Announcing the release of IRTPRO version 1.2

SSI has enjoyed great success over the years in the development and publishing of statistical software and is proud to announce another important step in its evolution as a software publishing company.

For many years SSI has had a widely distributed suite of four IRT (Item Response Theory) programs called Bilog-MG, Multilog, Parscale and Testfact. However, with growth and success also come new users and new expectations. Despite their success, these IRT programs are far from perfect measured by today's software standards. Furthermore, they share a great deal of similarities and overlapping functionalities. Thus, it became evident that a single, state-of-the-art software product must be developed to replace these four programs.

In an effort to meet the growing demands of our user community, SSI has developed a new software product, called IRTPRO, which is on the cutting edge of current technology. The program has been tested extensively on the Microsoft Windows platform with Windows7, Vista and XP operating systems. Important sections of the IRTPRO numeric engine have been parallelized to run on multiple cores simultaneously.

The development of IRTPRO was supported by an NIH SBIR contract HHSN-2612007-00013C**,** titled: MEDPRO: Developing Item Response Theory Software for Outcome and Behavioral Measurements.

Technical Description

IRTPRO imports data from a variety of statistical software packages as well as importing data from fixed format data (**.fixed**), comma-separated (**.csv**), tab-delimited (usually **.txt**), and Excel (**.xls**) files. Whatever the original format, the imported data are saved to an IRTPRO data file with extension **.ssig** that is displayed visually as a spreadsheet, similar in appearance to an Excel spreadsheet.

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) [with which equality constraints includes the one-parameter logistic (1PL)
- Three-parameter logistic (3PL)
- Graded
- Generalized Partial Credit
- Nominal

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. That being said, alternative computational methods may be used, each of which provides best performance for some combinations of dimensionality and model structure:

- Bock-Aitkin (BAEM)
- Bifactor EM
- Generalized Dimension Reduction EM
- Adaptive Quadrature (ADQEM)
- Metropolis-Hastings Robbins-Monro (MHRM)

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
- Expected *a posteriori* (EAP) for summed scores

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. [Most often, if there is only one group, the population latent variable mean(s) and variance(s) are fixed 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, with accurate item parameter error variance-covariance matrices computed using a Supplemented EM (SEM) algorithm. 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) and the Bayesian Information Criterion (BIC) 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, a M_2 statistic is also computed. Diagnostic statistics include generalizations for polytomous responses of the local dependence (LD) statistic and the *SS-X*^{2} item-fit statistic.

IRTPRO supports both model-based and data-based graphical displays. Model-based graphs, currently available for unidimensional IRT models only, are trace lines, information curves, combined trace lines-information curves, total information, and test characteristic curves.

Documentation

SSI has written a user's guide (available in PDF format) and online help file that can be accessed when the program is launched.

The user's guide has been written to introduce item response theory (IRT) models to researchers new in this field. It also serves as a guide to researchers who are already familiar with the existing IRT programs

distributed by Scientific Software International and are upgrading to a program that has an easy to use graphical users interface (GUI) and can handle multidimensional models. In this guide the focus is on the "how to" part of IRT.

Chapter 2 provides a short description of the GUI, since the examples in the remaining chapters further illustrate the features of the user's interface. IRTPRO uses its own data format, displayed in spreadsheet form. Data may be imported from a long list of statistical software packages and spreadsheet programs. Chapter 3 deals with data import and manipulation and Chapter 4 deals with the calculation of traditional summed-score statistics.

Chapters 5 to 7 deals with the estimation (calibration) of IRT models. Chapter 5 is concerned with the fitting of unidimensional models and Chapter 6 deals with multiple groups and differential item functioning (DIF). In Chapter 7 we describe how IRTPRO handles exploratory and confirmatory factor analysis models. This chapter also contains examples illustrating the fit of bifactor and one and two-tier testlet response theory models.Unlike classical test theory, IRT does not in general base the estimate of the respondent's ability (or other attribute) on the number-correct (NC) or summed score. To distinguish IRT scores from their classical counterparts, we refer to them as "scale" scores. The computation of IRT scale scores in IRTPRO may be done using one of the three methods discussed in Chapter 8.

