTIVE PERSONS BECOME COLUMNS
NN = 18           ; ACTIVE ITEMS BECOME ROWS
ITEM1 = 1
XWIDE = 1
NAME1 = 37
NAMLEN = 16           ; ends at 52
CODES = 01
STKEEP = Y
; Add here from original control file: C:\WINSTEPS\examples\exam1.txt
&END
Richard M  1
...... (more item labels)
Helen   F  35
END NAMES
11111111111111111111111111111111 1-4           1
...... (more data records)
00000000000000000000000000000000 4-1-3-4-2-1-4 18

342. Unobserved and dropped categories

If you have data in which a category is not observed, then you must make an assertion about the unobserved category. There are several options:

For intermediate categories: either
(a) this category will never be observed (this is called a "structural zero"). Generally, these categories are
collapsed or recoded out of the rating scale hierarchy. This happens automatically with STKEEP=No. or (b) this category didn't happen to be observed this time (an "incidental" or "sampling" zero). These categories can be maintained in the rating scale hierarchy (using STKEEP=Yes), but are estimated to be observed with a probability of zero.

For extreme categories:
(a) if this category will never be observed, the rating scale is analyzed as a shorter scale. This is the Winsteps standard.
(b) if this category may be observed, then introduce a dummy record into the data set which includes the unobserved extreme category, and also extreme categories for all other items except the easiest (or hardest) item. This forces the rare category into the category hierarchy.
(c) If an extreme (top or bottom) category is only observed for persons with extreme scores, then that category will be dropped from the rating (or partial credit) scales. This can lead to apparently paradoxical or incomplete results. This is particularly noticeable with ISGROUPS=0.

In order to account for unobserved extreme categories, a dummy data record needs to be introduced. If there is a dropped bottom category, then append to the data file a person data record which has bottom categories for all items except the easiest, or if the easiest item is in question, except for the second easiest.

If there is a dropped top category, then append to the data file a person data record which has top categories for all items except the most difficult, or if the most difficult item is in question, except for the second most difficult.

This extra person record will have very little impact on the relative measures of the non-extreme persons, but will make all categories of all items active in the measurement process.

If it is required to produce person statistics omitting the dummy record, then use PSELECT= to omit it, and regenerate Table 3.

### 343. User-friendly rescaling

Transforming logits into other units using UIMEAN=, UPMEAN=, USCALE=. These can be more meaningful for particular applications, see Chapter 8 of BTD. Anchor values are treated according to UANCHOR=.

Example 1: CHIPS are a useful transformation, in which 1 logit = 4.55 CHIPS. In this user-scaling system, standard errors tend to be about 1 CHIP in size. The recommended control variable settings are:

- USCALE = 4.55
- UIMEAN = 50
- UDECIM = 1
- MRANGE = 50

The probability structure of this relationship is:

<table>
<thead>
<tr>
<th>Probability of Success</th>
<th>Difference between Person Ability Measure and Item Difficulty Measure in CHIPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>.10</td>
<td>-10</td>
</tr>
<tr>
<td>.25</td>
<td>-5</td>
</tr>
<tr>
<td>.50</td>
<td>0</td>
</tr>
<tr>
<td>.75</td>
<td>5</td>
</tr>
<tr>
<td>.90</td>
<td>10</td>
</tr>
</tbody>
</table>

Example 2: WITs are one tenth the size of CHIPS, enabling the elimination of decimals from your output tables.

- USCALE = 45.5
- UIMEAN = 500
- UDECIM = 0
- MRANGE = 500

Example 3: You want the lowest reportable person measure to be 0 and the highest to be 100. Looking at Table 20, you see the extreme values are -4.53 and +5.72. You have not used USCALE= and UMEAN=.
USCALE = (wanted range) / (current range)
USCALE = (100 - 0) / (5.72 - -4.53) = 100 / 10.25 = 9.76
UMEAN = (wanted low) - (current low * USCALE) = 0 - (-4.53 * 9.76) = 44.20

Required values are:
USCALE = 9.76
UMEAN = 44.20
UDECIM = 0 to show no decimal places in report

Example 4: You want the lowest reportable person measure to be 0 and the highest to be 100. Looking at Table 20, you see the extreme values are -4.53 and +5.72. The current values in the output are USCALE=1 and UMEAN=0.

