Graduate
Management
Admission
Council®

The MEDPRO Project: An SBIR Project for a Comprehensive IRT and CAT Software System—IRT Software

David Thissen

The University of North Carolina at Chapel Hill

Presented at the U.S. Government Supported CAT Programs and Projects Paper Session, June 2, 2009

Abstract

IRTPRO (Item Response Theory for Patient-Reported Outcomes) is an entirely new application for item calibration and test scoring using IRT. IRTPRO implements algorithms for maximum likelihood estimation of item parameters (item calibration) for several unidimensional and multidimensional item response theory (IRT) models for dichotomous and polytomous item responses. In addition, the software provides computation of goodness-of-fit indices, statistics for the diagnosis of local dependence and for the detection of differential item functioning (DIF), and IRT scaled scores. This paper illustrates the use, and some capabilities, of the software.

Acknowledgments

IRTPRO (Item Response Theory for Patient-Reported Outcomes) is a component of the MEDPRO Project, which also includes CATPRO (software for the administration and scoring of computerized adaptive tests). MEDPRO is being developed by Scientific Software International, Inc., in collaboration with Assessment Systems Corporation, and has been funded by SBIR contract HHSN-2612007-00013C with the National Cancer Institute of the National Institutes of Health. The IRTPRO development team is directed by Stephen du Toit, and includes Mathilda du Toit, Li Cai, Shyan Lam, and Johan du Toit, with advice from R. Darrell Bock, Robert Gibbons, and David Thissen. Presentation of this paper at the 2009 Conference on Computerized Adaptive Testing was supported in part with funds from GMAC®.

Citation

Thissen, D. (2009). The MEDPRO project: An SBIR project for a comprehensive IRT and CAT software system—IRT software. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing.* Retrieved [date] from www.psych.umn.edu/psylabs/CATCentral/

Author Contact

**David Thissen, Department of Psychology, UNC-CH, CB#3270, Davie Hall,
Chapel Hill, NC 27599-3270, U.S.A. Email: dthissen@email.unc.edu**

The MEDPRO Project: An SBIR Project for a Comprehensive IRT and CAT Software System—IRT Software

The IRTPRO (Item Response Theory for Patient-Reported Outcomes) component of the MEDPRO Project is an entirely new application for item calibration and test scoring using IRT. Beta release of this software is anticipated to occur in late 2009 through the middle of 2010, with full release to follow soon thereafter.

Features

Item response theory (IRT) models for which item calibration and scoring are implemented in IRTPRO are based on unidimensional and multidimensional [confirmatory factor analysis (CFA) or exploratory factor analysis (EFA)] versions of the following widely-used response functions:

- Two-parameter logistic (2PL) (Birnbaum, 1968) [with which equality constraints includes the one-parameter logistic (1PL) (Thissen, 1982)]
- Three-parameter logistic (3PL) (Birnbaum, 1968)
- Graded (Samejima, 1969; 1997)
- Generalized partial credit (Muraki, 1992, 1997) (as a special case of the nominal model)
- Nominal (Bock, 1972, 1997; Thissen, Cai, & Bock, 2010)

These item response models may be mixed in any combination within a test or scale, and any (optional) user-specified equality constraints among parameters, or fixed values for parameters, may be specified.

IRTPRO implements the method of maximum likelihood (ML) for item parameter estimation (item calibration), or it computes maximum *a posteriori* (MAP) estimates if (optional) prior distributions are specified for the item parameters. Alternative computational methods may also be used, each of which provides best performance for some combinations of dimensionality and model structure; these include:

- Bock-Aitkin (BAEM) (Bock & Aitkin, 1981)
- Bi-factor EM (Gibbons & Hedeker, 1992; Gibbons et al., 2007)
- Two-tier EM (Cai, in press-a)
- Adaptive quadrature (ADQEM) (Schilling & Bock 2005)
- Metropolis-Hastings Robbins-Monro (MH-RM) (Cai, 2010, in press-b)

The computation of IRT scale scores in IRTPRO may be done using any of the following methods:

- Maximum *a posteriori* (MAP) for response patterns
- Expected *a posteriori* (EAP) for response patterns (Bock & Mislevy, 1982)
- Expected *a posteriori* (EAP) for summed scores (Thissen & Orlando, 2001; Thissen, Nelson, Rosa, & McLeod, 2001)

Data structures in IRTPRO may categorize the item respondents into groups, and the population latent variable means and variance-covariance matrices may be estimated for multiple groups (Mislevy, 1984, 1985). [Most often, if there is only one group, the population latent variable mean(s) and variance(s) are fixed (usually at 0 and 1) to specify the scale; for multiple groups, one group is usually denoted the "reference group" with standardized latent values.]

To detect differential item functioning (DIF), IRTPRO uses Wald tests, modeled after a proposal by Lord (1977), but with accurate item parameter error variance-covariance matrices computed using the supplemented EM (SEM) algorithm (Cai, 2008).

Depending on the number of items, response categories, and respondents, IRTPRO reports several varieties of goodness-of-fit and diagnostic statistics after item calibration. The values of –2loglikelihood, Akaike information criterion (AIC) (Akaike, 1974) and the Bayesian information criterion (BIC) (Schwarz, 1978) are always reported. If the sample size sufficiently exceeds the number of cells in the complete cross-classification of the respondents based on item response patterns, the overall likelihood ratio test against the general multinomial alternative is reported. For some models, the M_2 statistic (Maydeu-Olivares & Joe, 2005, 2006; Cai, Maydeu-Olivares, Coffman, & Thissen, 2006) is also computed. Diagnostic statistics include generalizations for polytomous responses of the local dependence (LD) statistic described by Chen and Thissen (1997) and the $S\text{-}X^2$ item-fit statistic suggested by Orlando and Thissen (2000, 2003).

The operation of IRTPRO is best illustrated with the examples below.

Example 1—Unidimensional Analysis of the Affect Adjective Check List (AACL)

The Affect Adjective Check List (AACL) (Zuckerman, 1980) involves 21 adjectives; the first 11 are called the "anxiety-plus" adjectives, and the final 10 words are "anxiety-minus" adjectives. All 21 adjectives are listed in the following table:

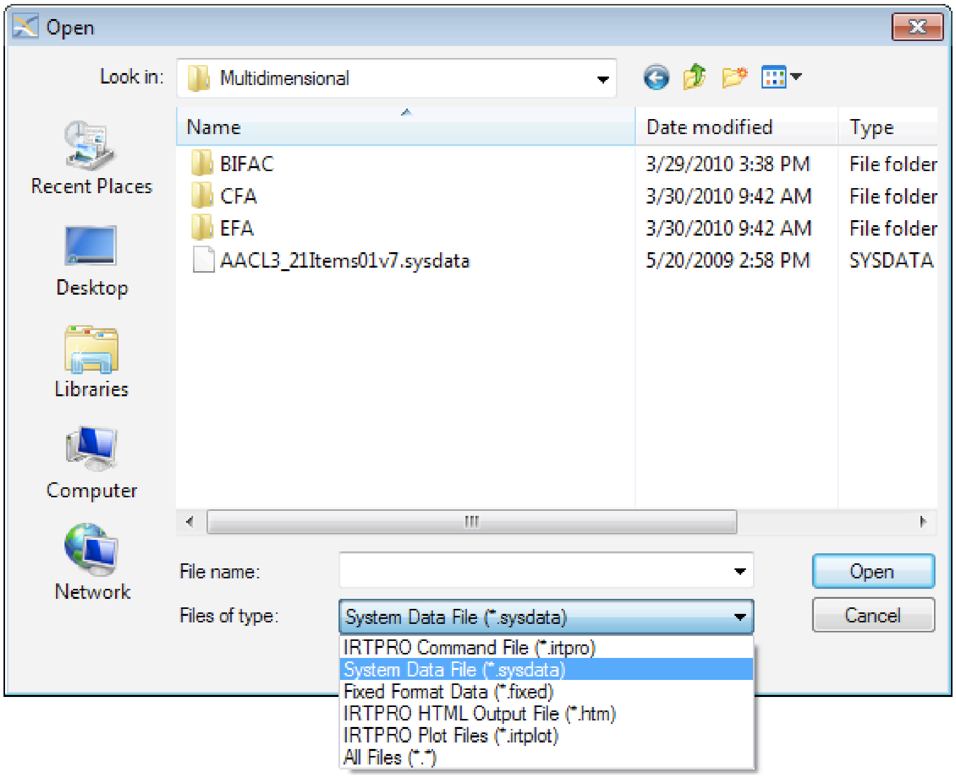
Anxiety-plus		Anxiety-minus	
1. Afraid	7. Shaky	12. Calm	17. Loving
2. Desperate	8. Tense	13. Cheerful	18. Pleasant
3. Fearful	9. Terrified	14. Contented	19. Secure
4. Frightened	10. Upset	15. Happy	20. Steady
5. Nervous	11. Worrying	16. Joyful	21. Thoughtful
6. Panicky			

To collect the data[1] analyzed here, the adjectives were framed with the instructions "Please indicate whether or not the adjective listed describes how you feel today, beginning with the time you woke up this morning." Anxiety-plus words are scored 1 if checked, and anxiety-minus words are scored 1 if not checked. In the first of two analyses of these data, a unidimensional 2PL model is fitted to the entire 21-item set, and the diagnostic statistics are examined.

The data have already been imported into IRTPRO and saved as the file "AACL_21Items01v7.sysdata" in the "C:\IRTPRO Examples\Multidimensional" folder.

[1] Data (N=290) from the 1988 "Affect Adjective Check List (CAPS-ANXIETY module)", hdl:1902.29/CAPS-ANXIETY Odum Institute [Distributor] (http://arc.irss.unc.edu/dvn/).

To begin the analysis of these data, start IRTPRO and select “Open” under the “File” menu; then navigate to the “C:\IRTPRO Examples\Multidimensional” folder, and select “Files of type:” “System Data File (*.sysdata)” in the open file dialog:


Open
Recent Places
Desktop
Libraries
Computer
Network
Open
Cancel
IRTPRO Command File (*.irtpro)
System Data File (*.sysdata)
Fixed Format Data (*.fixed)
IRTPRO HTML Output File (*.htm)
IRTPRO Plot Files (*.irtplot)
All Files (*.*)

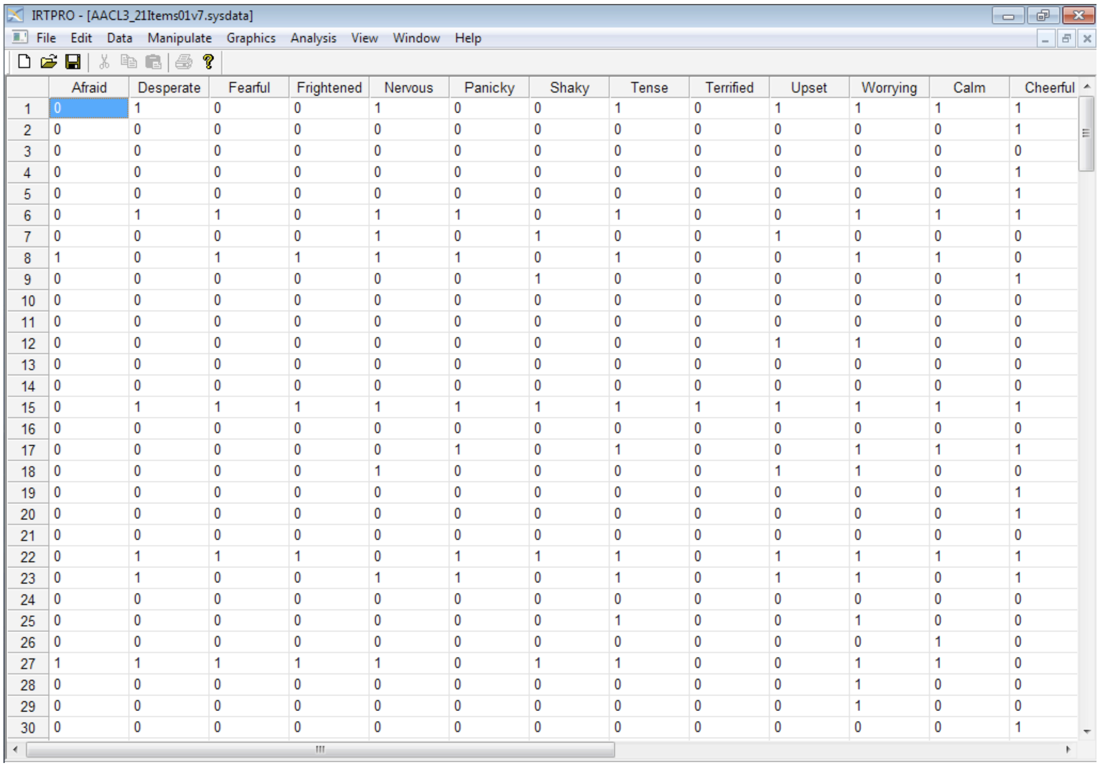
Look in: Multidimensional

Name	Date modified	Type
BIFAC	3/29/2010 3:38 PM	File folder
CFA	3/30/2010 9:42 AM	File folder
EFA	3/30/2010 9:42 AM	File folder
AACL3_21Items01v7.sysdata	5/20/2009 2:58 PM	SYSDATA