Graphics are often a useful data-exploring technique through which the researcher may familiarize her- or him with the data. IRTPRO offers both model-based and data-based graphs. The Model-based graphs discussed in Chapter 9 cover item- and test- characteristic curves; information and total information curves and are available for unidimensional IRT models only. In the case of the data-based graphs presented in Chapter 10, IRTPRO distinguishes between univariate and bivariate graphs. Univariate graphs are particularly useful to obtain an overview of the characteristics of a variable. However, they do not necessarily offer the tools needed to explore the relationship between a pair of variables.

For most unidimensional and bifactor IRT models parameter estimation can be done effectively selecting the Bock-Aitkin EM algorithm (the default estimation method). In the case of multidimensional models, the method of estimation depends to a large extend on the number of dimensions of the model to be fitted. A general rule is that two-dimensional models can be handled effectively using Bock-Aitkin or adaptive quadrature. For three- to four-dimensional models, the estimation methods of choice are adaptive quadrature and MH-RM. Higher dimensional models are handled most effectively using MH-RM. Chapter 11 provides a short description of the options available for each of these estimation methods.

Each analysis created by the GUI produces a syntax file, essentially being a record of a user's selections from the sequence of dialogs. If a syntax file is opened, IRTPRO automatically fills the relevant GUI dialogs that can be viewed and modified. These aspects are dealt with in Chapter 12. The user's guide provides users with an extensive set of references relevant to the various topics addressed.

Example: Unidimensional analysis of four "self monitoring" items

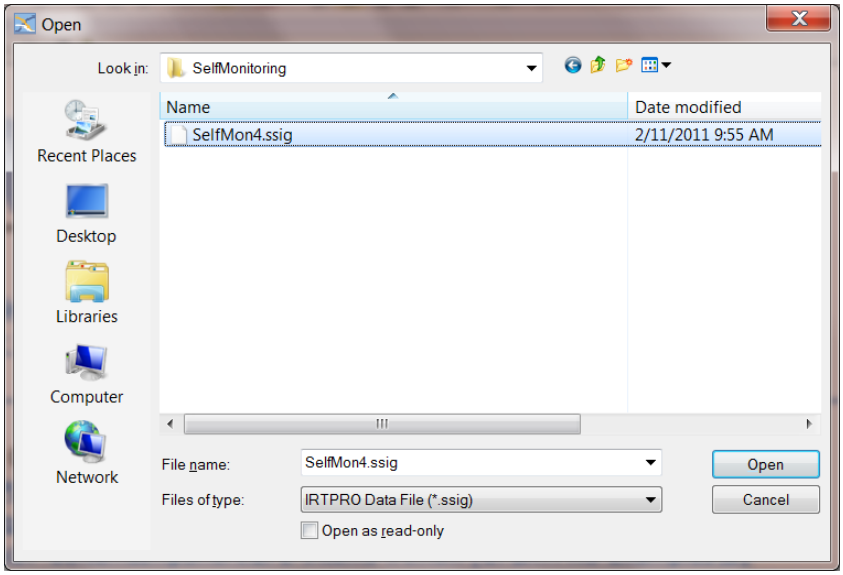
Thissen & Steinberg (2009) describe IRT model fitting for the responses of 393 undergraduate students to four items of the Self-Monitoring Scale (SMS; Snyder, 1974; modified by Snyder and Gangestad, 1986). The data used is from the 1988 "Self-Monitoring Scale (CAPS-SELFMON, SELF_MONIT and SELF_MONIT_PAPER module)", hdl:1902.29/CAPS-SELFMON Odum Institute Dataverse. They consider a subset of the data for the following four items:

- SelfMon8: I have considered being an entertainer. (T)
- SelfMon13: I have never been good at games like charades or improvisational acting. (F)
- SelfMon18: I would probably make a good actor. (T)
- SelfMon20: In different situations and with different people, I often act like a very

