1. Introducing IRTPRO

1.1 IRTPRO features

IRTPRO (Item Response Theory for Patient-Reported Outcomes) is an entirely new application for item calibration and test scoring using IRT.

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)
- 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. 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) (Bock & Aitkin, 1981)
- Bifactor EM (Gibbons & Hedeker, 1992; Gibbons *et al*., 2007; Cai, Yang & Hansen in press)
- Generalized Dimension Reduction EM (Cai, 2010-a)
- Adaptive Quadrature (ADQEM) (Schilling & Bock 2005)
- Metropolis-Hastings Robbins-Monro (MHRM) (Cai, 2010-b, 2010-c)

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 & Thissen (1997) and the $SS\text{-}X^2$ item-fit statistic suggested by Orlando & Thissen (2000, 2003).

1.2 Organization of the user's guide

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.

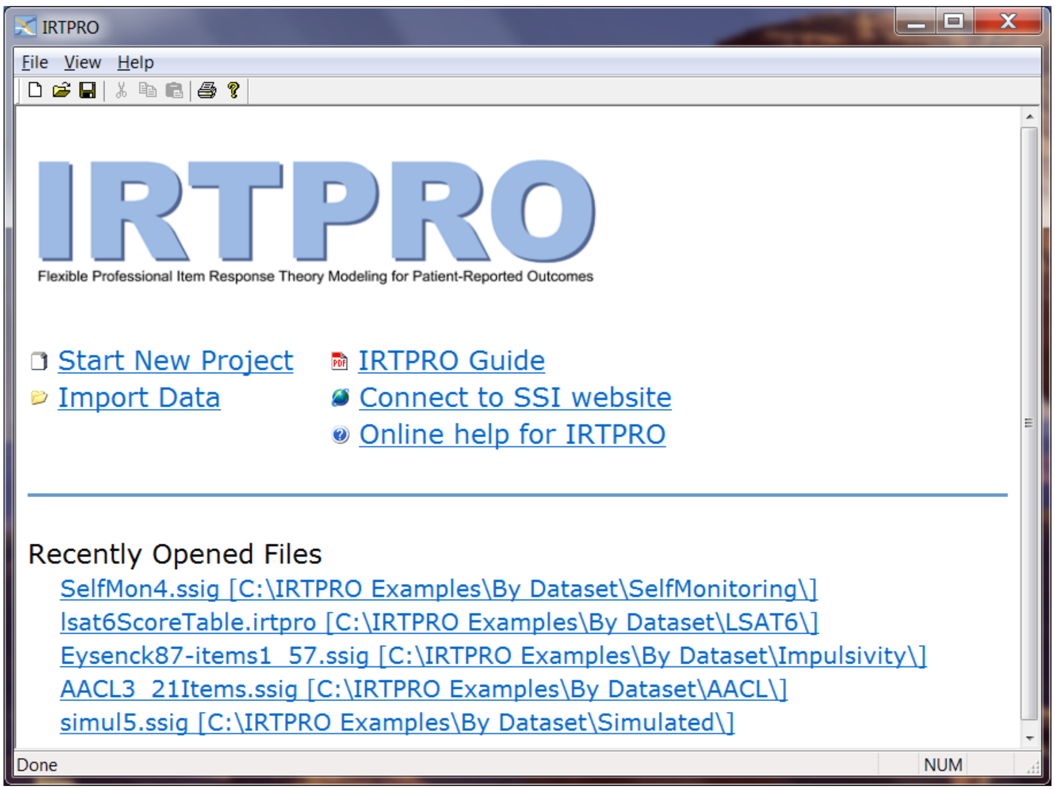
Each analysis created by the GUI produces a syntax file, essentially being a record of a user's selections from the sequence of dialogs. If a syntax file is opened, IRTPRO automatically fills the relevant GUI dialogs that can be viewed and modified. These aspects are dealt with in Chapter 12.

2. Graphical users interface

2.1 Introduction

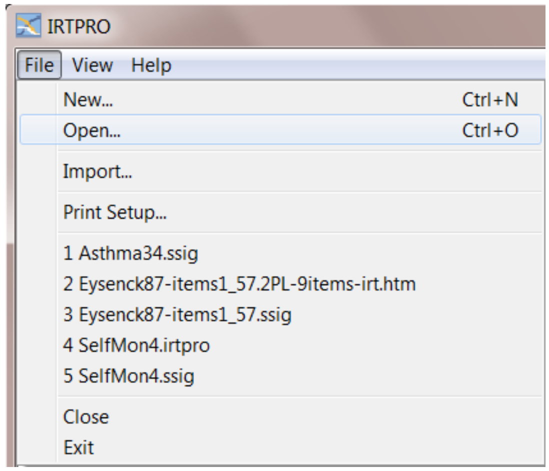
In this chapter, the main features of the IRTPRO graphical users interface (GUI) are summarized. The examples in Chapter 3 through to Chapter 10 were all created via the GUI. Use of the available menus and dialogs are discussed in detail in these chapters.

When IRTPRO is launched (typically by clicking on the IRTPRO icon on the computer desktop), a page is opened containing clickable links to recently used files, the **Import Data** menu, online help documentation and to gain access to the SSI website. At his stage, the main menu bar displays the **File**, **View** and **Help** options.

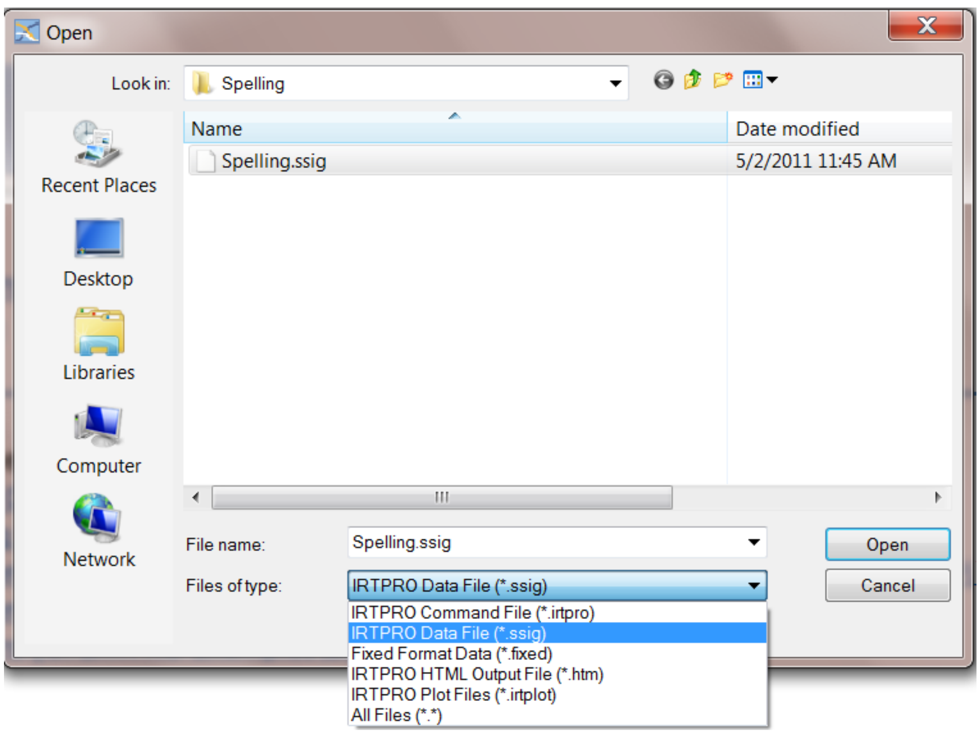


2.2 Opening an IRTPRO data (ssig) file

By clicking the **File** button, the drop-down menu shown below is activated.



By selecting the **Open** option, a standard **Open** dialog box is displayed.



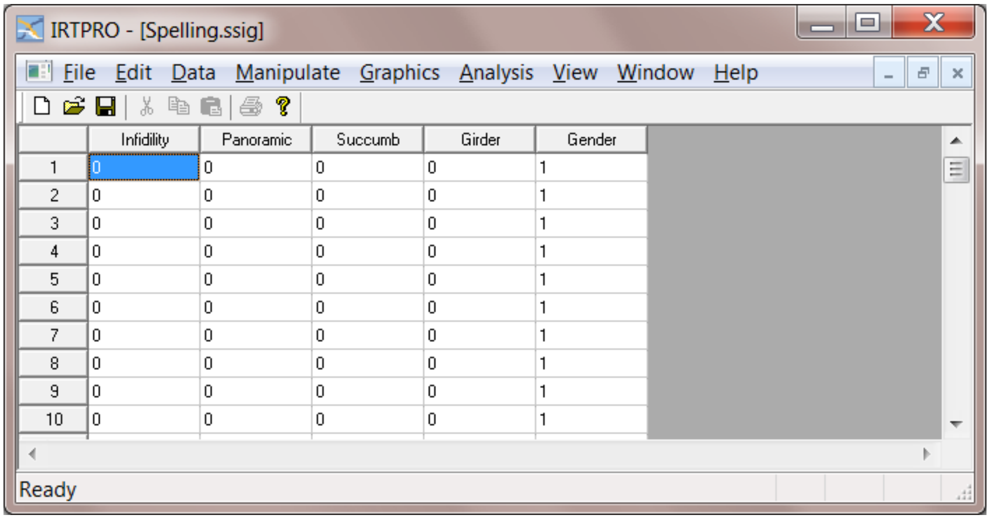
There are five main file types that IRTPRO can open, these being:

- An IRTPRO command (syntax) file with extension **.irtpro** (See Chapter 12)
- An IRTPRO data file with extension **.ssig** (See Chapters 4 to 10)
- Fixed format data with extension **.fixed** (See Chapter 3)
- An IRTPRO HTML output file with extension **.htm** (See Chapters 4 to 8)
- An IRTPRO plot file with extension **.irtplot** (See Chapter 9)

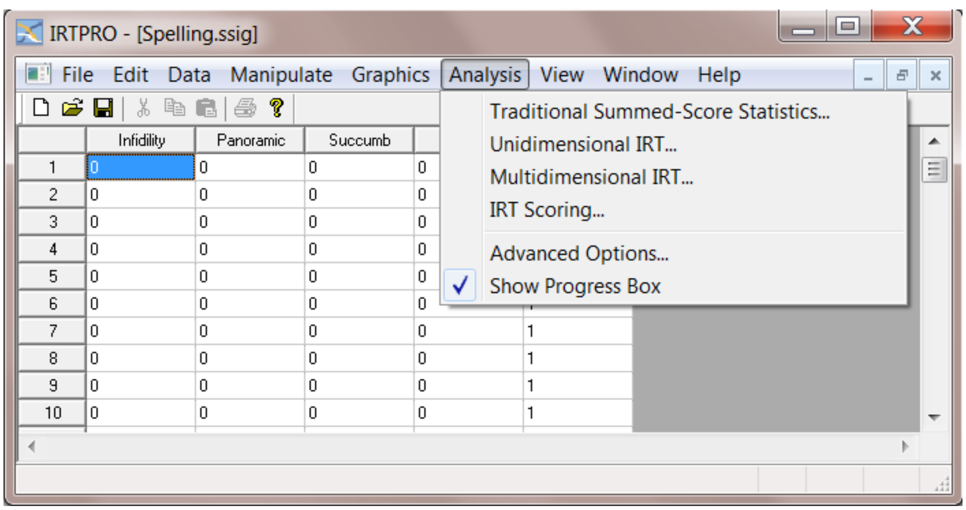
A file with extension **.ssig** refers to an IRTPRO data file and is typically created by importing data from a statistical software package such as SPSS or SAS or a spreadsheet program such as Excel. The import of data into **.ssig** format is dealt with in Chapter 3.

2.3 Spreadsheet main menu bar

IRTPRO data files are displayed in spreadsheet form.



Once a file of type **.ssig** is opened, the main-menu bar displays several options. For example, by clicking the **Analysis** button the drop-down menu shown below is obtained.

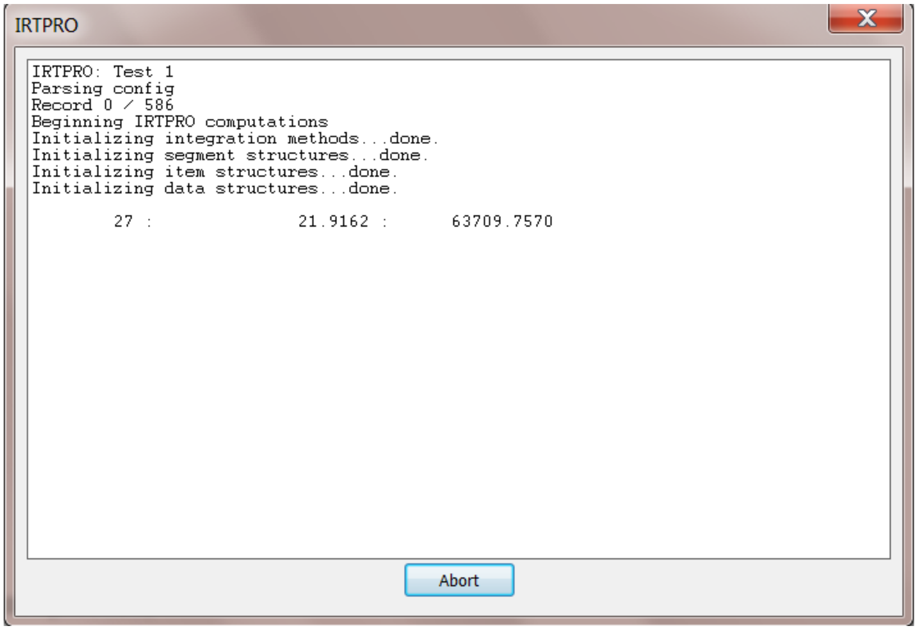


There are four main types of analyses, namely:

- Traditional summed-score statistics (See Chapter 4)
- Unidimensional IRT (See Chapter 5 and 6)
- Multidimensional IRT (See Chapter 7)
- IRT scoring (See Chapter 8)

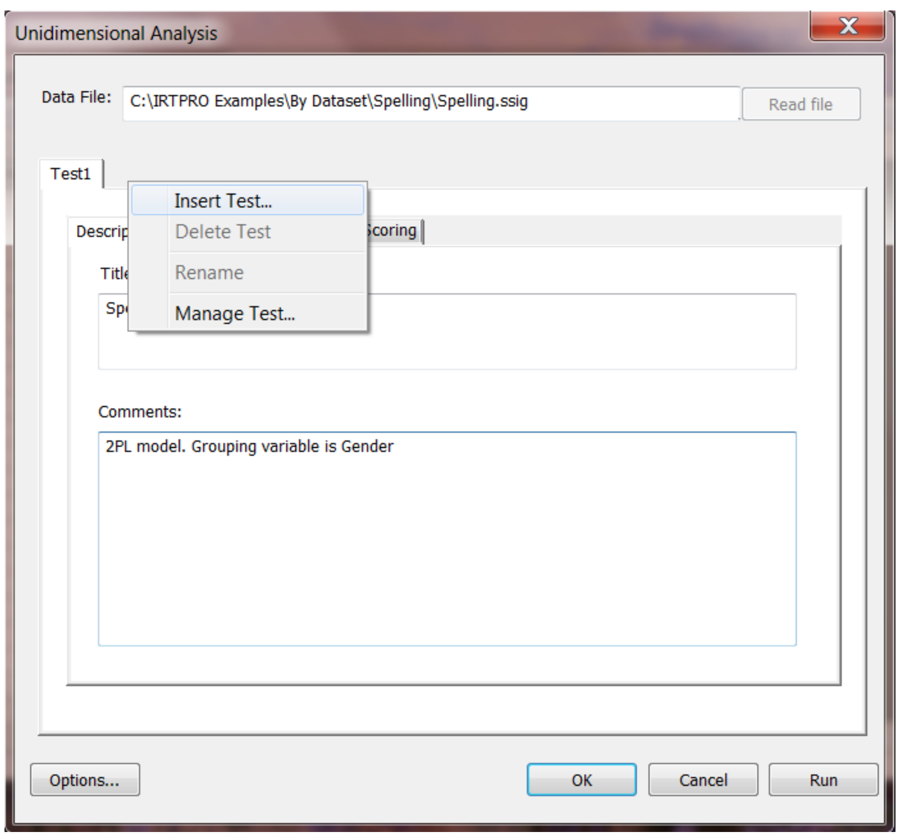
There are two additional items on the **Analysis** list, these being **Advanced Options…** (see Section 2.9) and **Show Progress Box**. By selecting the **Show Progress Box** option (the default), various results of the analysis are displayed, enabling the user to visually determine if the analysis is still running and what progress has been made.

The image below is a screen shot of the progress box for an analysis that is partially completed.



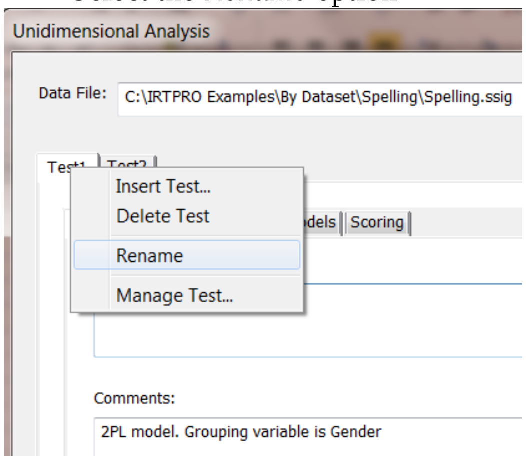
2.4 Test Tabs

Regardless of the type of analysis specified, several tests (analyses) may be created using the same IRTPRO dataset. To insert a new test, right-click on the right-hand side of a current test to insert a new test tab.

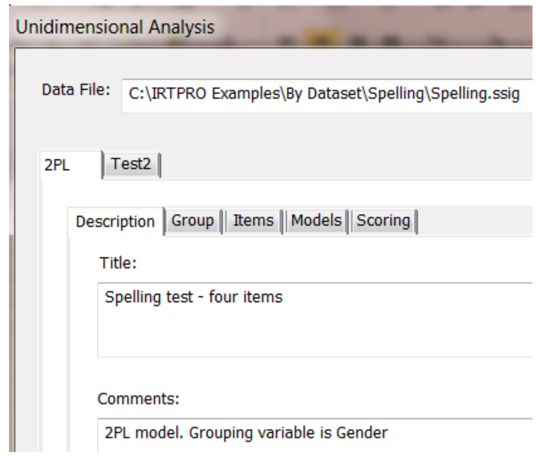


By right clicking on a test tab, the test may be renamed or deleted. The sequence of steps to rename the first test tab to 2PL is shown below.

Select the **Rename** option



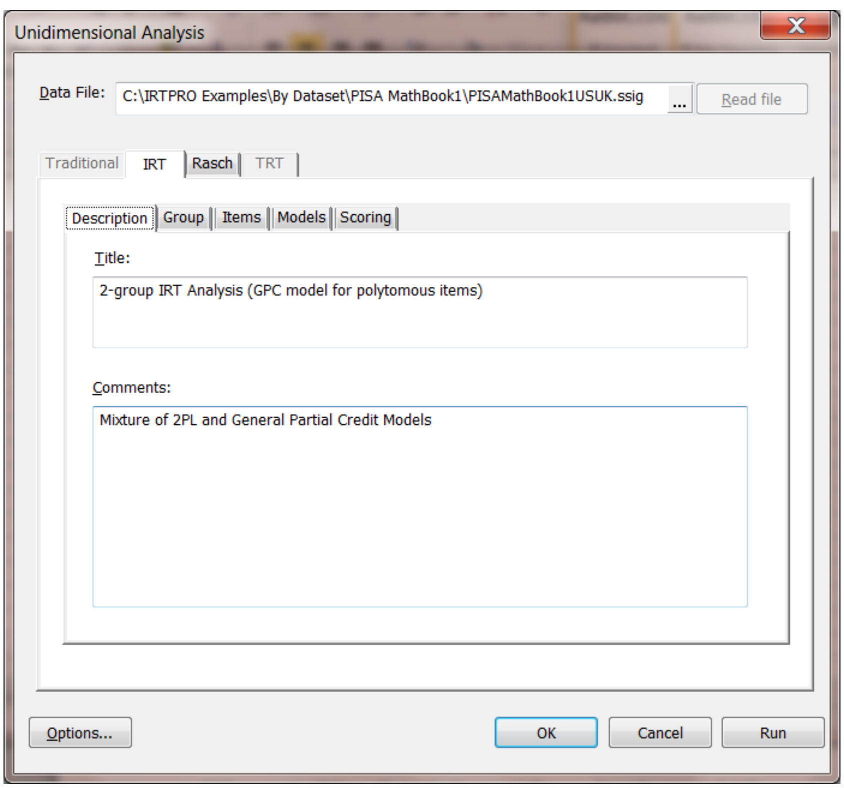
Enter 2PL



2.5 The Description, Group and Items tabs

When a traditional statistics, unidimensional, or multidimensional analysis is requested via the **Analysis** option, the first three tabs displayed in the corresponding analysis window are **Description**, **Group** and **Items**. Each of the **Description**, **Group** and **Items** dialogs will be briefly discussed in Sections 2.5.1 to 2.5.3.

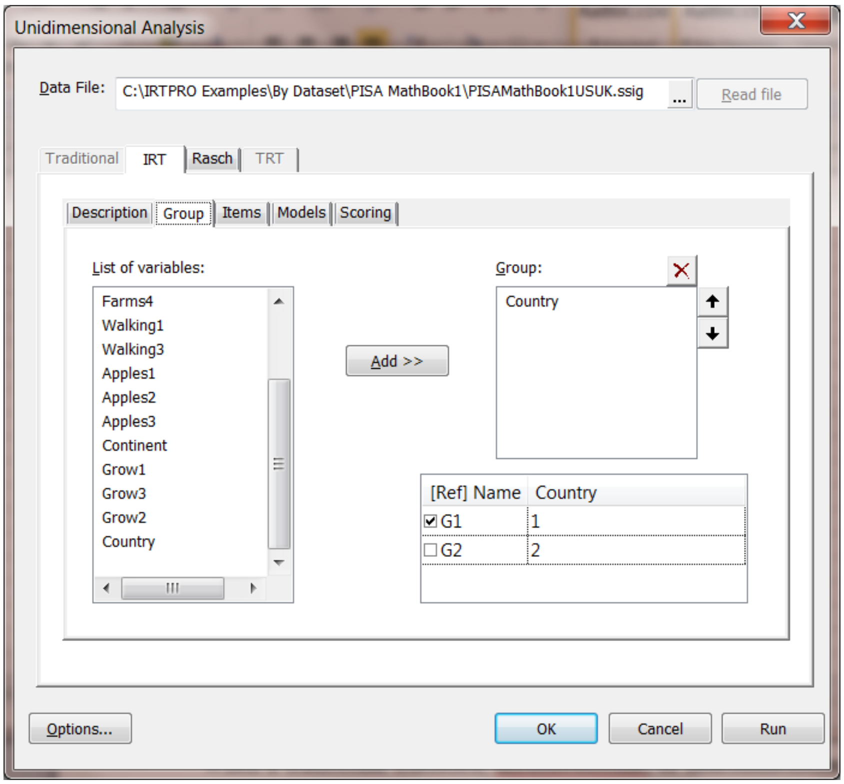
2.5.1 The Description tab



The **Description** tab has two text boxes that are used to enter a title (description) and optional comments for each test tab. Shown below is the **Description** tab for a **Unidimensional Analysis** based on the test named **IRT**.

2.5.2 The Group tab

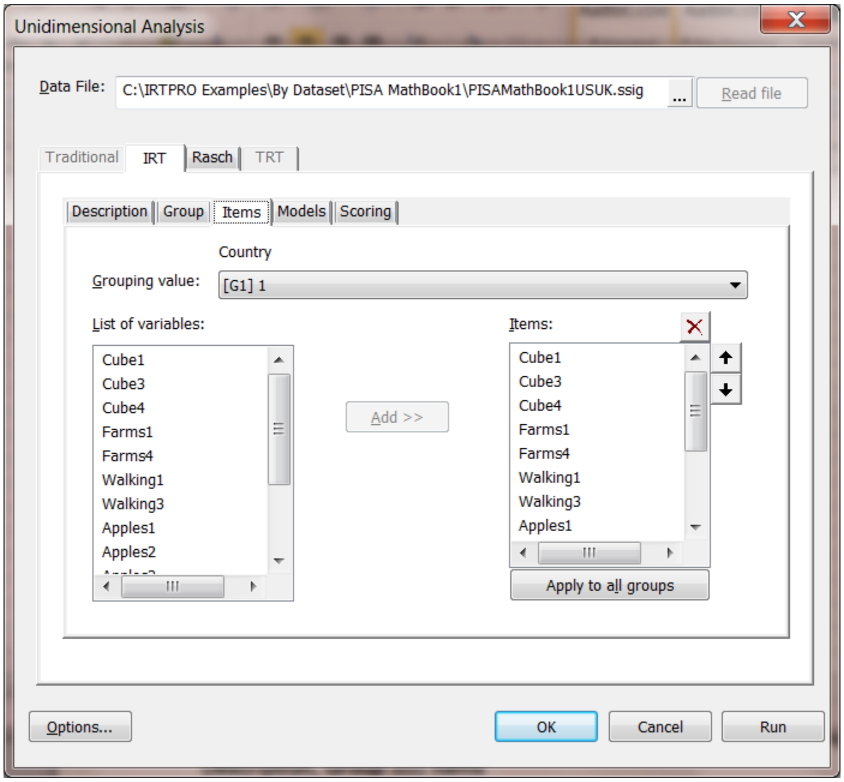
The **Group tab** allows one to select one or more grouping variable(s) from the **List of Variables:** text box. In the dialog shown below, the variable Country was selected as the grouping variable. By default, the first group is selected as the reference group. However, the **Group** dialog box allows the user to select any other group as the reference. Examples of the use of the **Group tab** are given in Chapters 6 and 7.



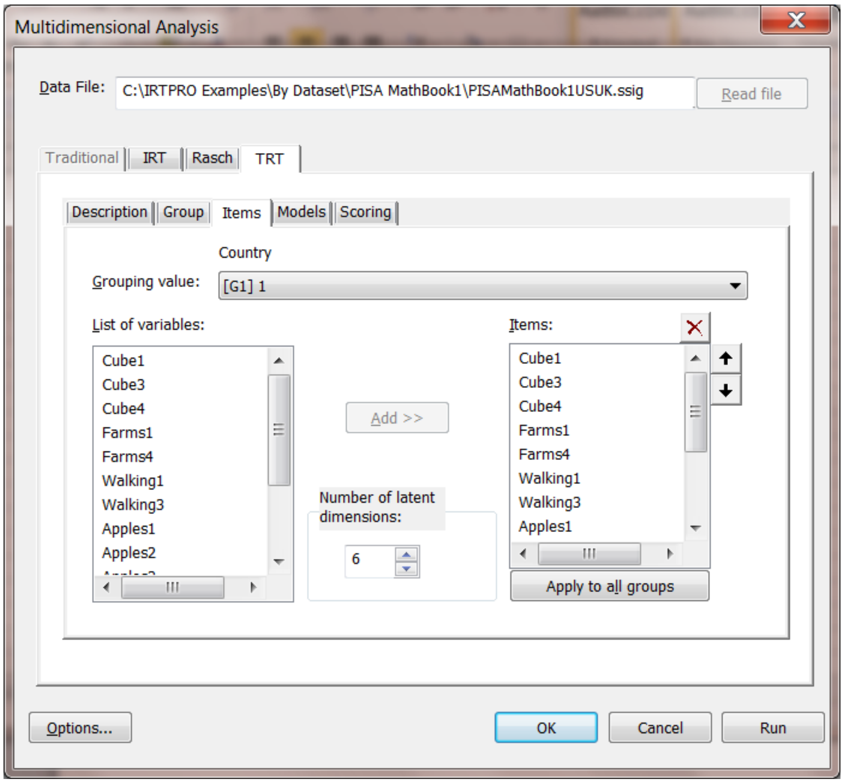
2.5.3 The Items tab

The **Items tab** dialog box for a traditional statistics or unidimensional IRT analysis is shown below. Items can be selected for each group from the **List of variables:** column and adding it to the **Items:** column.

In most practical applications, a multiple group analysis is based on the selection of the same set of items for each group. If this situation applies, the user selects the items from the **List of variables:** for the first group and then clicks on the **Apply to all groups** button to make the same selection for all groups. See Chapters 4 to 6 for examples that illustrate, amongst other, the use of the **Items tab**.



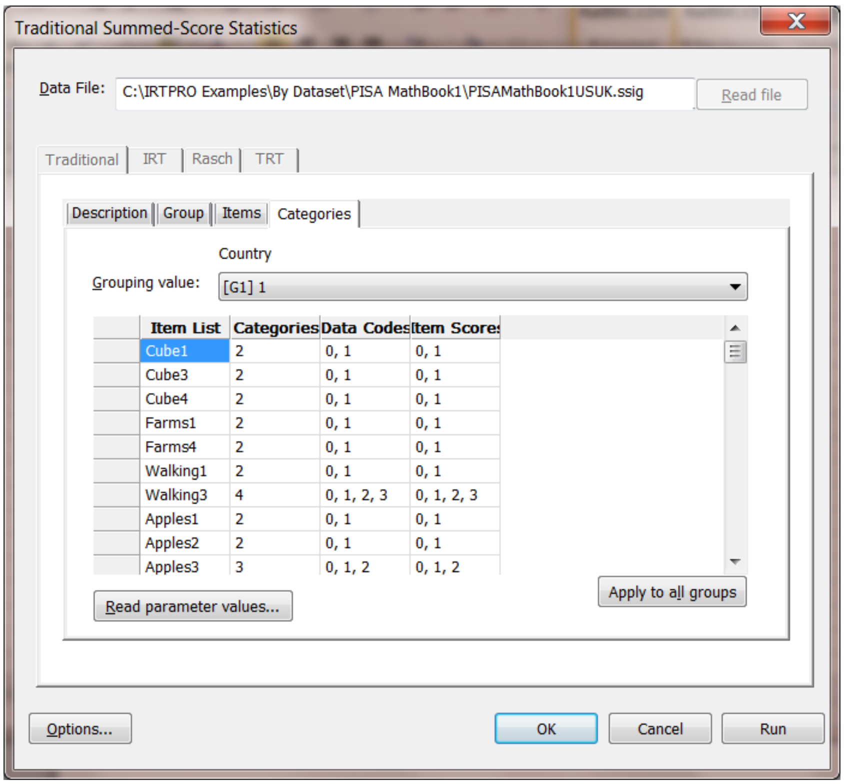
The **Items tab** dialog for a multidimensional IRT analysis is shown below.



The only difference between this dialog and the corresponding one for a traditional statistics or unidimensional IRT analysis is the presence of the text box **Number of latent dimensions:**. Note that the number of latent dimensions must be specified by the user. Examples that illustrate the **Items tab** dialog for multidimensional IRT analyses are given in Chapter 7.

2.6 The Categories tab

When a traditional summed-score statistics analysis is requested via the **Analysis** option, the fourth (and last) tab displayed in the corresponding analysis window, is the **Categories tab**. The dialog associated with the selection of this tab displays the default item scores associated with each of the selected items. A user may change these scoring values by selecting a cell and then right-clicking on the selected cell to display the **Recode Item Scores…** option as demonstrated in Section 2.7.1.