USCALE = (previous USCALE) * (wanted range) / (current range) = 1* (100 - 0)/ (5.72 - -4.53) = 1 * 100 / 10.25 = 9.76
UMEAN = (wanted low) - (current low - previous UMEAN) * (wanted range)/(current range) = 0 - (-4.53 - 0) *100/10.25 = 44.20
UDECIM = 0 to show no decimal places in report

Double checking, when previous UMEAN=0, USCALE=1:
low value = (current low) * (USCALE) + (UMEAN) = (-4.53 * 9.76) + 44.20 = -0.01
high value = (current high) * (USCALE) + (UMEAN) = (5.72 * 9.76) + 44.20 = 100.02

Example 5: You want the lowest reportable person measure to be 100 and the highest to be 900. Looking at Table 20, you see the extreme values are -4.53 and +5.72. Looking at the second page of output, you see the current values are USCALE=1 and UMEAN=0.

USCALE = (previous USCALE) * (wanted range: 900 - 100) / (reported range: 5.72 - -4.53) = 1 * 800 / 10.25 = 78.05
UMEAN = (wanted low) - (reported low - previous UMEAN) * (wanted range)/(reported range) = 100 - (-4.53 - 0) *800/10.25 = 453.56
UDECIM = 0 to show no decimal places in report

Example 6: You want norm-referenced user-scaling, such that the person mean is 0.0, and the person sample standard deviation is 1.0.

In a standard analysis, set:
UDECIM=4
USCALE=1
UMEAN=0

Look at Table 18

<table>
<thead>
<tr>
<th>ENTRY</th>
<th>RAW</th>
<th>NUMBER</th>
<th>SCORE</th>
<th>COUNT</th>
<th>MEASURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAN</td>
<td>6.7</td>
<td>14.0</td>
<td>.0</td>
<td>2.2202</td>
<td></td>
</tr>
<tr>
<td>S.D.</td>
<td>2.4</td>
<td>.0</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Set (either in a new analysis, or using the "Specification" pull-down menu
USCALE = 1/person S.D. = 1/2.2202 = 0.4504
UMEAN = - person mean/person S.D. = - (.3728)/2.2202 = 0.1679

Look at Table 18

<table>
<thead>
<tr>
<th>ENTRY</th>
<th>RAW</th>
<th>NUMBER</th>
<th>SCORE</th>
<th>COUNT</th>
<th>MEASURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAN</td>
<td>6.7</td>
<td>14.0</td>
<td>.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.D.</td>
<td>2.4</td>
<td>.0</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example 7: You want to give your pass-fail point the value 400 and 100 to the lowest reported measure.
Inspecting your output you see that the pass-fail point is at 1.5 logits, and that -3.3 logits is the lowest reported measure.

Then 400-100 new units = 1.5 - (-3.3) logits, so
USCALE = 300 / 4.8 = 62.5
UMEAN = 400 - (1.5) * 62.5 = 306.25
Then:  1.5 logits = 306.25 + 1.5*62.5 = 400
-3.3 logits = 306.25 - 3.3*62.5 = 100

344. Using a word processor or text editor

If WordPad does not work on your computer or you don’t like it, then change word processor

a) Input files: all lines in your control and data files follow DOS text conventions. This means that files created with a Word Processor, such as "Word Perfect", must be saved as "DOS-text with line breaks" or "ASCII" files.

1. Lines must not contain tabs or word processor codes.
2. Lines cannot overflow onto the next line, except for data records which are processed using the FORMAT= or MFORMS= control variables.
3. Lines must end with DOS or ASCII Carriage Return and Line Feed codes.