File name:

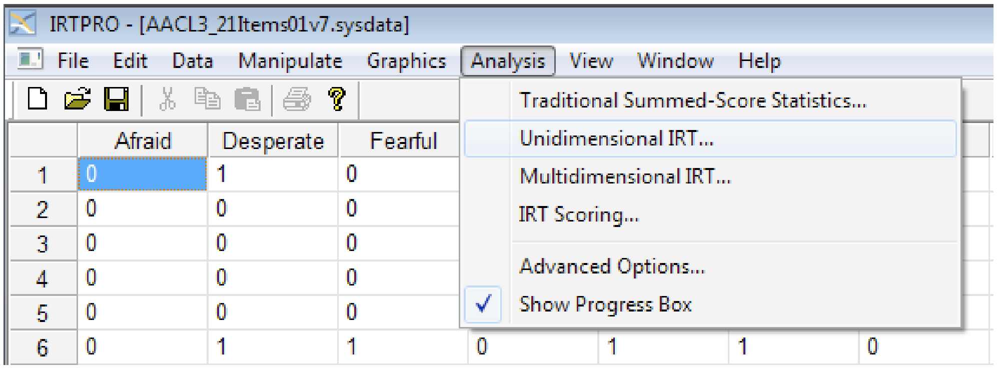
Files of type: System Data File (*.sysdata)

Opening the file “AACL_21Items01v7.sysdata” presents the data as a spreadsheet:



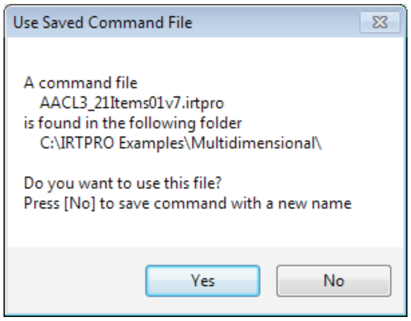
	Afraid	Desperate	Fearful	Frightened	Nervous	Panicky	Shaky	Tense	Terrified	Upset	Worrying	Calm	Cheerful
1	0	1	0	0	1	0	0	1	0	1	1	1	1
2	0	0	0	0	0	0	0	0	0	0	0	0	1
3	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	1
5	0	0	0	0	0	0	0	0	0	0	0	0	1
6	0	1	1	0	1	1	0	1	0	0	1	1	1
7	0	0	0	0	1	0	1	0	0	1	0	0	0
8	1	0	1	1	1	1	0	1	0	0	1	1	0
9	0	0	0	0	0	0	1	0	0	0	0	0	1
10	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	1	1	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	1	1	1	1	1	1	1	1	1	1	1	1
16	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	1	0	1	0	0	1	1	1
18	0	0	0	0	1	0	0	0	0	1	1	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	1
20	0	0	0	0	0	0	0	0	0	0	0	0	1
21	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	1	1	1	0	1	1	1	0	1	1	1	1
23	0	1	0	0	1	1	0	1	0	1	1	0	1
24	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	1	0	0	1	0	0
26	0	0	0	0	0	0	0	0	0	0	0	1	0
27	1	1	1	1	1	0	1	1	0	0	1	1	0
28	0	0	0	0	0	0	0	0	0	0	1	0	0
29	0	0	0	0	0	0	0	0	0	0	1	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	1

To initiate the unidimensional IRT analysis, select "Unidimensional IRT …" under the "Analysis" menu:



	Afraid	Desperate	Fearful				
1	0	1	0				
2	0	0	0				
3	0	0	0				
4	0	0	0				
5	0	0	0				
6	0	1	1	0	1	1	0

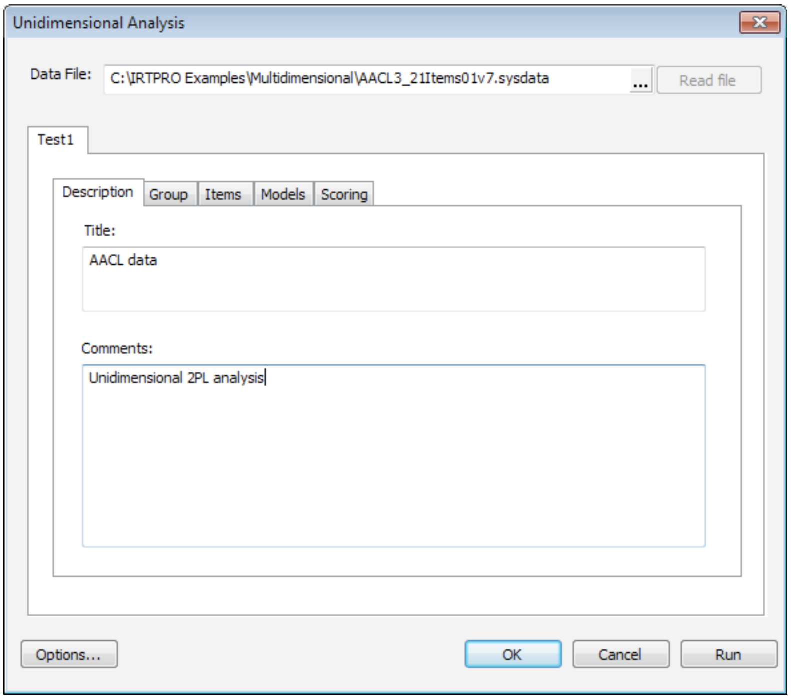
In this case there is already a saved command file called "AACL_21Items01v7.irtpro" in that folder, so IRTPRO asks if it should use the saved command file. Because a separate command file is desired to do a unidimensional analysis (the stored commands are for a multidimensional analysis) the user responds "No".


Use Saved Command File

A command file
AACL3_21Items01v7.irtpro
is found in the following folder
C:\IRTPRO Examples\Multidimensional\

Do you want to use this file?
Press [No] to save command with a new name

The Unidimensional Analysis dialog opens, and title and comments are entered:

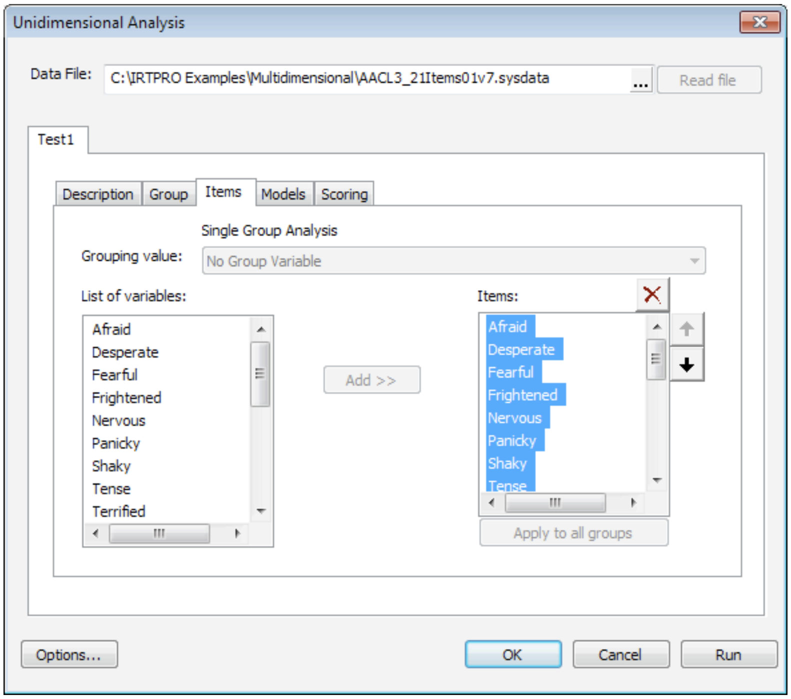

Unidimensional Analysis

Data File: C:\IRTPRO Examples\Multidimensional\AACL3_21Items01v7.sysdata

Title: AACL data

Comments: Unidimensional 2PL analysis

The “Group” tab may be skipped, because there is only one group in this analysis. The “Items” tab is used to select all the items in this case:



Data File: C:\IRTPRO Examples\Multidimensional\AACL3_21Items01v7.sysdata

Single Group Analysis

Grouping value: No Group Variable

List of variables:

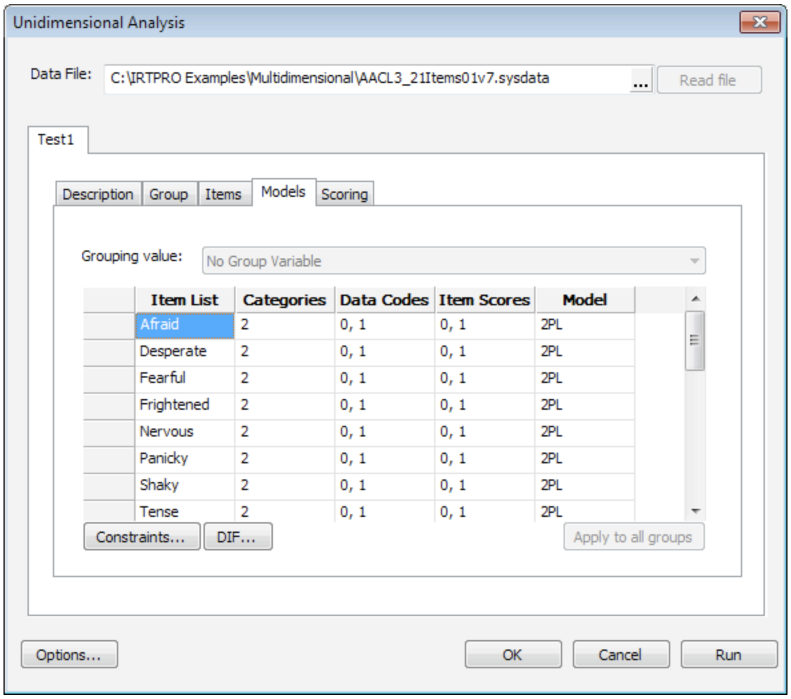
Afraid
Desperate
Fearful
Frightened
Nervous
Panicky
Shaky
Tense
Terrified

Items:

Afraid
Desperate
Fearful
Frightened
Nervous
Panicky
Shaky
Tense

There are several ways to select all of the items: (1) click on any item, and then press <Ctrl>A to select all, and then either (a) drag the set from the "List of variables" box to the "Items" box, or (2) double-click each item in turn, or (3) select each item and press the "Add >>" button.

To see more details, selection of the "Models" tab shows the list of items, their data codes, the translation of those codes into response categories, and the model selected:

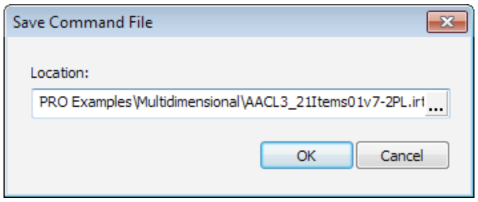


Data File: C:\IRTPRO Examples\Multidimensional\AACL3_21Items01v7.sysdata

Grouping value: No Group Variable

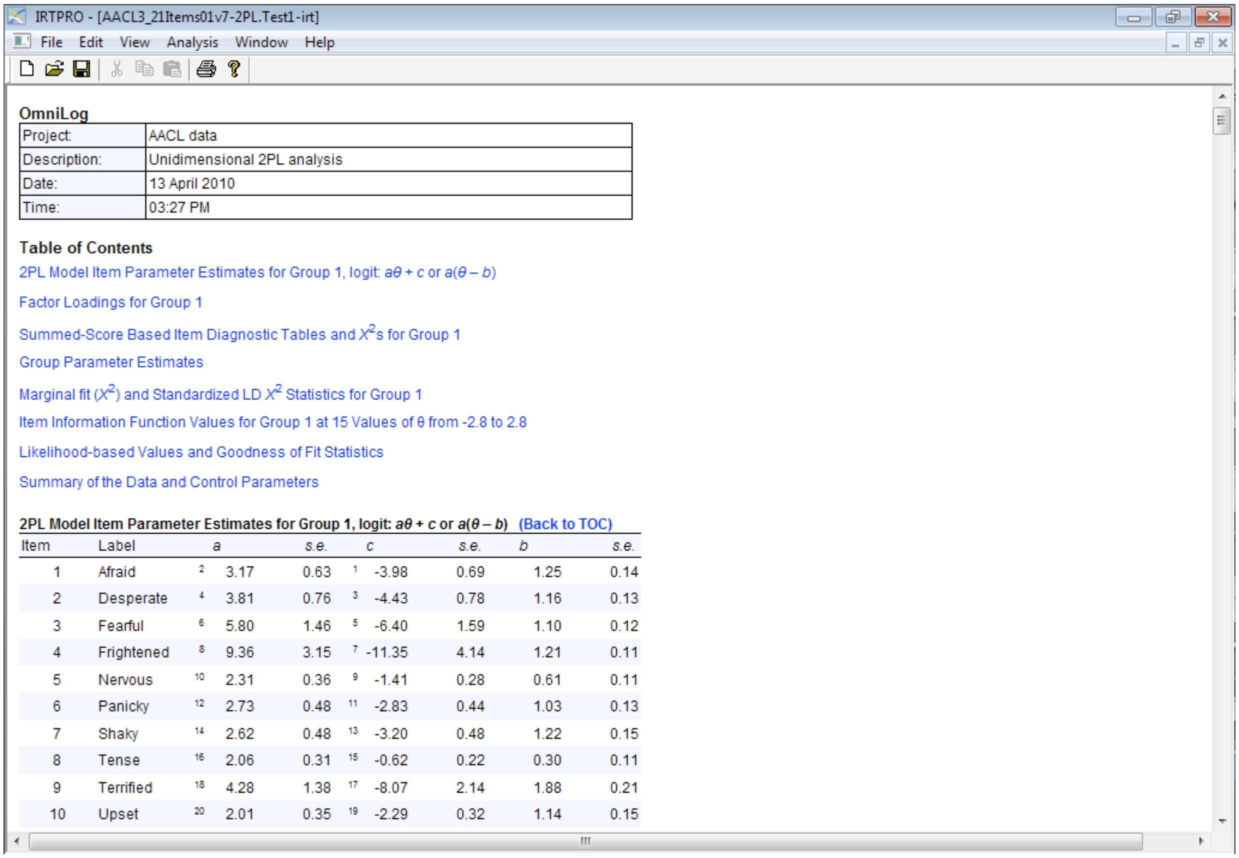
Item List	Categories	Data Codes	Item Scores	Model
Afraid	2	0, 1	0, 1	2PL
Desperate	2	0, 1	0, 1	2PL
Fearful	2	0, 1	0, 1	2PL
Frightened	2	0, 1	0, 1	2PL
Nervous	2	0, 1	0, 1	2PL
Panicky	2	0, 1	0, 1	2PL
Shaky	2	0, 1	0, 1	2PL
Tense	2	0, 1	0, 1	2PL

Clicking the “Run” button brings up a dialog box to specify the filename for the command file:



Location: PRO Examples\Multidimensional\AACL3_21Items01v7-2PL.irt

The results are then produced, after a wait of only a few seconds for this small problem; other problems may require more time.



OmniLog

Project:	AACL data
Description:	Unidimensional 2PL analysis
Date:	13 April 2010
Time:	03:27 PM

Table of Contents

2PL Model Item Parameter Estimates for Group 1, logit: $a\theta + c$ or $a(\theta - b)$ (Back to TOC)

Item	Label		*a*	*s.e.*		*c*	*s.e.*	*b*	*s.e.*
1	Afraid	2	3.17	0.63	1	-3.98	0.69	1.25	0.14
2	Desperate	4	3.81	0.76	3	-4.43	0.78	1.16	0.13
3	Fearful	6	5.80	1.46	5	-6.40	1.59	1.10	0.12
4	Frightened	8	9.36	3.15	7	-11.35	4.14	1.21	0.11
5	Nervous	10	2.31	0.36	9	-1.41	0.28	0.61	0.11
6	Panicky	12	2.73	0.48	11	-2.83	0.44	1.03	0.13
7	Shaky	14	2.62	0.48	13	-3.20	0.48	1.22	0.15
8	Tense	16	2.06	0.31	15	-0.62	0.22	0.30	0.11
9	Terrified	18	4.28	1.38	17	-8.07	2.14	1.88	0.21
10	Upset	20	2.01	0.35	19	-2.29	0.32	1.14	0.15

The table of contents lists (blue) hyperlinks that can be used to navigate the output. To return to the table of contents from any part of the output file, click on the “(Back to TOC)” hyperlink that appears at the right of the heading for most output tables.

The table of parameter estimates is shown at the top of the following page. Note that the table caption indicates that the logit is $a\theta + c$ or $a(\theta - b)$; that means that the 2PL model trace line is expressed as

$$T = \frac{1}{1+\exp[-(a\theta+c)]} = \frac{1}{1+\exp[-a(\theta-b)]} \;,$$

in which the first form is called “slope-intercept” with parameters *a* (the slope, or discrimination) and *c* (the intercept). That is the form in which the model parameters are estimated. The values of the derived parameter *b* (the threshold) are also printed in the table.

Also note especially that there is no “1.7” or “*D*” anywhere in the model. IRTPRO parameter estimates *for all models* are *always* in the “logistic metric” (in Bilog terminology).[2]

2PL Model Item Parameter Estimates for Group 1, logit: $a\theta + c$ or $a(\theta - b)$ (Back to TOC)

Item	Label		*a*	*s.e.*		*c*	*s.e.*	*b*	*s.e.*
1	Afraid	2	3.17	0.63	1	-3.98	0.69	1.25	0.14
2	Desperate	4	3.81	0.76	3	-4.43	0.78	1.16	0.13
3	Fearful	6	5.80	1.46	5	-6.40	1.59	1.10	0.12
4	Frightened	8	9.36	3.15	7	-11.35	4.14	1.21	0.11
5	Nervous	10	2.31	0.36	9	-1.41	0.28	0.61	0.11
6	Panicky	12	2.73	0.48	11	-2.83	0.44	1.03	0.13
7	Shaky	14	2.62	0.48	13	-3.20	0.48	1.22	0.15
8	Tense	16	2.06	0.31	15	-0.62	0.22	0.30	0.11
9	Terrified	18	4.28	1.38	17	-8.07	2.14	1.88	0.21
10	Upset	20	2.01	0.35	19	-2.29	0.32	1.14	0.15
11	Worrying	22	2.08	0.32	21	0.05	0.21	-0.02	0.10
12	Calm	24	1.78	0.28	23	-1.32	0.23	0.74	0.13
13	Cheerful	26	1.05	0.19	25	-0.75	0.16	0.71	0.17
14	Contended	28	1.84	0.28	27	-0.97	0.22	0.53	0.12
15	Happy	30	1.70	0.27	29	-1.19	0.22	0.70	0.13
16	Joyful	32	1.20	0.20	31	0.65	0.17	-0.54	0.15
17	Loving	34	0.69	0.16	33	-0.52	0.14	0.75	0.24
18	Pleasant	36	1.67	0.31	35	-2.27	0.29	1.36	0.19
19	Secure	38	1.99	0.30	37	-1.03	0.23	0.52	0.11
20	Steady	40	2.16	0.35	39	-1.77	0.29	0.82	0.13
21	Thoughtful	42	1.02	0.22	41	-1.67	0.20	1.63	0.30

When feasible, IRTPRO computes the value of an updated version of the M_2 statistic proposed by Maydeu-Olivares and Joe (2005, 2006). That statistic is based on the one- and two-way marginal tables of the complete cross-classification of the respondents based on their response patterns. In this case, the value of that statistic indicates that the unidimensional model does not fit these data very well:

Statistics based on one- and two-way marginal tables			
M_2	Degrees of freedom	Probability	RMSEA
1208.12	189	0.0001	0.14
Note: M_2 is based on full marginal tables.			
Note: Model-based weight matrix is used.			

IRTPRO also computes (approximately) standardized LD X^2 statistics based on the local

[2] To be comparable to normal ogive discrimination parameters, the IRTPRO estimates of the *a* parameters could be divided by 1.7.

dependence statistic proposed by Chen and Thissen (1997). These begin as basically (approximately) χ^2-distributed statistics comparing the observed and expected frequencies in the two-way marginal tables for each item pair. To make the values roughly comparable among items that might have different numbers of response categories, (approximately) *z*-scores are computed by subtracting the degrees of freedom from those (approximately) χ^2-distributed statistics, and dividing by the square root of twice the degrees of freedom.

In this example, those statistics yield a clear suggestion of multidimensionality: In the table below, note the cluster of values that are printed in red for items 12-20 (the anxiety-minus words). The values are printed in red if the observed covariation between responses to a pair of items exceeds that predicted by the model, and in blue if the observed covariation is less than fitted. Thus, a cluster of red values indicates a cluster of items that may measure an un-modeled dimension. Because these are approximately standardized statistics, values that exceed 10.0 are also very, very large and unexpected.

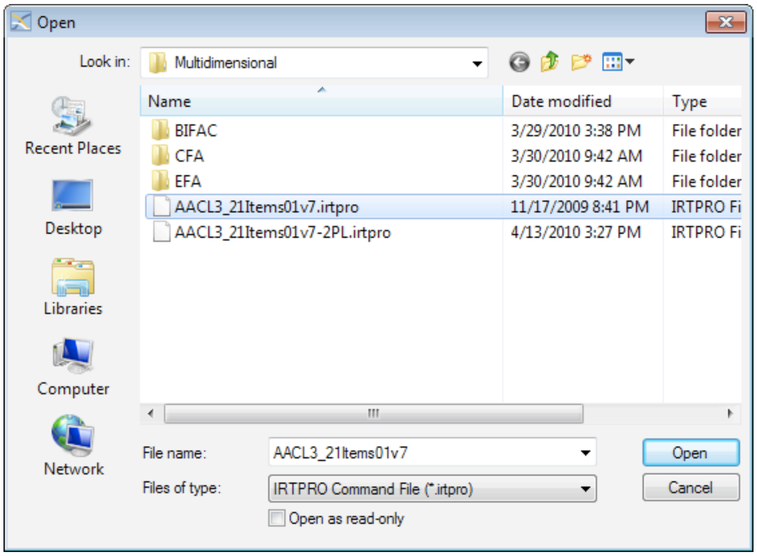
Marginal fit (X^2) and Standardized LD X^2 Statistics for Group 1 **(Back to TOC)**

	Marginal										
Item	X^2	1	2	3	4	5	6	7	8	9	10
1	0.0										
2	0.0	1.1									
3	0.0	0.6	0.9								
4	0.3	-0.4	3.0	0.6							
5	0.2	2.3	-0.1	-0.3	0.2						
6	0.0	-0.5	-0.6	2.6	5.3	1.0					
7	0.0	0.5	2.6	1.3	-0.2	-0.4	1.0				
8	0.2	-0.5	0.4	-0.5	0.0	6.0	1.8	-0.5			
9	0.8	0.0	-0.1	----	-0.1	----	0.7	1.0	----		
10	0.0	-0.1	-0.1	0.2	0.2	-0.2	-0.7	-0.2	-0.6	0.2	
11	0.1	-0.2	1.8	-0.5	0.5	2.1	4.1	-0.4	11.0	----	2.2
12	0.1	2.3	-0.5	1.3	0.6	-0.3	-0.1	-0.4	0.0	0.4	-0.1
13	0.0	6.3	-0.3	6.4	4.4	4.4	0.3	1.6	1.3	1.3	1.1
14	0.1	4.8	0.9	5.4	6.8	4.0	2.6	2.5	0.5	7.4	2.8
15	0.1	5.2	2.0	6.9	4.1	5.2	3.9	3.6	3.9	1.8	1.4
16	0.0	0.9	0.7	1.6	1.9	2.7	1.6	1.0	3.7	----	-0.3
17	0.0	9.1	1.9	6.4	6.5	5.4	0.7	0.7	4.1	1.4	0.5
18	0.0	11.0	-0.6	3.9	5.1	4.3	-0.3	0.1	0.8	0.5	2.3
19	0.1	4.5	2.9	2.4	2.6	3.2	2.5	0.5	1.7	----	0.9
20	0.1	1.4	1.1	1.7	1.2	5.7	1.2	-0.1	1.1	0.7	2.0
21	0.0	3.1	1.5	0.5	0.4	3.4	1.0	-0.0	3.9	0.3	2.4

	Marginal										
Item	X^2	11	12	13	14	15	16	17	18	19	20
11	0.1										
12	0.1	-0.5									
13	0.0	2.6	-0.6								
14	0.1	0.0	5.3	2.8							
15	0.1	2.0	-0.6	30.3	11.4						
16	0.0	5.6	-0.6	19.9	8.3	16.7					
17	0.0	5.0	-0.6	14.6	2.0	10.4	27.0				
18	0.0	2.0	-0.3	12.3	0.5	1.0	3.9	7.5			
19	0.1	0.1	-0.5	-0.6	3.6	-0.0	-0.2	1.8	0.8		
20	0.1	0.0	-0.4	-0.5	0.0	-0.6	-0.5	-0.6	0.6	9.9	
21	0.0	2.4	0.3	4.3	-0.5	1.2	1.4	12.1	4.1	1.5	2.2

Example 2—Multidimensional Analysis of the Affect Adjective Check List (AACL)