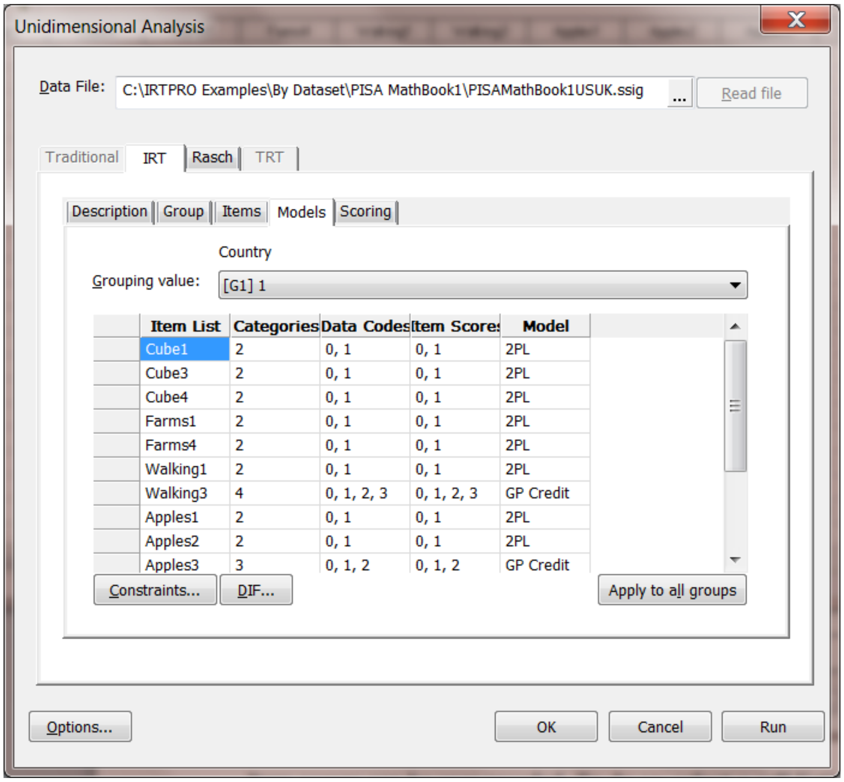
2.7 The Models tab

When a unidimensional IRT, or multidimensional IRT analysis is requested via the **Analysis** option, the fourth tab displayed in the corresponding analysis window, is the **Models tab**. The dialog associated with selection of this tab displays the default models associated with the items and allow a user to change the model type and scoring values of the items.

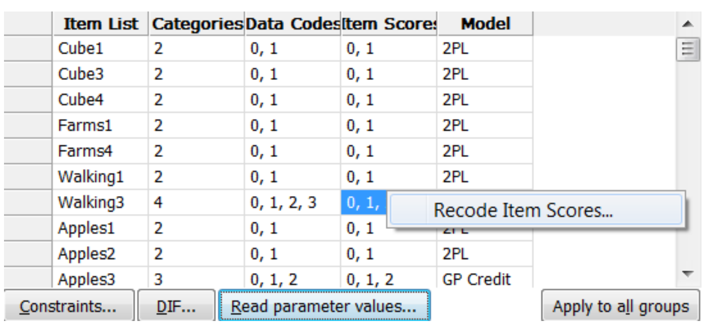
The dialogs for unidimensional and multidimensional IRT differ somewhat in functionality. These differences will be briefly discussed in Sections 2.7.1 and 2.7.2.

2.7.1 The Models tab, Unidimensional IRT Analysis

The **Models** dialog displays, for each group, five columns of information, namely an item list, the number of categories (distinct values) for each item, the data codes (values) extracted from the IRTPRO dataset, the item scores (coded as 0, 1, 2,… where 0 corresponds to the smallest data code value, etc.), and the model selected. For an item with two categories, the default model is the 2PL model and for an item with more than two categories, the default is the Graded model.

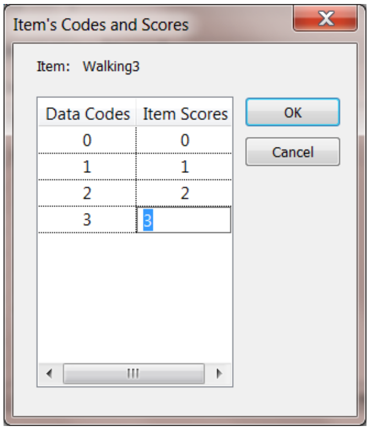


Item scores can be user-recoded. To do so, select a cell listing the scores to be changed. By right-clicking on the selected cell, the **Recode Item Scores…** option is displayed.

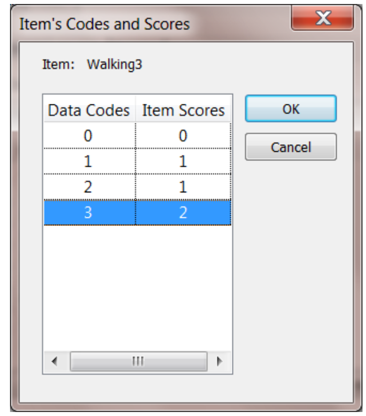


Selection of this option opens, for each of the groups, an **Item's Codes and Scores** dialog. By double-clicking on an **Item Score:** cell, the relevant cell may be edited and a new value entered. The screenshots below show the recoding of the scores for the item Walking3 from (0, 1, 2, 3) to (0, 1, 1, 2).

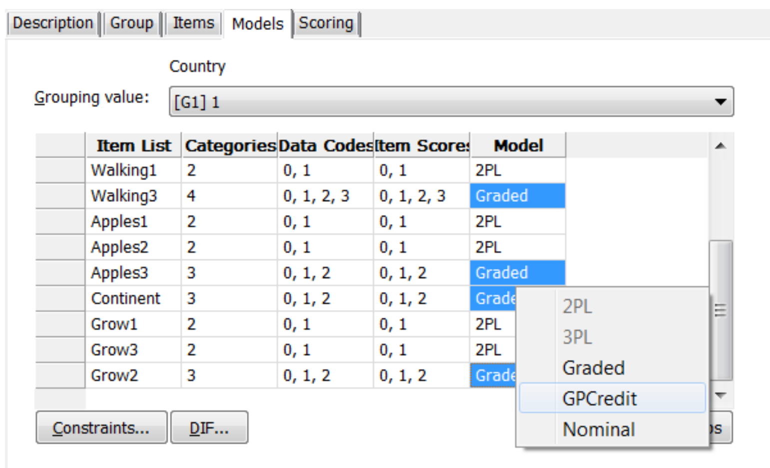
Double click on row



Edit number and click **OK**

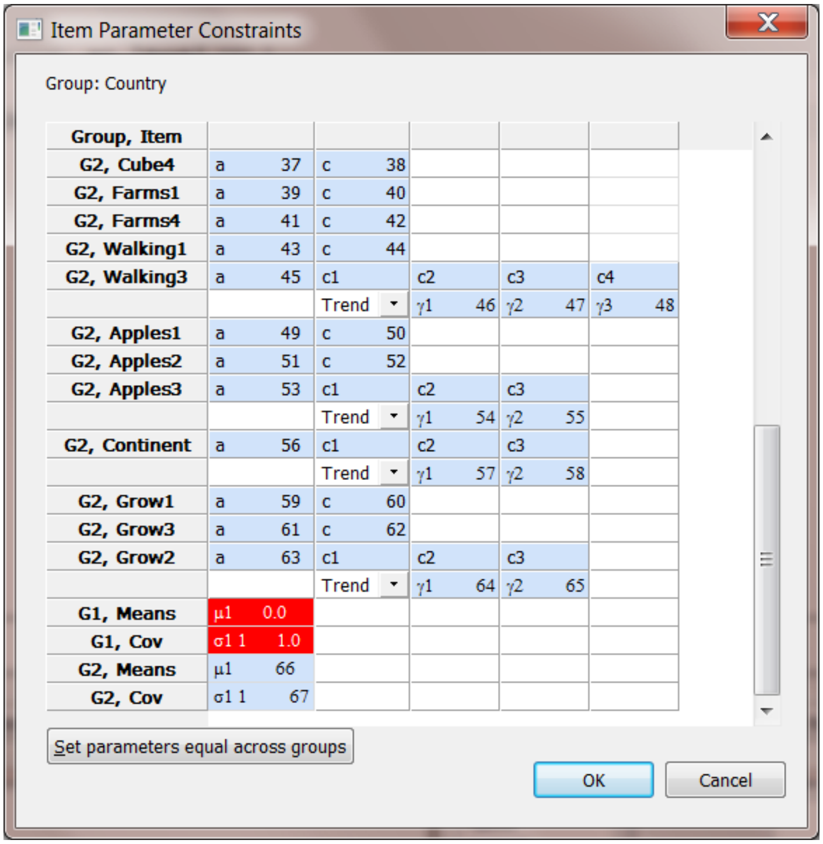


The user may also change the default model type. This is accomplished by selecting cell(s) that display a similar model type that needs to be changed. Right-click on any of the selected cells to display a drop down list of available models and make a selection. See Section 6.2.2 for an example that illustrates this function.

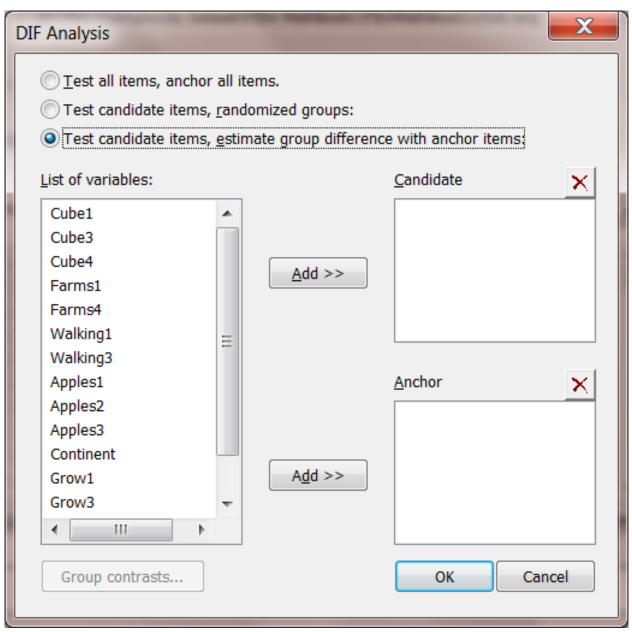


At the bottom of the **Models** dialog there are three buttons, labeled **Constraints**…, and **DIF...**, respectively. The latter button gives access to a dialog for entering parameter values or reading them from a file. Typically, these values are used to score a set of items that were previously calibrated, see Chapter 8 for more details.

By clicking on the **Constraints**… button an **Item Parameter Constraints** window is invoked. Use of this window allows the user to fix or free parameters or to set selected parameters equal. Examples to illustrate the use of the **Item Parameter Constraints** window are given in Sections 5.2.2, 5.2.3, 6.1, 6.2.1, 6.2.2, 6.2.3, and 7.3.



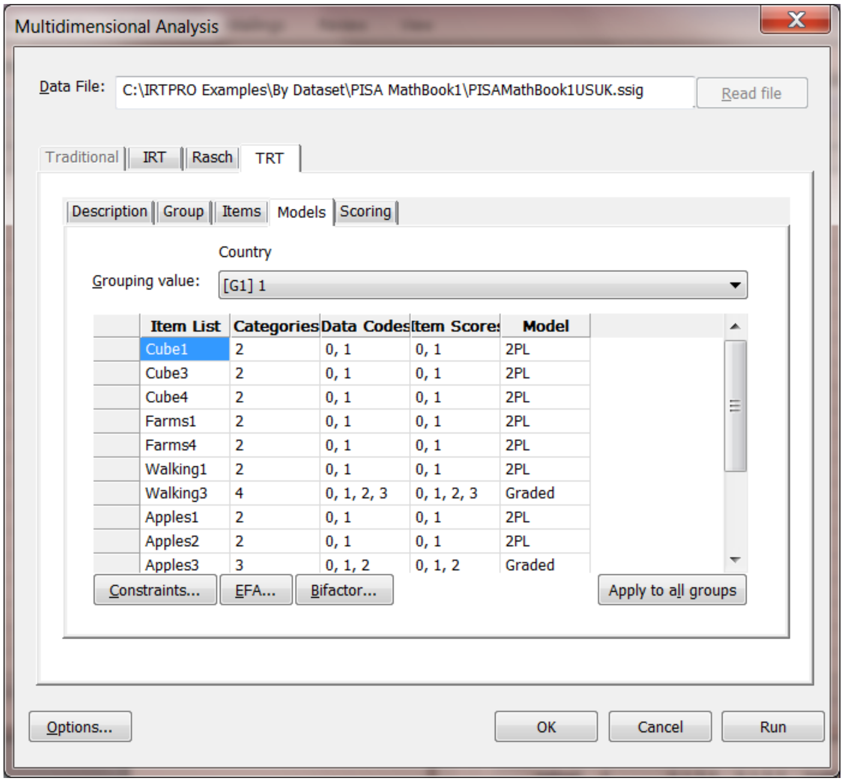
The DIF button (differential item functioning) is enabled when the analysis entails multiple groups. An example that illustrates the use of the **DIF Analysis** dialog is given in Section 6.1.



2.7.2 The Models tab, Multidimensional IRT Analysis

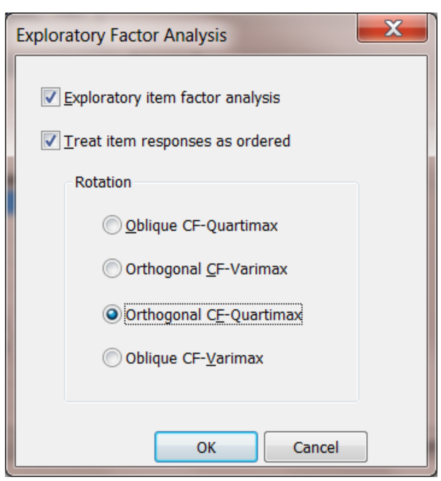
The **Models tab** for a multidimensional analysis has exactly the same functionality than that described in the previous section for the unidimensional case, except that the buttons below the

Multidimensional Analysis window are labeled **Constraints…**, **EFA...**, and **Bifactor...**, where **EFA** denotes exploratory factor analysis and **Bifactor** denotes a bifactor analysis.

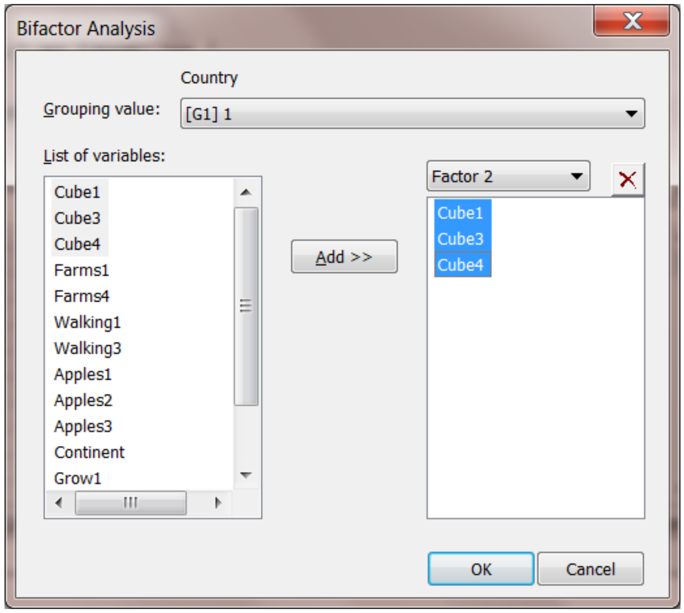


Clicking the **EFA...** button activates the **Exploratory Factor Analysis** dialog shown below. To verify that the user intends to specify EFA, the **Exploratory item factor analysis** box is checked. Additionally, a selection of one of the four available rotation methods can be made. Section 7.2.1 gives an example of an exploratory factor analysis.

The reader should note that once the **EFA...** option is selected, the **Constraints...** option is no longer available, since IRTPRO automatically sets up the constraints in this case.

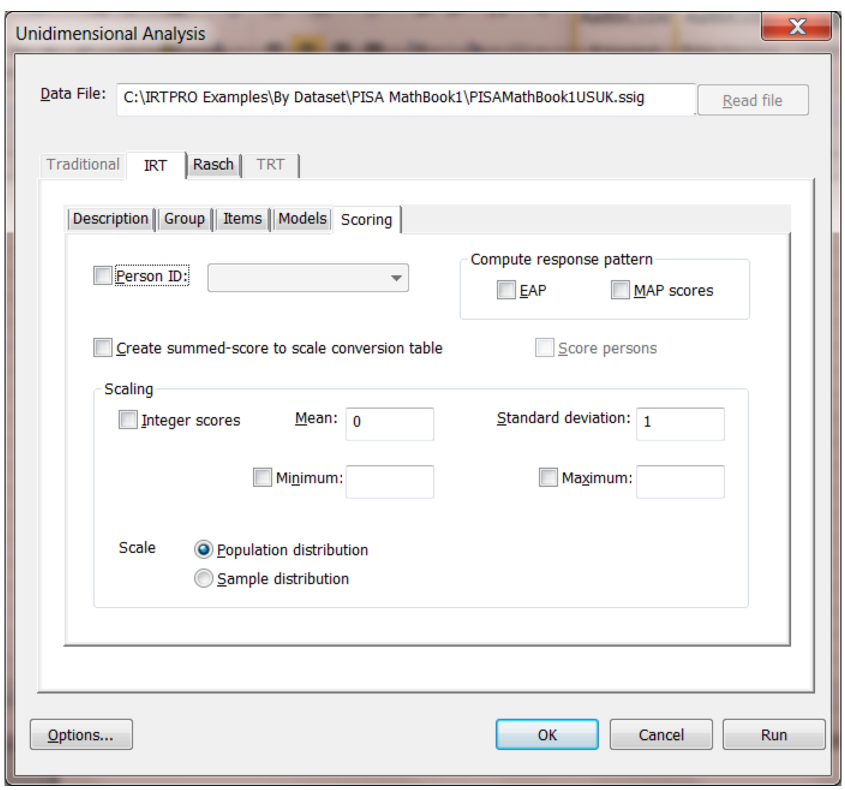


The **Bifactor…** option provides access to the **Bifactor Analysis** dialog that allows the user to select items associated with specific factors. An example of a bifactor analysis is given in Section 7.2.2.



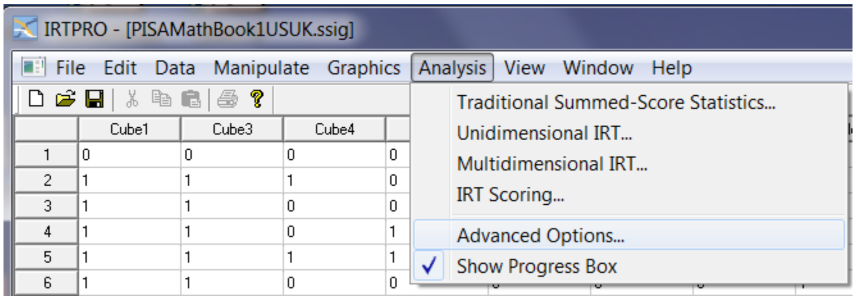
2.8 The Scoring tab

When a unidimensional IRT, or multidimensional IRT analysis is requested via the **Analysis** option, the last tab displayed in the corresponding analysis window, is the **Scoring tab**. Examples of the use of the **Scoring** dialog is given in Chapter 8.

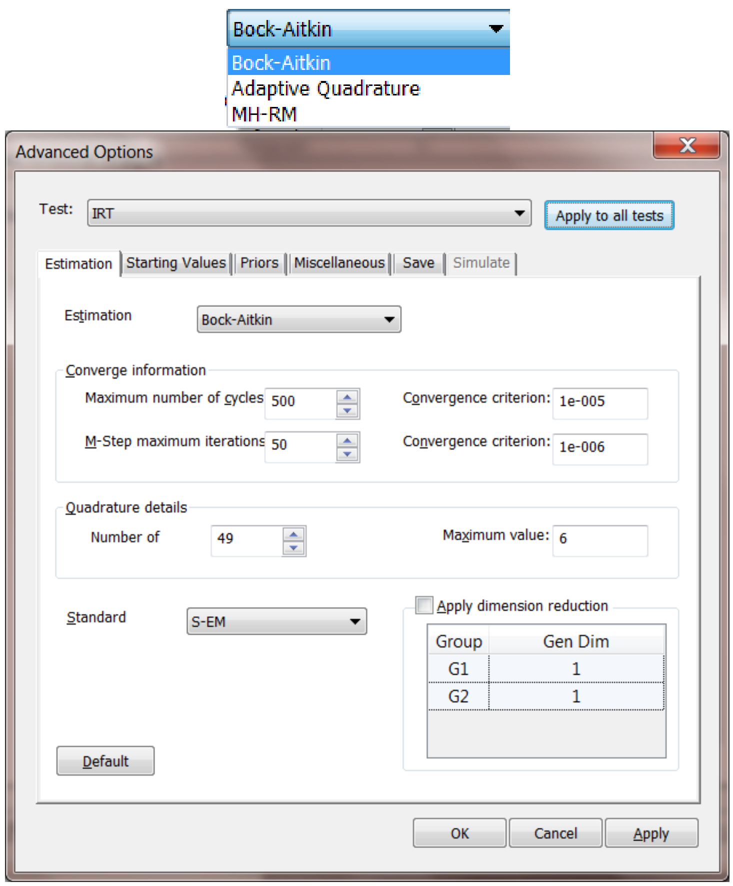


2.9 Advanced options window

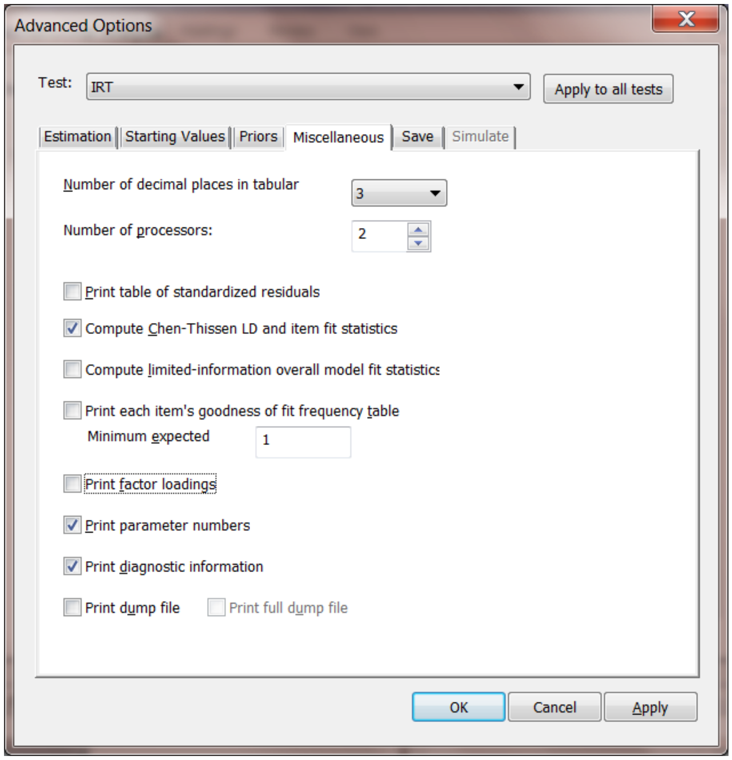
The **Advanced options** window can be accessed using the **Analysis, Advanced Options…** selection via the main-menu bar, or alternatively, by clicking the **Options…** button (lower right-hand corner of an **Analysis** window).



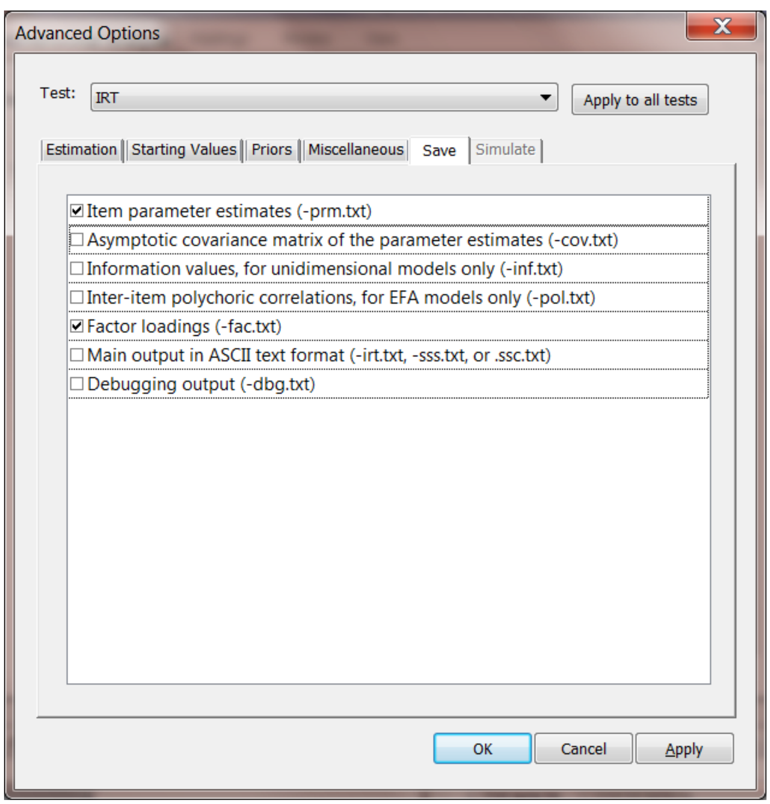
This window currently has five active tabs, these being **Estimation, Starting Values, Priors, Miscellaneous,** and **Save**. The estimation window is shown below and makes provision for three estimation methods that are described in Chapter 11:



The **Miscellaneous** dialog is used to control printout of results, and the number of processors to be used.

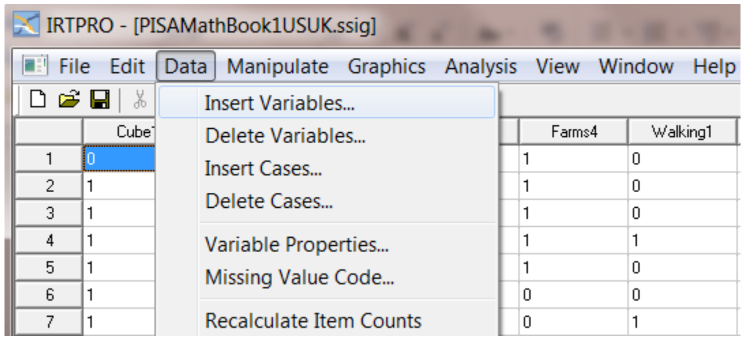


The **Save** dialog is used to request the print-out of results to specific files. See Chapter 8 for an example that illustrates the use of this dialog.



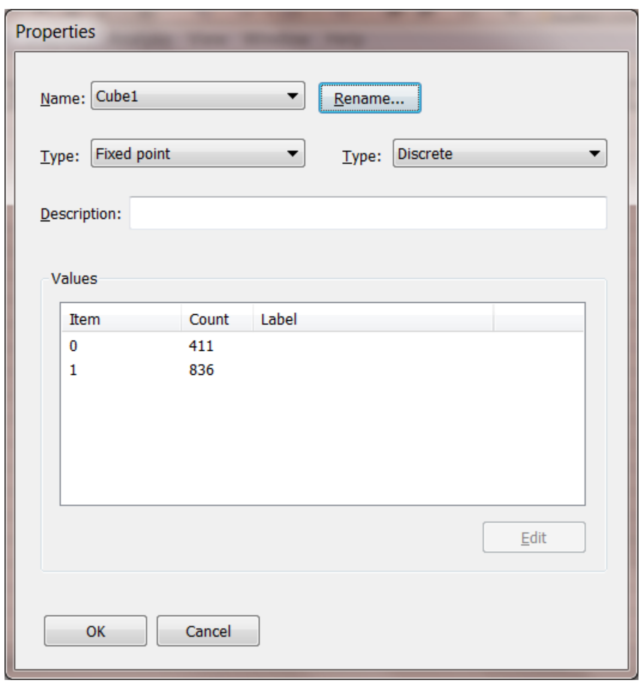
2.10 The Data menu

The **Data** option (main menu-bar) enables one to insert or delete variables and/or cases from the IRTPRO data file that is currently open. In addition, the drop-down menu makes provision for the renaming of variables (**Variable Properties…** option) and for entering a missing value code.

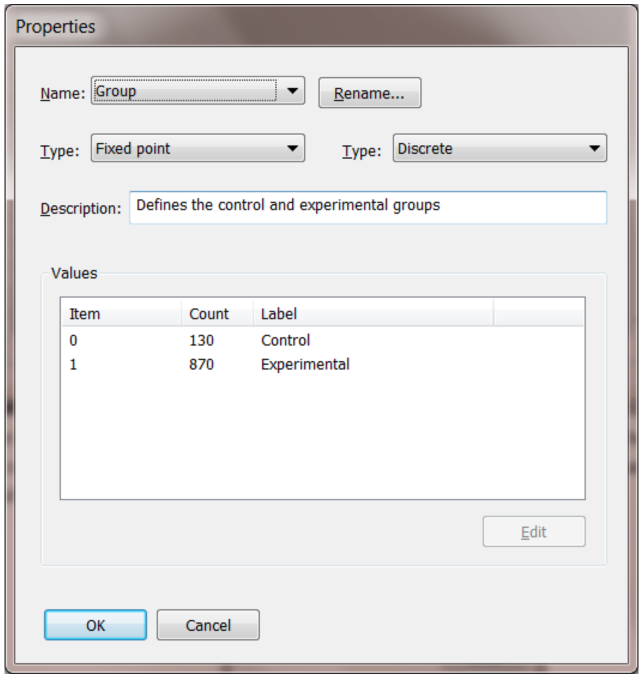


2.10.1 The Variable Properties… Option

The **Properties** dialog displays the distinct values (data codes) for each item together with the frequency counts.

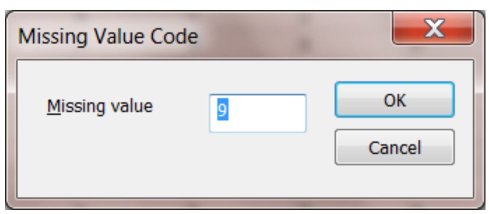


Variables may be renamed; a description of each item may be entered along with descriptive names for corresponding to the numeric values. For example 0 = Experimental, 1 = Control.



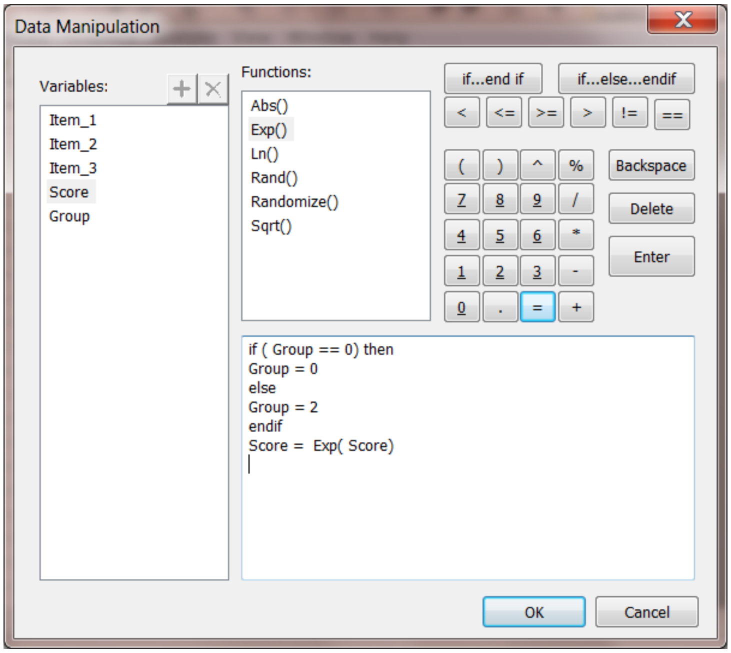
2.10.2 The Missing Value Code… Option

The **Missing Value Code…** option allows one to assign a missing value code by entering the appropriate value in the **Missing Value** text box. The value entered is accepted if the **OK** button is clicked. In this case, the user must use the **File, Save** option to ensure that this change to the dataset definitions is permanent.