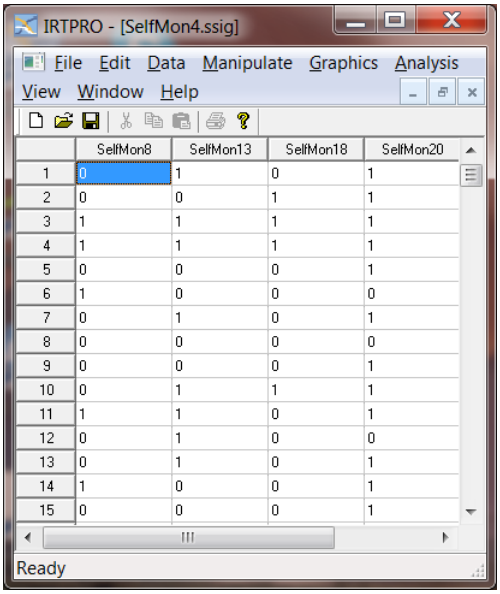
different person. (T)

The high self-monitoring response to each of the items above (T or F in parentheses after each item) is coded 1 and the other response is coded 0.

To begin, we use the **Open** file dialog under the **File** menu of IRTPRO and navigate to the **C:\IRTPRO Examples\By Dataset\SelfMonitoring** folder. However, here we change the **Files of type:** selection from its default **IRTPRO Command File (*.irtpro)** to **IRTPRO Data File (*.ssig)** in the open file dialog, and open the file **SelfMon4.ssig**.

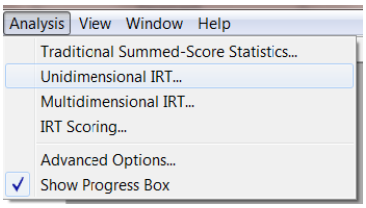


The first 15 cases of this data are shown below.

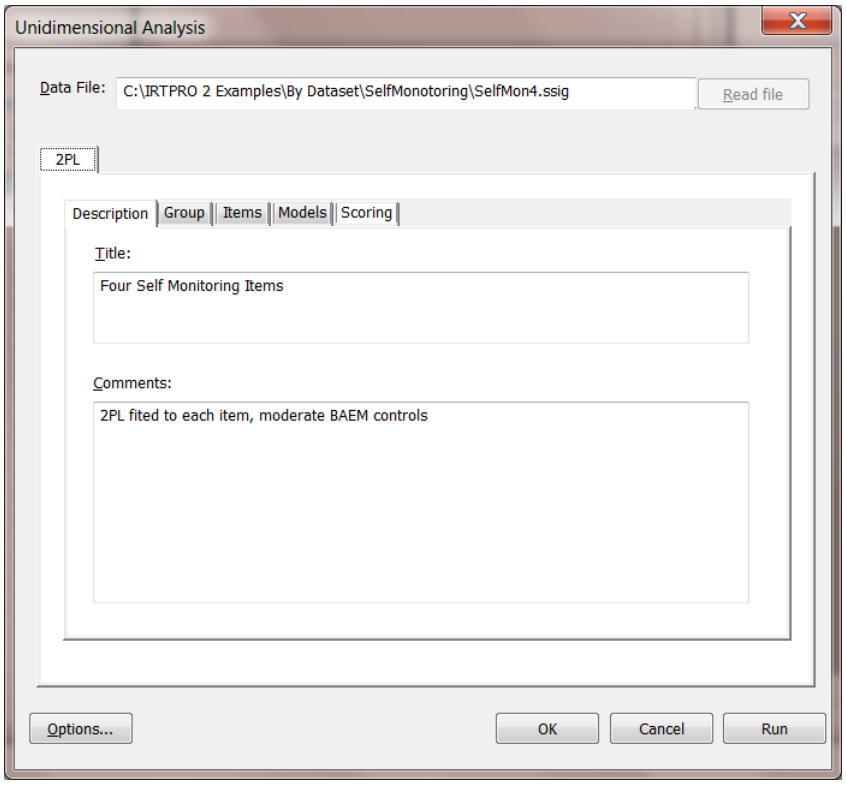


	SelfMon8	SelfMon13	SelfMon18	SelfMon20
1	0	1	0	1
2	0	0	1	1
3	1	1	1	1
4	1	1	1	1
5	0	0	0	1
6	1	0	0	0
7	0	1	0	1
8	0	0	0	0
9	0	0	0	1
10	0	1	1	1
11	1	1	0	1
12	0	1	0	0
13	0	1	0	1
14	1	0	0	1
15	0	0	0	1