To obtain a better fit, we consider a two-dimensional model that fits one factor (latent variable) for the “anxiety-plus” items and a second (correlated) factor for the “anxiety-minus” items. This model is the one that was already set up in the AACL_21Items01V7.irtpro command file. To use this command file, one closes the output and data windows in IRTPRO and uses the “Open” command under the “File” menu to open it:



Look in: Multidimensional

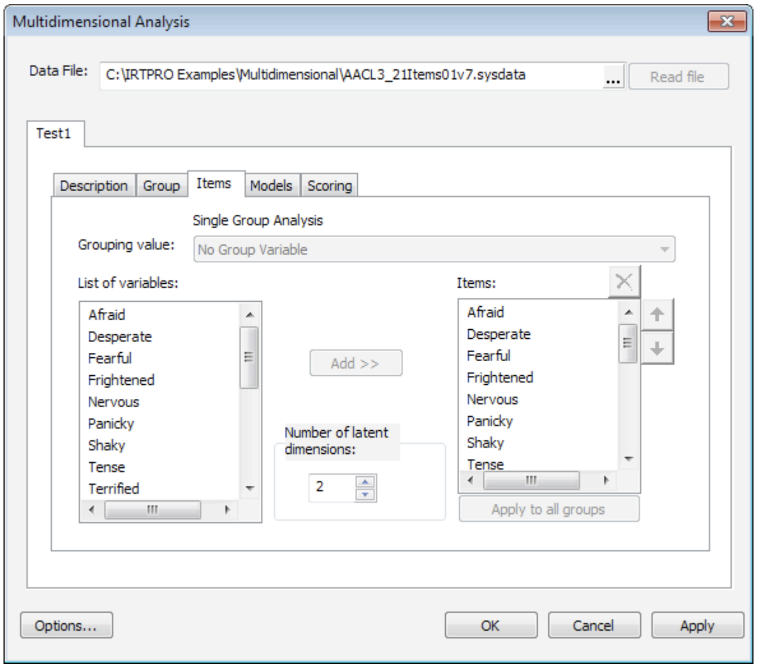
Name	Date modified	Type
BIFAC	3/29/2010 3:38 PM	File folder
CFA	3/30/2010 9:42 AM	File folder
EFA	3/30/2010 9:42 AM	File folder
AACL3_21Items01v7.irtpro	11/17/2009 8:41 PM	IRTPRO Fi
AACL3_21Items01v7-2PL.irtpro	4/13/2010 3:27 PM	IRTPRO Fi

File name: AACL3_21Items01v7

Files of type: IRTPRO Command File (*.irtpro)

□ Open as read-only

To see how the model has been set up, select “Multidimensional IRT…” under the “Analysis” menu. The “Items” tab shows that all of the items have been selected, and that the number of latent dimensions is 2:



Data File: C:\IRTPRO Examples\Multidimensional\AACL3_21Items01v7.sysdata

Single Group Analysis

Grouping value: No Group Variable

List of variables:

Afraid
Desperate
Fearful
Frightened
Nervous
Panicky
Shaky
Tense
Terrified

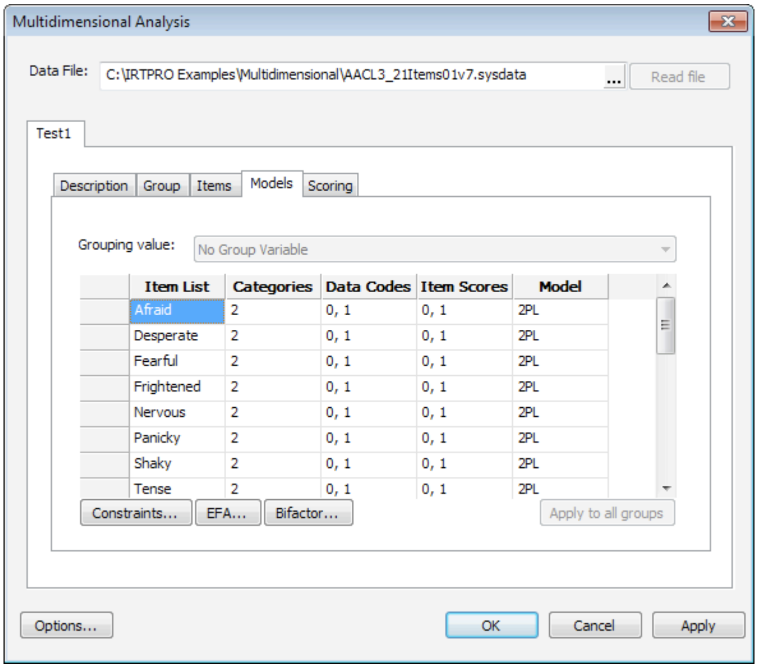
Items:

Afraid
Desperate
Fearful
Frightened
Nervous
Panicky
Shaky
Tense

Number of latent dimensions: 2

The particular two-dimensional model to be fitted to these data is a “simple structure” confirmatory factor analysis (CFA) model, which has non-zero slopes (or factor loadings) for the first eleven “anxiety-plus” adjectives on the first factor, non-zero slopes (or factor loadings) for the final ten “anxiety-minus” adjectives on the second factor, and zero slopes (or loadings) for the other combinations. In addition, the correlation between the two latent variables (for the “anxiety-plus” adjectives and the “anxiety-minus” adjectives) is estimated.

To see these constraints in the IRTPRO graphical user interface (GUI), click on the “Models” tab,

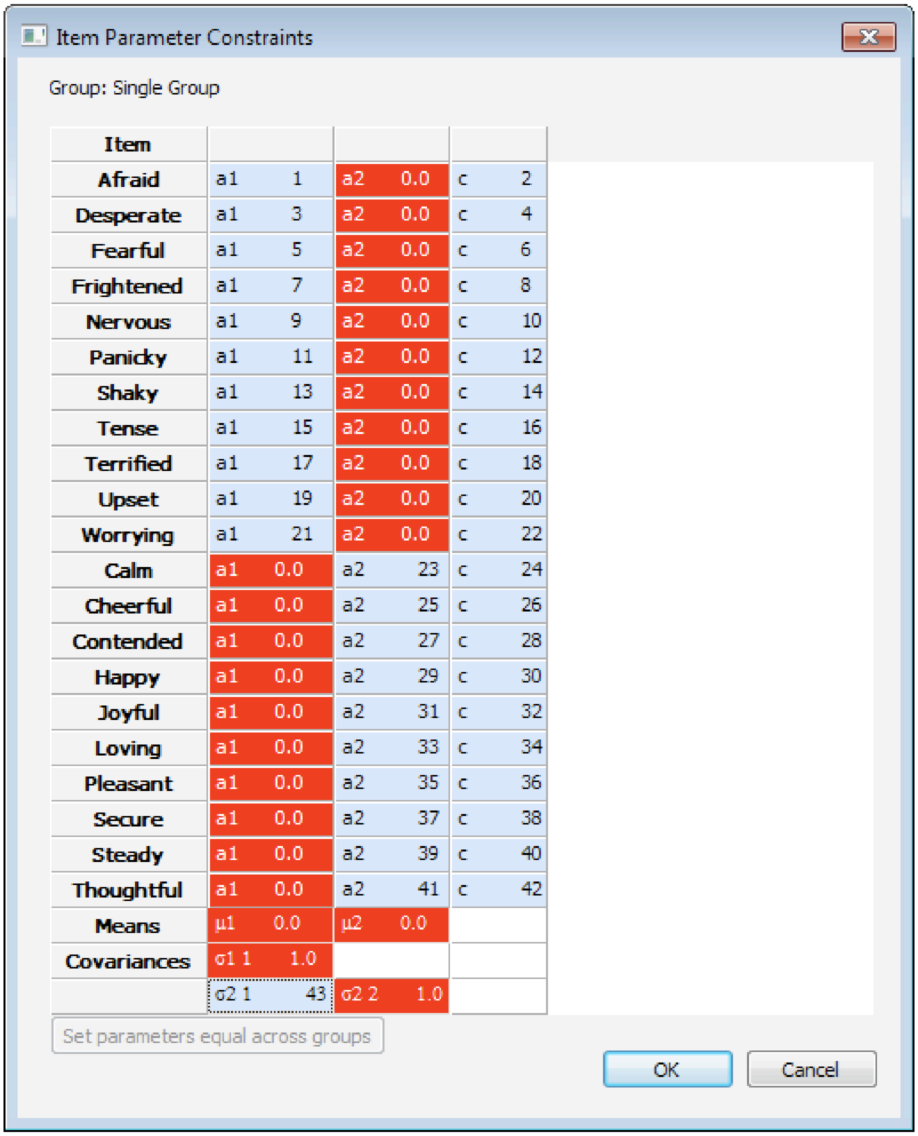


Data File: C:\IRTPRO Examples\Multidimensional\AACL3_21Items01v7.sysdata

Grouping value: No Group Variable

Item List	Categories	Data Codes	Item Scores	Model
Afraid	2	0, 1	0, 1	2PL
Desperate	2	0, 1	0, 1	2PL
Fearful	2	0, 1	0, 1	2PL
Frightened	2	0, 1	0, 1	2PL
Nervous	2	0, 1	0, 1	2PL
Panicky	2	0, 1	0, 1	2PL
Shaky	2	0, 1	0, 1	2PL
Tense	2	0, 1	0, 1	2PL

and then within that tab, on the “Constraints” button beneath the item list, and that shows the “Item Parameter Constraints” window,


Item Parameter Constraints
Set parameters equal across groups
OK
Cancel

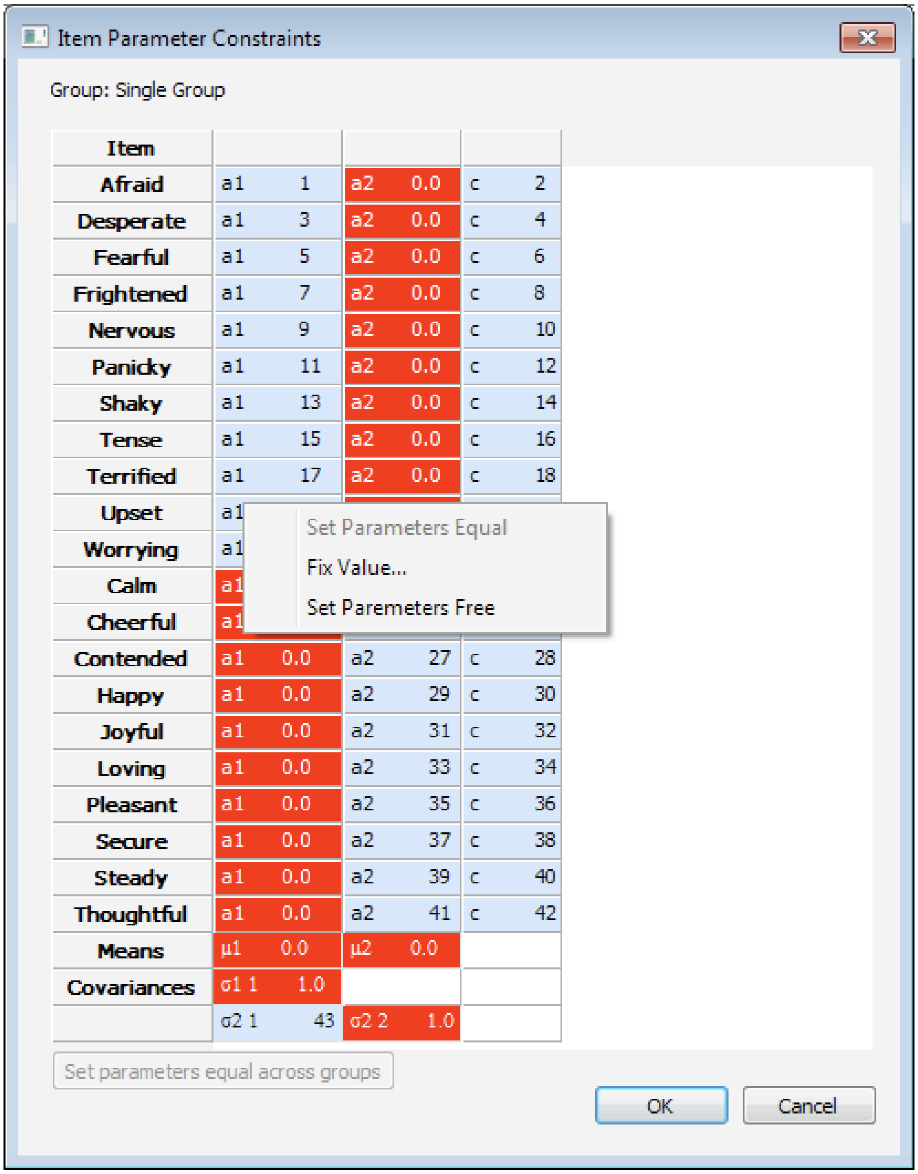
Group: Single Group

Item			
Afraid	a1 1	a2 0.0	c 2
Desperate	a1 3	a2 0.0	c 4
Fearful	a1 5	a2 0.0	c 6
Frightened	a1 7	a2 0.0	c 8
Nervous	a1 9	a2 0.0	c 10
Panicky	a1 11	a2 0.0	c 12
Shaky	a1 13	a2 0.0	c 14
Tense	a1 15	a2 0.0	c 16
Terrified	a1 17	a2 0.0	c 18
Upset	a1 19	a2 0.0	c 20
Worrying	a1 21	a2 0.0	c 22
Calm	a1 0.0	a2 23	c 24
Cheerful	a1 0.0	a2 25	c 26
Contended	a1 0.0	a2 27	c 28
Happy	a1 0.0	a2 29	c 30
Joyful	a1 0.0	a2 31	c 32
Loving	a1 0.0	a2 33	c 34
Pleasant	a1 0.0	a2 35	c 36
Secure	a1 0.0	a2 37	c 38
Steady	a1 0.0	a2 39	c 40
Thoughtful	a1 0.0	a2 41	c 42
Means	µ1 0.0	µ2 0.0	
Covariances	σ1 1 1.0		
	σ2 1 43	σ2 2 1.0	

The “Item Parameter Constraints” window lists the items in the leftmost column, and then, for each item, the model parameters are indicated symbolically (*a* for slopes, and *c* for intercepts). Integers in blue cells with the parameter symbol indicate the numbers for parameters that will be estimated. Real values in the red cells indicate fixed parameter values.