2.11 The Data Manipulation window

This window makes provision for the recoding of variables. Suppose, for example, that the variable Group is coded 0, 3 and 4 and that we want to recode these values so that 0 = 1; 3 and 4 = 2. This recoding is accomplished by clicking the **if…else…endif** button. Variable names can be entered by double-clicking on a variable name, or dragging it to the appropriate position in the recode window. The last statement shows the recoding of the variable Score to Score = exp(Score).


Data Manipulation
Variables:
Item_1
Item_2
Item_3
Score
Group
Functions:
Abs()
Exp()
Ln()
Rand()
Randomize()
Sqrt()
if...end if
if...else...endif
Backspace
Delete
Enter
if (Group == 0) then
Group = 0
else
Group = 2
endif
Score = Exp(Score)
OK
Cancel

3. Data import and manipulation

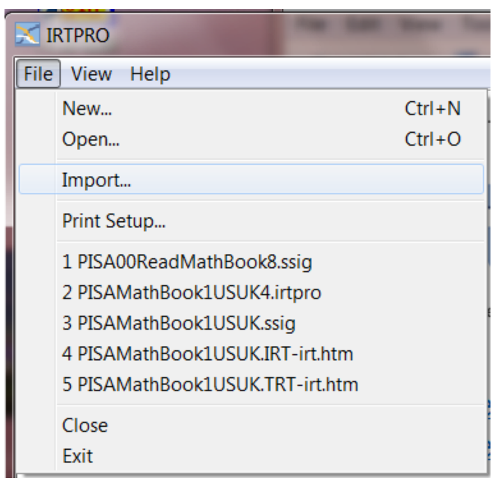
There are many ways to import data into IRTPRO for analysis. In this chapter, we briefly describe procedures for data-import from three commonly used formats: the **.sav** file format used by SPSS, fixed-format ASCII data, and comma-delimited text files.

In all cases, data are "imported" into IRTPRO (fixed-format input uses the **Open** command, but the effect is the same), and then re-saved as an IRTPRO system data (**.ssig**) file that is subsequently opened for analysis.

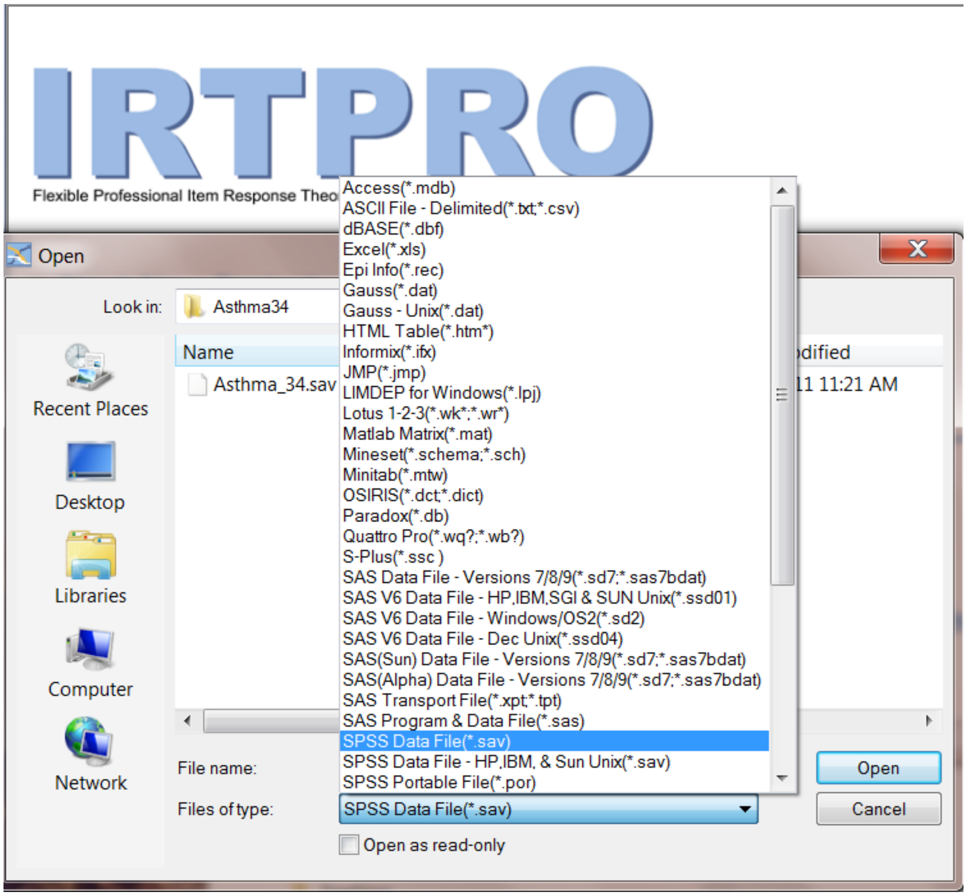
3.1 Importing Data from SPSS .sav Files

SPSS **.sav** files represent one example of many proprietary formats from which IRTPRO can import data.

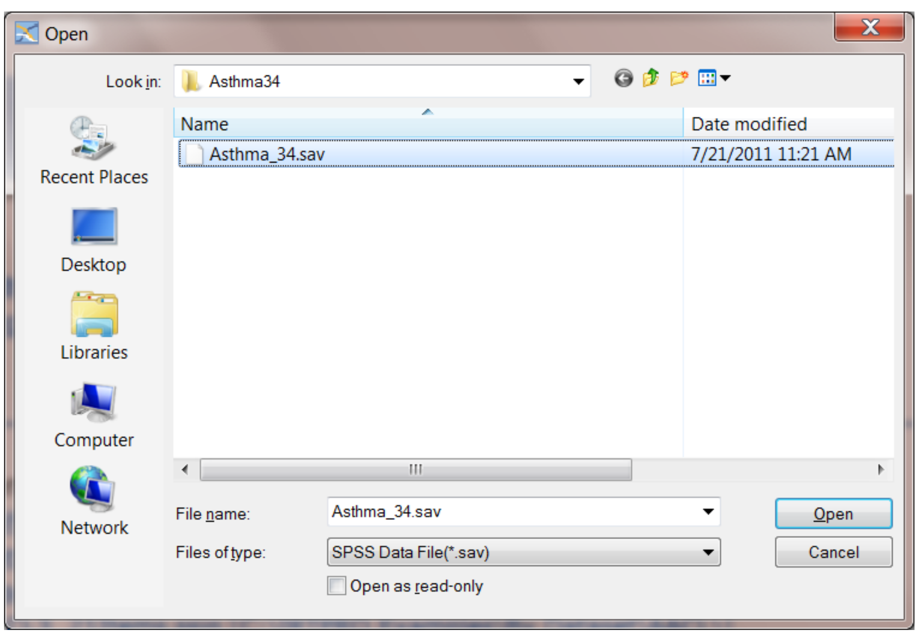
To begin the data-import process, one starts IRTPRO and selects **Import…** under the **File** menu:



This brings up a standard **Open File** dialog; in the lower center is a pop-up menu from which the user may select one of a large number of formats.

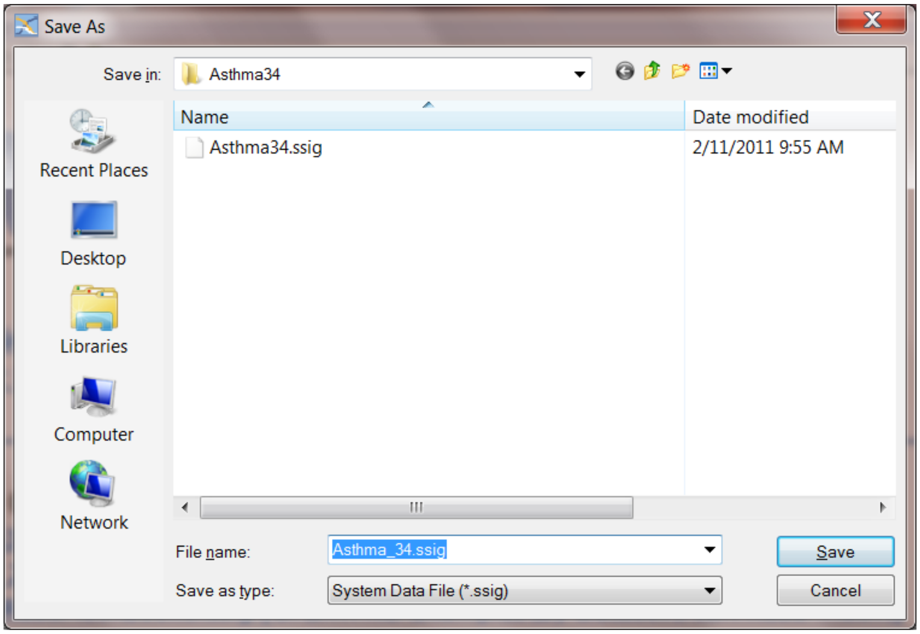


Here we select **SPSS Data File (*.sav)**; then, after navigating to the folder that contains the **.sav** file from which we wish to import data, we **Open** the file:

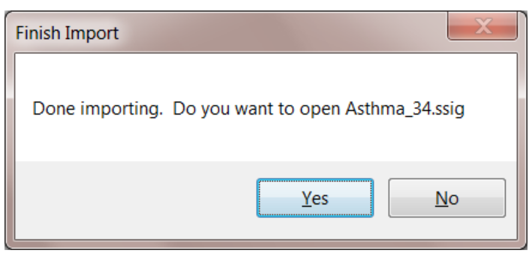


The next thing that happens is a standard **Save As** dialog appears, which has as its default to save the data as a **.ssig** file with the same name as the **.sav** file (in the case of this example, **Asthma_34**).

The user may (optionally) change the first part of the name; however, the extension should remain **.ssig**. Click the **Open** button to start the data import process.



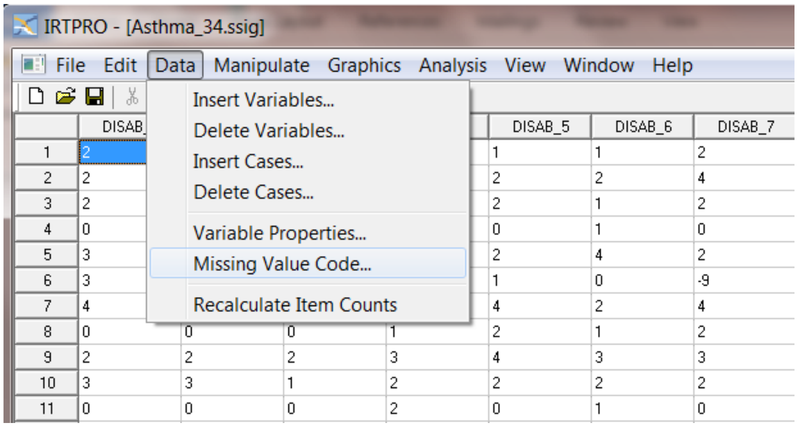
After one clicks **Save** the file is saved as a **.ssig** file, and the user has the opportunity to **Open** it to begin the analysis:



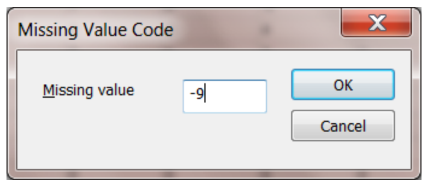
If one clicks **Yes,** the file opens.

A very important first thing to do, the first time a new **.ssig** file is opened, is to enter missing the code that represents missing data in the dataset. This code must be numeric; there can only be a single missing data code, common to all of the items; and the missing data code cannot also be a valid item response code for any item. For data coded 0, 1, 2, 3, …, it is common to use -9 as the missing value code. Note that the default missing value code in IRTPRO is -1.

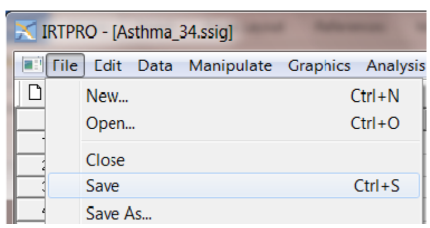
To set the missing value code, select the **Missing Value Code…** entry under the **Data** menu:



That brings up a **Missing Value Code** dialog into which the user may enter the code, and click on **OK**.



After that is done, it is important to **Save** the **.ssig** file:

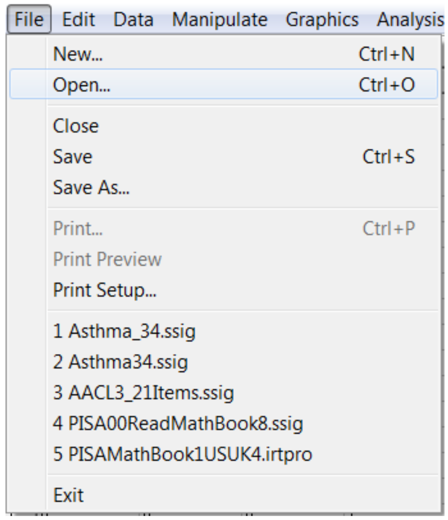


Once the missing value code has been set, and the **.ssig** file has been saved, the missing data code will be stored within the **.ssig** file and IRTPRO will "remember" the code in subsequent uses of the data.

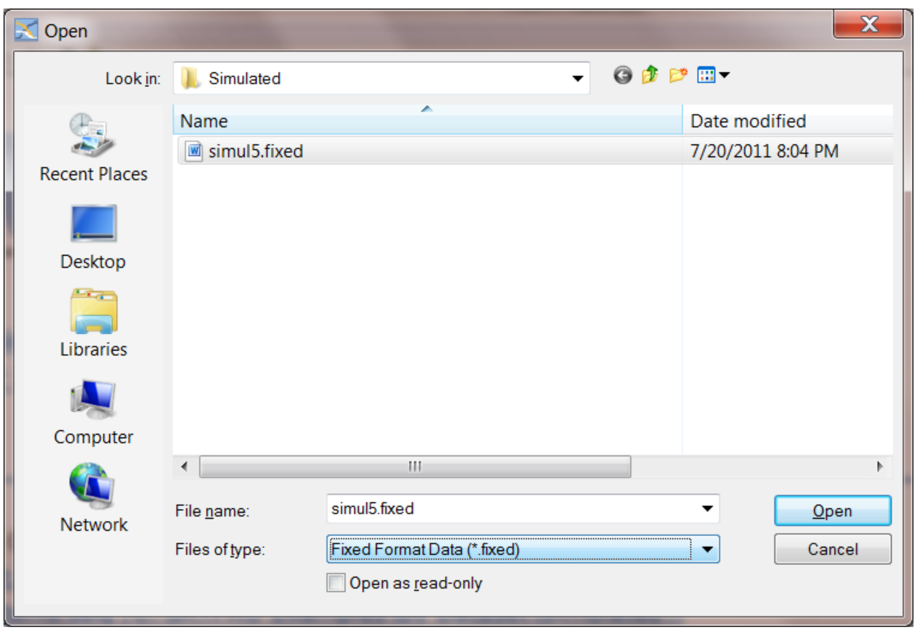
The file is now ready, and the user may proceed with analyses as described in Chapters 4 to 5.

3.2 Opening Fixed-Format Data Files

To bring in data from a fixed-format file, there is a slightly different procedure. It begins with **Open under the File** menu:

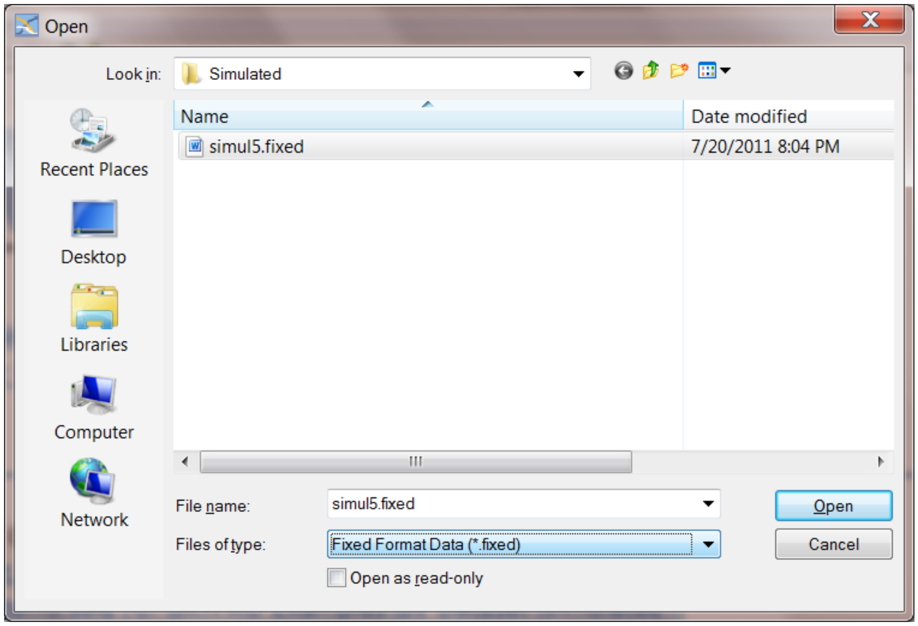


which brings up a standard **Open File** dialog. In the lower center of this dialog, the user selects **Fixed Format Data (*.fixed)** from the pop-up menu, identifying **Files of type:**

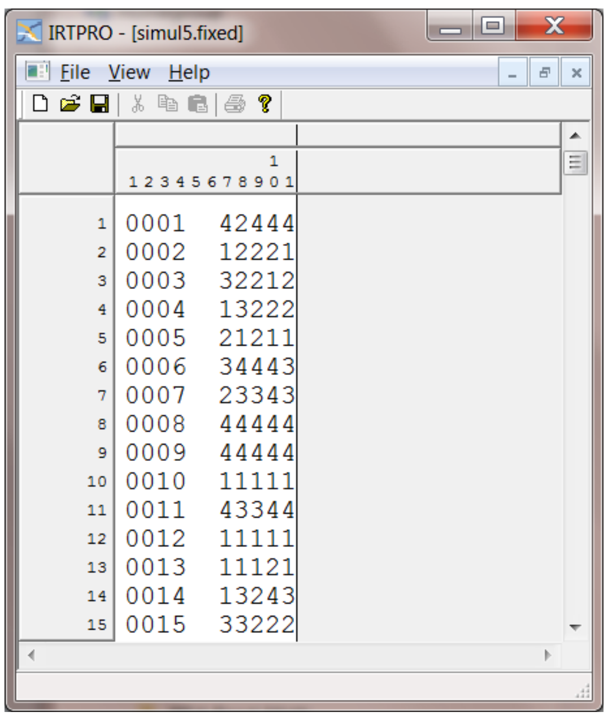


Then one opens the file; here we use as an example the file **simul5.fixed** stored in the folder **IRTPRO Examples\By Dataset\Simulated.** The data consists of five multiple category items. The simulated data represents 1000 examinees drawn at random from a population with mean ability score of 0.0 and standard deviation of 1.0.

Note that it is necessary that the fixed-format data file has the extension **.fixed**.

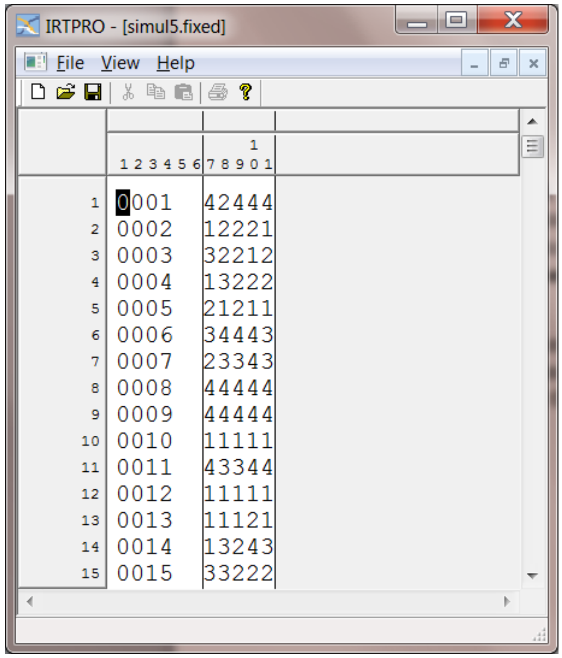


After the user clicks **Open**, an image of the file appears on the screen:

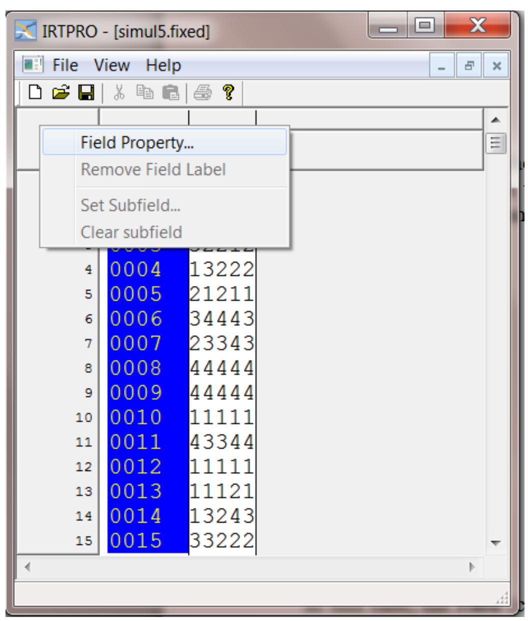


In the file **simul5.fixed**, there is a **Case Number** variable in columns 1 to 4 (its values are 0001 to 1000), and item responses for five items, each of which is in a one-column field, in columns 7 to 11. To bring those data into IRTPRO as an **.ssig** file, the user must indicate the division of the file into (sets of) columns, or **Fields**, and assign names to the variables.

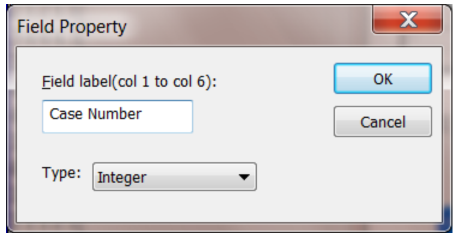
To indicate that columns 1 to 6 should be separated from columns 7 to 11, the user double-clicks between the small **6** and **7** in the gray column-header; after that is done, a vertical line appears between columns 1 to 6 and the subsequent columns:



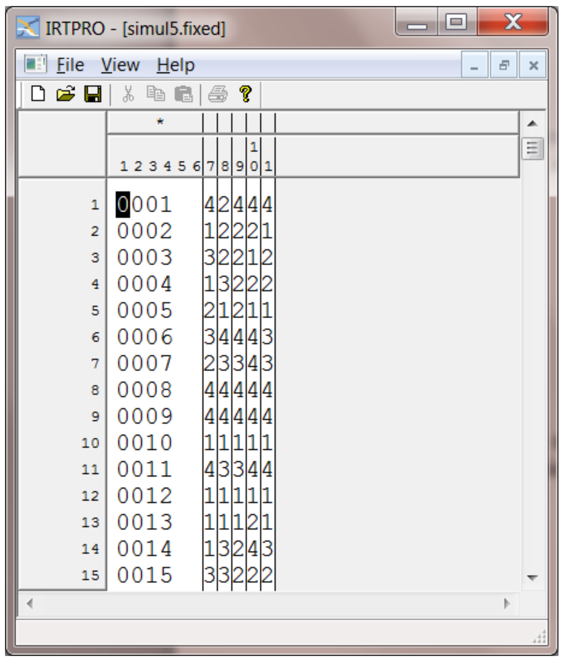
After that is accomplished, there is a small rectangular box above the column-header numbers 1 to 6. A right-click within that box brings up a menu within which the user selects the entry **Field Property …** to give a name to this **Field**:



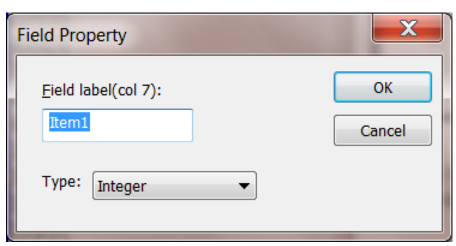
In this case, the **Field** (columns 1 to 6) contains the data for the Case Number variable, so we give it the label Case Number, and click **OK**:



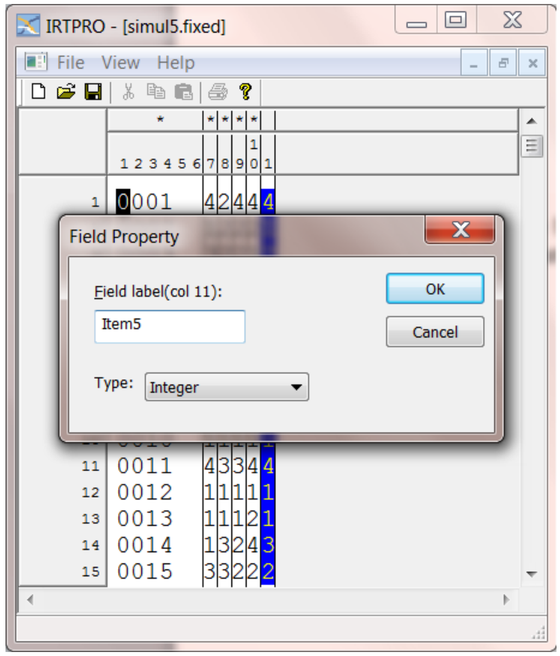
Then we move to the right, and double-click between the column headers 7 and 8; then 8 and 9 then 9 and 10; and then 10 and 11 to get the vertical separation lines shown below.



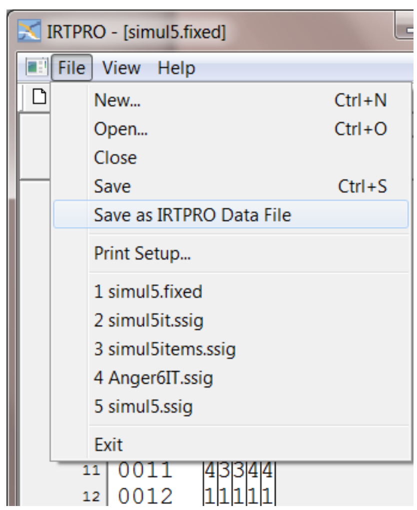
Once this is done, we right-click on the empty gray rectangle above the column heading **7**; that again brings up the **Field Property** dialog. In this case, we enter the label Item1 and click **OK**:



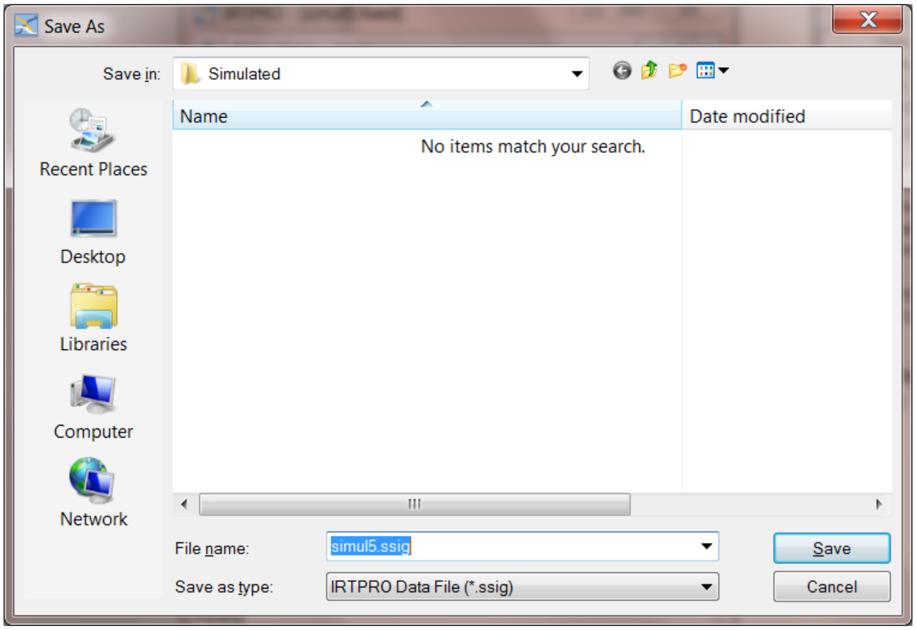
Repeat this procedure by right-clicking, in turn, on the empty gray rectangles above the column headings **8**, **9**, **10** and **11** and enter the item names Item2, Item3, Item4, and Item5 respectively.



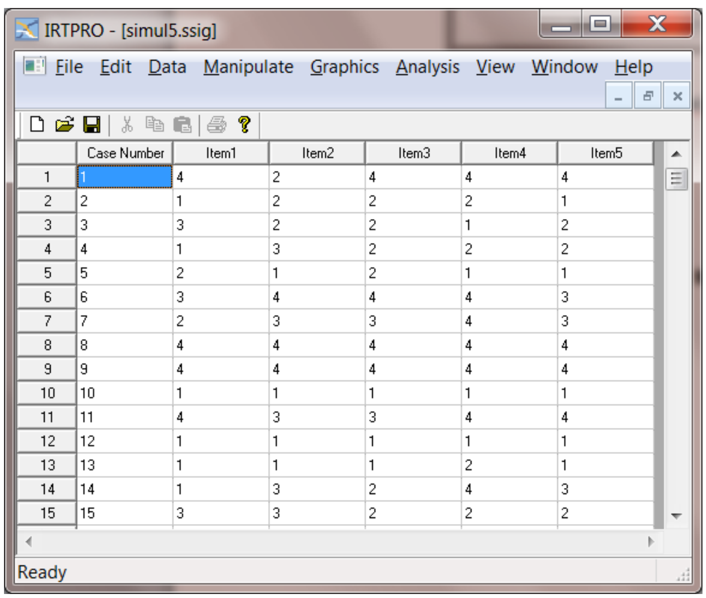
Once the **OK** button is clicked (see image above) after entering the last item name, each rectangle will be marked by an * symbol, and the **File** menu becomes active. We select **Save as IRTPRO Data File** from the **File** menu:



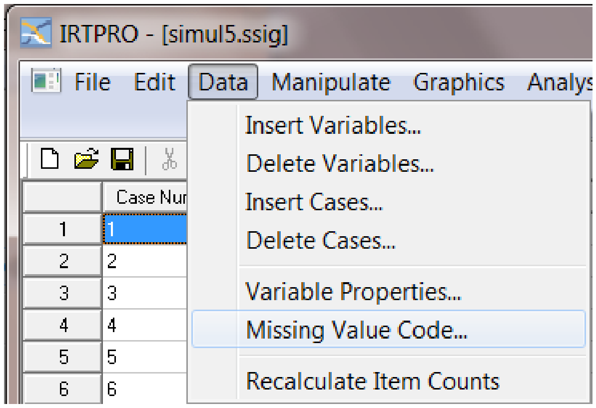
This brings up the standard **Save As** dialog, and we save the file as **Simul5.ssig** (or whatever name we might prefer, with the extension**.ssig**):



In this case, unlike **Import**, the new **.ssig** file opens immediately.