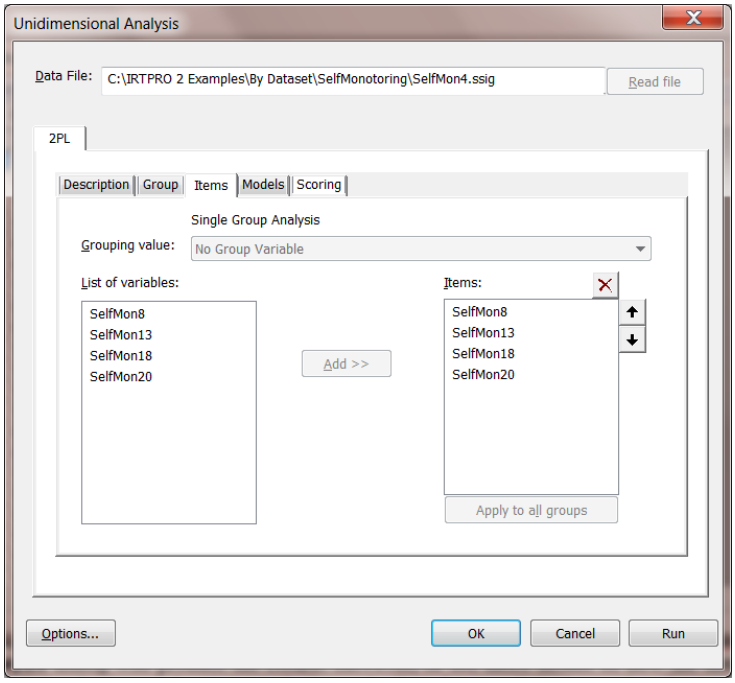
To set up the analyses, select the **Unidimensional IRT…** option from the **Analysis** menu to invoke the unidimensional analysis widow.



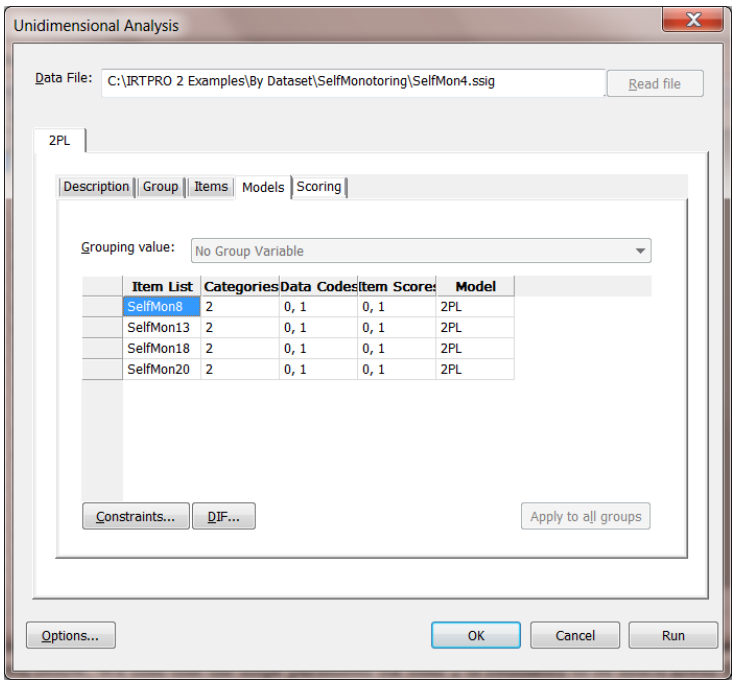
This window has five tabs called **Description**, **Group**, **Items**, **Models** and **Scoring**. Start with the default tab **Description** and provide a title and comments in the appropriate text boxes as shown below. Note that the default name for the current analysis is Test1. As will be shown later, more tests based on the same dataset may be inserted and each of these can be renamed to something that may be more suitable. In the present case, right-click on the Test1 tab and rename it to 2PL.