The elements of the mean vector and covariance matrix of the latent variables are also model parameters; they are shown at the bottom of the “Item Parameter Constraints” window. In this example, the means and variances are fixed (at 0.0 and 1.0, respectively) to standardize the two latent variables. The covariance between those two standardized variables ($\sigma_{2\,1}$) is estimated—that is the correlation between the two latent variables.

This model setup was already specified in the AACL_21Items01V7.irtpro file. However, had it not been, pop-up menus for cell in the “Item Parameter Constraints” window may be used to “Fix” or “Free” item parameters as desired to specify any particular model:



Item			
Afraid	a1 1	a2 0.0	c 2
Desperate	a1 3	a2 0.0	c 4
Fearful	a1 5	a2 0.0	c 6
Frightened	a1 7	a2 0.0	c 8
Nervous	a1 9	a2 0.0	c 10
Panicky	a1 11	a2 0.0	c 12
Shaky	a1 13	a2 0.0	c 14
Tense	a1 15	a2 0.0	c 16
Terrified	a1 17	a2 0.0	c 18
Upset	a1		
Worrying	a1		
Calm	a1		
Cheerful	a1		
Contended	a1 0.0	a2 27	c 28
Happy	a1 0.0	a2 29	c 30
Joyful	a1 0.0	a2 31	c 32
Loving	a1 0.0	a2 33	c 34
Pleasant	a1 0.0	a2 35	c 36
Secure	a1 0.0	a2 37	c 38
Steady	a1 0.0	a2 39	c 40
Thoughtful	a1 0.0	a2 41	c 42
Means	µ1 0.0	µ2 0.0	
Covariances	σ1 1 1.0		
	σ2 1 43	σ2 2 1.0	

For convenience, blocks of item parameters may be selected using standard conventions, such as shift-clicking, so that constraints may be applied to several parameters at a time.

When the parameters are as desired, one clicks “OK” in the “Item Parameter Constraints” window, then “OK” in the “Multidimensional Analysis” dialog box, then one selects “Run—Test1” from the “Analysis” menu.

When the parameters have been estimated, the output appears. Among the item parameter estimates, there are two columns of slopes, labeled a_1 and a_2—those are slopes on the two latent dimensions. There are *c* (intercept) parameters, but no *b* (threshold) parameters, because the latter do not have meaning for multidimensional models.

2PL Model Item Parameter Estimates for Group 1, logit: aθ + *c* (Back to TOC)

Item	Label		*a*1	*s.e.*		*a*2	*s.e.*		*c*	*s.e.*
1	Afraid	2	3.47	0.69		0.00	-----	1	-4.33	0.75
2	Desperate	4	3.28	0.62		0.00	-----	3	-3.97	0.65
3	Fearful	6	5.08	1.21		0.00	-----	5	-5.76	1.26
4	Frightened	8	8.16	3.02		0.00	-----	7	-10.35	3.63
5	Nervous	10	3.14	0.52		0.00	-----	9	-1.69	0.35
6	Panicky	12	2.68	0.45		0.00	-----	11	-2.76	0.41
7	Shaky	14	2.27	0.40		0.00	-----	13	-2.92	0.40
8	Tense	16	2.93	0.46		0.00	-----	15	-0.71	0.27
9	Terrified	18	4.88	1.82		0.00	-----	17	-9.49	3.14
10	Upset	20	2.06	0.34		0.00	-----	19	-2.31	0.32
11	Worrying	22	3.71	0.76		0.00	-----	21	0.16	0.31
12	Calm		0.00	-----	24	1.51	0.25	23	-1.17	0.20
13	Cheerful		0.00	-----	26	2.16	0.35	25	-1.04	0.24
14	Contended		0.00	-----	28	2.94	0.49	27	-1.26	0.30
15	Happy		0.00	-----	30	3.60	0.70	29	-1.93	0.43
16	Joyful		0.00	-----	32	2.68	0.48	31	1.01	0.27
17	Loving		0.00	-----	34	1.50	0.24	33	-0.66	0.18
18	Pleasant		0.00	-----	36	2.66	0.49	35	-2.99	0.46
19	Secure		0.00	-----	38	2.13	0.34	37	-1.00	0.23
20	Steady		0.00	-----	40	2.01	0.34	39	-1.63	0.26
21	Thoughtful		0.00	-----	42	1.48	0.27	41	-1.90	0.24

To see the estimated correlation between the two latent variables, click on the entry "Group Latent Variable Means" in the table of contents; below the means IRTPRO lists the latent variable variance-covariance matrix:

Latent Variable Variance-Covariance Matrix for Group 1, (Back)

	θ_1	*s.e.*	θ_2	*s.e.*
	1.00	-----		
43	0.55	0.06	1.00	-----

The correlation between the latent variables that accounts for the covariation among the "anxiety-plus" and "anxiety-minus" adjectives is only 0.55. That value would have needed to be 1.0 for a unidimensional model to fit, which explains why the unidimensional model did not appear to fit well.

The M_2 statistic for this model suggests much better fit:

Statistics based on one- and two-way marginal tables			
M_2	Degrees of freedom	Probability	RMSEA
463.91	188	0.0001	0.07
Note: M_2 is based on full marginal tables.			
Note: Model-based weight matrix is used.			

The difference between the values of –2loglikelihood for the undimensional model (5058.51) and for this two-dimensional model (4748.47) may be interpreted as a χ^2-distributed statistic on 1 degree of freedom (because the undimensional model is nested within this two-dimensional model, and the two-dimensional model uses one more fitted parameter). That difference is 310.1, which is highly significant. So there is strong evidence that these data need a two-dimensional model.

With the two-dimensional model, the standardized LD X^2 statistics no longer suggest strong residual covariance. There are few extremely large values and no obvious red clusters:

Marginal fit (X^2) and Standardized LD X^2 Statistics for Group 1 (Back to TOC)

	Marginal										
Item	X^2	1	2	3	4	5	6	7	8	9	10
1	0.0										
2	0.0	0.9									
3	0.0	0.4	-0.3								
4	0.0	-0.7	1.1	1.1							
5	0.0	0.2	-0.7	-0.6	-0.6						
6	0.0	-0.5	-0.7	1.8	5.0	-0.3					
7	0.0	0.1	0.6	-0.1	-0.7	-0.7	0.1				
8	0.0	0.2	-0.5	1.6	----	0.0	-0.0	-0.7			
9	0.1	0.0	-0.6	----	----	----	0.0	0.7	----		
10	0.0	0.3	0.1	0.2	0.1	-0.7	-0.7	0.1	-0.3	-0.5	
11	0.0	-0.2	0.4	----	----	-0.2	1.2	-0.6	0.4	----	-0.1
12	0.0	0.3	9.6	2.8	3.3	3.7	11.4	4.2	9.1	0.1	2.0
13	0.0	3.0	-0.6	2.1	1.1	2.2	-0.7	-0.3	0.2	0.1	-0.3
14	0.0	-0.7	0.6	-0.6	-0.7	-0.7	-0.5	-0.5	-0.2	0.6	-0.7
15	0.0	-0.2	-0.5	-0.5	-0.7	0.1	-0.6	-0.7	-0.1	-0.5	-0.7
16	0.0	-0.4	-0.7	-0.5	-0.2	0.6	-0.4	-0.6	1.5	----	-0.7
17	0.0	8.5	1.1	5.2	5.7	5.5	0.2	0.1	4.5	1.0	0.2
18	0.0	2.5	2.3	-0.7	-0.6	-0.1	0.7	0.0	-0.7	-0.6	-0.6
19	0.0	-0.5	0.7	1.6	1.0	-0.2	0.5	1.8	0.1	-0.6	0.5
20	0.0	1.4	3.1	3.4	3.6	-0.6	2.2	4.0	0.8	0.1	0.2
21	0.0	0.3	-0.6	-0.7	-0.7	0.6	-0.6	-0.6	1.1	-0.6	0.0

	Marginal										
Item	X^2	11	12	13	14	15	16	17	18	19	20
11	0.0										
12	0.0	5.5									
13	0.0	1.8	0.5								
14	0.0	-0.1	3.1	1.2							
15	0.0	-0.4	-0.1	3.9	-0.6						
16	0.0	3.7	0.5	1.9	-0.7	1.3					
17	0.0	6.3	2.6	0.1	1.1	-0.7	5.7				
18	0.0	-0.4	-0.7	0.6	0.8	1.8	-0.6	-0.5			
19	0.0	1.1	-0.7	5.9	-0.2	1.0	2.3	-0.2	-0.7		
20	0.0	1.5	0.4	4.5	-0.7	2.6	1.6	3.8	-0.7	9.8	
21	0.0	0.6	1.8	-0.7	1.4	-0.2	-0.5	2.0	-0.4	-0.6	-0.0

One value of the pairwise *LD* statistics that stands out is the 11.4 for items 6 and 12; those adjectives are "panicky" and "calm" (the latter reverse scored). It is likely that there is additional un-modeled local dependence between those two near-antonyms.

Example 3—Unidimensional Analysis of the State-Trait Anxiety Inventory (STAI)

This example examines item responses obtained from 517 undergraduate students at the University of Houston and the University of Arkansas who completed a 20-item anxiety questionnaire derived from the State-Trait Anxiety Inventory (STAI, Spielberger, 1983).[3]

For illustration of fitting the graded response model, six items are selected:

1. I feel calm.
2. I am tense.
3. I am regretful.
4. I feel at ease.
5. I feel anxious.
6. I feel nervous.

In these data, the responses were on a five-point unipolar Likert-type response scale: 1 = not at all, 2 = very little, 3 = somewhat, 4 = moderately, and 5 = very much.

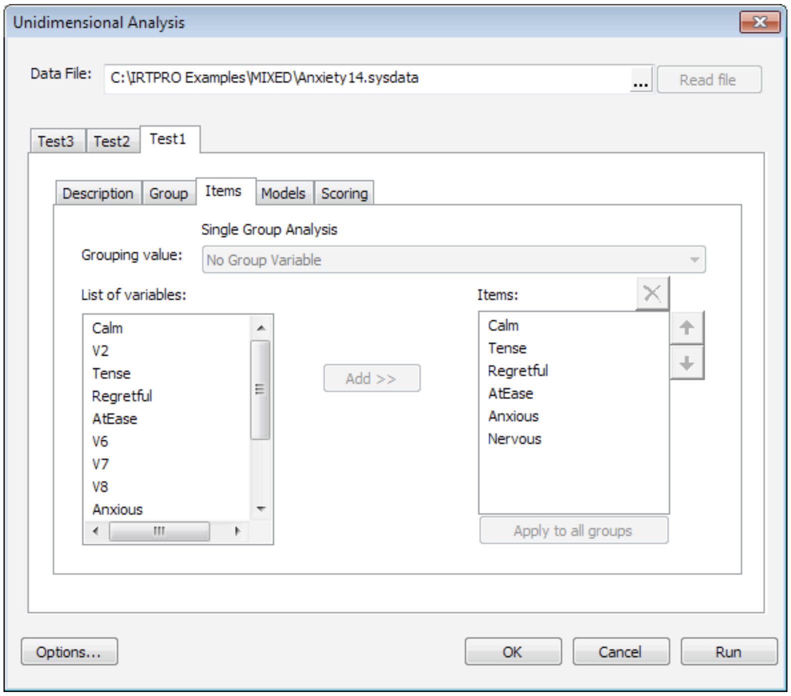
To see the data, the procedures described above are repeated to use the "Open" file dialog under the "File" menu of IRTPRO, navigate to the "C:\IRTPRO Examples\MIXED" folder, select "Files of type:" "System Data File (*.sysdata)" in the open file dialog, and open the file "Anxiety14.sysdata".