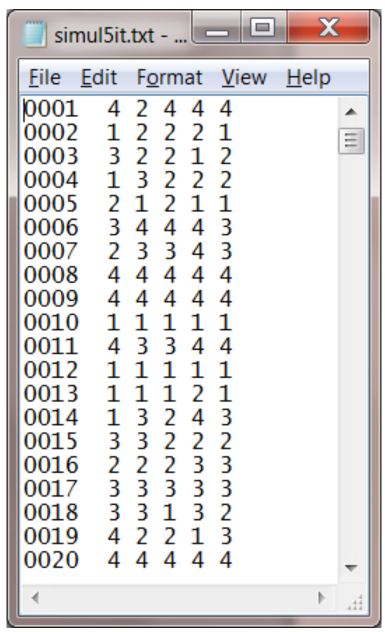


It is important to remember to set the **Missing Value Code**, if there are missing values in the data, as described in the previous section:



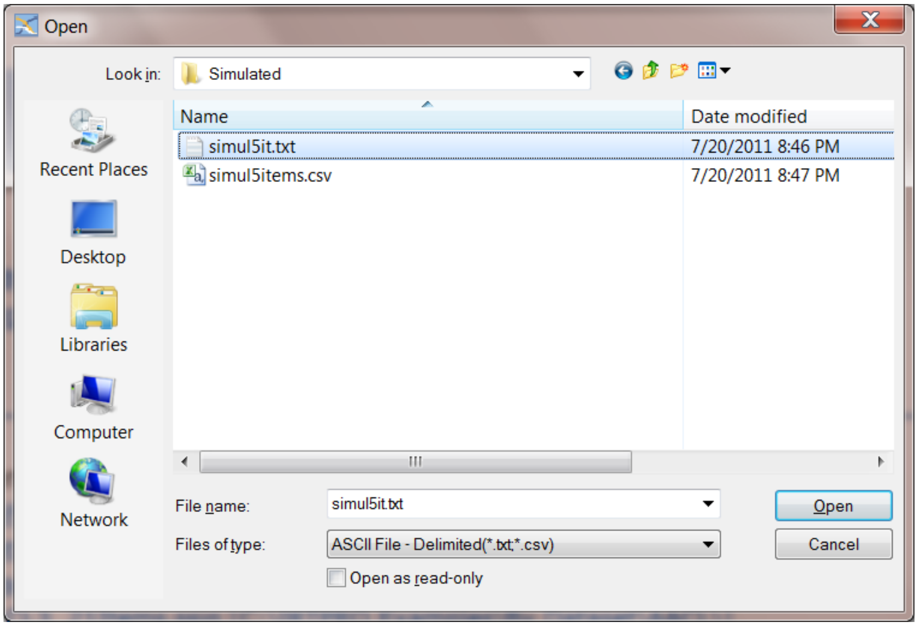
Importing space delimited fixed-format files

If there are spaces between the columns in a fixed format file, one can import the file directly if the file is saved with an extension **.txt**. As an illustration, consider the same simulated dataset used above, but in this instance saved with spaces between each variable:

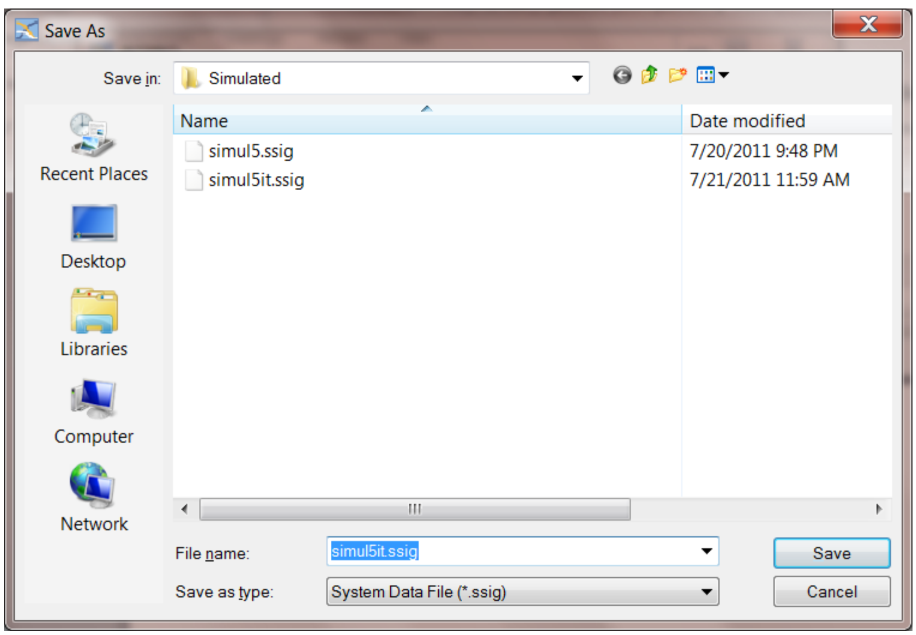



```
0001   4 2 4 4 4
0002   1 2 2 2 1
0003   3 2 2 1 2
0004   1 3 2 2 2
0005   2 1 2 1 1
0006   3 4 4 4 3
0007   2 3 3 4 3
0008   4 4 4 4 4
0009   4 4 4 4 4
0010   1 1 1 1 1
0011   4 3 3 4 4
0012   1 1 1 1 1
0013   1 1 1 2 1
0014   1 3 2 4 3
0015   3 3 2 2 2
0016   2 2 2 3 3
0017   3 3 3 3 3
0018   3 3 1 3 2
0019   4 2 2 1 3
0020   4 4 4 4 4
```

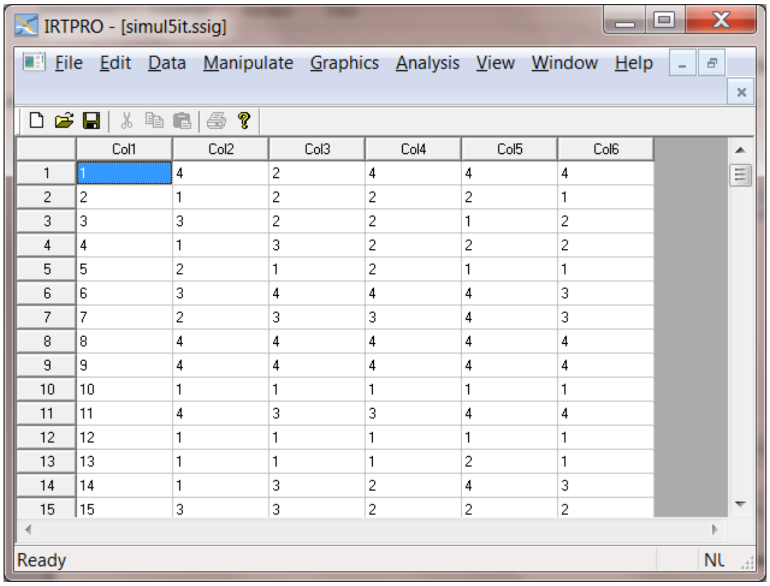
Use the **File, Import** option and select files of type (***.txt**, (***.csv**). Browse for the file **simul5it.txt** stored in the folder **IRTPRO Examples\By Dataset\Simulated** and click the **Open** button:



This action will prompt the user to save the IRTPRO data file:



A portion of this file is shown below.



The default column names are Col1, Col2,… . To rename, use the **Data, Variable Properties…** option.

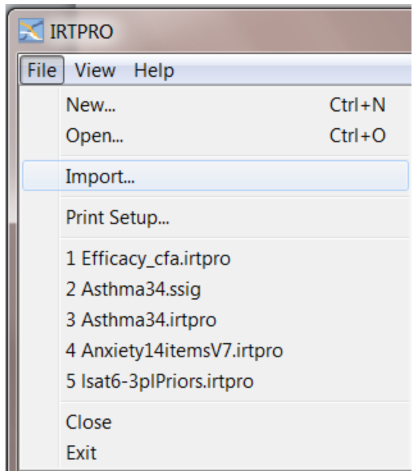
3.3 Importing Comma-delimited Data

Comma-delimited **.csv** files represent another commonly used format from which IRTPRO can import data.[1] [2]

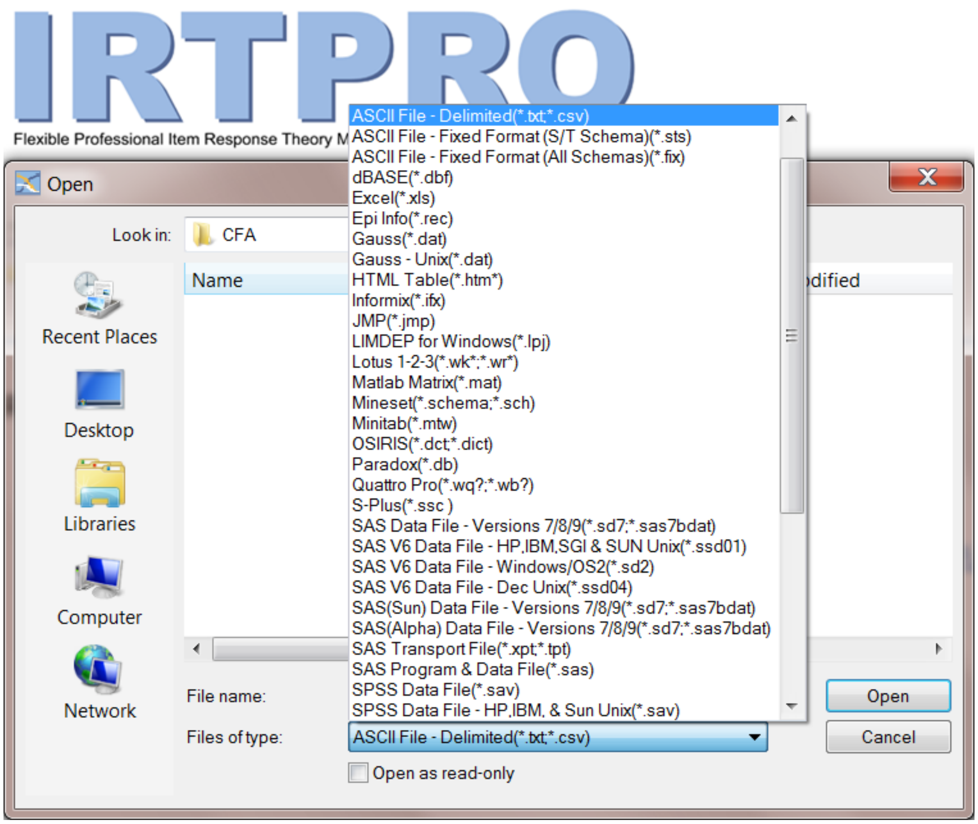
To begin the data-import process, one starts IRTPRO and selects **Import…** under the **File** menu:

1 While it is possible that IRTPRO may not properly open certain types of Excel ".xls" worksheet files, Excel will also save data as comma-delimited, and that can be used if the data are in an Excel-readable format.

2 While tab- or space-delimited data are also commonly used, IRTPRO cannot currently open those files. However, one can use a text editor to change tabs to commas, and then one has a comma-delimited file that IRTPRO can open.

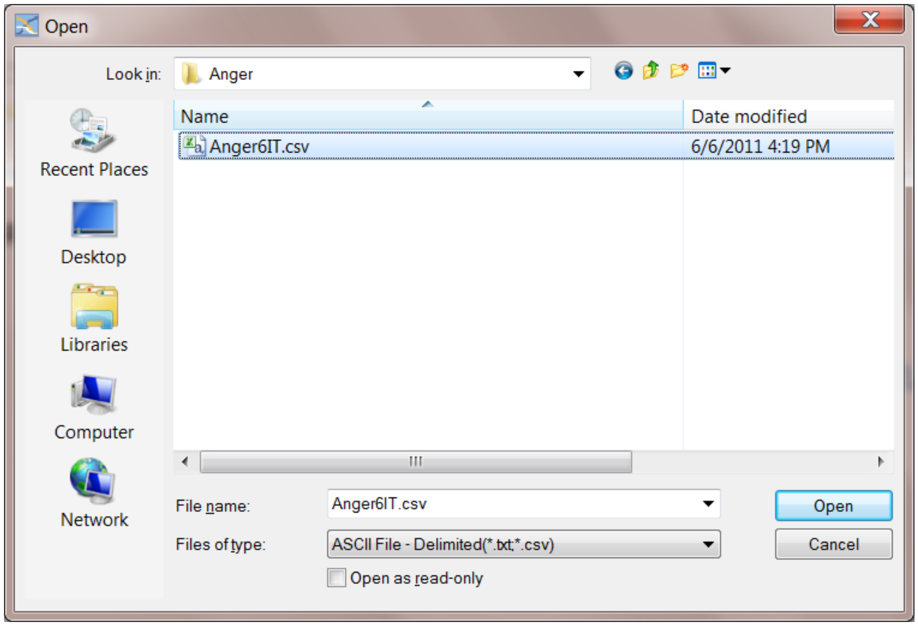


This brings up a standard **Open File** dialog; in the lower center is a pop-up menu from which the user may select one of a large number of formats:

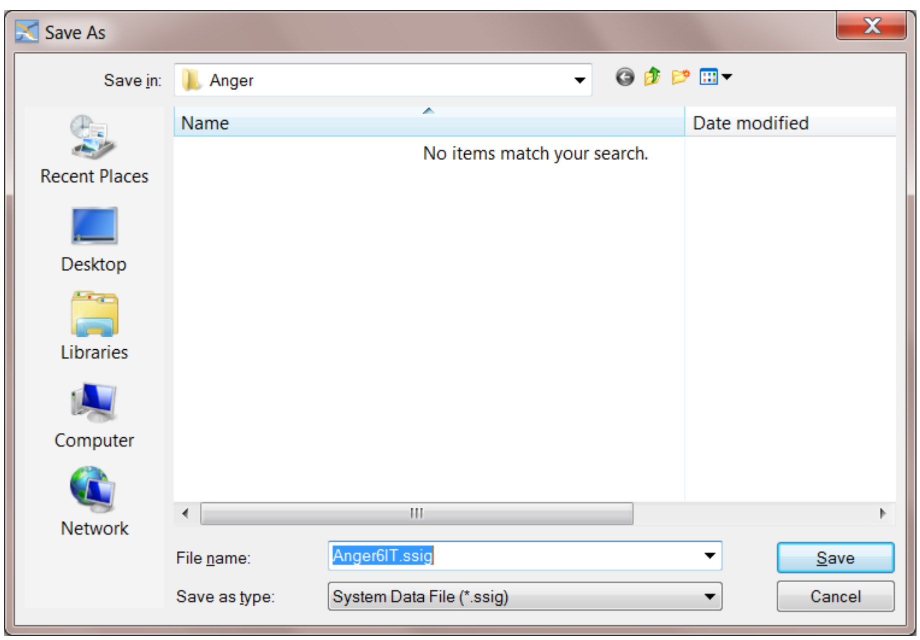


Here, we select **ASCII File – Delimited (*.txt,*.csv)**.

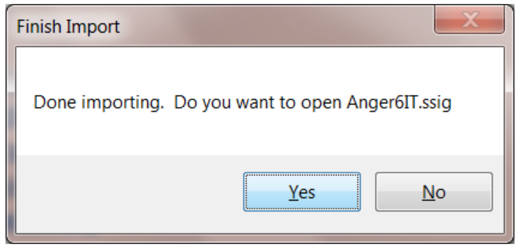
Then we navigate to the folder that contains the **.csv** file we wish to import, and **Open** it:



In this case, as an example, we are using the file **Anger6IT.csv** which contains the same data as the **Anger6IT.fixed** file used in the previous section, except that the data in **Anger6IT.csv** are comma-delimited, one line per observation, instead of in fixed columns. When we **Open** the file, a standard **Save As** dialog appears

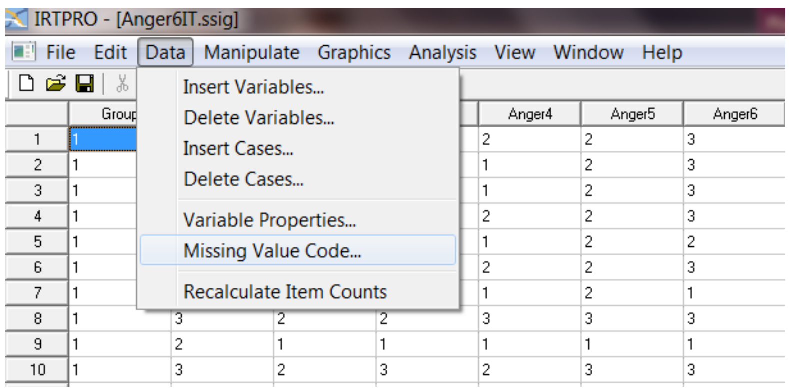


and we **Save** the file as **Anger6IT.ssig**. After one clicks **Save** the file is saved as a **.ssig** file, and the user has the opportunity to **Open** it to begin the analysis:



If one clicks **Yes,** the file opens.

It is again important to remember to set the **Missing Value Code** as described in Section 3.1.

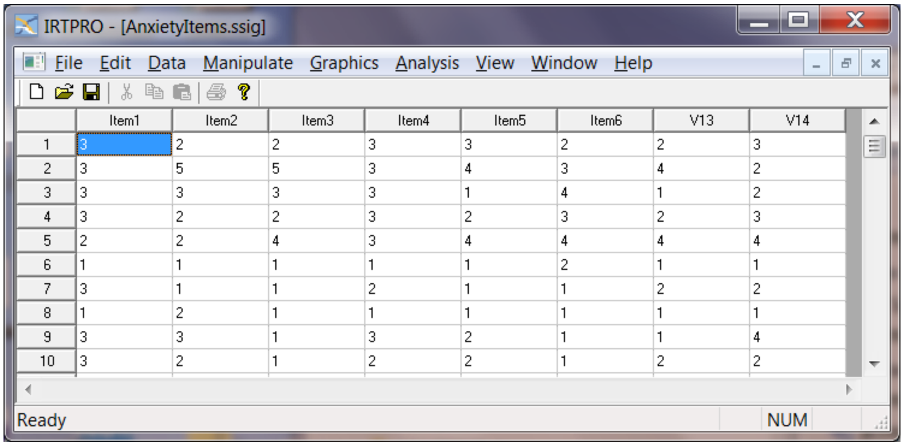


There are many other ways to "get data into" IRTPRO, but they are variations on the procedures described in this document. If you encounter difficulties opening a file of some particular format, please let us know. However, in the interim, a good work-around would be to re-write the file in one of the formats that IRTPRO *does* successfully read, and proceed from there.

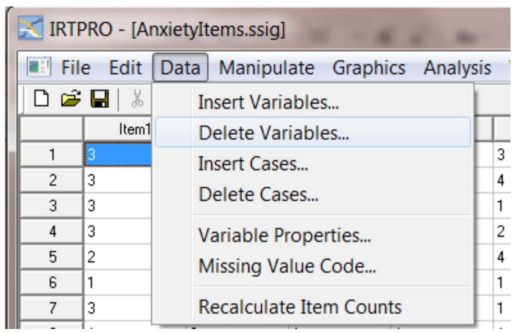
3.4 Data Manipulation: Data menu

3.4.1 Introduction

To demonstrate the data manipulation options available in IRTPRO, we use the dataset **AnxietyItems.ssig.** To see the data, use the **Open** file dialog under the **File** menu, navigate to the **C:\IRTPRO Examples\By Dataset\Anxiety14** folder, select **Files of type: IRTPRO Data File (*.ssig)** in the **Open File** dialog, and open the file **AnxietyItems.ssig**. There are eight variables and the first ten cases are shown below.

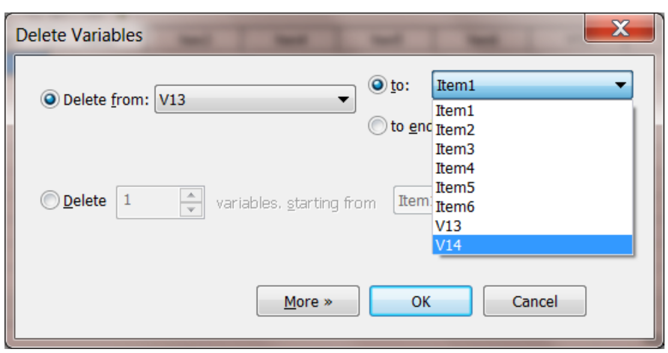


If the spreadsheet is the current window, the main menu-bar displays the **Data**, **Manipulate**, **Graphics** and **Analysis** options. The list of available options from the **Data** drop-down menu is next.

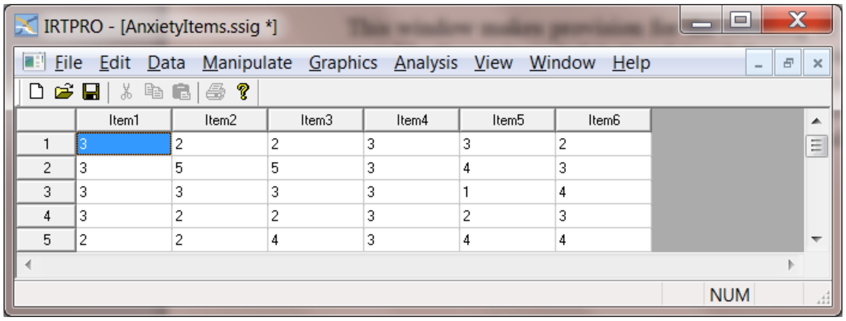


3.4.2 Delete variables or cases

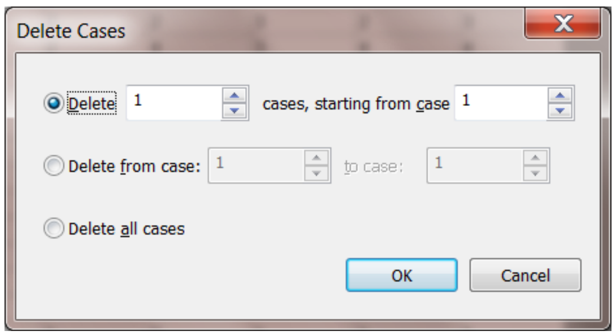
Selection of the **Data, Delete Variables…** option provides the user with access to the **Delete Variables** dialog. In the following demonstration, the variables V13 to V14 are deleted by selecting the **Delete from:** drop-down list and then the **Delete to:** drop-down list.



By clicking the **OK** button, the revised spreadsheet is displayed. These changes have not been made to the original data yet and therefore an asterisk (*) sign is appended to the file name, as shown in the top pane of the IRTPRO window. Use the **File, Save** option to make the changes permanent.

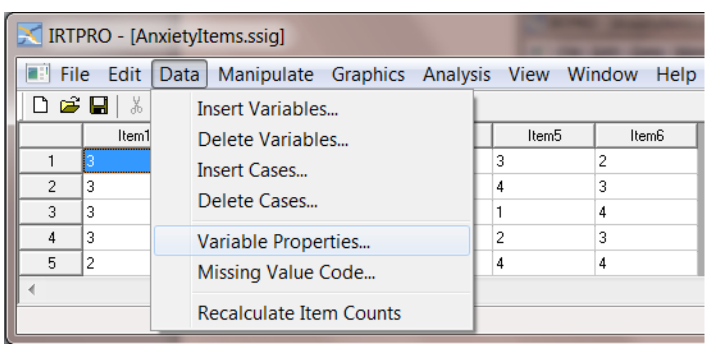


To delete cases from the data, select the **Data, Delete Cases…** option and make the required selections using the **Delete Cases** dialog.

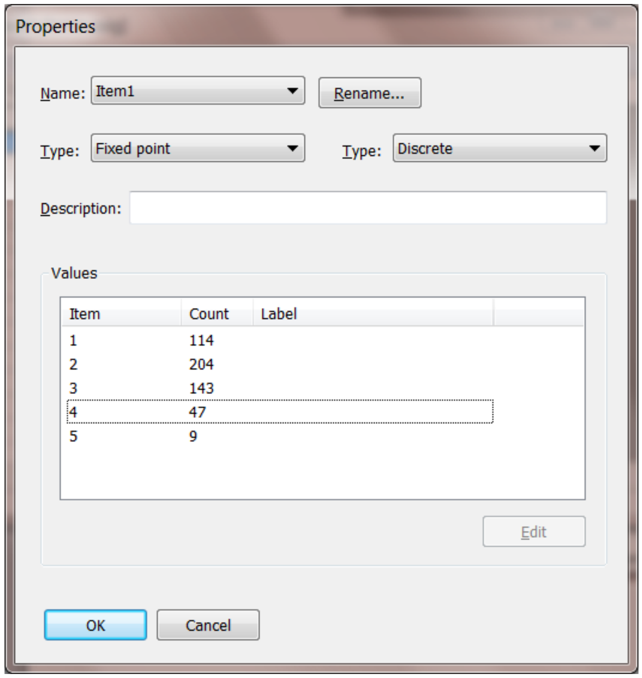


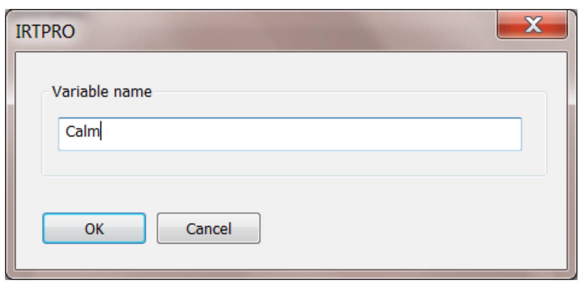
3.4.3 Renaming Variables

Next, we would like to rename the variable names Item1 to Item6. These names are to be replaced by Calm, Tense, Regretful, AtEase, Anxious, and Nervous. Select the **Variable Properties…** option from the **Data** menu to activate the **Properties** dialog.

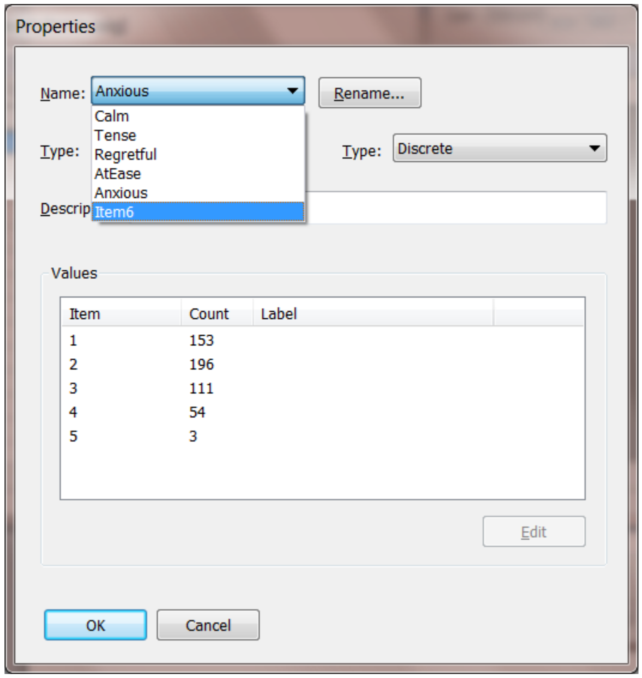


Starting with Item1 in the **Name:** drop-down list, click the **Rename…** button and change the name to Calm (see the two dialogs below).

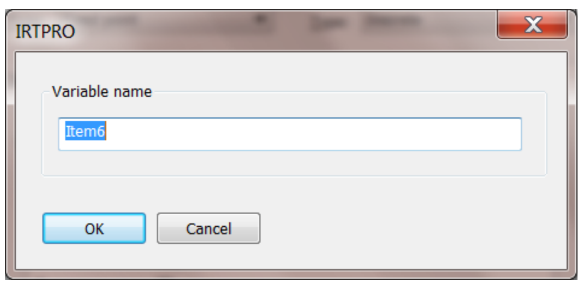




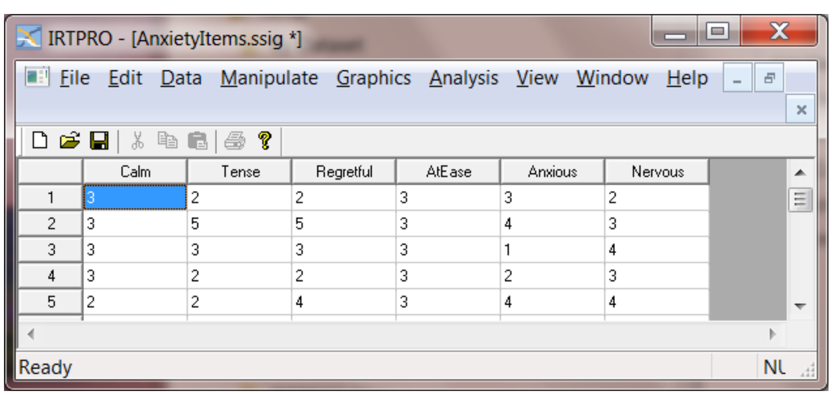
Click the **OK** button to return to the **Properties** dialog. Repeat the above procedure for Item2 to Item6.



Once the last variable has been renamed by using the **Variable name** text box, click the **OK** button to return to the **Properties** dialog.

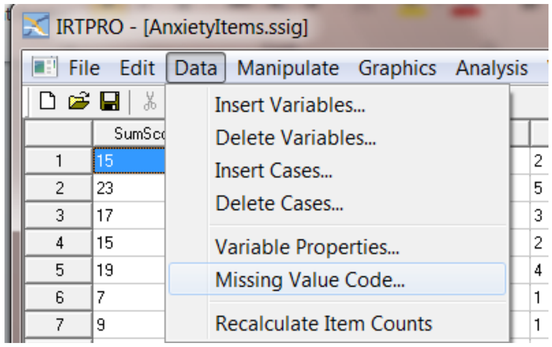


When the **Properties** dialog is displayed, use the **OK** button to display the revised spreadsheet and then use the **File, Save** option to make the changes to **AnxietyItems.ssig** permanent.

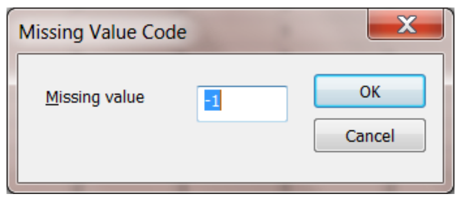


3.4.4 Missing value code

To set the missing value code, select the **Missing Value Code…** entry under the **Data** menu:



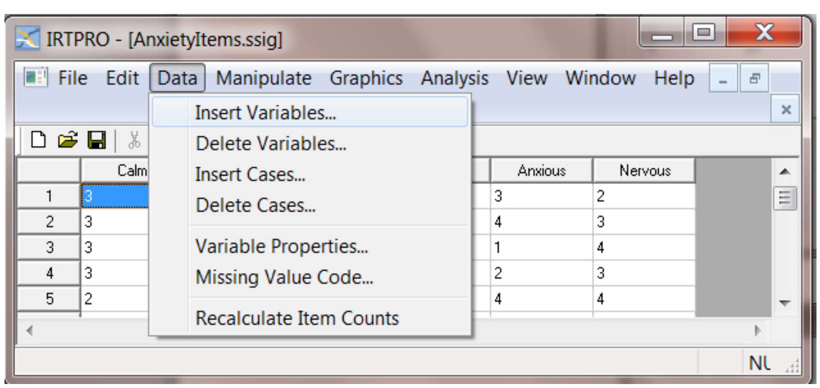
That brings up a **Missing Value Code** dialog into which the user may enter the code (-1 is the default, but is also the code for this data) and click on **OK**.