Since this data contains no grouping variable, the **Group** tab is skipped and we proceed to the **Items** tab, where all four items from the **List of variables** are selected. Then use the **Add** button to list these items under **Items**.



Because the **2PL** model is the default for dichotomous items, the entry of information for the analysis is now complete, and clicking on the **Run** button in the lower right of the **Unidimensional Analysis** dialog will produce the results. However, to see more details of how the data will be modeled, 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:



Click the **Run** button to run the 2PL analysis. Portions of the output file **SelfMon4.2PL-irt.htm** are shown below. We find that the slope parameter for item 2 is estimated to be much lower than those for the other three items:

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

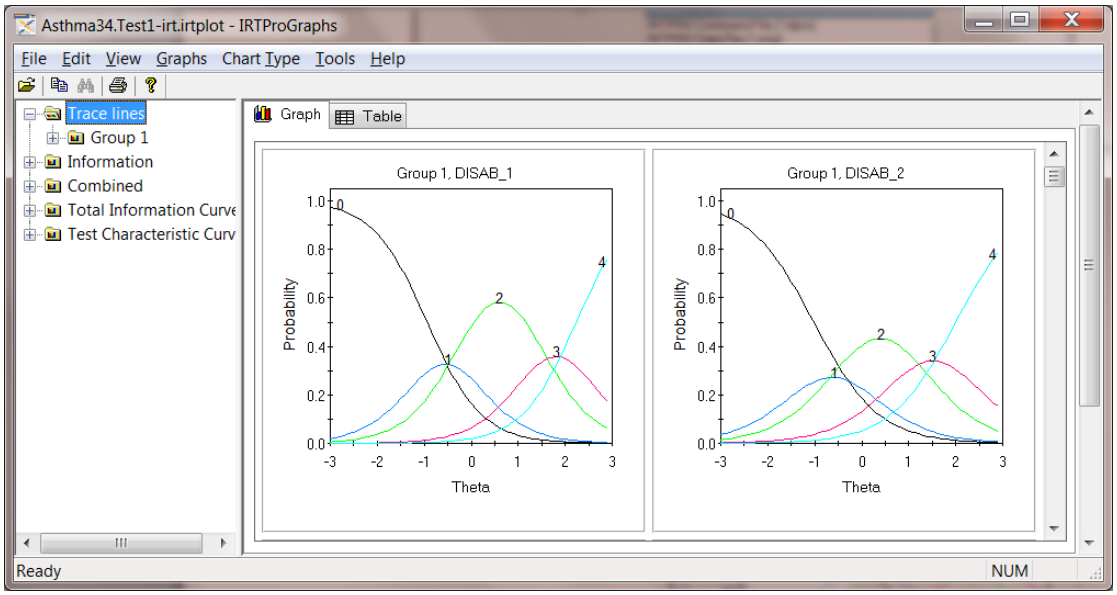
Item	Label	a		s.e.	c		s.e.	b	s.e.
1	SelfMon8	2	4.12	1.37	1	-0.09	0.28	0.02	0.07
2	SelfMon13	4	0.12	0.13	3	0.49	0.10	-4.23	4.67
3	SelfMon18	6	2.41	0.46	5	-0.91	0.22	0.38	0.08
4	SelfMon20	8	2.02	0.34	7	0.89	0.19	-0.44	0.09