Be particularly careful to instruct your Word Processor to allow more characters within each line than are present in the longest line in your control or data files. Then your Word Processor will not break long data or control lines into two or more text lines with "Soft Return" codes. These cause WINSTEPS to malfunction. Space for a large number of characters per line is obtained by specifying a very wide paper size and/or a very small type size to your Word Processor.

When using "Word Perfect" to edit control or data files, select the smallest available type size (often 20 cpi or 5 pt). Define and use a very wide (50 inch) page style. It does not matter whether your printer can actually print it. Always save control and data files as "DOS-text with line breaks" or ASCII files.

With WordStar, use "Non-Document" mode to avoid these difficulties.

b) Output files: when importing WINSTEPS output into a document file, the following options have proved useful:

Base Font - 17 cpi (or more) or 8 point (or smaller) or 132 characters per line (or more)
Left Justify
Page numbering
Margins: top = 1", bottom = 0.5", left = 1", right = 0"

345. Weighting items and persons

There are circumstances in which certain items are to be given more influence in constructing the measures than others. For instance, certain items may be considered critical to the demonstration of competence. Unweighted data is preferable for calibrating the Rasch items. This is because each observation is modeled to contribute one unit of independent statistical information. The effect of weighting is to distort the distribution of independent statistical information in the data.

Step 1. Analyze the data without weighting. Investigate misfit, construct validity etc.

Step 2. Weight the items. Compare the item calibrations with weighted and unweighted data to identify where there are discrepancies.

Though WINSTEPS supports several methods, IWEIGHT= is simplest.
Another approach is to replicate the data for particular items. This can be done with FORMAT= without changing the data file.
Items can also be rescored from say, 0-1 to 0-2, but this makes variable maps difficult to interpret.

The weights applied to items are persons are used in computing the measure estimates, standard errors and fit statistics. When using significance tests with weighting, normalize the weights so that the total amount of
independent statistical information in the data is not over- or under-inflated, i.e., when using PWEIGHT= with an observed sample size of N, multiply all PWEIGHT= values by N / (sum of all weights).

The standard is weights = 1.

When an item or person is weighted as 2, then the estimation acts as though that item or person appears twice in the data file.

When an item or person is weighted as 0, then that person does not influence the estimates, standard errors or fit statistics of other persons and items, but does have measure, standard error and fit statistics computed on all observations for itself. This is useful for evaluating pilot or off-dimensional items, or measuring idiosyncratic persons.

Weight Selection: On the output tables menu, these are the options for persons and/or items. When IWEIGHT= or PWEIGHT= are used in estimation, reports can be adjusted to reflect those weights or not. Weights of zero are useful for pilot items, variant items or persons with unusual characteristics. These can be reported exclusively or excluded from reports.

1) all items or persons are reported, with their weights (the standard). Tables 23 and 24 are computed as though all weights are 1.

2) items or persons with a weight of 0 are excluded from the reporting. Tables 23 and 24 are computed as though all weights are 1, but zero weights are omitted.

3) only items or persons with a weight of 0 are reported. Tables 23 and 24 are computed only from items or persons with a weight of 0.

4) all items or persons are reported as though they have a weight of 1.

346. Winsteps: history and steps

What is the origin of Winsteps and to what does "steps" refer? Winsteps is an outcome of this process of development:

In 1983, Benjamin D. "Ben" Wright of the University of Chicago and "Mike" Linacre released the first Rasch analysis program for personal computers. It was also the first to allow missing data.

1983: Microscale (on PCs). "Rasch scaling by microcomputer" - since MSDOS was limited to 8-character program names, the actual execution name was "MSCALE".

1987: Mscale (dichotomies and Andrich rating scales) + Msteps (for partial credit "steps"). Ben implemented the Microscale algorithm on a Unix minicomputer, but kept the PC execution name, "Mscale".
1989: Bigscale (back to PCs). Again under MS-DOS but with much larger datasets. Mike takes over development again.

1991: Bigsteps (the functionality of Msteps was included). Ben interpreted this to mean "big steps forward in social science measurement"

1998: Winsteps (when made Windows native). Ben interpreted this to mean "winning steps forward in social science measurement"

When talking about Rasch measurement, Ben used "step" to mean:

(a) the category number counting up from 0 at the bottom. The bottom step, for dichotomies or polytomies, was the lowest category, always numbered 0. Ben would talk about going up and down the steps as one moved up and down the latent variable.