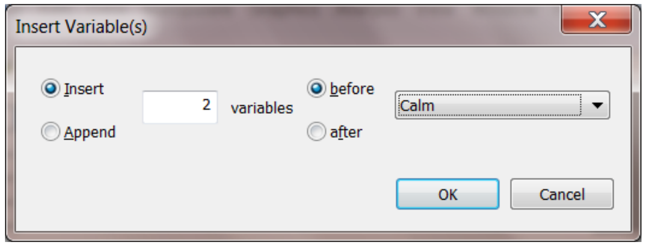
After that is done, it is important to **Save** the **.ssig** file by using the **File, Save** option. Once the missing value code has been set, and the **.ssig** file has been saved, the missing data code will be stored within the **.ssig** file and IRTPRO will "remember" the code in subsequent uses of the data.

3.4.5 Insert variables or cases

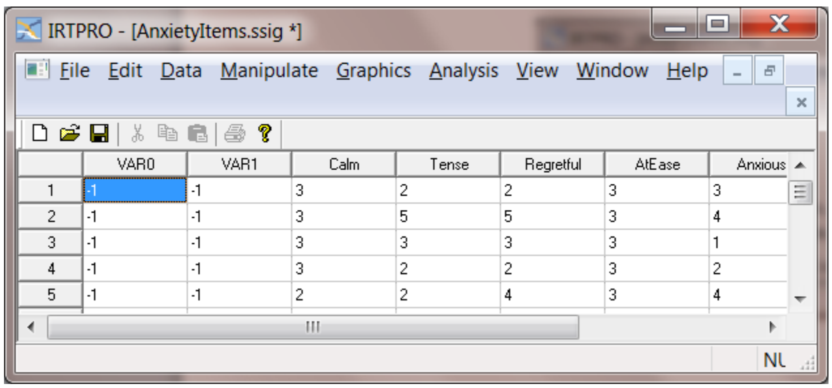
Suppose that we want to insert two new variables into **AnxietyItems.ssig** before the item Calm and then rename the new variables to SumScore and CalmRecoded. To proceed, select the **Data, Insert Variables…** option.



Selection of this option activates the **Insert Variable(s)** dialog. Make the selections shown below and click **OK**.



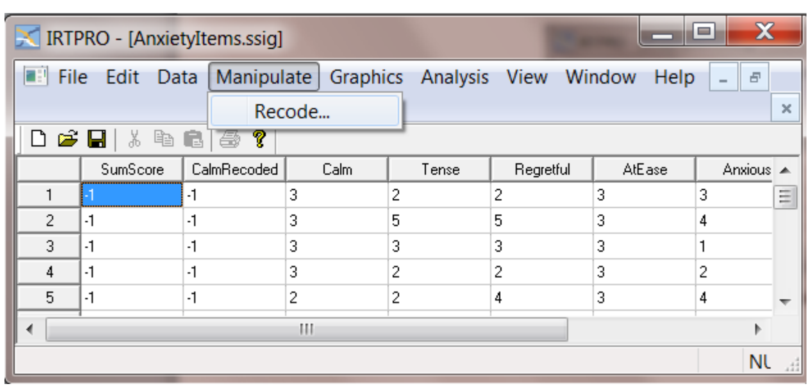
The revised spreadsheet is displayed with default variable names VAR0 and VAR1 and with all the corresponding data cells filled with the missing value code. Use the **File, Save** option to make the changes to **AnxietyItems.ssig** permanent.



Rename VAR0 to SumScore and VAR1 to CalmRecoded as explained in Section 3.4.3.

3.5 Data Manipulation: Manipulate menu

Currently, the only option available from the **Manipulate** menu, is the **Recode…** option as shown. This option is selected in what follows.

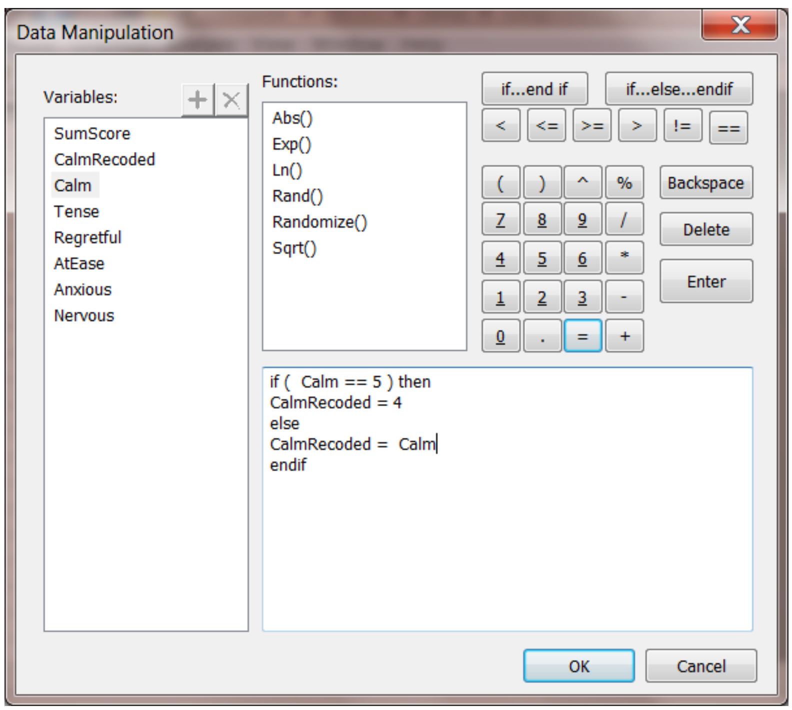


3.5.1 Recoding item scores

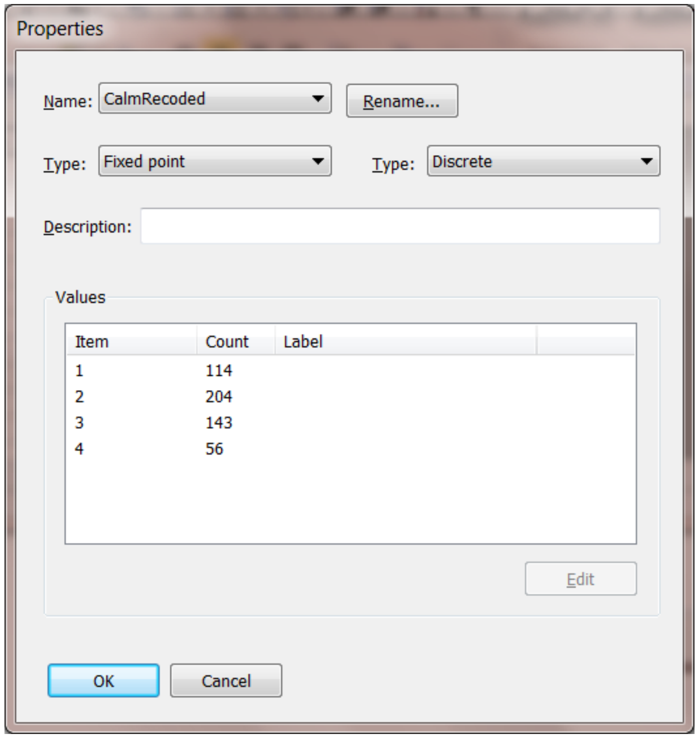
Suppose, for example, that we want to define a new variable called CalmRecoded by combining the fourth and fifth categories of the item Calm. In Section 3.4.3 the **Properties** dialog showed that the five distinct values of Calm are 1, 2, 3, 4, and 5. Therefore, we want to recode these values so that, for the new variable CalmRecoded 5 = 4 and all the remaining data values remain unchanged. This recoding is accomplished by, selection of the **Manipulate, Recode…** option to invoke the **Data Manipulation** window.

When using the if ( ) statement, follow the next rules:

1. Click with mouse pointer within the ( ) brackets, then double click on Calm or drag Calm to within the ( ) brackets.
2. Click on the appropriate operator from the following list:
    - **<** (Less than)
    - **<=** (Less than or equal to)
    - **>=** (Greater than or equal to)
    - **>** (Greater than)
    - **!=** (Not equal to)
    - **==** (Equal to; must be to = symbols)

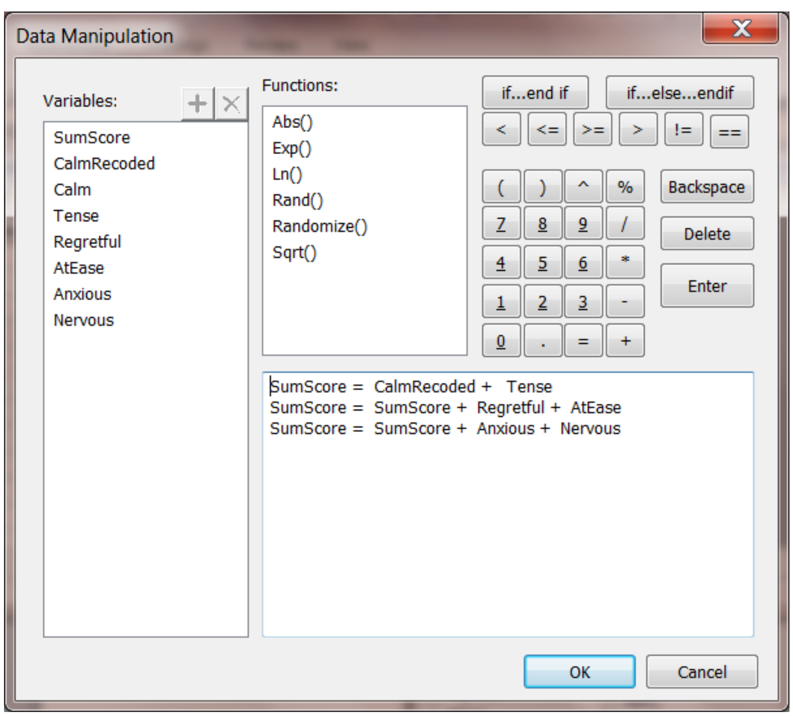


Click **OK**, then save the data file and select **Data, Properties…** from the main menu-bar to verify that CalmRecoded has four categories.

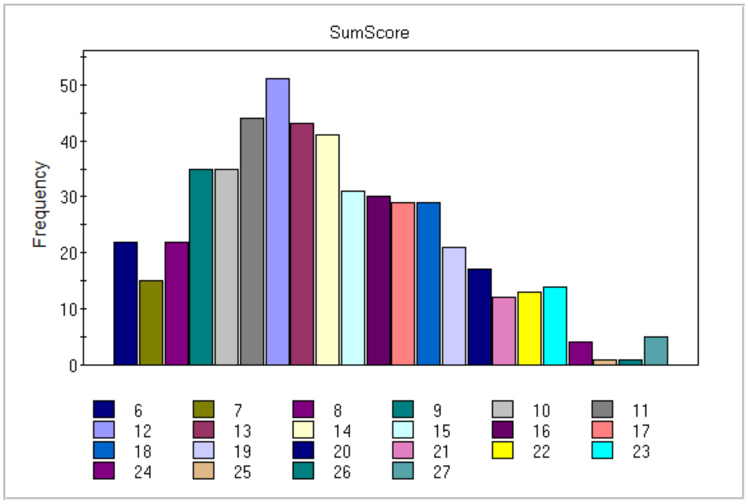


3.5.2 Calculating the sum of two or more variables

Suppose that the new variable SumScore equals the sum of the six items, CalmRecoded and Tense to Nervous. In the illustration below we used three statements. After the first statement is entered, use the **Enter** button to advance to the next line. Variables are entered onto the compute window by either double-clicking or dragging.



Click the **OK** button and use the **File,Save** option to make the changes to the file **AnxietyItems.ssig** permanent. The distribution of the SumScore values (see Section 10.2 of Chapter 10 to learn how to obtain this univariate bar chart) is shown below.



4. Traditional statistics

IRTPRO can compute a set of traditional summed-score-based statistics that are useful in checking data before an IRT analysis, and interpreting IRT results. We illustrate this feature with traditional summed-score-based statistics for 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 purposes, six items are selected:

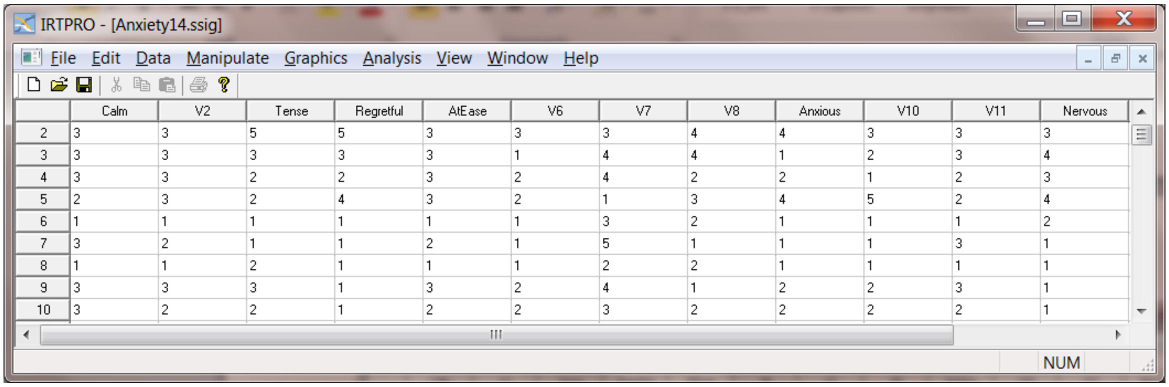
- I feel calm.
- I am tense.
- I am regretful.
- I feel at ease.
- I feel anxious.
- I feel nervous.

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.

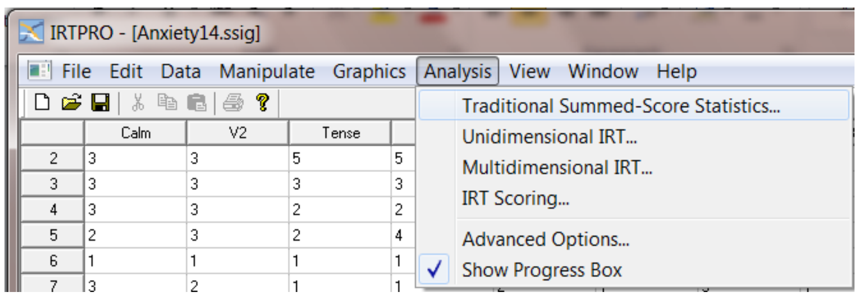
To see the data, use the **Open** file dialog under the **File** menu of IRTPRO, navigate to the **C:\IRTPRO Examples\Traditional** folder, select **Files of type: IRTPRO Data File (*.ssig)** in the **Open** file dialog, and open the file **Anxiety14.ssig**.

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. A portion of the spreadsheet is shown below.

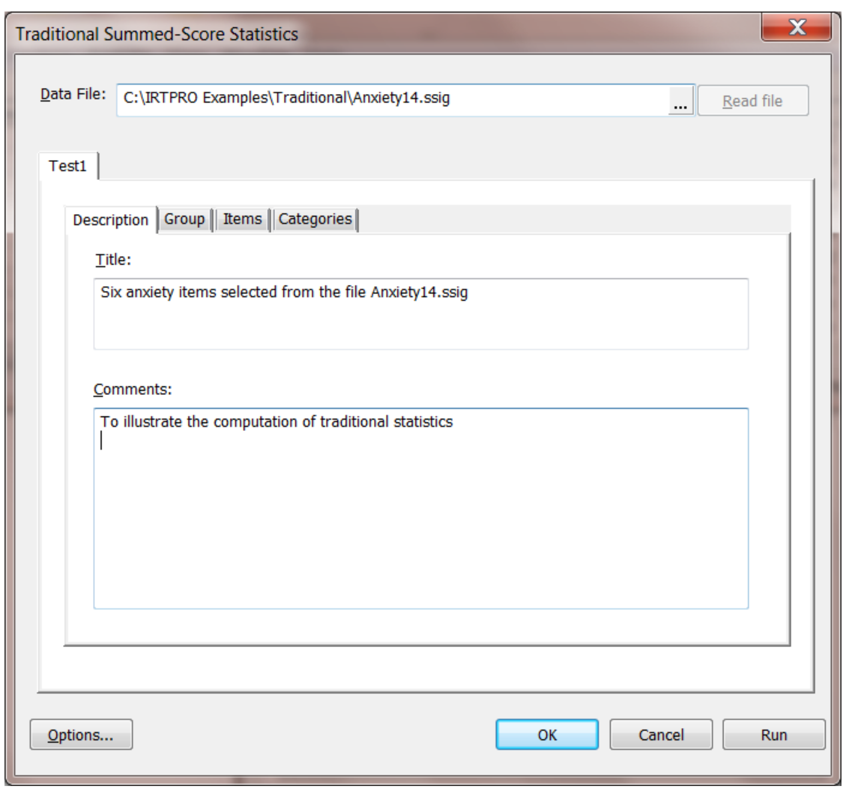
3 Thanks to Lynne Steinberg for these data, which are described more completely by Thissen & Steinberg (2009).



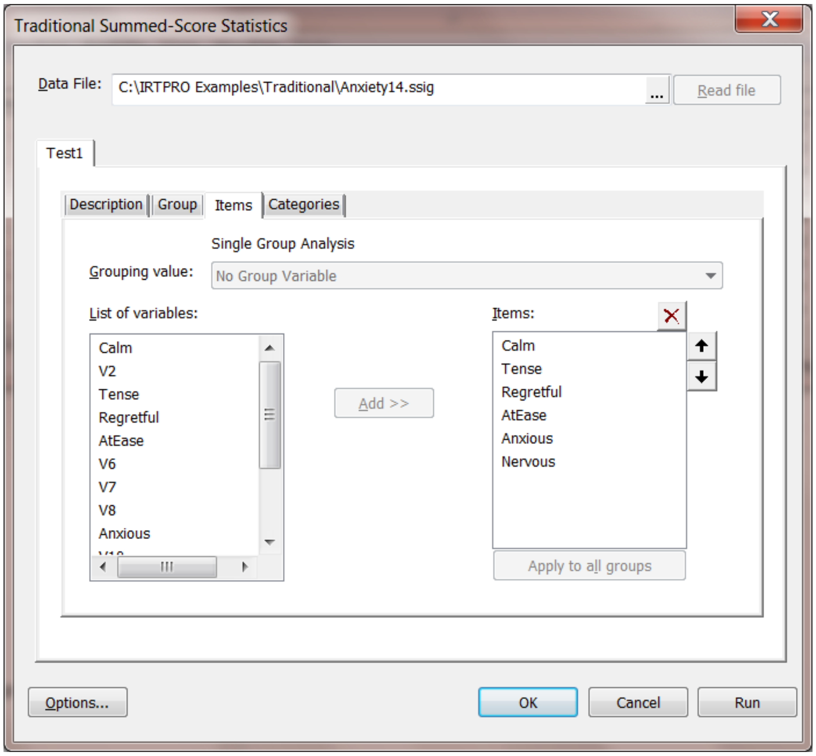
To view the statistics for these data, select **Traditional Summed-Score Statistics ...** from the **Analysis** menu.



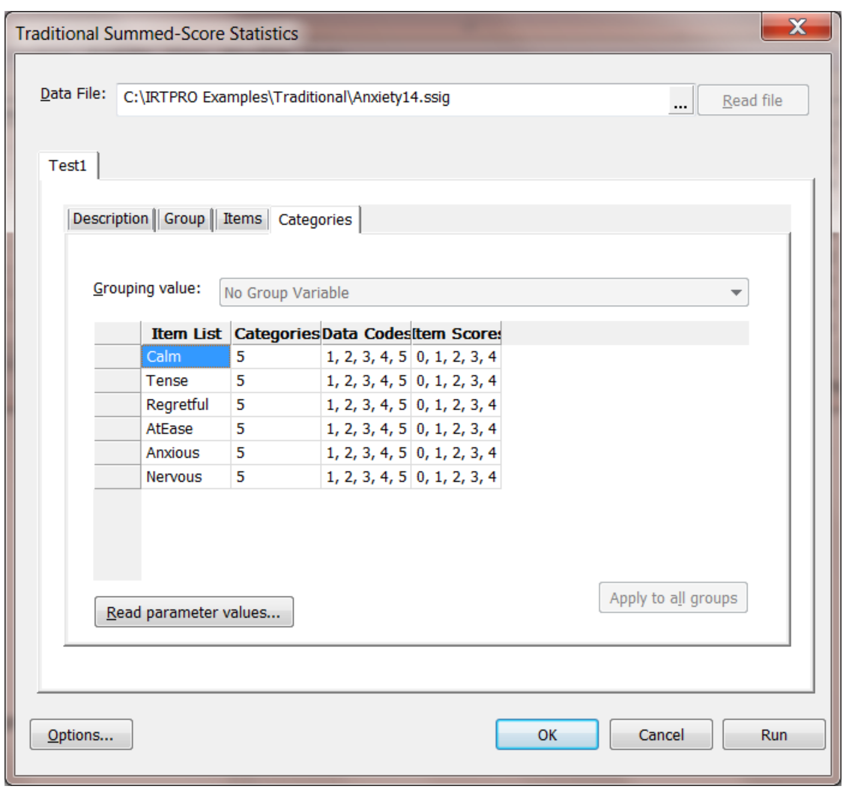
The **Traditional Statistics** dialog appears, and the user enters the title and any desired comments in the **Description** tab as shown below.



Since there is only one group, we proceed to the **Items** tab and select the six items in question:



IRTPRO computes the number of categories and associated values for each item. By clicking the **Categories** tab, these values are displayed as shown next.



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

Table of Contents

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

Coefficient alpha: 0.8425
Complete data N: 515

Coefficient Alpha, calculated using listwise deletion (if there are missing values in the data) is 0.8425 and in this case, is based on a sample size of 515 complete cases. The table below is a summary of the Coefficient Alpha, if each item in turn is deleted. For example, if item 2 is deleted, the reliability coefficient based on the remaining 5 items equals 0.8059.

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

Item	Response Average	Std. Dev.	With Item Deleted Item-Total Correlation	Coefficient 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

The tables for Item 1 to Item 6 below give the frequency count for each category of an item as well as the number of missing values for the item in question. Note that Item 4 and Item 6 have one missing value each and that this is reflected in, for example, the differences between the frequency distribution for Item 1 and the corresponding frequency distribution for listwise complete data for Item 1.

Item 1	Calm Category:	0	1	2	3	4	(Back) 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	

Each of the tables also contains a set of average scores and standard deviations. Those are the average and standard deviations for the summed score (totaled over all items) for the subsets of persons that selected each response for the item reported in a table. "Good" graded items should have higher average summed scores associated with higher-numbered responses.

Item	Tense						(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	Regretful						(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	AtEase						(Back)
4	Category:	0	1	2	3	4	Missing
Frequencies:		84	195	163	66	8	1
For listwise-complete data:							
Frequencies:		84	195	163	65	8	
Average (wtd) Score:		2.49	6.24	9.83	13.85	17.25	
Std. Dev. (wtd) Score:		2.23	2.72	3.29	3.54	5.70	

Item	Anxious						(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	Nervous						(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	

The summary of data below shows that the sample size (before list wise deletion) equals 517 and that the traditional statistics analysis is based on six items.

Summary of the Data and Control Parameters (Back to TOC)

Sample Size	517
Number of Items	6

5. Unidimensional analysis

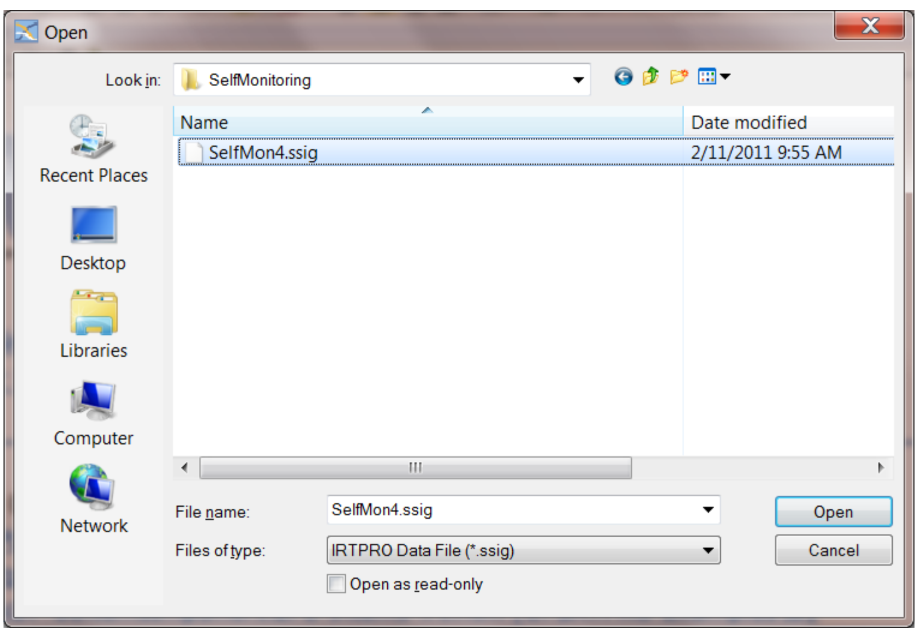
5.1 Unidimensional analysis of four "self monitoring" items

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:

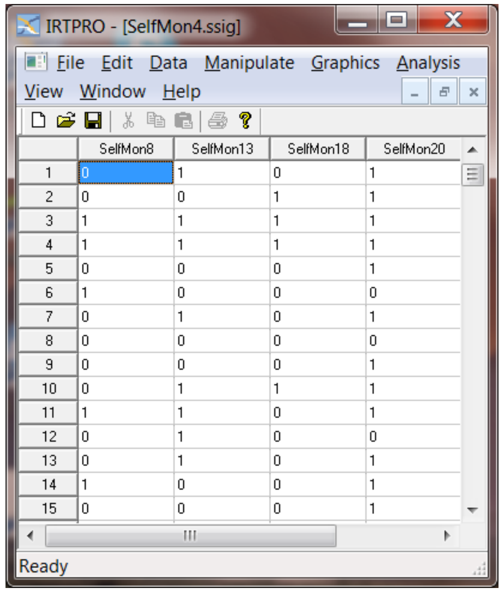
- SelfMon8: I have considered being an entertainer. (T)
- SelfMon13: I have never been good at games like charades or improvisational acting. (F)
- SelfMon18: I would probably make a good actor. (T)
- SelfMon20: In different situations and with different people, I often act like a very different person.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

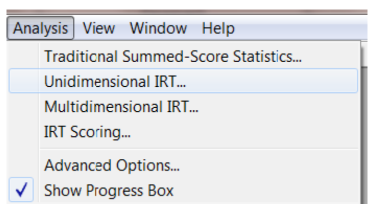
To begin, we use the **Open** file dialog under the **File** menu of IRTPRO and navigate to the **C:\IRTPRO Examples\By Dataset\SelfMonitoring** folder. However, here we change the **Files of type:** selection from its default **IRTPRO Command File (*.irtpro)** to **IRTPRO Data File (*.ssig)** in the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



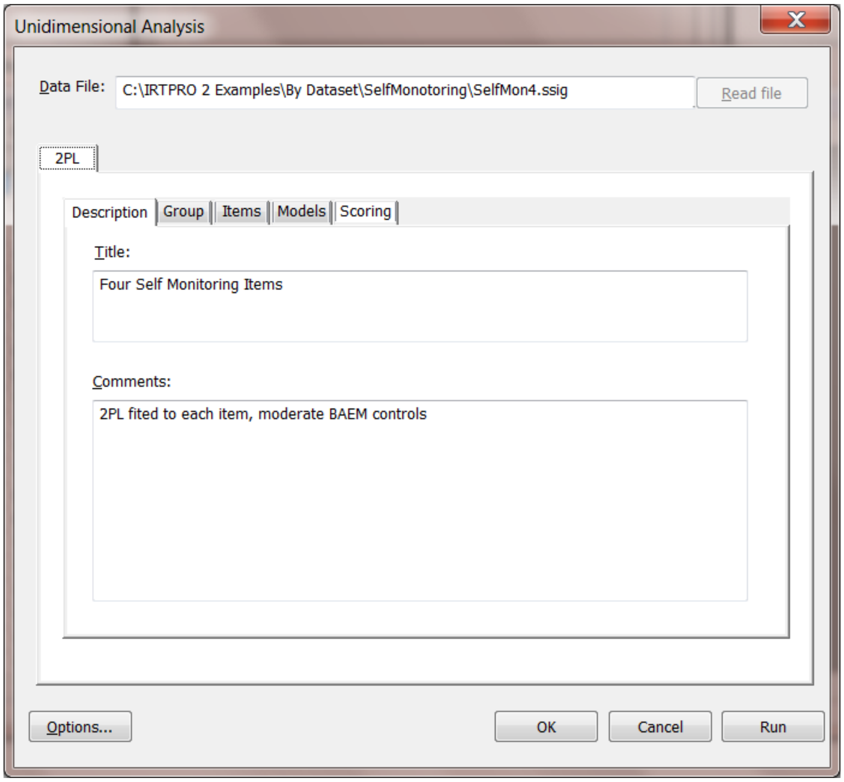
The first 15 cases of this data are shown below.



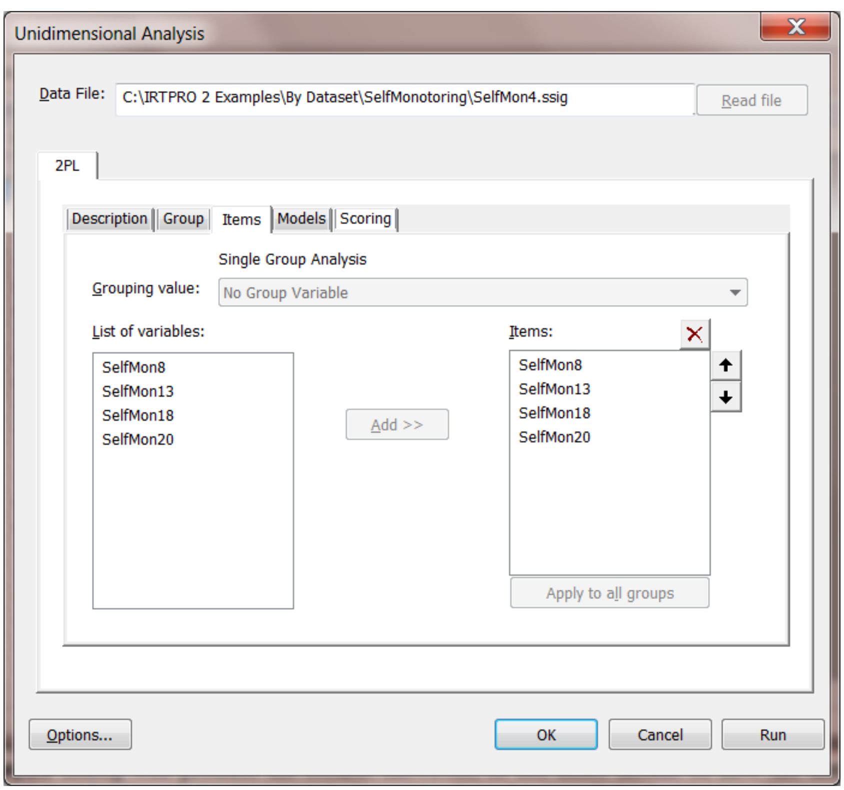
To set up the analyses, select the **Unidimensional IRT…** option from the **Analysis** menu to invoke the unidimensional analysis widow.