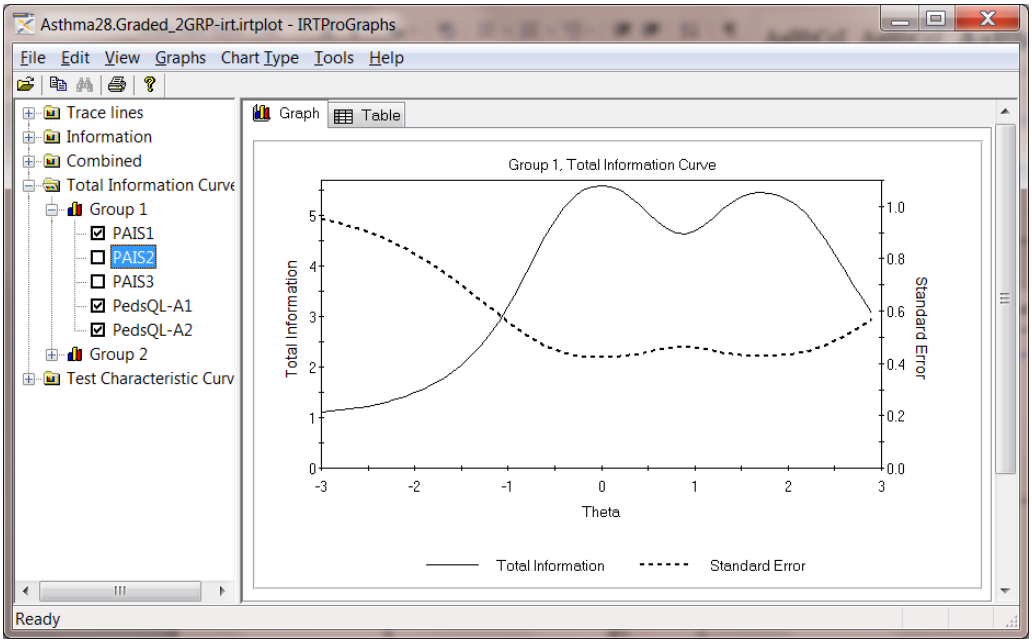
If we click on the entry **Likelihood-based Values and Goodness of Fit Statistics** in the table of contents for this 2PL fit to these items, we find that the model appears to fit very well:

Statistics based on the full item x item x ... classification			
G^2	Degrees of freedom	Probability	RMSEA
5.03	7	0.6570	0.00

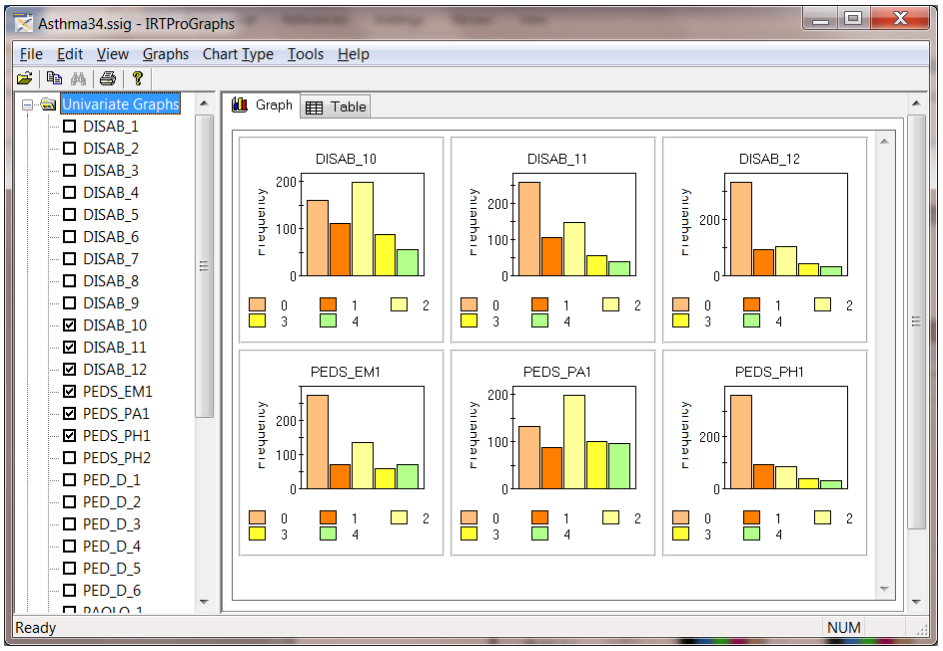
Statistics based on one- and two-way marginal tables			
M_2	Degrees of freedom	Probability	RMSEA
3.17	2	0.2063	0.04
Note: M_2 is based on full marginal tables.			
Note: Model-based weight matrix is used.			