(b) the location of the transition from one category to the next higher category on the latent variable. Now called the Rasch-Andrich threshold for polytomies and the item difficulty for dichotomies.

(c) the process of moving from one category to the next as one's amount of the latent variable changes. A low negative threshold below a category indicates that the category is easy to step into as one moves up the latent variable. A high positive threshold below a category indicates a category that is hard to step into. So "disordered" thresholds around a category (high below, low above) indicate a category that is "hard to step into and easy to step out of" as one moves up the latent variable, i.e., a narrow category. The extreme of this is an infinitely-narrow, i.e., unobserved, category. It is infinitely hard to step into and infinitely easy to step out of.

347. **Wordpad or text editor**

The default text editor used by Winsteps is WordPad, a Microsoft Accessory. If WordPad is not available, NotePad is used.

**How Winsteps decides what Text Editor to use:**

(a) On its first installation, or if you delete the file Winsteps.ini in your Winsteps directory, Winsteps looks on your C: drive for Wordpad.exe.

If there is more than one version of Wordpad.exe on your computer, Winsteps may try to use the wrong one. The Output Tables may not display, and you may receive messages like: "The application has failed to start because MFCANS32.dll was not found. Reinstalling the application may fix the problem".

Use Windows "Find" or "Search" to look for Wordpad.exe. If there is more than one found, the correct one is probably in "Accessories". Double-click the WordPad entries until one launches. Go to (e) below and check that the correct one is referenced in Winsteps.ini

(b) If WordPad is found automatically, then its path is placed in Winsteps.ini as the Editor.

(c) If WordPad is not found, then the Editor associated with .txt files is used. If there is none, **NotePad** is used.

(d) If file Winsteps.Ini exists, then Winsteps uses the Text Editor or Word Processor assigned by Editor= in Winsteps.Ini

(e) You can examine Winsteps.ini from the Edit pull-down menu, "Edit initial settings", or using a word processor or text editor to look at file Winsteps.ini in the Winsteps directory.

**Using your own text editor:**

You can make WINSTEPS use your own text editor or word processor by changing the Editor in Edit initial settings or editing Winsteps.ini in your Winsteps folder. Put the full pathname of your editor in place of WordPad, e.g., for
You can find the full path to your word processor by doing a "Find" or "Search" for it from the "Start" menu.

A useful replacement for Wordpad is the shareware program **TextPad**. This has many more features.

**348. Data vanishes**

Winsteps runs, but not all your data are reported. Please check:
- Your control and data files look correct. Use the Edit menu.
- \*ITEM1\* points to the first column of responses
- CODES= has all valid data codes.
- CUTLO= and CUTHI= are not specified.
- ISELECT=, PSELECT=, IDFILE=, PDFILE=, IDELETE=, PDELETE= are as you want.

**349. Display too big**

If your screen resolution is low, you may not see all the Winsteps window on your screen.

1. Start Menu => Settings => Control Panel => Display => Settings => Screen Resolution => More
2. For dialog boxes, right-click on the screen. This will bring up this dialog box:

![Dialog Box](image)

Click on "Left" to move the over-sized dialog box left, etc. Click "Home" to position it in the top left-hand corner of your screen.

**350. File misread problems**

Some file formatting problems are reported with error messages. Others simply mislead Winsteps.

**General rules:**
(a) All control and data files must be in "DOS TEXT" or ASCII format.
(b) To check if your control specifications have been correctly interpreted, look at Table 0, at the end of your "output" file.
(c) To check if your data have been correctly interpreted, produce an RFILE= file and look at the response strings, particularly for column alignment.

**Common Problems and their Diagnoses**

1) **Message:** VARIABLE UNKNOWN OR VALUE INVALID:

Diagnosis: A line in your control file is not of the format "variable = value". This may be because your control file is in not in DOS Text or ASCII format. Return to your word processor or WordPad and save your control file as a DOS Text file. Then rerun.