While this file contains responses to fourteen items, only the six items listed above have meaningful variable names ("Calm", "Tense", and so on). The other variables are named V2, V6, V7, etc., and will not be used here.

To initiate the unidimensional IRT analysis, select "Unidimensional IRT …" under the "Analysis" menu. Because there is already a saved command file called "Anxiety14.irtpro" in that folder, IRTPRO asks if it should use the saved command file; respond "Yes" to use the commands in that file.

The "Items" tab in the "Test1" tab in the "Unidimensional Analysis" dialog shows that only the six items to be analyzed have been selected, as shown at the top of the following page:

[3] Thanks to Lynne Steinberg for these data, which are described more completely by Thissen & Steinberg (2009).



Unidimensional Analysis

Data File: C:\IRTPRO Examples\MIXED\Anxiety14.sysdata

Single Group Analysis

Grouping value: No Group Variable

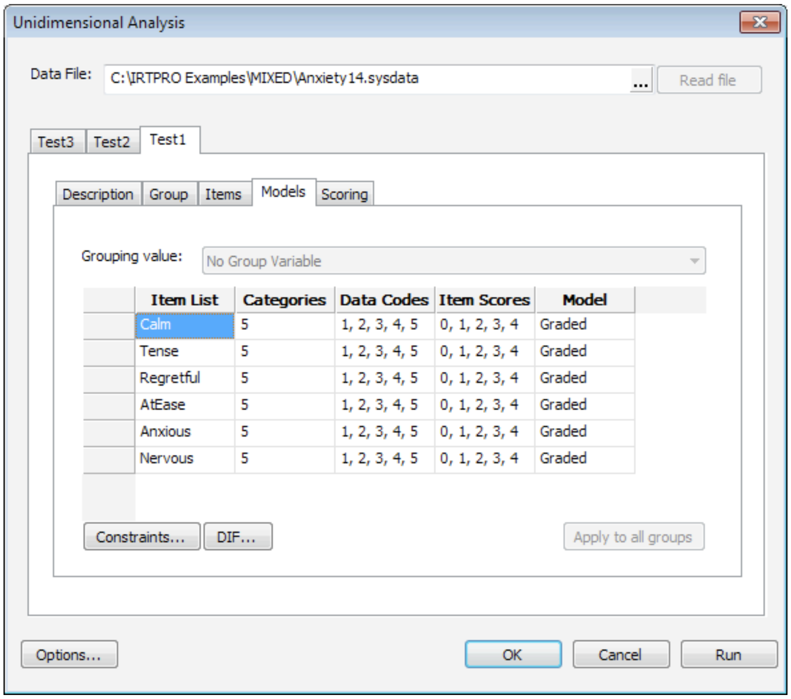
List of variables:

Calm
V2
Tense
Regretful
AtEase
V6
V7
V8
Anxious

Items:

Calm
Tense
Regretful
AtEase
Anxious
Nervous

Selection of the “Models” tab in that dialog shows that each item has five response categories; the data codes 1, 2, 3, 4, 5 have been automatically given item scores (model category values) 0, 1, 2, 3, 4, and Samejima’s (1969, 1997) graded model has been selected:



Unidimensional Analysis

Data File: C:\IRTPRO Examples\MIXED\Anxiety14.sysdata

Grouping value: No Group Variable

Item List	Categories	Data Codes	Item Scores	Model
Calm	5	1, 2, 3, 4, 5	0, 1, 2, 3, 4	Graded
Tense	5	1, 2, 3, 4, 5	0, 1, 2, 3, 4	Graded
Regretful	5	1, 2, 3, 4, 5	0, 1, 2, 3, 4	Graded
AtEase	5	1, 2, 3, 4, 5	0, 1, 2, 3, 4	Graded
Anxious	5	1, 2, 3, 4, 5	0, 1, 2, 3, 4	Graded
Nervous	5	1, 2, 3, 4, 5	0, 1, 2, 3, 4	Graded

If one clicks the “Run” button in the lower right hand corner of the dialog, the output appears.

The first table of “Graded Model Item Parameter Estimates” lists the slopes (*a*) and intercepts (*c*); the second table lists the slopes (*a*) and thresholds (*b*):[4]

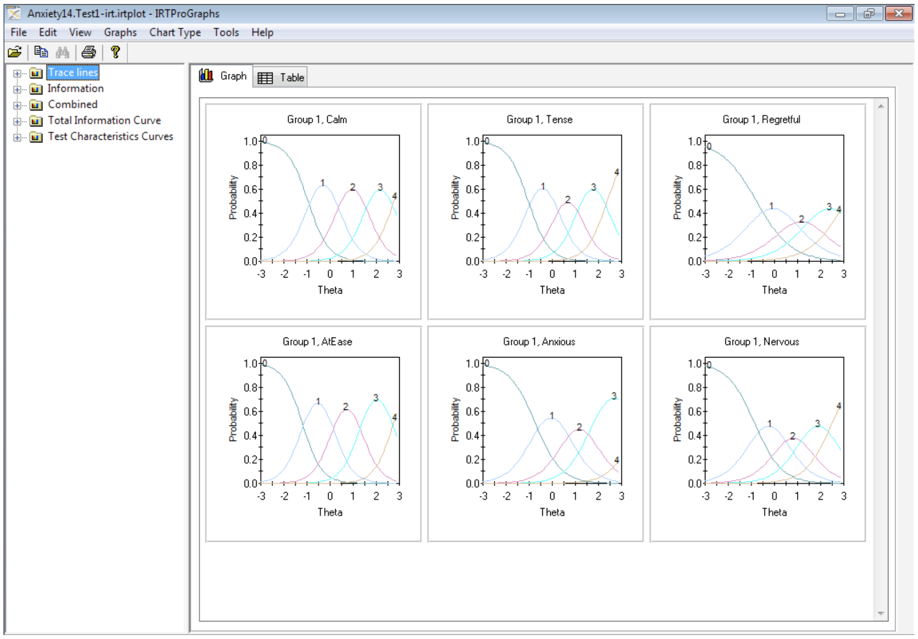
Graded Model Item Parameter Estimates, logit: *aθ* + *c*

Item	Label		*a*	*s.e.*		*c*1	*s.e.*		*c*2	*s.e.*		*c*3	*s.e.*		*c*4	*s.e.*
1	Calm	5	2.29	0.21	1	2.17	0.20	2	-0.80	0.16	3	-3.57	0.28	4	-6.31	0.51
2	Tense	10	2.26	0.19	6	2.34	0.20	7	-0.46	0.16	8	-2.60	0.22	9	-5.37	0.39
3	Regretful	15	1.33	0.13	11	1.03	0.13	12	-0.85	0.12	13	-2.22	0.16	14	-4.08	0.28
4	AtEase	20	2.42	0.22	16	2.89	0.24	17	-0.32	0.16	18	-3.19	0.26	19	-6.63	0.55
5	Anxious	25	1.80	0.16	21	1.31	0.15	22	-1.13	0.15	23	-3.07	0.22	24	-6.60	0.62
6	Nervous	30	1.71	0.15	26	1.46	0.15	27	-0.61	0.13	28	-2.20	0.18	29	-4.28	0.29

Graded Model Item Parameter Estimates for Group 1, logit: *a*(*θ* - *b*) (Back to TOC)

Item	Label		*a*	*s.e.*	*b*1	*s.e.*	*b*2	*s.e.*	*b*3	*s.e.*	*b*4	*s.e.*
1	Calm	5	2.29	0.21	-0.95	0.09	0.35	0.07	1.56	0.11	2.76	0.22
2	Tense	10	2.26	0.19	-1.04	0.09	0.20	0.07	1.15	0.09	2.38	0.17
3	Regretful	15	1.33	0.13	-0.77	0.11	0.64	0.10	1.67	0.16	3.08	0.29
4	AtEase	20	2.42	0.22	-1.20	0.09	0.13	0.07	1.32	0.10	2.75	0.21
5	Anxious	25	1.80	0.16	-0.73	0.09	0.63	0.08	1.70	0.13	3.67	0.39
6	Nervous	30	1.71	0.15	-0.85	0.10	0.36	0.08	1.28	0.11	2.50	0.20

To see the graded model trace lines graphically, when the output file “Anxiety14.Test1-irt.htm” is in the IRTPRO viewer window, select “Graphs” under the “Analysis” menu, and a separate program “IRTPROgraphs” starts and shows various graphics that may be selected using a left side navigation bar:



[4] There are slight numerical differences (larger in the standard errors than in the parameter estimates) between the results obtained with IRTPRO and the Multilog (du Toit, 2003) estimates reported for these items by Thissen & Steinberg (2009). These differences are due to differences in numerical quadrature, and the facts that IRTPRO usually converges to more decimal places, and computes much more accurate standard errors, than did Multilog.

When finished with the graphics, the “IRTPROgraphs” program may be closed using the “X” in the upper right-hand corner of the window; control returns to the main IRTPRO window that has remained running behind the graphics application.

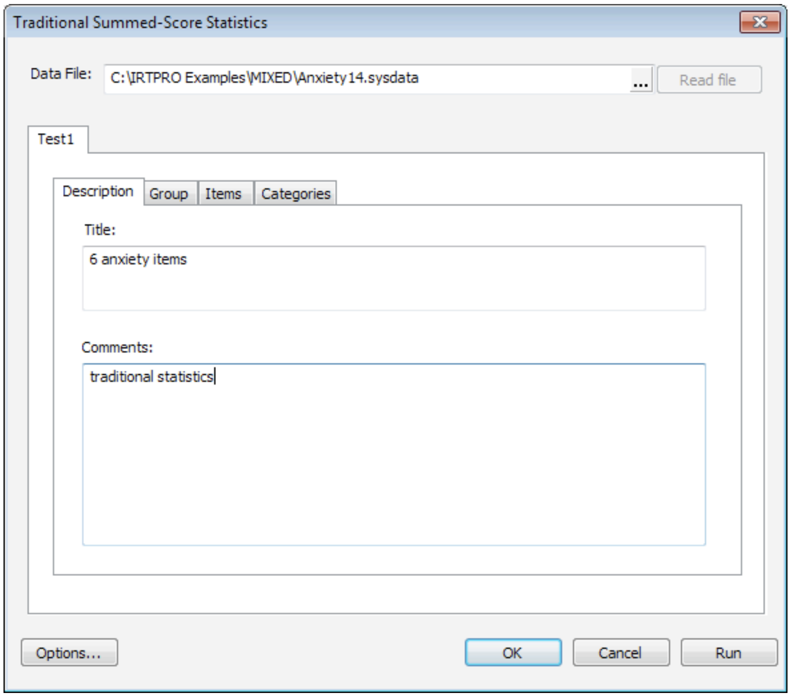
Before leaving these data, note as a first aside that the Anxiety14.irtpro file also includes set-ups to fit these same data with Muraki’s generalized partial credit model, and with Bock’s nominal model. To compute and view the output for those models, one can use the “Window” menu to return to the “Anxiety14.sysdata” spreadsheet window, and then the “Analysis” menu to select “Unidimensional IRT …”.

There, if one selects the “Test2” tab, and the “Models” tab within that, it shows that the “GP Credit” model has been selected for all items from the pull down menu that is obtained by right-clicking an item’s, or a selected group of items’, “Model” cells. If the “Run” button is clicked in that dialog, the results appear for that model. Similarly, the “Test3” tab yields results for the nominal model.

As a second aside, note that IRTPRO can also compute a set of traditional summed-score-based statistics that are useful in checking data before an IRT analysis and interpreting IRT results. To view these statistics for these data, close the output and data (spreadsheet) windows, re-open the data file, and then select “Traditional Summed-Score Statistics …” from the “Analysis” menu.

Because there is already a command file called “Anxiety14.irtpro” in that folder, IRTPRO asks if it should use the saved command file. Because a separate command file is desired to compute traditional statistics, the user responds “No”.