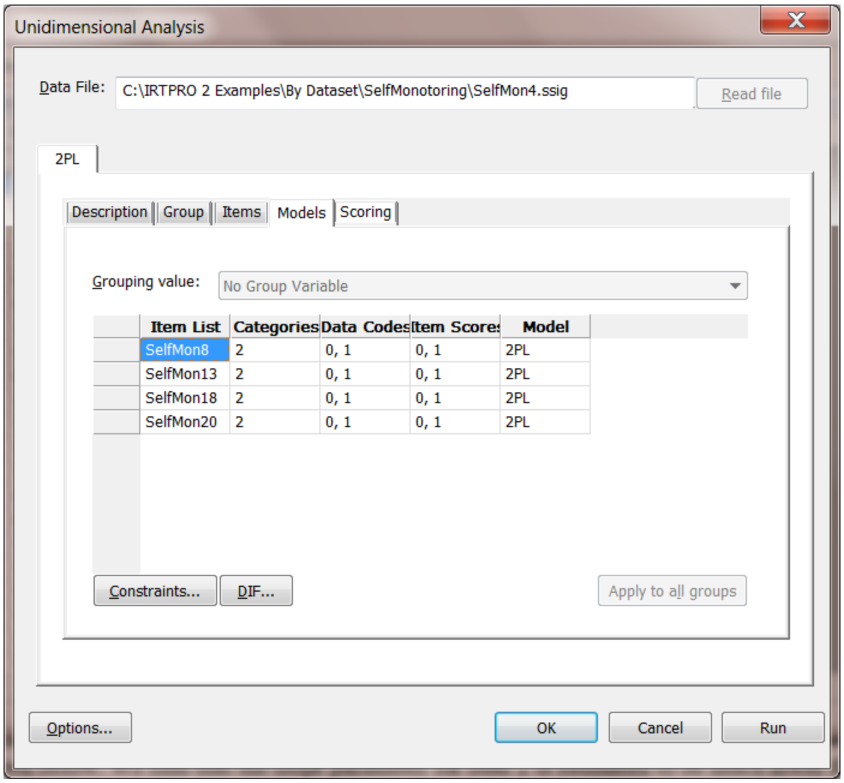
This window has five tabs called **Description**, **Group**, **Items**, **Models** and **Scoring**. Start with the default tab **Description** and provide a title and comments in the appropriate text boxes as shown below. Note that the default name for the current analysis is `Test1`. As will be shown 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

Note that the values in the table above are not exactly the same as those reported by Thissen & Steinberg (2009) for two reasons: (1) Thissen & Steinberg (2009) tabulate the slopes divided by 1.7, or in the so-called "normal metric", to compare to normal ogive slopes, and (2) even if that is corrected, there are slight numerical differences between the results, due to different numerical quadrature used for the Newton-Raphson estimation in R that Thissen & Steinberg (2009) used vs. the EM estimation used here in IRTPRO.

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.			

The statistics based on the full item classification can rarely be computed in applications IRT, because they require that the sample is sufficiently large to "fill" all of the cells of 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	44	44.72	-0.11	-1.07	0.65
0	0	0	1	25	25.07	-0.01	-0.44	0.49
0	0	1	0	4	3.55	0.24	-0.35	0.47
0	0	1	1	4	6.35	-0.94	0.04	0.43
0	1	0	0	65	64.87	0.02	-1.02	0.63
0	1	0	1	40	39.11	0.15	-0.41	0.48
0	1	1	0	6	5.60	0.17	-0.33	0.46
0	1	1	1	12	10.48	0.48	0.06	0.43
1	0	0	0	7	5.90	0.45	-0.02	0.43
1	0	0	1	25	20.10	1.12	0.36	0.46
1	0	1	0	2	3.88	-0.96	0.44	0.47
1	0	1	1	38	39.42	-0.24	1.02	0.62
1	1	0	0	9	9.67	-0.22	0.00	0.43
1	1	0	1	29	34.44	-0.97	0.39	0.46
1	1	1	0	8	6.72	0.50	0.47	0.48
1	1	1	1	75	73.10	0.25	1.07	0.63

The G^2 value (5.03, on 7 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 6 five-alternative items 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 & 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 & Thissen (2000, 2003). For the 2PL fit to these four items, these statistics are tabulated as shown in the following table.

S-X2 Item Level Diagnostic Statistics

Item	Label	X^2	d.f.	Probability
1	SelfMon8	0.51	2	0.7748
2	SelfMon13	0.23	2	0.8909
3	SelfMon18	1.86	2	0.3950
4	SelfMon20	0.23	2	0.8902

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-X2(2) = 0.5 , p = 0.7748 **(Back)**

	Category 0		Category 1	
Score	Observed	Expected	Observed	Expected
0	44	45.1	7	5.9
1	94	95.6	36	34.4
2	50	48.5	75	76.5
3	12	10.9	75	76.1

Item 2 S-X2(2) = 0.2 , p = 0.8909 (Back)

	Category 0		Category 1	
Score	Observed	Expected	Observed	Expected
0	44	44.5	65	64.5
1	36	35.3	55	55.7
2	31	29.6	49	50.4
3	38	39.6	75	73.4

Item 3 S-X2(2) = 1.9 , p = 0.3950 (Back)

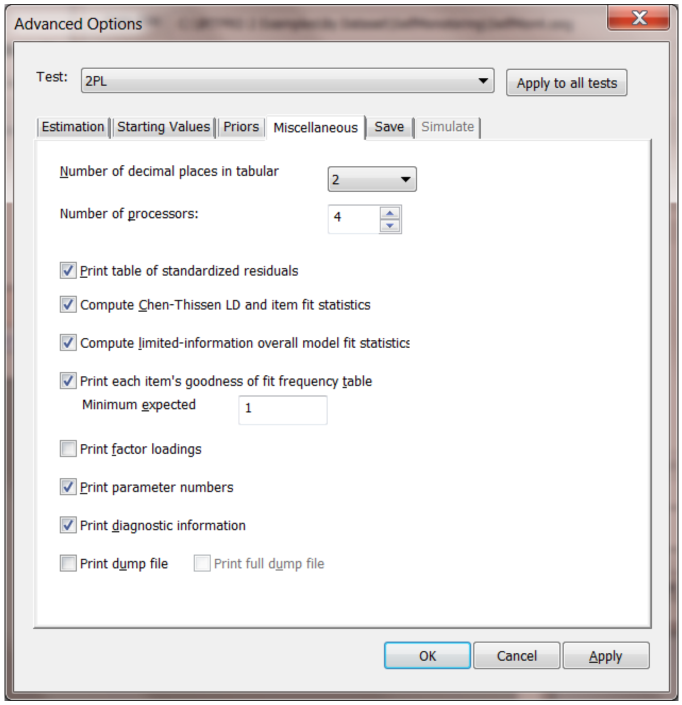
	Category 0		Category 1	
Score	Observed	Expected	Observed	Expected
0	44	44.5	4	3.5
1	97	93.5	12	15.5
2	74	72.4	58	59.6
3	29	33.3	75	70.7

Item 4 S-X2(2) = 0.2 , p = 0.8902 (Back)

	Category 0		Category 1	
Score	Observed	Expected	Observed	Expected
0	44	44.2	25	24.8
1	76	77.0	69	68.0
2	17	17.8	79	78.2
3	8	7.0	75	76.0

The tables shown 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. Printing of optional results is control by clicking the **Options** button (bottom left-hand corner of the **Unidimensional Analysis** window) and then selecting the **Advanced Options, Miscellaneous** dialog:



5.2 Hybrid model fitted to Eysenck Extraversion Scale data

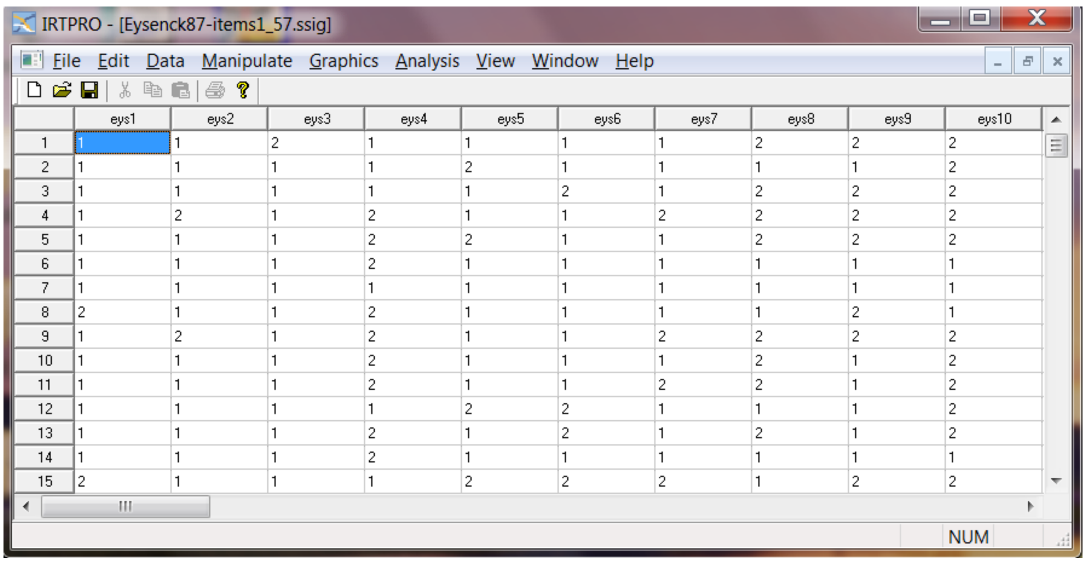
To illustrate the way the 1PL and 2PL models fit data, and their use in item analysis, we consider the responses to a subset of items from the Eysenck Personality Inventory Form A Extraversion scale (Eysenck and Eysenck, 1969). Responses of 189 undergraduate students to nine items are used to illustrate models in this section. The Extraversion scale has been divided into two subscales by Revelle, Humphreys, Simon, and Gilliland (1980). The items under consideration comprise the "impulsivity" subscale and are shown below. The response (yes or no) is keyed in the direction of impulsivity.

Item #	Question	Key
10	Would you do almost anything for a dare?	Y
5	Do you stop and think things over before doing anything?	N
8	Do you generally do and say things quickly without stopping to think?	Y
22	When people shout at you, do you shout back?	Y
9	Do you like doing things in which you have to act quickly?	Y
13	Do you often do things on the spur of the moment?	Y
3	Are you usually carefree?	Y
1	Do you often long for excitement?	Y
41	Are you slow and unhurried in the way you move?	N

The item response data are from the *Computer Administered Panel Survey* (CAPS), a survey sponsored by the Odum Institute for Research in Social Science at the University of North Carolina at Chapel Hill. Approximately 100 undergraduates participated in the study, which involved responding to a large number of questionnaires via computer terminals. Data are available for each academic year between 1983-84 and 1987-88.

The Eysenck Personality Inventory was among the scales administered in 1987 and 1988. The data (for all of the questionnaires) remain publicly available at the Odum Institute's website. In this section, we use the data for the two academic years 1987 and 1988[4] to illustrate the models of interest here.

Below is an IRTPRO spreadsheet presentation of the data showing the first 15 cases for the items eys1 to eys10.



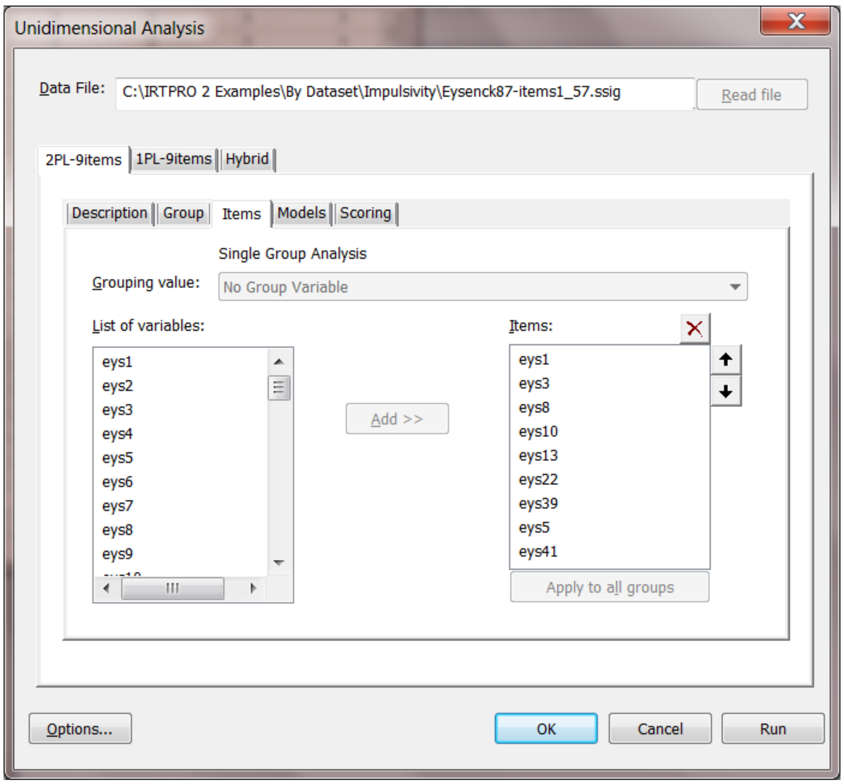
	eys1	eys2	eys3	eys4	eys5	eys6	eys7	eys8	eys9	eys10
1	1	1	2	1	1	1	1	2	2	2
2	1	1	1	1	2	1	1	1	1	2
3	1	1	1	1	1	2	1	2	2	2
4	1	2	1	2	1	1	2	2	2	2
5	1	1	1	2	2	1	1	2	2	2
6	1	1	1	2	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	1	1
8	2	1	1	2	1	1	1	1	2	1
9	1	2	1	2	1	1	2	2	2	2
10	1	1	1	2	1	1	1	2	1	2
11	1	1	1	2	1	1	2	2	1	2
12	1	1	1	1	2	2	1	1	1	2
13	1	1	1	2	1	2	1	2	1	2
14	1	1	1	2	1	1	1	1	1	1
15	2	1	1	1	2	2	2	1	2	2

4 Data from the 1988 "Eysenck Personality Inventory Form A (CAPS-EYSENCK module)", http://hdl.handle.net/1902.29/CAPS-EYSENCK Odum Institute [Distributor] V1 [Version].

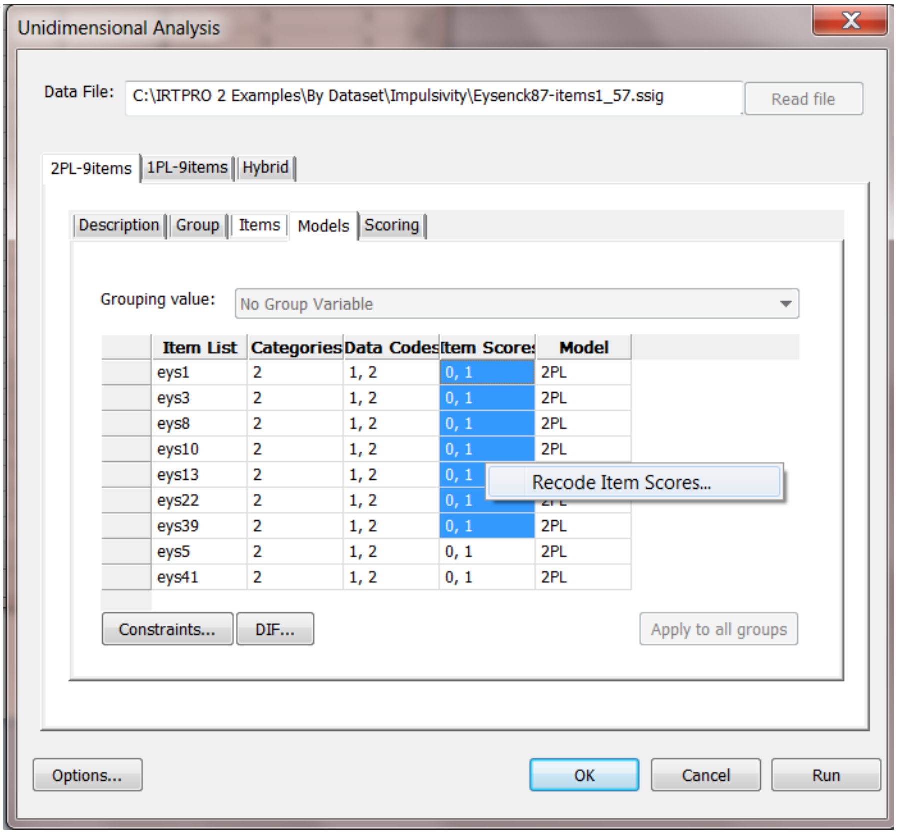
5.2.1 The 2PL Model

IRT item analysis for these data begins with fitting the 2PL model. In this example, use is made of the IRTPRO data file stored as **Eysenck87-items1_57.ssig** in the folder **IRTPRO Examples\By Dataset\Impulsivity**.

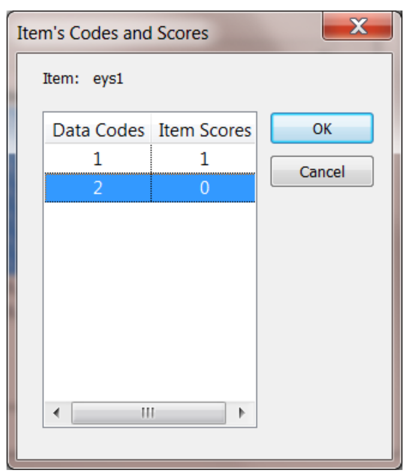
From the main menu bar, we select the **Analysis**, **Unidimensional IRT…** option. Click on the right-hand side of the **Test1** tab and insert a second test. Next right-click on the **Test1** tab and rename it to **2PL-9Items**. Likewise, rename the **Test2** tab to **1PL-9Items**. Starting with the **2PL-9Items** tab, fill in a title and optional comments on the **Description** window, then select the **Items** tab and select the items Eys1, Eys3, Eys8, Eys10, Eys13, Eys22, Eys39, Eys5 and Eys41.



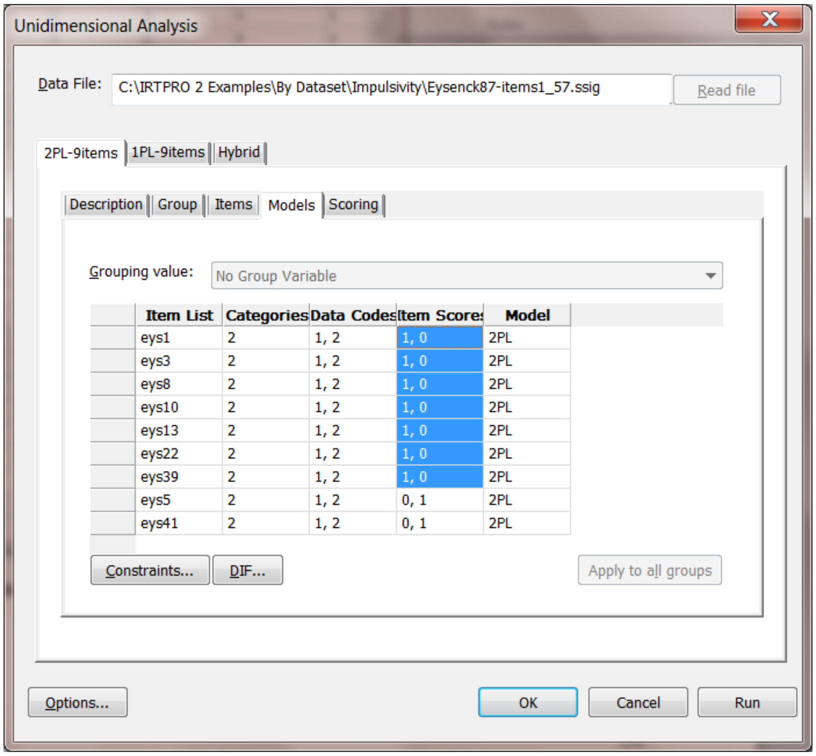
By clicking on the **Models** tab of the **Unidimensional Analysis** window, the list of items selected and model type is displayed. In the IRTPRO data, 1 = Yes and 2 = No. Therefore, (see Section 5.2) all of the items except Eys5 and Eys41 need to be reverse keyed on the Models dialog. The recoding is accomplished by selecting the first seven cells in the **Item Score** column. Right-click on any of the selected cells and then click the **Recode Item Scores…** option.



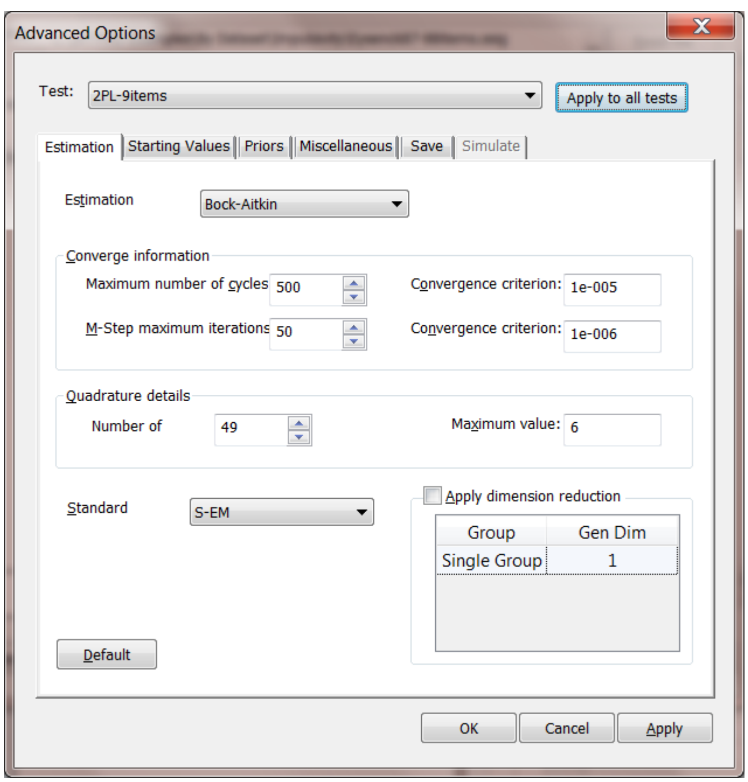
Selection of the **Recode Item Scores …** option results in the display of the **Item's Codes and Scores** dialog where the **Item Scores** are changed in accordance with the answer key.



When done, click **OK** to obtain the revised **Models** dialog shown next.



In this example, as in Section 5.1, we use the implementation of the Bock-Aitkin (1981) EM algorithm to compute the maximum likelihood (ML) estimates of the parameters. By clicking the **Options** button (see left bottom above), the **Advanced Options** window is displayed. The default display is the estimation settings, shown below. One can use this dialog to set convergence criteria to values deemed suitable for a given analysis.



Click **OK** to close the **Advanced Options** window, and then click the **Run** button to start the estimation procedure.

A selection of parts of the output is listed below. The first part shown lists the item parameter estimates for the 2PL model. The *a* (slope, or discrimination) parameter estimates vary from 0.13 up to 3.7. However, we note that the corresponding standard errors vary from 0.19 up to 7.13; such large values are attributable to the small sample size (for IRT), less than 200 respondents.

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	eys1	2 1.18	0.37	1 1.99	0.34	-1.68	0.39
2	eys3	4 0.75	0.41	3 1.24	0.21	-1.65	0.81
3	eys8	6 1.07	0.36	5 -0.43	0.21	0.40	0.19
4	eys10	8 0.83	0.42	7 -1.68	0.32	2.03	0.79
5	eys13	10 3.70	7.13	9 2.11	3.81	-0.57	0.13
6	eys22	12 0.52	0.24	11 0.19	0.18	-0.37	0.42
7	eys39	14 0.76	0.27	13 0.60	0.21	-0.79	0.41
8	eys5	16 0.82	0.39	15 -1.60	0.37	1.95	0.66
9	eys41	18 0.13	0.19	17 0.56	0.16	-4.16	6.04

Summed-Score Based Item Diagnostic Tables and X^2s for Group 1 (Back to TOC)

S-X2 Item Level Diagnostic Statistics

Item	Label	X^2	d.f.	Probability
1	eys1	7.66	5	0.1757
2	eys3	10.66	5	0.0584
3	eys8	11.29	5	0.0458
4	eys10	3.59	5	0.6111
5	eys13	6.63	3	0.0844
6	eys22	9.74	6	0.1359
7	eys39	14.92	5	0.0107
8	eys5	12.15	5	0.0327
9	eys41	13.29	5	0.0208

The value of the M_2 goodness-of-fit statistic (Maydeu-Olivares & Joe, 2005, 2006; Cai, Maydeu-Olivares, Coffman, & Thissen, 2006) reported in the printout below indicates some lack of fit (M_2 =47.45, 27 *d.f.*, p = 0.009); however the associated RMSEA value (0.06) suggests this may be due to a limited amount of "model error"; there must be some error in any strong parametric model.

Likelihood-based Values and Goodness of Fit Statistics (Back to TOC)

Statistics based on the loglikelihood	
-2loglikelihood:	1918.13
Akaike Information Criterion (AIC):	1954.13
Bayesian Information Criterion (BIC):	2012.48

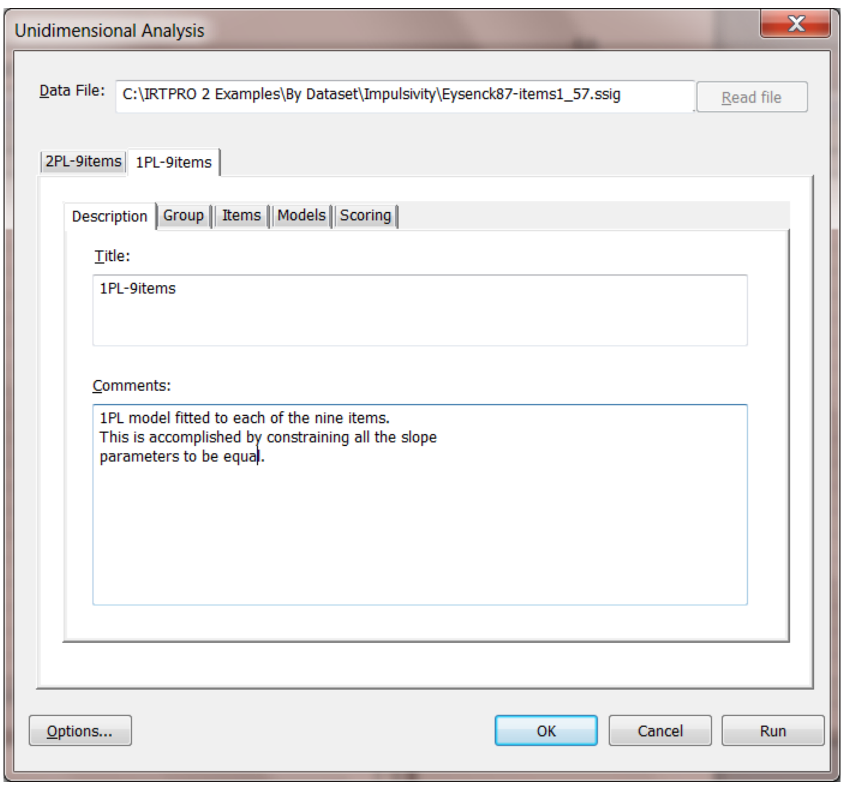
Statistics based on the full item x item x ... classification
The table is too sparse to compute the general multinomial goodness of fit statistics.

Statistics based on one- and two-way marginal tables			
M_2	Degrees of freedom	Probability	RMSEA
47.45	27	0.0088	0.06
Note: M_2 is based on full marginal tables.			
Note: Model-based weight matrix is used.			

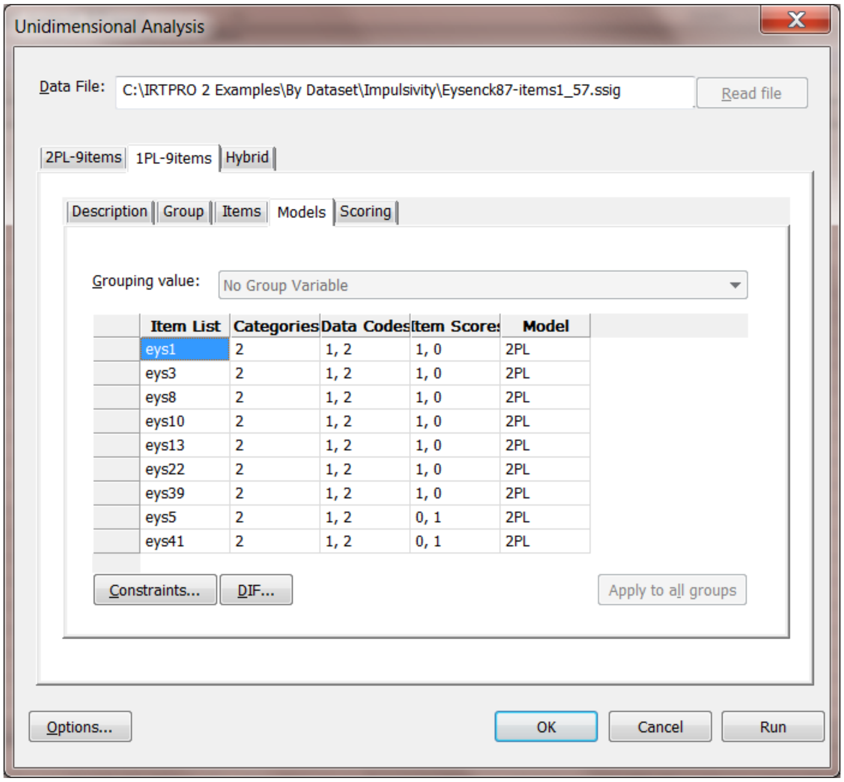
Before interpreting the variability among the slopes as indicative of reliable differences among the items' association with the latent variable being measured (impulsivity), we fit the 1PL model to the data to use the difference between the goodness of fit of the two models to compute a test of significance of the variation among the 2PL slopes.

5.2.2 The 1PL Model

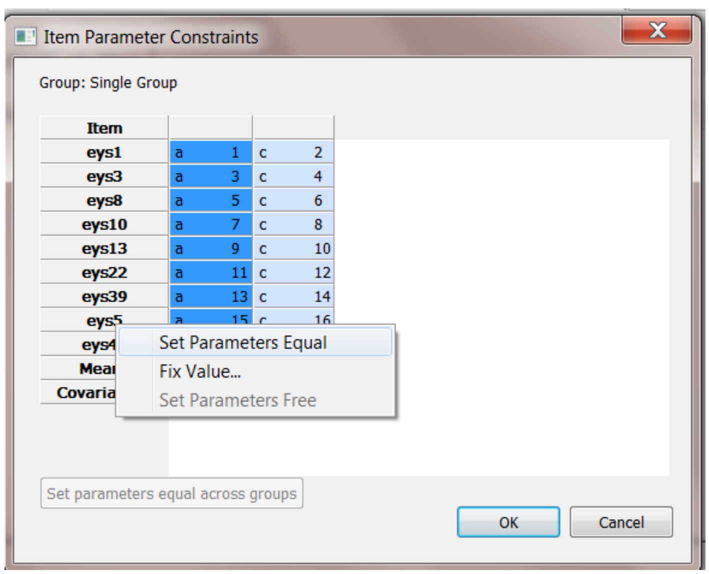
By closing the output file with extension **.htm** we return to the **Analysis** option on the main menu bar and click on the **1PL-9items tab**. Start with the **Description** tab and enter a title and comments.