2A) **Message:** PROBLEM: BLANK LINE NEEDED AT END OF: filename

Diagnosis: The last line of file "filename" does not end with a line-feed (LF) code (so does not meet strict Windows specifications). This can be fixed by adding a blank line to the end of that file. You can use **WordPad** to do this by editing the file from the Winsteps .
2B) Message:
Use last entry in File menu to add final blank line to:
filename
Then press enter key, or enter a different file name:

Diagnosis: The last line of file "filename" does not end with a line-feed (LF) code (so does not meet strict Windows specifications). This can be fixed by adding a blank line to the end of that file. You can use WordPad to do this by access the file from the Winsteps File menu. It will probably be the last file listed there. You could also copy the file, add a blank last line, and then type in the file name in Winsteps.

3) Blank characters disappear, causing misalignment of data or control instructions.

Diagnosis:
(a) These "blank" characters could be ASCII "13" (CR) codes. Some word processors treat them as blanks, some as end-of-line codes, and some ignore them, making them "disappear". Winsteps ignores them. WordPad reports them as squares. Replace them with blanks.
(b) These "blank" characters could be ASCII "9" (TAB) codes. Word processors expand these into several blanks. Winsteps treats them as one blank. For consistency, globally replace Tab characters with blanks.

4) Data or control lines disappear.

Diagnosis:
(a) Word processors automatically wrap long lines across several lines. Winsteps does not. Make sure that each line ends with an "end of line" code.
(b) Some word processors use ASCII "13" (CR) as an end-of-line code. Winsteps ignores these codes. Use ASCII "10" (LF) as the end-of-line code. Press the Alt key, while typing 10 on the numeric keypad to generate this code.

5) Data or control lines are split across several lines.

Diagnosis: You may have edited your control or data files with a word processor that wrapped one line across several lines. When you saved this file in DOS text format, each line was then saved as several lines. Re-edit the file, reduce the point size, maximize the page size (e.g., landscape) until each line only occupies one line. Then save the file again in DOS text or ASCII format.

6) Message: BEYOND CAPACITY
Diagnosis: Your data file contains more person records than can be analyzed in one run. Some records have been bypassed.

Data sets that exceed program capacity to analyze in one run offer opportunities, as well as challenges. There are several strategies.
I. Analyze a sample of the data. Use this to produce anchor values. Then, using the anchor values, run all the data one section at a time.
II. Analyze a sample of the data. Analyze another sample. Compare results to identify instability and compute reasonable anchor values. Remember that small random changes in item calibrations have negligible effect on person measures.

To select a sample of your data, use the FORMAT= statement. See the example on pseudo-random person selection on page.

There are versions of Winsteps that support more persons. Contact www.winsteps.com for details.

7) File read errors
Reading Control Variables..
Input in process..
Opening: c:\input-file.txt
Your data file may be “read only” and missing an end-of-line indicator on the last data record. Copying and paste your data file as a data file with ordinary file access privileges. Then reference this file in your `DATA=` specification.

### 351. If Winsteps does not work

1) Repeat the installation process. It will not delete any of your data files.
2) Check that WINSTEPS runs on an example dataset, e.g., exam1.txt
3) If the program will not run, or produces implausible results:
   a) It may be a Windows-related problem, see [www.winsteps.com/problems.htm](http://www.winsteps.com/problems.htm)
   b) There may not be enough disk space for work files, see "Not enough disk space".
   c) There may not be sufficient RAM memory to execute. See "Not enough memory".
   d) It may be a WordPad problem, see Changing your Word Processor setting.
4) If you still have problems, use the comment form at [www.winsteps.com](http://www.winsteps.com)

### 352. Initialization fails

You’ve downloaded and run WinstepsInstall.exe or MinistepInstall.exe or a similar installation routine. It ran, and then automatically started Winsteps or Ministep, but now ...