The “Traditional Statistics” dialog appears, and the user enters the title and any desired comments in the “Description” tab:

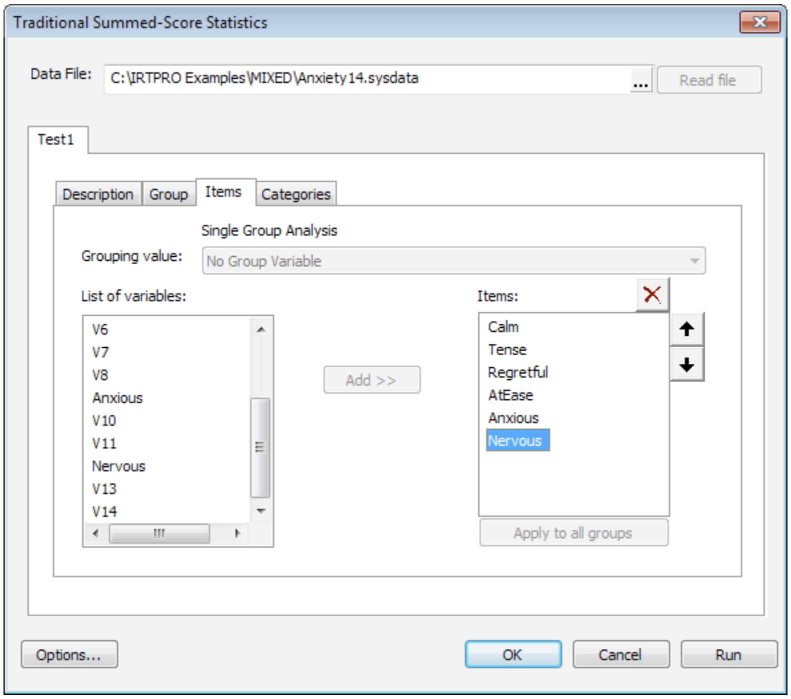


Data File: C:\IRTPRO Examples\MIXED\Anxiety14.sysdata

Title: 6 anxiety items

Comments: traditional statistics

Then the six items are selected:



Traditional Summed-Score Statistics

Data File: C:\IRTPRO Examples\MIXED\Anxiety14.sysdata

Single Group Analysis

Grouping value: No Group Variable

List of variables:
V6
V7
V8
Anxious
V10
V11
Nervous
V13
V14

Items:
Calm
Tense
Regretful
AtEase
Anxious
Nervous

When "Run" is clicked, the output appears, excerpts of which are on the following page.

Item and (Weighted) Summed-Score Statistics for Group 1 (Back to TOC)

Coefficient alpha: 0.8425
Complete data N: 515

The following Statistics are Computed only for the Listwise-Complete Data:

			With Item Deleted	
	Response		Item-Total	Coefficient
Item	Average	Std. Dev.	Correlation	α
1	1.287	0.966	0.6340	0.8148
2	1.443	1.067	0.6746	0.8059
3	1.212	1.114	0.5325	0.8354
4	1.452	0.960	0.6580	0.8105
5	1.146	0.980	0.6199	0.8172
6	1.357	1.147	0.6261	0.8165

Item	ItemLabl						(Back)
1	Category:	0	1	2	3	4	Missing
	Frequencies:	114	204	143	47	9	0
	For listwise-complete data:						
	Frequencies:	114	203	143	46	9	
	Average (wtd) Score:	3.30	6.85	10.26	14.83	16.78	
	Std. Dev. (wtd) Score:	2.61	2.99	3.35	3.33	5.12	

Item	ItemLabl						(Back)
2	Category:	0	1	2	3	4	Missing
	Frequencies:	104	188	131	77	17	0
	For listwise-complete data:						
	Frequencies:	104	188	130	77	16	
	Average (wtd) Score:	2.87	6.41	9.82	13.18	17.00	
	Std. Dev. (wtd) Score:	2.44	2.50	3.38	3.29	3.27	

Item	ItemLabl						(Back)
3	Category:	0	1	2	3	4	Missing
	Frequencies:	162	178	99	60	18	0
	For listwise-complete data:						
	Frequencies:	162	178	97	60	18	
	Average (wtd) Score:	4.20	7.66	9.70	13.15	16.22	
	Std. Dev. (wtd) Score:	3.08	3.19	3.98	3.41	3.86	

…

Item	ItemLabl						(Back)
5	Category:	0	1	2	3	4	Missing
	Frequencies:	153	196	111	54	3	0
	For listwise-complete data:						
	Frequencies:	152	196	110	54	3	
	Average (wtd) Score:	3.64	7.64	10.92	14.06	19.00	
	Std. Dev. (wtd) Score:	2.82	2.93	3.52	3.89	2.65	

Item	ItemLabl						(Back)
6	Category:	0	1	2	3	4	Missing
	Frequencies:	140	169	112	73	22	1
	For listwise-complete data:						
	Frequencies:	140	168	112	73	22	
	Average (wtd) Score:	3.46	6.97	10.10	12.77	15.86	
	Std. Dev. (wtd) Score:	2.47	2.79	3.54	3.25	4.14	

Example 4—Unidimensional Analysis of Four "Neurotic Symptoms" Items

In his contribution to the American Soldier series, Lazarsfeld (1950, pp. 441*ff*) presented data from four questions he called a "neurotic inventory;" the responses were from a sample of 1,000 soldiers. [Stouffer et al. (1950) did not really seem to know what to call these items; Lazarsfeld called them a "neurotic inventory" but the index to the volume lists the example under "psychosomatic symptoms."] The questions originally involved several response alternatives, but the data were dichotomized before they were tabulated. The four questions were:

NI1. Have you ever been bothered by pressure or pains in the head?

NI2. Have you ever been bothered by shortness of breath when you were not exercising or working hard?

NI3. Do your hands ever tremble enough to bother you?

NI4. Do you often have trouble in getting to sleep or staying asleep?

For items 1-3, the responses "Yes, often" or "Yes, sometimes" or No answer (omission) are coded 1; for item 4, only "Very often" or No answer (omission) are coded 1; other responses are coded 0.

To begin the analysis of these data, select "Open" under the "File" menu; then navigate to the folder that contains the data file "ArmyNeuro.sysdata". Select "Files of type:" "System Data File (*.sysdata)" in the open file dialog, and open "ArmyNeuro.sysdata".

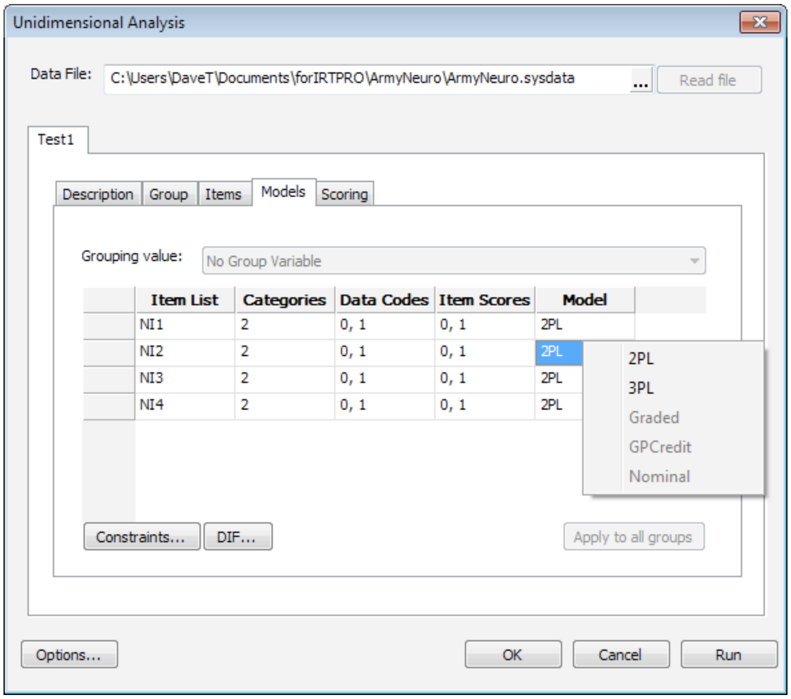
Then select "Unidimensional IRT …" under the "Analysis" menu, and enter a title and any desired comments in text boxes in the "Description" tab.

In the "Items" tab, add all four items to the test box.

In the "Models" tab, all four items list the 2PL model in the rightmost column, by default. There is evidence from previous analyses, not shown here, that the 2PL model fits these data poorly; that might be because for item 2, shortness of breath, there may be causes other than anxiety; for example, asthma or allergies. So some proportion of soldiers who exhibit no other "neurotic symptoms" may endorse the item about "shortness of breath."

The 3PL model includes a third parameter that fits just this kind of probability of endorsement, unrelated to the trait being measured.

To fit item 2 with the 3PL model, right-click the "Model" cell for item 2, and change the model in the pull-down menu from 2PL to 3PL as shown:



Data File: C:\Users\DaveT\Documents\forIRTPRO\ArmyNeuro\ArmyNeuro.sysdata

Grouping value: No Group Variable

Item List	Categories	Data Codes	Item Scores	Model
NI1	2	0, 1	0, 1	2PL
NI2	2	0, 1	0, 1	2PL
NI3	2	0, 1	0, 1	2PL
NI4	2	0, 1	0, 1	2PL

When the "Run" button us clicked, the output appears, with the items' parameters tabulated separately by item type. First, there is a table of the parameters for the three 2PL items:

2PL Model Item Parameter Estimates for Group 1, logit: *aθ* + *c* or *a*(*θ* – *b*) (Back to TOC)

Item	Label		*a*	s.e.		*c*	s.e.	*b*	s.e.
1	NI1	2	1.68	0.14	1	-2.64	0.16	1.57	0.09
3	NI3	7	1.90	0.12	6	-0.45	0.08	0.24	0.04
4	NI4	9	1.85	0.12	8	0.45	0.08	-0.24	0.04

3PL Model Item Parameter Estimates for Group 1, logit: *aθ* + *c* or *a*(*θ* – *b*) (Back to TOC)

Item	Label		*a*	s.e.		*c*	s.e.	*b*	s.e.		logit *g*	s.e.	*g*	s.e.
2	NI2	5	7.08	0.97	4	-5.90	0.86	0.83	0.03	3	-1.96	0.12	0.12	0.01

There is a separate table (above) for the 3PL fit to item 2. In addition to the *a*, *c*, and *b* parameters of the 2PL model, this table lists the logit of the *g* parameter of the 3PL model, which is estimated, and the *g* parameter itself, which is derived from the logit.

The 3PL model is expressed as

$$T = g + \frac{1-g}{1+\exp[-(a\theta+c)]} = g + \frac{1-g}{1+\exp[-a(\theta-b)]} \quad .$$

Especially note that there is no "1.7" or "*D*" anywhere in the model. IRTPRO parameter estimates *for all models* are *always* in the "logistic metric" (in Bilog terminology).

This model fits the data reasonably well, as shown by the goodness-of-fit statistics:

Statistics based on the full item x item x ... classification

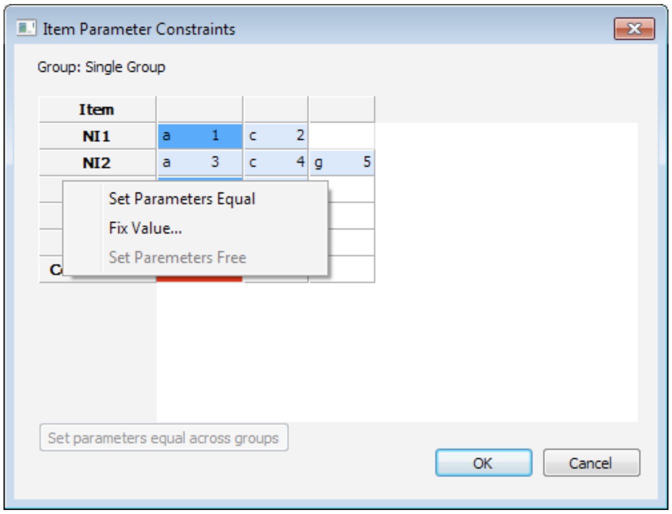
G^2	Degrees of freedom	Probability	RMSEA
12.51	6	0.0514	0.03

Statistics based on one- and two-way marginal tables

M_2	Degrees of freedom	Probability	RMSEA
9.40	1	0.0022	0.09

Note: M_2 is based on full marginal tables.
Note: Model-based weight matrix is used.