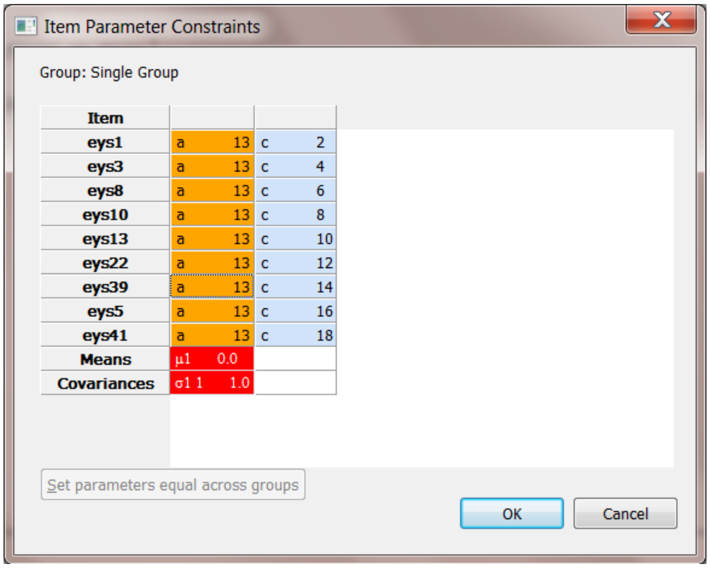
Select the same nine items that were chosen for the 2PL model and then click the **Models** tab and recode the first seven items as discussed in the previous section:



In the **Models** window, click the **Constraints…** button to obtain the **Item Parameter Constraints** window shown below. Next, select all nine items by clicking on them while holding down the **Shift** or **Control** key and right-clicking from any of the selected cells.



From the drop-down menu, select **Set Parameters Equal** to obtain the required constraints on the nine slope parameters.



As shown below, the ML estimate of the (single, common, equal) *a* parameter for all nine items is 0.83, with a standard error of 0.10; the standard error has become much smaller than the slope standard errors for the 2PL model because the data from all nine items is used to estimate the single common slope.

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	eys1	9 0.83	0.10	1 1.80	0.22	-2.16	0.34
2	eys3	9 0.83	0.10	2 1.26	0.19	-1.52	0.27
3	eys8	9 0.83	0.10	3 -0.40	0.17	0.49	0.21
4	eys10	9 0.83	0.10	4 -1.68	0.20	2.02	0.32
5	eys13	9 0.83	0.10	5 0.93	0.18	-1.12	0.24
6	eys22	9 0.83	0.10	6 0.21	0.16	-0.25	0.20
7	eys39	9 0.83	0.10	7 0.61	0.17	-0.73	0.21
8	eys5	9 0.83	0.10	8 -1.60	0.20	1.93	0.31
9	eys41	9 0.83	0.10	10 0.64	0.17	-0.77	0.21

Summed-Score Based Item Diagnostic Tables and X2s for Group 1 (Back to TOC)

S-X2 Item Level Diagnostic Statistics

Item	Label	X^2	d.f.	Probability
1	eys1	5.93	5	0.3144
2	eys3	10.53	5	0.0614
3	eys8	12.54	6	0.0509
4	eys10	3.83	5	0.5747
5	eys13	16.27	5	0.0061
6	eys22	11.41	5	0.0437
7	eys39	14.94	5	0.0106
8	eys5	11.70	5	0.0390
9	eys41	25.41	5	0.0001

The primary purpose of fitting the 1PL model is to obtain the value of -2loglikelihood to use with the corresponding value from the 2PL fit to test the significance of the variation among the *a* parameters in the 2PL model.

Likelihood-based Values and Goodness of Fit Statistics [(Back to TOC)]

Statistics based on the loglikelihood	
-2loglikelihood:	1943.71
Akaike Information Criterion (AIC):	1963.71
Bayesian Information Criterion (BIC):	1996.13

For these data, the likelihood ratio test of the significance of variation among the slope parameters is computed as the difference between -2loglikelihood for the 1PL and 2PL models, 1943.71 - 1918.13 = 25.58, which is distributed as χ^2 on 8 d.f., $p = 0.0012$. The significance of that test statistic leads to the conclusion that there is some reliable difference in discrimination for these nine items; further data analysis can be used to identify that variation.

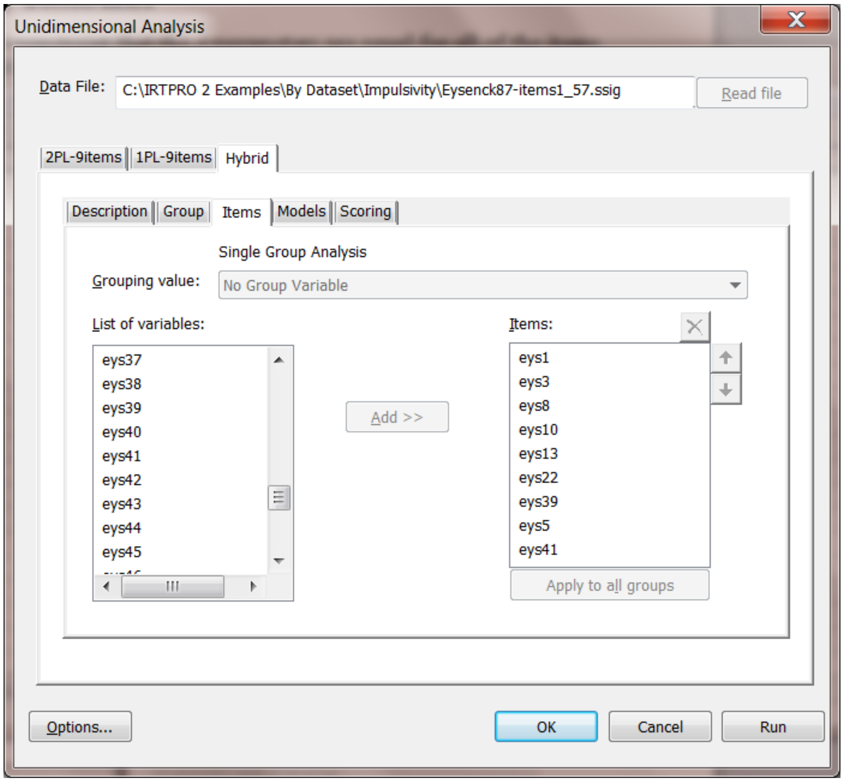
Under the hypothesis of perfect model fit, the $S\text{-}X^2$ statistics are approximately distributed as χ^2 values with the tabulated degrees of freedom; significant values indicate lack of fit. Because a very strong model such as the 2PL rarely fits perfectly, one expects some (slightly) significant values, because the model is not perfect. The statistics tabulated for the 2PL fit illustrate this point: Four of the nine values are significant at the $p = 0.05$ level; however, none have $p < 0.01$. For the 1PL model, however, the value of the $S\text{-}X^2$ for item 41 is very large: 25.41 on 5 d.f., $p < 0.0001$. That value suggests that closer inspection of the underlying frequency table is warranted.

We notice that the reason for the large value of the $S\text{-}X^2$ diagnostic statistic for this item is that for low summed scores on the other items (0-4), we observe more "no" responses than expected, while for higher summed scores on the other items (5-8) we observe more "yes" responses than expected. This pattern suggests that the 1PL fitted value of the slope (a) parameter for item 41 is too high, producing expected values that are too high for low scores, and too low for high scores. Having noted this large $S\text{-}X^2$ diagnostic statistic for this item, with the pattern of observed and expected values, and also that item 41 has (by far) the lowest estimated a parameter for the 2PL fit, we re-fit the data with a special model that is a hybrid between the 1PL and 2PL models.

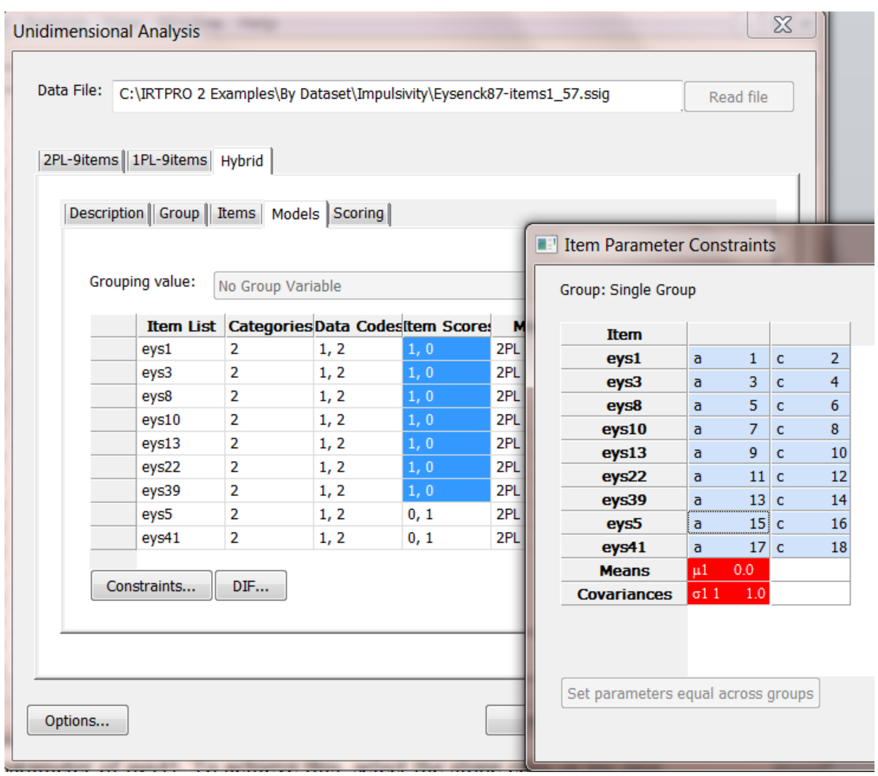
5.2.3 A Special Model with Equality Constraints

The special hybrid model imposes the constraint that the *a* parameters are equal for all of the items except eys41; however, it permits eys41 to have its own slope estimate, which is lower.

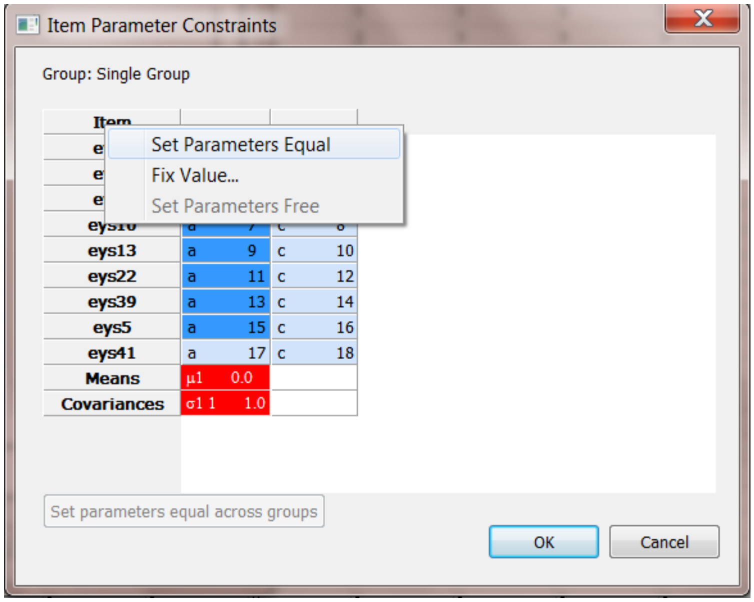
Select the **Analysis**, **Unidimensional IRT…** option and right-click next to the **1PL-9items** tab to insert a **Test3** tab. Right-click on this tab and rename it to **Hybrid**. Select the same nine items that were used in Sections 5.2.1 and 5.2.2.



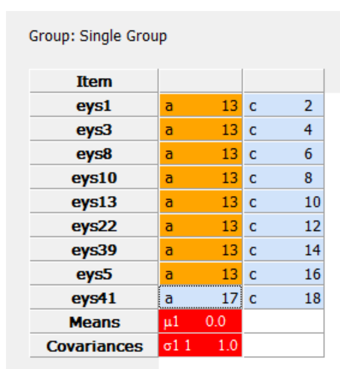
Proceed to the **Models** tab and recode the item scores of the first seven items from (0,1) to (1,0). See Section 5.2.1 where the steps to do the recoding is discussed. Click the **Constraints…** button to obtain the following display.



In the special model, we want to constrain the slope parameters of the first eight items to be equal, but freely estimate the slope parameter of eys41. To achieve this, select the slope cells of the first eight items, then right-click to obtain the **Set Parameters Equal**, **Fix value…**, or **Set Parameters Free** options.



By selecting the **Set Parameters Equal** option, the constraints screen changes to reflect that the first eight items have equal slope and that the slope of the ninth item is estimated freely, as shown below.



The item parameter estimates, standard errors, and goodness of fit statistics for this model are as follows.

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	eys1	[9] 0.98	0.12	[1] 1.87	0.22	-1.92	0.29
2	eys3	[9] 0.98	0.12	[2] 1.32	0.19	-1.35	0.23
3	eys8	[9] 0.98	0.12	[3] -0.42	0.17	0.43	0.18
4	eys10	[9] 0.98	0.12	[4] -1.75	0.22	1.79	0.27
5	eys13	[9] 0.98	0.12	[5] 0.97	0.18	-1.00	0.21
6	eys22	[9] 0.98	0.12	[6] 0.22	0.17	-0.22	0.17
7	eys39	[9] 0.98	0.12	[7] 0.64	0.18	-0.65	0.19
8	eys5	[9] 0.98	0.12	[8] -1.67	0.21	1.71	0.27
9	eys41	[11] 0.11	0.20	[10] 0.55	0.15	-5.26	10.03

Summed-Score Based Item Diagnostic Tables and X2s for Group 1 **(Back to TOC)**

S-X2 Item Level Diagnostic Statistics

Item	Label	X^2	d.f.	Probability
1	eys1	6.23	5	0.2860
2	eys3	10.67	5	0.0582
3	eys8	11.34	5	0.0450
4	eys10	3.43	5	0.6348
5	eys13	13.97	5	0.0157
6	eys22	13.11	5	0.0223
7	eys39	14.94	5	0.0106
8	eys5	12.85	5	0.0247
9	eys41	13.23	5	0.0213

From the S-X^2 item level diagnostic statistics given above it follows that six of the nine values are significant at the p = 0.05 level. However, none is significant at the p = 0.01 level, suggesting these statistics may indicate real, but negligible, misfit.

Likelihood-based Values and Goodness of Fit Statistics **(Back to TOC)**

Statistics based on the loglikelihood	
-2loglikelihood:	1931.41
Akaike Information Criterion (AIC):	1953.41
Bayesian Information Criterion (BIC):	1989.07

Likelihood ratio tests indicate that this model fits significantly better than the 1PL model (χ^2 = 1943.71 - 1931.41 = 12.3, 1 d.f., p = 0.0005), but it does not fit significantly worse than the 2PL model (χ^2 = 1931.41 - 1918.13 = 12.3, 7 d.f., p = 0.066). The values of the M_2 overall goodness of fit statistic and its associated RMSEA are approximately the same for this hybrid model as they were for the 2PL model, although the special model estimates seven fewer parameters (2 a's and 9 c's, as opposed to 9 a's and 9 c's). Therefore, the combined considerations of goodness of fit and parsimony suggest the use and interpretation of the special hybrid model.

Inspection of the underlying summed-score by item response tabulations for these statistics confirms that. The printout below shows the underlying summed-score by item response tabulation for the S-X^2 item level diagnostic statistics for the special hybrid model fit to eys1 and eys39. The table for item 1 (eys1) shows perfectly good fit, with a non-significant value of S-X^2. The summed-score by item response table for item 7 (eys39), has a "significant" S-X^2 value of 14.9 on 5 d.f. (p = 0.0106). We note that even in the table for eys39, the worst-fitting of the nine items, there is no particular tendency for higher observed than expected values to occur in blocks or "runs" that might indicate bad fit of the trace line, as we previously observed with eys41 fitted with the 1PL model.

Item 1 S-X2(5) = 6.2 , p = 0.2860 **(Back)**

	Category 0		Category 1	
Score	Observed	Expected	Observed	Expected
0	1	1.3	2	1.7
1	2	2.9	6	5.1
2	9	8.7	22	22.3
3	6	5.1	18	18.9
4	6	7.7	44	42.3
5	7	3.6	25	28.4
6	1 1	2.1 2.8	26	24.9
7	0	0.6	11	10.4
8	0	0.1	3	2.9

Item 7 S-X2(5) = 14.9 , p = 0.0106 (Back)

Score	Category 0				Category 1			
	Observed		Expected		Observed		Expected	
0	1		0.7		0		0.3	
1	4	5	6.5	7.2	6	6	3.5	3.8
2		9		8.5		6		6.5
3		22		17.0		14		19.0
4		10		16.8		34		27.2
5		17		11.1		20		25.9
6		6		7.0		25		24.0
7	1	1	2.0	2.3		11		10.0
8	0		0.3			3		2.7

Chen & Thissen (1997) proposed the LD X^2 statistic, computed by comparing the observed and expected frequencies in each of the two-way cross tabulations between responses to each item and each of the other items. These diagnostic statistics are (approximately) standardized χ^2 values (that is, they are approximately z-scores) that become large if a pair of items indicates local dependence, that is, if data for that item pair indicates a violation of the local independence assumption that is the essence of the item response model. The printout below shows the pairwise values of the standardized LD X^2 statistics for this impulsivity subscale fitted with the special model intermediate between the 1PL and 2PL models; italic entries indicate positive LD, while roman entries indicate negative LD. All of the values are relatively small[5], indicating no evidence of LD, and suggesting that the model fits satisfactorily.

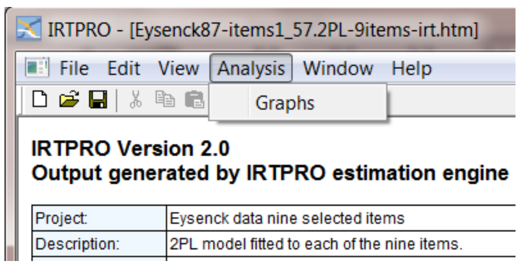
[5] Because the standardized LD X2 statistic is only approximately standardized, and is known to be based on a statistics with a long-tailed (□2) distribution, we do not consider values larger than 2 or 3 to be large. Rather, we consider values larger than 10 large, indicating likely LD; values in the range 5-10 lie in a gray area, and may either indicate LD or they may be a result of sparseness in the underlying table of frequencies. In practice data analysts use inspection of the item content, as well as these statistics, to evaluate the presence of LD when it is indicated.

Marginal fit (X2) and Standardized LD X2 Statistics for Group 1 [(Back to TOC)](#)

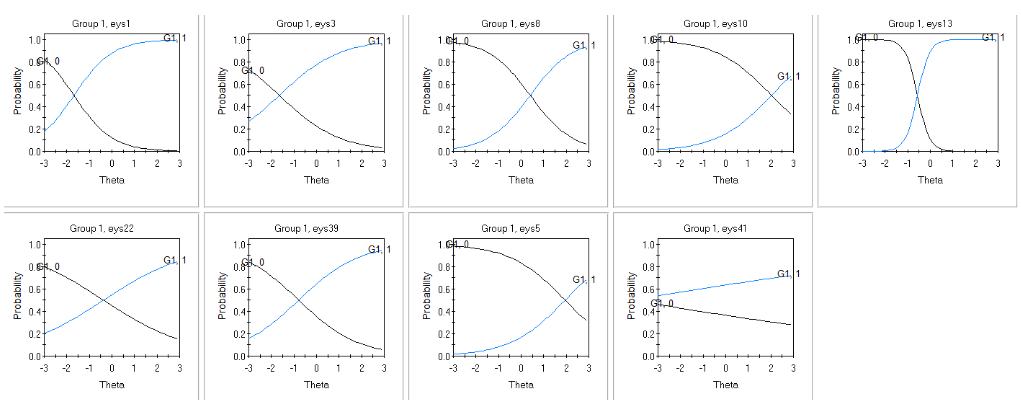
Item	Label	Marginal X^2	1	2	3	4	5	6	7	8
1	eys1	0.0								
2	eys3	0.0	-0.6							
3	eys8	0.0	-0.7	-0.4						
4	eys10	0.0	-0.2	1.2	-0.5					
5	eys13	0.0	4.6	-0.7	2.6	-0.6				
6	eys22	0.0	-0.5	2.7	-0.1	-0.2	-0.7			
7	eys39	0.0	-0.1	-0.7	0.6	-0.4	-0.1	-0.2		
8	eys5	0.0	6.6	-0.6	0.5	0.1	-0.0	2.2	-0.6	
9	eys41	0.0	-0.6	-0.6	1.3	-0.2	-0.4	0.0	-0.5	-0.6

5.2.4 Trace Lines and information curves for the 2PL model

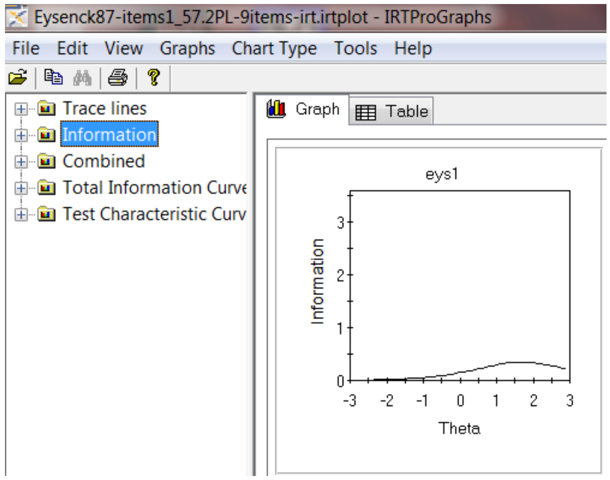
To obtain graphical representations of the trace lines and information curves for the nine impulsivity subscale items, open the output file containing the results of the 2PL analysis (**Eysenck87–items1_57.2PL–9items.htm**) and select **Analysis, Graphs** as shown below.



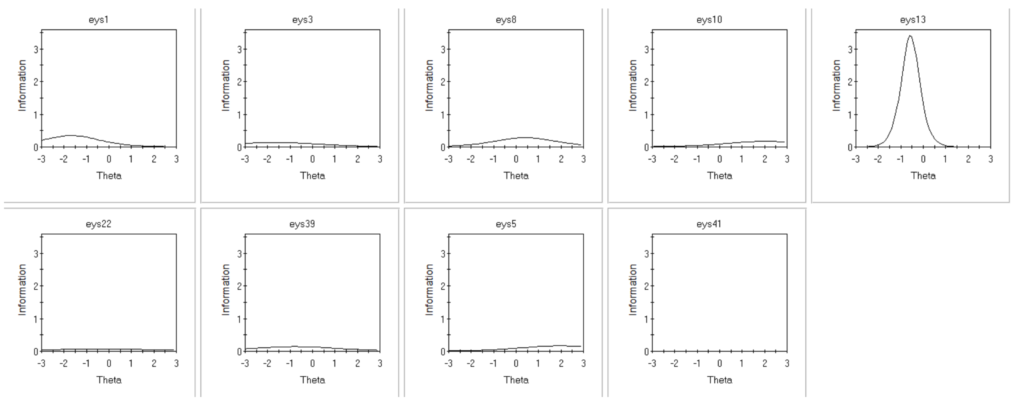
This action opens the **Graphics** window and as default, a display of the trace lines. The trace lines shown plot the probability of the "impulsive" responses for each item as a function of the underlying latent variable. The curves show that the nine items are spread to cover the range of the impulsivity continuum. For item 41, with its very low slope value (0.11), the probability of a "no" response changes very little across levels of impulsivity from lowest to highest.



To obtain the information curves, click on the **Information** option.



Below we show the information curves for each item. If we were constructing an "**impulsivity**" scale based on the IRT analyses, we would omit item 41 since it provides negligible information.



5.3 Analysis of the State-Trait Anxiety Inventory (STAI) data

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).[6]

For illustration of fitting the graded, partial credit, and nominal response models, six items are selected:

- I feel calm.
- I am tense.
- I am regretful.
- I feel at ease.
- I feel anxious.
- I feel nervous.

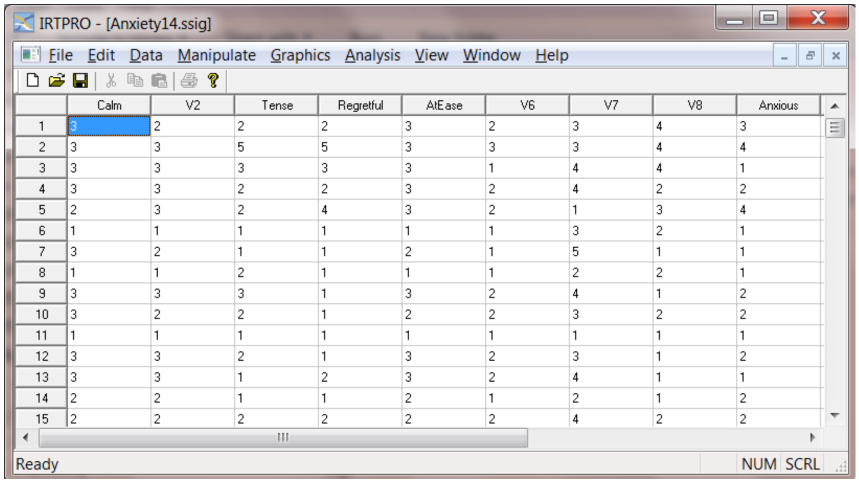
In these data, the responses were on a five-point unipolar Likert-type response scale: 1 = not at all, 2 = very little, 3 = somewhat, 4 = moderately, and 5 = very much.

5.3.1 Graded model

To see the data, use the **Open** file dialog under the **File** menu of IRTPRO, navigate to the **C:\IRTPRO Examples\By Dataset\Anxiety14** folder, select **Files of type: IRTPRO Data File (*.ssig)** in the **Open File** dialog, and open the file **Anxiety14.ssig**.

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.

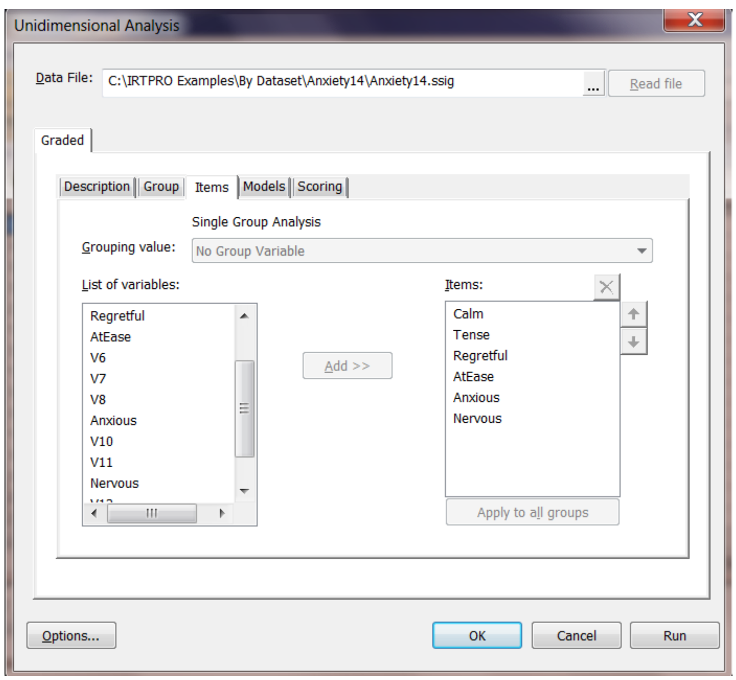
6 6 Thanks to Lynne Steinberg for these data, which are described more completely by Thissen & Steinberg (2009).



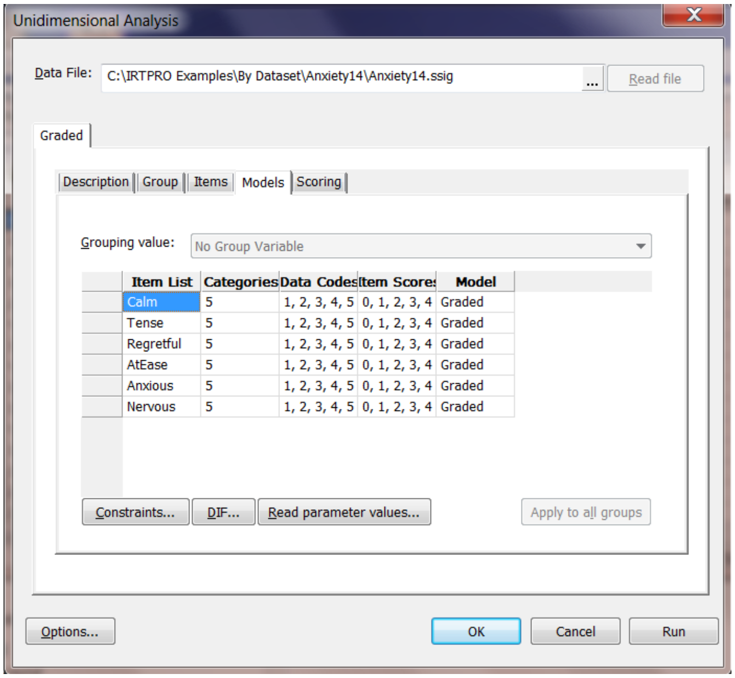
	Calm	V2	Tense	Regretful	AtEase	V6	V7	V8	Anxious
1	3	2	2	2	3	2	3	4	3
2	3	3	5	5	3	3	3	4	4
3	3	3	3	3	3	1	4	4	1
4	3	3	2	2	3	2	4	2	2
5	2	3	2	4	3	2	1	3	4
6	1	1	1	1	1	1	3	2	1
7	3	2	1	1	2	1	5	1	1
8	1	1	2	1	1	1	2	2	1
9	3	3	3	1	3	2	4	1	2
10	3	2	2	1	2	2	3	2	2
11	1	1	1	1	1	1	1	1	1
12	3	3	2	1	3	2	3	1	2
13	3	3	1	2	3	2	4	1	1
14	2	2	1	1	2	1	2	1	2
15	2	2	2	2	2	2	4	2	2

To initiate the unidimensional IRT analysis, select **Unidimensional IRT …** under the **Analysis** menu.

After adding a title and (optional) comments, click the **Items** tab and select the six items as shown next.



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:



If we click the **Run** button in the lower right hand corner of the dialog, we see the output. 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*).

Note that 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.