**It "hangs" with the message: "Constructing Winsteps.ini ..."**

So then you kill Winsteps or Ministep from the “Ctrl+Alt+Delete” program box

*Is this the scenario?* If so, here’s what to do next:

You need to create a text file called “Winsteps.ini” with NotePad in the same directory that holds Winsteps.exe or Ministep.exe.

If you cannot do this manually, then this is the reason why Winsteps or Ministep is failing.

The format of the text file is:

```
Editor="C:\Program Files\Accessories\WORDPAD.EXE"
Excel="C:\Program Files\Microsoft Office\Office\EXCEL.EXE"
Filter=All Files (*.*)
Temporary directory="C:\TEMP"
Reportprompt=Yes
Welcome=Yes
```

Replace `C:\Program Files\Accessories\WORDPAD.EXE` with the path to *WordPad* on your machine.

Replace or remove the line `Excel=`

If you do not have “Wordpad.exe” on your computer, then put in `Editor="NOTEPAD.EXE"

Then start Winsteps or Ministep again. (Not the Install procedure).

Pull-down the Winsteps or Ministep "edit" menu immediately, look at “initial settings”. They should correspond to `Winsteps.ini`

Now proceed normally with *Starting Winsteps*

Other installation snags are solved at [www.winsteps.com/problems.htm](http://www.winsteps.com/problems.htm)

### 353. Not enough disk space

You need about twice as much work file space on your disk as the size of your data file. Double-click "My Computer" and right-click on the disk drive. The “properties” shows the amount of available disk
space.
Temporary work files are placed in the TEMP directory (see WINSTEPS.INI). The temporary table output files are placed temporarily in the current directory (reported on your screen when WINSTEPS starts). Delete unwanted files to give yourself more disk space, or log onto a different disk drive, with more available space, before executing WINSTEPS. Files with names "Z......ws.txt", "Z......ws.xls", "(numbers)ws.txt" and ".TMP" and files in the TEMP directory are work files. These can be deleted.

354. Not enough memory

WINSTEPS can use all available memory. If memory is exhausted, then WINSTEPS fails or proceeds very slowly. Terminate other programs. Due to "memory leakage", Windows gradually loses access to memory not properly released when programs terminate. Reboot your computer to free up this memory.

355. Plotting problems

The Plots menu calls Microsoft Excel to perform many plotting functions. This enables you to use all the functionality of Excel to customize the plots to your needs.

Many different versions of Excel have been published by Microsoft, and they have some incompatibilities. You may find that plotting fails with an error code, e.g., "1004".

The Winsteps-to-Excel interface is in a module in your Winsteps folder named winexcel.exe. Please replace this module with a module that matches your version of Excel. These are available from www.winsteps.com/problems.htm

356. Tab-delimited file problems

Question:
I'm having trouble reading files from SPSS or EXCEL. I save the file in EXCEL or SPSS as a text file and have tried saving in all formats (tab-delimited, comma, etc.). It always saves the data with tabs separating the column values. When I read the file in Winsteps, the tabs are translated into columns, thus producing every other row of empty columns. How do I get around this?

Answer:
There are numerous ways round this difficulty.

(a) Use Winsteps to format your SPSS .sav file into a Winsteps file (use the SPSS pull-down menu in Winsteps).

(b) Use the Winsteps "setup" routine to format your EXCEL data.
Start Winsteps "setup" (Setup pull-down menu in Winsteps) Copy the cells you want from EXCEL, paste them into the Data area of the Setup screen.

(c) Get EXCEL to write out a file without Tabs: "Save as" a ".prn" file. "Formatted text (space delimited)"
Before you do this, set all data columns widths to 1.

(d) Tell Winsteps the data are Tab separated: use DELIMITER= .... but this is tricky to get to work correctly.

(e) Edit your data file with WordPad. Replace all tabs with nothing. To "find" a Tab, highlight a Tab, copy it (ctrl+C), then paste it into the "find" box: ctrl+v.

(f) Edit your data file with WordPad. Replace all tabs with a space. To "find" a Tab, highlight a Tab, copy it (ctrl+C), then paste it into the "find" box: ctrl+v. Press the space bar in the "replace" box. Use XWIDE=2

(g) Use FORMAT= to pick up every other column. This is also tricky.