However, a better (apparent) fit can be obtained by setting the slope parameters for items 1, 3, and 4 to be equal. That can be done by creating an additional “Test2”, re-entering the title and comments in the “Description” tab, re-selecting all four items, and using the “Models” tab to fit item 2 with the 3PL item. Then use the “Constraints” window to set the slope parameters equal for items 1, 3, and 4, as shown below, where the cells for those *a* parameters have been control-clicked, and then right-clicked to bring up the pull-down menu that includes the option “Set Parameters Equal”:

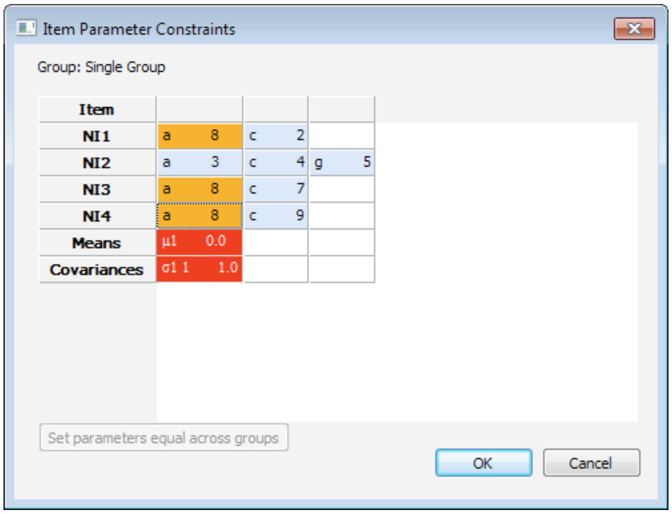


Item Parameter Constraints

Group: Single Group

Item			
NI1	a 1	c 2	
NI2	a 3	c 4	g 5

After those parameters are set equal, the “Item Parameter Constraints” window looks like this:



Item Parameter Constraints

Group: Single Group

Item			
NI1	a 8	c 2	
NI2	a 3	c 4	g 5
NI3	a 8	c 7	
NI4	a 8	c 9	
Means	µ1 0.0		
Covariances	σ1 1 1.0		

The revised item parameter estimates are as shown in the following two tables:

2PL Model Item Parameter Estimates for Group 1, logit: $a\theta + c$ or $a(\theta - b)$ (Back to TOC)

Item	Label		*a*	*s.e.*		*c*	*s.e.*	*b*	*s.e.*
1	NI1	7	1.81	0.06	1	-2.74	0.11	1.51	0.08
3	NI3	7	1.81	0.06	5	-0.44	0.08	0.24	0.04
4	NI4	7	1.81	0.06	6	0.44	0.08	-0.24	0.04

3PL Model Item Parameter Estimates for Group 1, logit: $a\theta + c$ or $a(\theta - b)$ (Back to TOC)

Item	Label		*a*	*s.e.*		*c*	*s.e.*	*b*	*s.e.*		logit *g*	*s.e.*	*g*	*s.e.*
2	NI2	4	7.07	0.96	3	-5.87	0.85	0.83	0.03	2	-1.98	0.12	0.12	0.01

The goodness-of-fit appears to be better (although this has been accomplished not so much by fitting the data better as by using fewer free parameters to accomplish the fit):

Statistics based on the full item x item x ... classification			
G^2	Degrees of freedom	Probability	RMSEA
13.08	8	0.1087	0.03

Statistics based on one- and two-way marginal tables			
M_2	Degrees of freedom	Probability	RMSEA
11.29	3	0.0103	0.05

Note: M_2 is based on full marginal tables.

Note: Model-based weight matrix is used.

The likelihood ratio test for equality of the three slopes can be computed in this case either as the difference between the values of G^2 for the constrained vs. the less constrained model (13.08 – 12.51 = 0.57) or as the difference between the values of –2loglikelihood reported for the two models (4336.80 – 4336.23 = 0.57). In either case, reference to the χ^2 distribution with 2 degrees of freedom indicates that the three slopes are not significantly different from each other. [While the values of G^2 are only available if statistics based on the full item x item x ... classification can be computed, the values of –2loglikelihood are always reported.]

The "Statistics based on the full item x item x ... classification" can rarely be computed in applications of IRT, because they require that the sample is sufficiently large to fill all of the cells of the cross-classification created by listing all response patterns. For four dichotomous items, that is a manageable 16 cells, and so IRTPRO tabulates the observed and expected frequencies (and some other values) for each response pattern as follows:

Response Pattern Observed and Expected Frequencies, Standardized Residuals, EAPs and SDs for Group 1 (Back to TOC)

Item:				Frequencies		Standard		
1	2	3	4	Observed	Expected	Residual	EAP[θ|u]	SD[θ|u]
0	0	0	0	285	279.58	0.38	-0.90	0.71
0	0	0	1	161	172.54	-0.97	-0.18	0.56
0	0	1	0	64	71.69	-0.94	-0.18	0.56
0	0	1	1	141	126.48	1.38	0.28	0.44
0	1	0	0	41	41.36	-0.06	-0.79	0.80
0	1	0	1	49	43.62	0.83	0.34	0.74
0	1	1	0	11	18.12	-1.69	0.34	0.74
0	1	1	1	110	108.56	0.15	1.07	0.54
1	0	0	0	9	7.18	0.68	-0.18	0.56
1	0	0	1	11	12.67	-0.47	0.28	0.44
1	0	1	0	8	5.26	1.20	0.28	0.44
1	0	1	1	14	17.92	-0.93	0.57	0.36
1	1	0	0	3	1.82	0.88	0.34	0.74
1	1	0	1	10	10.88	-0.27	1.07	0.54
1	1	1	0	8	4.52	1.64	1.07	0.54
1	1	1	1	75	77.80	-0.33	1.60	0.58

Note that none of the standardized residuals, $(o-e)/\sqrt{e}$, are unusually large when considered as approximately *z*-scores.

This tabulation also includes the values of EAP[θ] and the posterior standard deviation SD[θ] for each response pattern; however, those are more routinely computed in the "Scoring" section of IRTPRO, which is not described here.

The G^2 value (13.08, on 8 *d.f.*) reported above is obtained by using the likelihood-ratio chi-square to compare the observed and expected frequencies in that table. In this case, those values do not differ by more than is expected by sampling error.

For scales that involve more items, or more response categories, the complete cross-classification is often too large to "fill;" for example, for the six 5-alternative items considered in the previous example, the full cross-classification would have 5^6 = 15,625 cells, which can only be sparsely populated by a few hundred respondents. So in that case, as well as this one, an updated version of the M_2 statistic proposed by Maydeu-Olivares and Joe (2005, 2006) may be used as a proxy for G^2. The M_2 statistic is based on the one- and two-way marginal tables of the complete cross-classification; those subtables are easier to “fill” with reasonable sample sizes.

IRTPRO can also compute a trace line diagnostic statistic for each item, which is a generalization for polytomous responses of the $S\text{-}X^2$ item-fit statistic suggested by Orlando and Thissen (2000, 2003). For the 2PL fit to these four items, these statistics are tabulated as follows:

$S\text{-}X^2$ Item Level Diagnostic Statistics

Item	Label	X^2	*d.f.*	Probability
1	NI1	1.45	2	0.4859
2	NI2	1.10	1	0.2947
3	NI3	2.32	2	0.3151
4	NI4	5.40	2	0.0672

The summary table (above) indicates that for all four of these items, the trace lines have been fitted sufficiently well that the model-expected proportions responding 0 and 1 match the observed data well.

For each item, the complete tables printed below have one row for each summed score for the “other items” (for each of these four items, the summed score on the “other items” ranges 0-3), and within those scores the observed and expected frequencies are tabulated. The entries are printed in blue if the observed frequency exceeds the expected frequency, and in red if too few responses are observed.

Item 1 $S\text{-}X^2(2) = 1.4$, p = 0.4859 (Back)

	Category 0		Category 1	
Score	Observed	Expected	Observed	Expected
0	285	286.6	9	7.4
1	266	269.4	22	18.6
2	201	198.0	32	35.0
3	110	107.8	75	77.2

Item 2 $S\text{-}X^2(1) = 1.1$, p = 0.2947 (Back)

	Category 0		Category 1	
Score	Observed	Expected	Observed	Expected
0	285	284.0	41	42.0
1	234	237.1	63	59.9
2	160	155.0	128	133.0
3	14	16.7	75	72.3

Item 3 $S\text{-}X^2(2) = 2.3$, p = 0.3151 (Back)

	Category 0		Category 1	
Score	Observed	Expected	Observed	Expected
0	285	277.8	64	71.2
1	211	221.1	160	149.9
2	63	59.9	132	135.1
3	10	10.4	75	74.6

Item 4 $S\text{-}X^2(2) = 5.4$, p = 0.0672 (Back)

	Category 0		Category 1	
Score	Observed	Expected	Observed	Expected
0	285	275.8	161	170.2
1	114	125.0	201	190.0
2	22	24.2	134	131.8
3	8	4.6	75	78.4

The tables above illustrate good fit. In cases in which the model fits poorly, deviations between observed and expected may be large, and there may be long "runs" of red (or blue) in columns, as the trace line is "over" or "under" due to some model misspecification. (The latter can really only be observed in longer tests.)

Because the "full tables" for $S\text{-}X^2$ can be very large, printing them is optional.

Conclusions

Only a few of the features of IRTPRO have been described in this brief presentation. Taken as a whole, this software represents a significant step forward that will permit more thorough IRT analyses of health outcomes scales, educational tests, and a variety of other measurement instruments.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, *Statistical theories of mental test scores* (pp. 392-479). Reading, MA: Addison-Wesley.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, *37*, 29-51.

Bock, R. D. (1997). The nominal categories model. In W. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 33-50). N.Y.: Springer.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: an application of the EM algorithm. *Psychometrika*, *46*, 443-459.

Bock, R. D. & Lieberman, M. (1970). Fitting a response model for *n* dichotomously scored items. *Psychometrika*, *35*, 179-197.

Bock, R. D. & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement, 6*, 431-444.

Cai, L. (2008). SEM of another flavour: Two new applications of the supplemented EM algorithm. *British Journal of Mathematical and Statistical Psychology*, 61, 309-329.

Cai, L. (2010). High-dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. *Psychometrika*, 75, 33-57.

Cai, L. (in press-a). A two-tier full-information item factor analysis model with applications. *Psychometrika*.

Cai, L. (in press-b). Metropolis-Hastings Robbins-Monro Algorithm for Confirmatory Item Factor Analysis. *Journal of Educational and Behavioral Statistics*.

Cai, L., Maydeu-Olivares, A., Coffman, D.L., & Thissen, D. (2006). Limited information goodness-of-fit testing of item response theory models for sparse 2^p tables. *British Journal of Mathematical and Statistical Psychology*, 59, 173-194.

Chen, W.-H., & Thissen, D. (1997). Local dependence indices for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, *22*, 265-289.

du Toit, M. (2003). *IRT from SSI*. Lincolnwood, IL: Scientific Software International.

Gibbons, R.D., Bock, R. D., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., Kupfer, D. J., Frank, E., Grochocinski, V. J., & Stover, A. (2007). Full-information item bifactor analysis of graded response data. *Applied Psychological Measurement*, *31*, 4-19.

Gibbons, R. D., & Hedeker, D. (1992). Full-information item bi-factor analysis. *Psychometrika, 57*, 423-436.

Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A. Clausen, *Measurement and Prediction* (pp. 362-412). New York: Wiley.

Lord, F. M. (1977). A study of item bias, using item characteristic curve theory. In Y. H. Portinga (Ed.), *Basic problems in cross-cultural psychology* (pp. 19-29). Amsterdam : Swets and Zeitlinger.

Maydeu-Olivares, A., & Joe, H. (2005). Limited and full information estimation and testing in 2^n contingency tables: A unified framework. *Journal of the American Statistical Association*, *100*, 1009–1020.

Maydeu-Olivares, A. & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika*, *71*, 713-732.

Mislevy, R. (1984). Estimating latent distributions. *Psychometrika*, *49*, 359-381.

Mislevy, R. J. (1985). Estimation of latent group effects. *Journal of the American Statistical Association*, *80*, 993-997.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16*, 159-176.

Muraki, E. (1997). A generalized partial credit model. In W. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 153-164). N.Y.: Springer.

Orlando, M., & Thissen, D. (2000). Likelihood-based item fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, *24*, 50-64.

Orlando, M. & Thissen, D. (2003). Further investigation of the performance of $S\text{-}X^2$: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*, *27*, 289-298.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph*, No. 17, *34*, Part 2.

Samejima, F. (1997). Graded response model. In W. van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 85-100). N.Y.: Springer.

Schilling, S., & Bock, R. D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika, 70*, 533–555.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.

Spielberger, C.D. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Stouffer, S. A., Guttman, L., Suchman, E.A., Lazarsfeld, P. F. Star, S. A. , & Clausen, J. A. (1950). *Measurement and Prediction*. New York: Wiley.

Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika*, *47*, 201-214.

Thissen, D., Cai, L., & Bock, R.D. (2010). The nominal categories item response model. In M.L. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models* (pp. 43-75). New York, NY: Routledge.

Thissen, D., Nelson, L., Rosa, K., & McLeod, L. D. (2001). Item response theory for items scored in more than two categories. In D. Thissen & H. Wainer (Eds.), *Test Scoring* (Pp. 141-186). Mahwah, NJ: Lawrence Erlbaum Associates.

Thissen, D., & Orlando, M. (2001). Item response theory for items scored in two categories. In D. Thissen & H. Wainer (Eds.), *Test Scoring* (pp. 73-140). Mahwah, NJ: Lawrence Erlbaum Associates.

Thissen, D. & Steinberg, L. (2009). Item response theory. In R. Millsap & A. Maydeu-Olivares (Eds.), *The Sage handbook of quantitative methods in psychology* (pp. 148-177). London: Sage Publications.

Zuckerman, M. (1980). The development of an affect adjective check list for the measurement of anxiety. *Journal of Consulting Psychology*, *24*, 457-462.