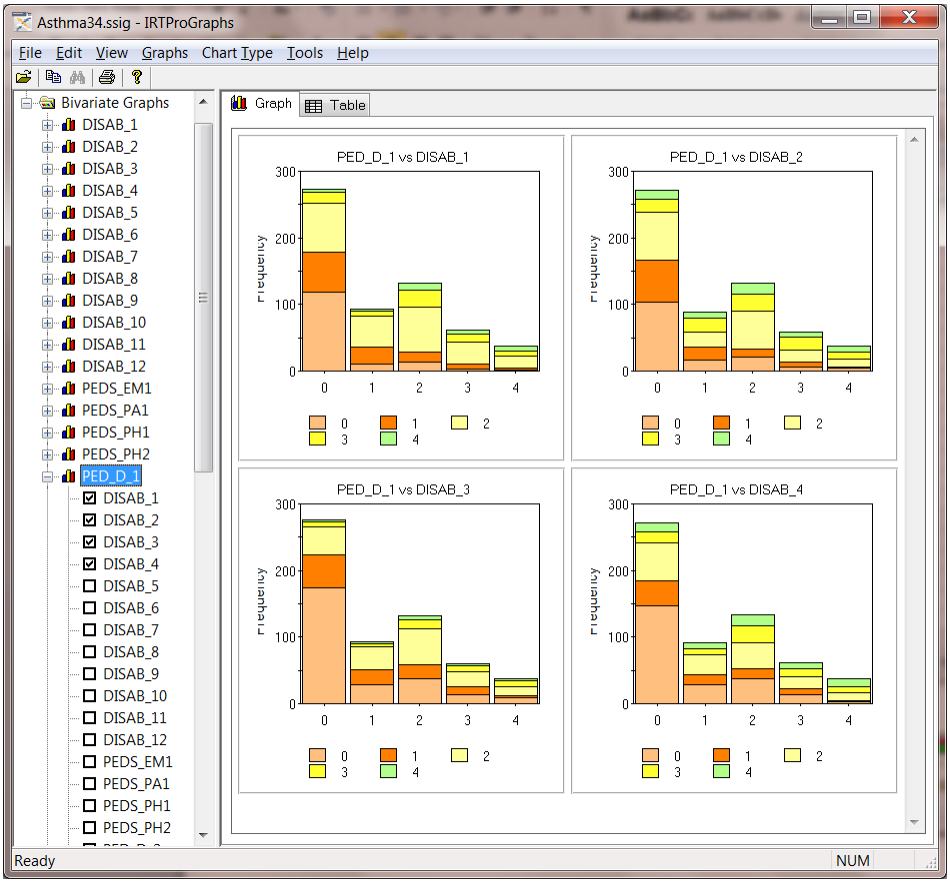
Examples of Graphical Displays: Model-Based


Asthma34.Test1-irt.irtplot - IRTProGraphs
File Edit View Graphs Chart Type Tools Help
Trace lines
Group 1
Information
Combined
Total Information Curve
Test Characteristic Curv
Graph
Table
Group 1, DISAB_1
Group 1, DISAB_2
Probability
Theta
Ready
NUM


Asthma28.Graded_2GRP-irt.irtplot - IRTProGraphs
File Edit View Graphs Chart Type Tools Help
Trace lines
Information
Combined
Total Information Curve
Group 1
PAIS1
PAIS2
PAIS3
PedsQL-A1
PedsQL-A2
Group 2
Test Characteristic Curv
Graph
Table
Group 1, Total Information Curve
Total Information
Standard Error
Theta
Total Information
Standard Error
Ready

Examples of Graphical Displays: Data-Based





Cost and ordering information\

The software can be ordered from our online website https://www.ssicentral.biz/default.aspx and will be delivered electronically. Single user licenses, concurrent licenses, and rental licenses (6 or 12 months) are available. All pricing information is available on the site.

A printable PDF copy of the IRTPRO User's Guide is distributed with the full and rental editions of IRTPRO 2.1 for Windows.

Single user licenses for BILOG-MG 3, MULTILOG 7, PARSCALE 4, and TESTFACT 4 for Windows can be upgraded to single user licenses for IRTPRO for Windows.