357. Winsteps problems and error codes

For Windows-related and other problems, see www.winsteps.com/problems.htm
Winsteps expects its input files to be in MS-DOS text file format, i.e., readable by NotePad. If the file appears not to be, this is displayed:

**File may not have CR+LF line breaks**
The file is not in standard "text" file format. This can be reported for files created on a Mac.
**Solution:** Open the file with WordPad. Check that the file looks correct. Then "Save as type: Text Document - MS-DOS Format"

**VARIABLE=VALUE EXPECTED : \{rtf1\}....**
Your control file is in RTF, not TXT format. Double-click on the file, and "Save as" "text with line breaks" or or "MS-DOS format"

**NUMERIC VALUE EXPECTED: \üþT I T L E   =**
Your control file is in Unicode, not ANSI coding. Double-click on the file, and "Save as" with "Encoding" "ANSI" or "MS-DOS format"

**NUMERIC VALUE EXPECTED: ý T I T L E   =**
Your control file is in Unicode Blgendiam, not ANSI coding. Double-click on the file, and "Save as" with "Encoding" "ANSI" or "MS-DOS format"

**NUMERIC VALUE EXPECTED: ïn¿TITLE =**
Your control file is in UTF-8, not ANSI coding. Double-click on the file, and "Save as" with "Encoding" "ANSI" or "MS-DOS format"

**UNRECOGNIZED VARIABLE (CHECK SPELLING): Facets = 3**
This can occur when a Facets control file is submitted to Winsteps.

Error codes are reported by Winsteps for which no automatic action is possible. See Winsteps Help.

358. **Winsteps SPSS error codes**

If you encounter one of these when running Winsteps and don't understand the reason, please contact [www.winsteps.com](http://www.winsteps.com). See Winsteps Help
Control Variable Index

&END end of control variables (required)..................................................................................................................180
&INST start of control variables (ignored)..................................................................................................................180
@Fieldname = name for location in person or item label...........................................................................................179
ALPHANUM= alphabetic numbering................................................................................................................................69
ASCII= output only ASCII characters = Yes................................................................................................................69
ASYMPTOTE= item upper and lower asymptotes = No..................................................................................................70
BATCH= Running in batch mode = No............................................................................................................................71
BYITEM= display graphs for items = Yes.........................................................................................................................74
CATREF= reference category for Table 2 = (item measure)...............................................................................................74
CFILE= scored category label file ..................................................................................................................................
CHART= output graphical plots in Tables 10 & 13-15 = Yes..............................................................................................76
CLFILE= codes label file...................................................................................................................................................
CODES= valid data codes = 01...........................................................................................................................................
CONVERGE= select convergence criteria = Either ...........................................................................................................
CSV= comma-separated values in output files = No...........................................................................................................81
CURVES= curves for Tables 2 & 21 = 111, all........................................................................................................................82
CUTHI= cut off responses with high probability of success = 0, no...........................................................................
CUTLO= cut off responses with low probability of success = 0, no............................................................................
DATA= name of data file = (data at end of control file)..................................................................................................
DELIMITER= data field delimiters = " ", fixed fields........................................................................................................83
DIF= columns within person label for Table 30 = $S$1W1................................................................................................85
DISCRIMINATION= report item discrimination = No......................................................................................................86
DISFILE= category/distractor/option count file.............................................................................................................87
DISTRT= output option counts in Tables 10 & 13-15 = Yes............................................................................................87
DPF= columns within item label for Table 31 = $S$1W1...................................................................................................88
EDFILE= edit data file....................................................................................................................................................88
END LABELS end of control variables ............................................................................................................................89
END NAMES end of control variables..............................................................................................................................89
EQFILE= code equivalences........................................................................................................................................89
EXTRSC= extreme score correction for extreme measures = 0.3...............................................................................89
FITHIGH= higher bar in charts = 0, none.........................................................................................................................90
FITI= item misfit criterion = 2.0.......................................................................................................................................91
FITLOW= lower bar in charts = 0, none............................................................................................................................91
FITP= person misfit criterion = 2.0................................................................................................................................91
FORMAT= reformat data................................................................................................................................................91
FORMFD= the form feed character = ^, MS-DOS standard ............................................................................................97
FRANGE= half-range of fit statistics on plots = 0, auto-size.............................................................................................97
G0ZONE= One-to-zero zone = 50 %.................................................................................................................................102
G1ZONE= Zero-to-one zone = 50 %................................................................................................................................102
GRFILE= probability curve coordinate output file........................................................................................................98
GROUPS= assigns items to groups = " ", all in one grouping...........................................................................................98
GRPFROM= location of GROUPS= = No, before &END................................................................................................101
GUFILE= Guttmanized response file............................................................................................................................102
HEADER= display or suppress subtable headings = Yes..............................................................................................102
HIADJ= correction for top rating scale categories = 0.25...............................................................................................103
HLLINES= heading lines in output files = Yes................................................................................................................103
IFILE= item anchor file................................................................................................................................................103
IANCHQU= anchor items interactively = No.....................................................................................................................105
ICORFILE= item residual correlation file....................................................................................................................105
IDELETE= item one-line item deletion..........................................................................................................................106
IDELQU= delete items interactively = No.......................................................................................................................106
IDFILE= item deletion file..............................................................................................................................................107
IDROPEXTREME= drop items with extreme scores = No............................................................................................109
IFILE= item output file................................................................................................................................................109
ILFILE= item label file = (after &END)..........................................................................................................................111
IMAP= item label on item maps Tables 1 & 12 = (whole label).......................................................................................111
INUMB= label items by sequence numbers = No........................................................................................................111
QUOTED = quote marks around labels = Yes .................................................................154
RCONV = score residual at convergence = 0.1 ..........................................................155
REALSE = inflate S.E. of measures for misfit = No ....................................................155
RESPREC = response recoding (with NEWSCORE = or KEYn =) ..................................155
RESFROM = location of RESREC = No, before &END ..............................................157
RFILE = scored response file ....................................................................................157
SAVE = item structure anchor file ..............................................................................158
SAITEM = item numbers in SFILE = and SAFILE = (with one ISGROUPS =) = No ......163
SANCHQ = anchor structures interactively = No ........................................................163
SCOREFILE = person score file .................................................................................164
SDELQU = delete structure interactively = No ..........................................................165
SEPARATOR = data field delimiters = " ", fixed fields .................................................83
SFFILE = structure output file ..................................................................................166
SFILE = simulated data file .....................................................................................167
SPFILE = supplementary control file .......................................................................168
STBIAS = correct for JMLE statistical estimation bias = No ......................................168
STKEEP = keep non-observed intermediate categories in structure = Yes ...............169
T1P# = number of items summarized by "#" symbol in Table 1 = (auto-size) ............170
T1P# = number of persons summarized by "#" symbol in Table 1 = (auto-size) .........170
TABLES = output tables ..........................................................................................170
TARGET = estimate using information-weighting = No ............................................171
TFILE = list of Tables to be output ..........................................................................171
TITLE = title for output listing = (control file name) .................................................172
TOTALSCORE = show total scores with extreme observations = No .......................173
UANCHOR = anchor values supplied in user-scaled units = Yes ...............................173
UCOUNT = anchor scale value of 1 logit = 1 .............................................................174
UDECIMALS = number of decimal places reported = 2 .........................................174
UIMEAN = the mean or center of the item difficulties = 0 .......................................174
UPMEAN = the mean or center of the item difficulties = 0 .......................................174
USCALE = the user-scaled value of 1 logit = 1 ..........................................................175
W300 = Output files in Winsteps 3.00 format = No ....................................................176
WXACT = Wilson-Hilferty exact normalization = Yes ..............................................176
XFILE = analyzed response file ..............................................................................177
XML = consistent - almost unbiased - estimation = No ...........................................178
XWIDE = columns per response = 1 ........................................................................179