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

Item	Label	a		s.e.	c1		s.e.	c2		s.e.	c3		s.e.	c4		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

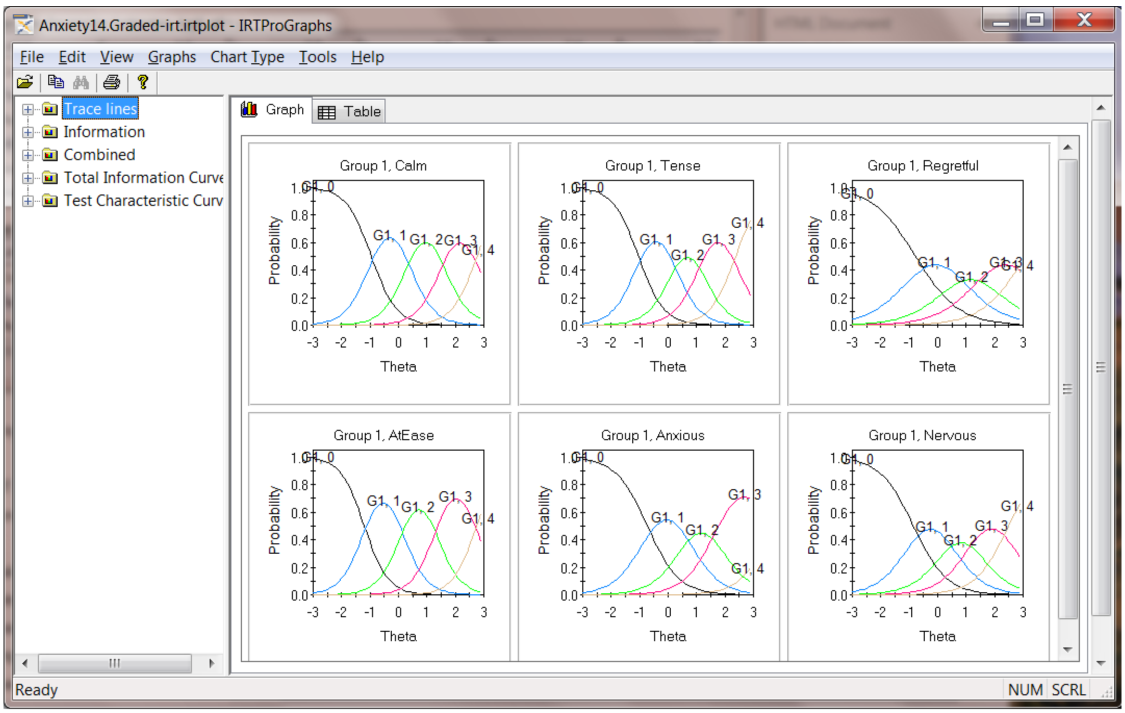
Graded Model Item Parameter Estimates for Group 1, logit: *a*(*θ* - *b*) (Back to TOC)

Item	Label	a		s.e.	b1	s.e.	b2	s.e.	b3	s.e.	b4	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

Likelihood-based Values and Goodness of Fit Statistics **(Back to TOC)**

Statistics based on the loglikelihood	
-2loglikelihood:	7522.70
Akaike Information Criterion (AIC):	7582.70
Bayesian Information Criterion (BIC):	7710.14

To see the graded model trace lines graphically, when the output file **Anxiety14.Test1-irt.htm** is in the IRTPRO viewer window, we may 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:

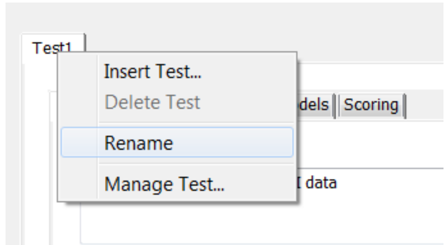


When finished looking at the graphics, the **IRTPROGraphs** program may be closed using the **X** in the upper right hand corner of the window, and we return to the main IRTPRO window that has remained running behind the graphics application.

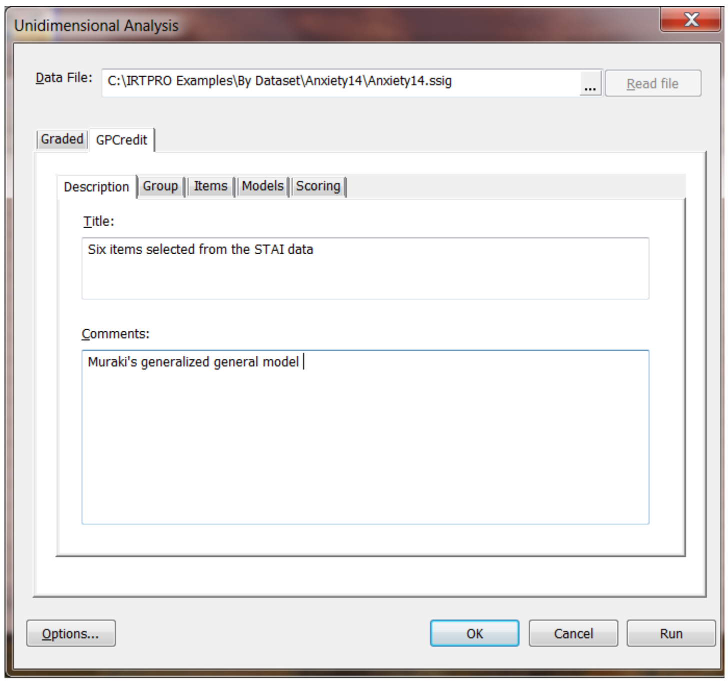
We leave it to the user as an exercise to page through the rest of the output, which shows reasonably good fit of the graded IRT model to these six items.

5.3.2 Muraki's generalized partial credit model fitted to the State Trait Anxiety Inventory (STAI) data

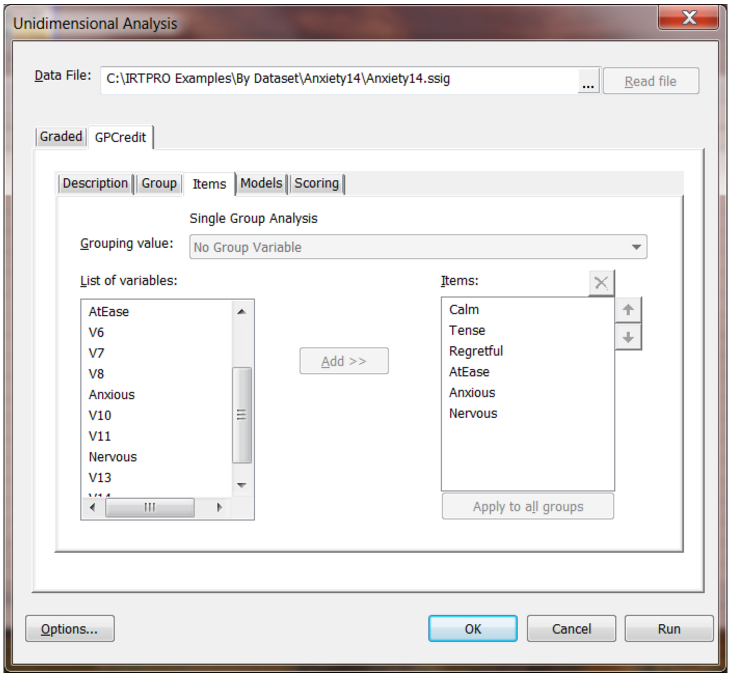
Continuing with the example from the previous section, we add an additional test in order to fit the generalized partial credit model. By right-clicking to the right of the **Test1** tab, a pop-up menu is obtained that enable the user to enter a new test, delete a test or rename an existing test, the default test names being Test1, Test2, …



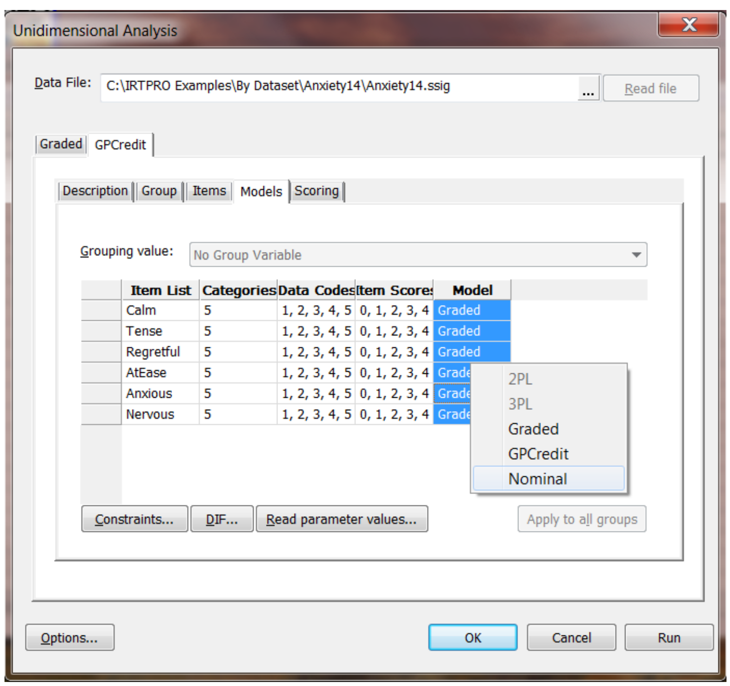
Right-click on each of the test tabs to invoke the **Insert Test…**, **Delete Test**, etc. drop-down menu. Click the **Rename** button and replace Test1 with Graded and Test2 with GPCredit, the latter being more descriptive of the type of model to be fitted to the item response data.



Use the **Items** tab in the **GPCredit** tab in the **Unidimensional Analysis** dialog to select the six items to be analyzed:



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 as default. To change the default model to the general partial credit model, select **Graded** for each item under the **Model** column and right click to obtain a pop-up menu showing a list of models appropriate for the selected items.



Select **GPCredit** from this list to obtain the new set of models shown below.

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

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

Table of Contents

The first table of Partial Credit Model Item Parameter Estimates lists the slopes (*a*) and intercepts (*c*); the second table lists the slopes (*a*) and thresholds (*b*):

GPC Model Item Parameter Estimates, logit: $a[k(\theta - b) + \Sigma d_k]$

Item	Label	a		s.e.	b	s.e.	d_1	d_2	s.e.	d_3	s.e.	d_4	s.e.	d_5	s.e.
1	Calm	1	1.69	0.19	0.90	0.11	0.00	1.82	0.12	0.50	0.09	-0.65	0.12	-1.67	0.20
2	Tense	6	1.63	0.18	0.66	0.10	0.00	1.66	0.11	0.36	0.09	-0.38	0.10	-1.64	0.16
3	Regretful	11	0.76	0.09	1.11	0.14	0.00	1.59	0.17	0.18	0.19	-0.19	0.23	-1.58	0.31
4	AtEase	16	1.87	0.22	0.74	0.11	0.00	1.92	0.12	0.58	0.09	-0.54	0.11	-1.96	0.21
5	Anxious	21	1.22	0.14	1.37	0.22	0.00	1.98	0.20	0.64	0.20	-0.10	0.22	-2.52	0.53
6	Nervous	26	1.05	0.12	0.80	0.10	0.00	1.46	0.13	0.30	0.13	-0.24	0.16	-1.52	0.22

This part of the printout is followed by a portion of the printout that provides slopes, scoring function contrasts, and intercept contrasts for the Nominal model. These values were used to obtain the GPC parameter estimates listed above. The relationship between this printout and the GPC item parameter estimates is described by Thissen, Cai, & Bock (2010) and the interested reader is referred to this publication for details.

Nominal Model Slopes and Scoring Function Contrasts for Group 1, logit: (a skθ + ck); s = Tα (Back to TOC)

Item	Label	a	s.e.	Contrasts	α_1	s.e.	α_2	s.e.	α_3	s.e.	α_4	s.e.
1	Calm	1 1.69	0.19	Trend	1.00	-----	0.00	-----	0.00	-----	0.00	-----
2	Tense	6 1.63	0.18	Trend	1.00	-----	0.00	-----	0.00	-----	0.00	-----
3	Regretful	11 0.76	0.09	Trend	1.00	-----	0.00	-----	0.00	-----	0.00	-----
4	AtEase	16 1.87	0.22	Trend	1.00	-----	0.00	-----	0.00	-----	0.00	-----
5	Anxious	21 1.22	0.14	Trend	1.00	-----	0.00	-----	0.00	-----	0.00	-----
6	Nervous	26 1.05	0.12	Trend	1.00	-----	0.00	-----	0.00	-----	0.00	-----

Nominal Model Scoring Function Values for Group 1, logit: (a skθ + ck); s = Tα (Back to TOC)

Item	Category	s_1	s_2	s_3	s_4	s_5
1	Calm	0.00	1.00	2.00	3.00	4.00
2	Tense	0.00	1.00	2.00	3.00	4.00
3	Regretful	0.00	1.00	2.00	3.00	4.00
4	AtEase	0.00	1.00	2.00	3.00	4.00
5	Anxious	0.00	1.00	2.00	3.00	4.00
6	Nervous	0.00	1.00	2.00	3.00	4.00

Nominal Model Intercept Contrasts for Group 1, logit: (a skθ + ck); c = Tγ (Back to TOC)

Item	Label	Contrasts	γ_1	s.e.	γ_2	s.e.	γ_3	s.e.	γ_4	s.e.
1	Calm	Trend	2 -1.51	0.22	3 4.04	0.44	4 0.13	0.13	5 0.13	0.08
2	Tense	Trend	7 -1.07	0.18	8 3.54	0.36	9 0.02	0.10	10 0.26	0.07
3	Regretful	Trend	12 -0.85	0.11	13 1.53	0.22	14 0.00	0.10	15 0.18	0.08
4	AtEase	Trend	17 -1.37	0.27	18 4.90	0.57	19 -0.03	0.14	20 0.23	0.08
5	Anxious	Trend	22 -1.67	0.32	23 3.54	0.63	24 -0.33	0.23	25 0.34	0.12
6	Nervous	Trend	27 -0.84	0.14	28 2.03	0.26	29 -0.03	0.10	30 0.18	0.08

The next part the printout lists summed-scored based item diagnostics, marginal fit (χ^2) and standardized LD χ^2 statistics. See Section 5.2.3 for an interpretation of these results.

Summed-Score Based Item Diagnostic Tables and X2s for Group 1 (Back to TOC)

S-X2 Item Level Diagnostic Statistics

Item	Label	X^2	d.f.	Probability
1	Calm	36.37	36	0.4527
2	Tense	59.84	39	0.0175
3	Regretful	54.60	49	0.2695
4	AtEase	35.83	35	0.4309
5	Anxious	51.47	39	0.0870
6	Nervous	56.92	47	0.1520

Marginal fit (X2) and Standardized LD X2 Statistics for Group 1 (Back to TOC)

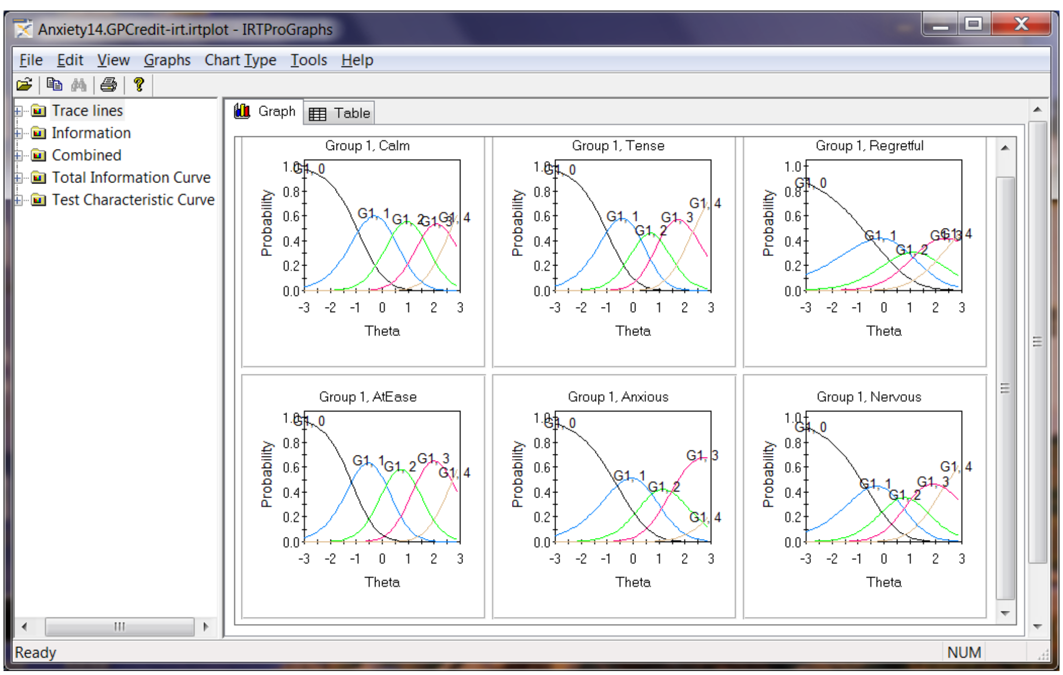
Item	Marginal X^2	1	2	3	4	5
1	0.1					
2	0.1	3.1				
3	0.0	0.3	1.3			
4	0.1	12.2	3.1	1.0		
5	0.0	1.3	3.0	2.9	0.9	
6	0.0	0.5	-1.4	4.1	1.1	7.3

The last part of the printout shown here are likelihood-based statistics.

Likelihood-based Values and Goodness of Fit Statistics (Back to TOC)

Statistics based on the loglikelihood	
-2loglikelihood:	7554.61
Akaike Information Criterion (AIC):	7614.61
Bayesian Information Criterion (BIC):	7742.05

To see the partial credit model trace lines graphically, when the output file **Anxiety14.GPCredit-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:

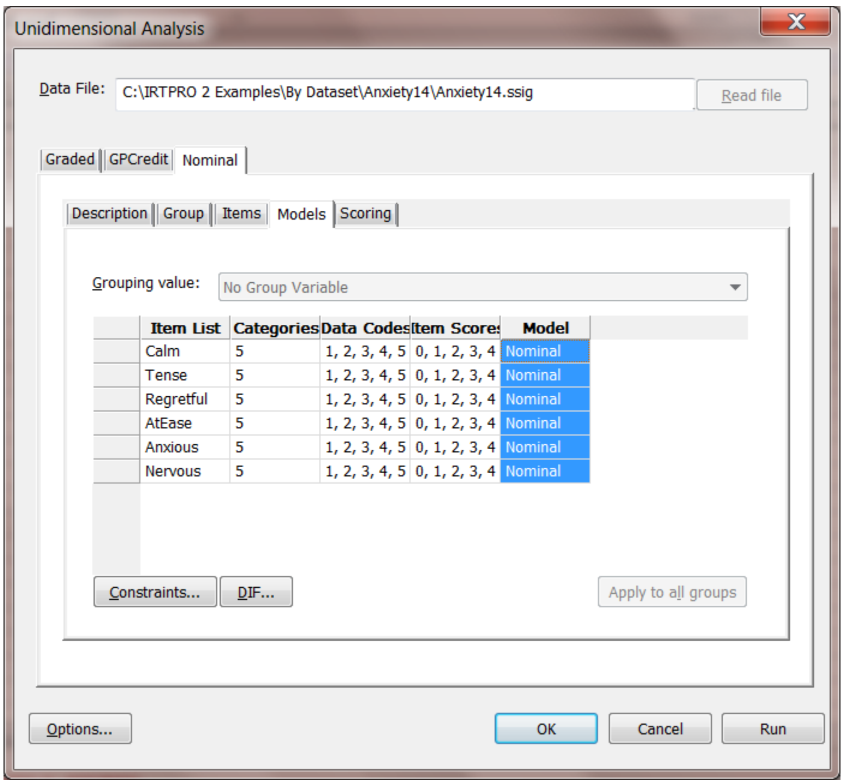


When finished with the graphics, the **IRTPROGraphs** program may be closed using the **X** in the upper right hand corner of the window; control returns to the main IRTPRO window that has remained running behind the graphics application.

5.3.3 Nominal model example

In this section, we fit Bock's nominal model to these data. To compute and view the output for this model, use the **Window** menu to return to the **Anxiety14.ssig** spreadsheet window, and then the **Analysis** menu to select **Unidimensional IRT …**.

Insert **Test3** by right-clicking next to the **GPCredit** tab, then right-click on this tab to rename **Test3** to **Nominal**. Follow the instructions in Sections 5.3.1 and 5.3.1. Select the same six items. Select all the Graded cells in the **Models** window, then right-click to change Graded to Nominal. Click the **Run** button to run the analysis.



A portion of the output is shown next.

Nominal Model Slopes and Scoring Function Contrasts for Group 1, logit: (a skθ + ck); s = Tα **(Back to TOC)**

Item	Label	a	s.e.	Contrasts	α_1	s.e.	α_2	s.e.	α_3	s.e.	α_4	s.e.
1	Calm	[4] 1.46	0.23	Trend	1.00	-----	[1] 0.46	0.39	[2] -0.21	0.21	[3] 0.22	0.13
2	Tense	[12] 1.57	0.23	Trend	1.00	-----	[9] 0.27	0.34	[10] 0.06	0.13	[11] 0.06	0.09
3	Regretful	[20] 0.78	0.12	Trend	1.00	-----	[17] 0.11	0.36	[18] 0.18	0.17	[19] 0.18	0.14
4	AtEase	[28] 1.70	0.25	Trend	1.00	-----	[25] 0.42	0.34	[26] 0.03	0.14	[27] 0.09	0.09
5	Anxious	[36] 1.28	0.47	Trend	1.00	-----	[33] 0.21	0.92	[34] 0.29	0.26	[35] -0.04	0.15
6	Nervous	[44] 1.04	0.13	Trend	1.00	-----	[41] 0.44	0.30	[42] 0.12	0.13	[43] -0.03	0.11

Likelihood-based Values and Goodness of Fit Statistics [(Back to TOC)]

Statistics based on the loglikelihood	
-2loglikelihood:	7528.45
Akaike Information Criterion (AIC):	7624.45
Bayesian Information Criterion (BIC):	7828.35

The table below is a summary of the information criteria for the three models. Since both the AIC and BIC are the smallest for the Graded model, we conclude that this model provides the better fit of the data.

Comparison of Information criteria

Model	AIC	BIC
Graded	7582.70	7710.14
General Partial Credit	7614.61	7742.05
Nominal	7624.45	7828.35

Note that we cannot use the -2loglikelihood statistic to compare these models, since, for example, the Nominal model is not nested within the Graded model.

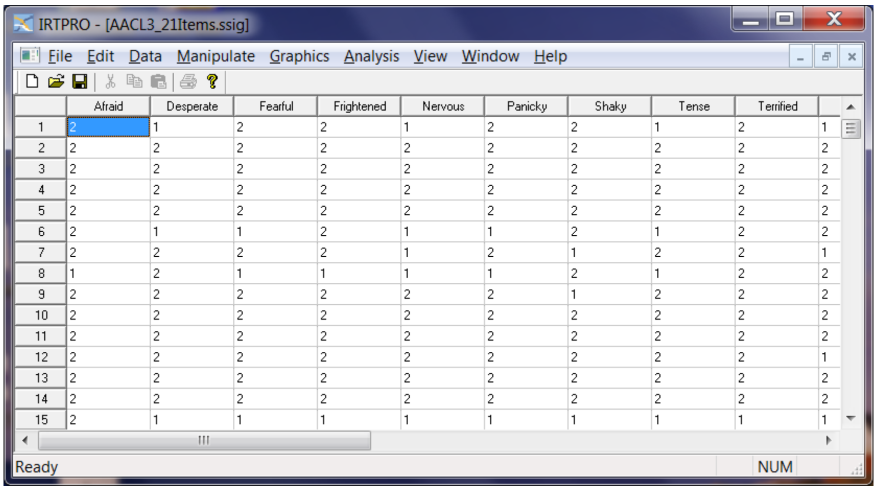
5.4 Unidimensional analysis of the Affect Adjective Check List (AACL)

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[7] analyzed here, the adjectives were framed with the instructions "Please indicate whether or not the adjective listed describes how you feel today, 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. A unidimensional 2PL model is fitted to the entire 21-item set, and the diagnostic statistics are examined.

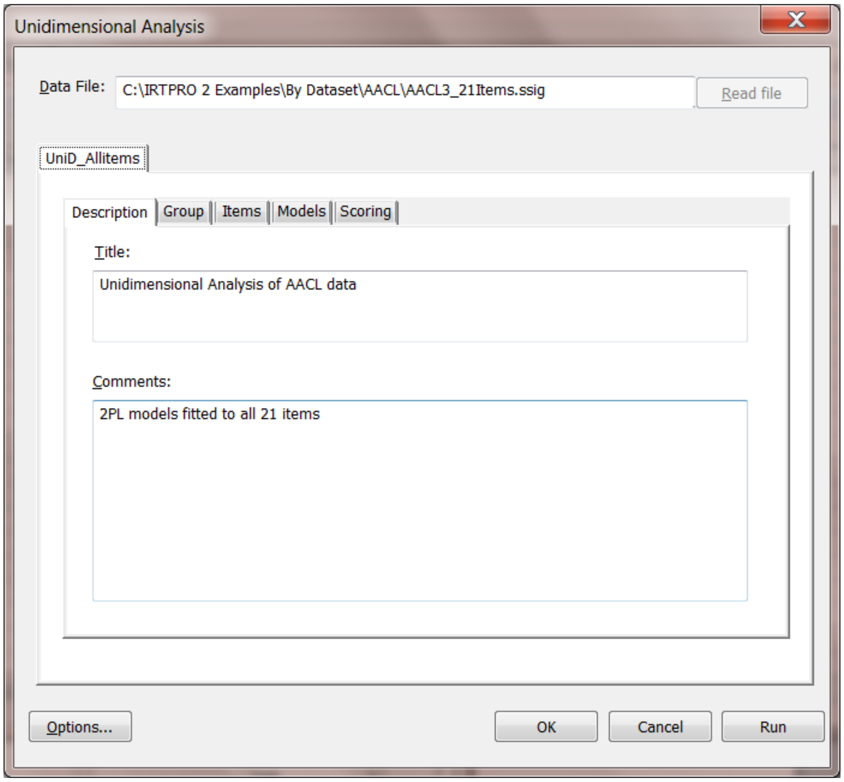
The data have already been imported into IRTPRO and saved as the file **AACL_21Items.ssig** in the **C:\IRTPRO Examples\by DataSet\AACL** folder installed with the software (see Chapter 3 for an example of importing data from other sources into IRTPRO). When the file **AACL_21Items.ssig** is opened, the data are displayed as a spreadsheet:



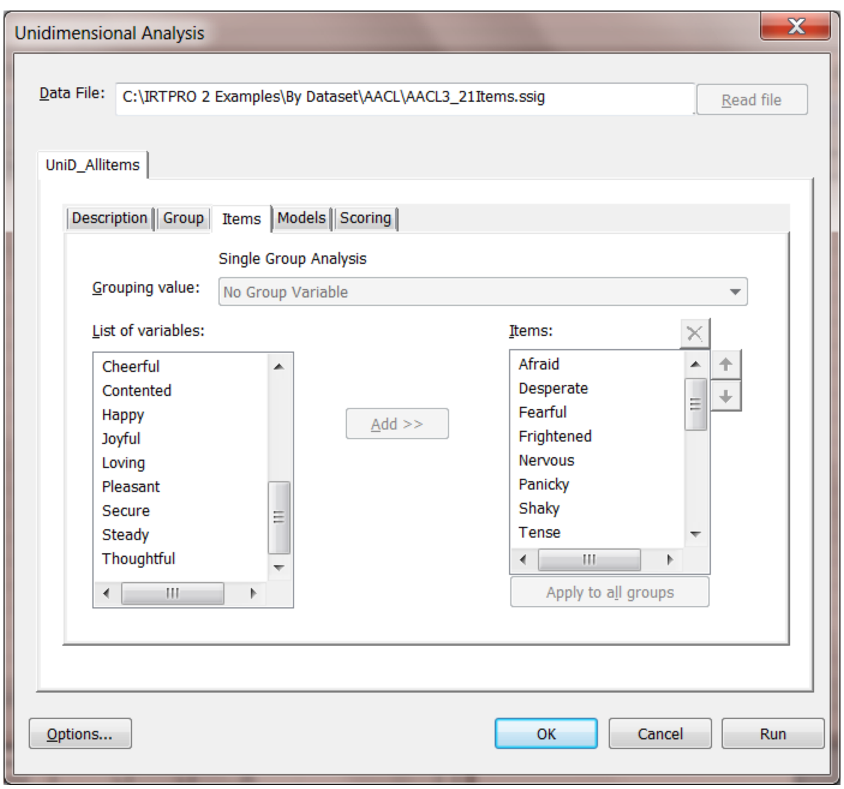
	Afraid	Desperate	Fearful	Frightened	Nervous	Panicky	Shaky	Tense	Terrified	
1	2	1	2	2	1	2	2	1	2	1
2	2	2	2	2	2	2	2	2	2	2
3	2	2	2	2	2	2	2	2	2	2
4	2	2	2	2	2	2	2	2	2	2
5	2	2	2	2	2	2	2	2	2	2
6	2	1	1	2	1	1	2	1	2	2
7	2	2	2	2	1	2	1	2	2	1
8	1	2	1	1	1	1	2	1	2	2
9	2	2	2	2	2	2	1	2	2	2
10	2	2	2	2	2	2	2	2	2	2
11	2	2	2	2	2	2	2	2	2	2
12	2	2	2	2	2	2	2	2	2	1
13	2	2	2	2	2	2	2	2	2	2
14	2	2	2	2	2	2	2	2	2	2
15	2	1	1	1	1	1	1	1	1	1

To initiate the unidimensional IRT analysis, select **Unidimensional IRT …** under the **Analysis** menu. The Unidimensional window opens with a **Test1** tab. Right-click on this tab and rename it to **UniD_Allitems**. Start by clicking on the **Description** tab in the **Unidimensional Analysis** window. Enter a title and description.

7 Data (N=290) from the 1988 "Affect Adjective Check List (CAPS-ANXIETY module)", hdl:1902.29/CAPS-ANXIETY, Odum Institute Dataverse.



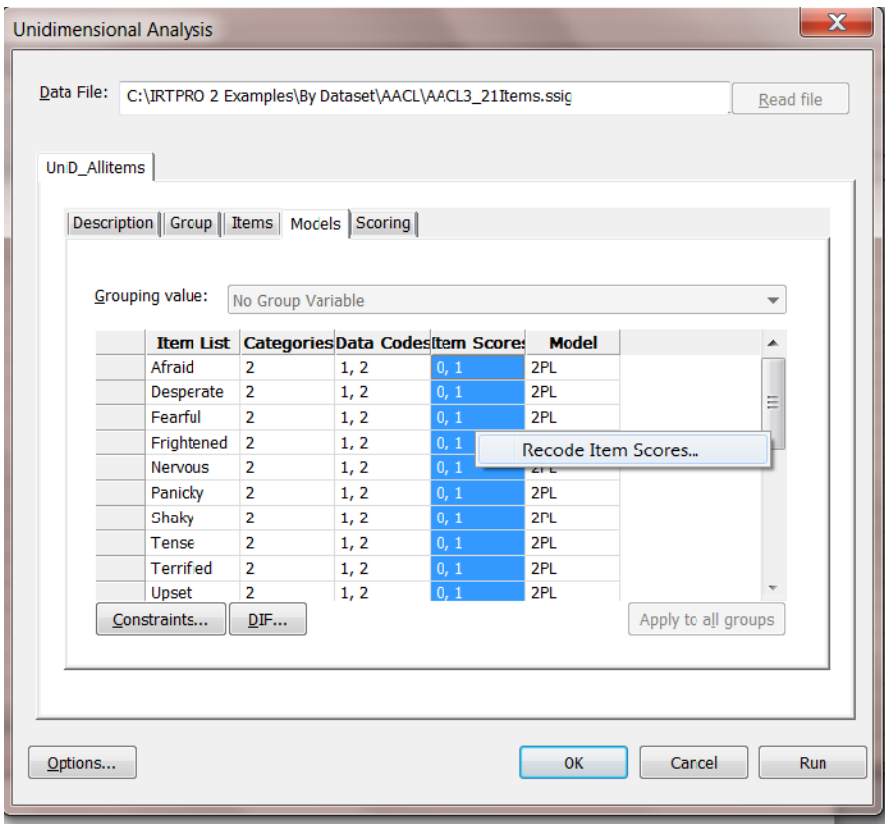
The **Group** tab may be skipped, because there is only one group in this analysis. The **Items** tab is used to select all the items in this case.



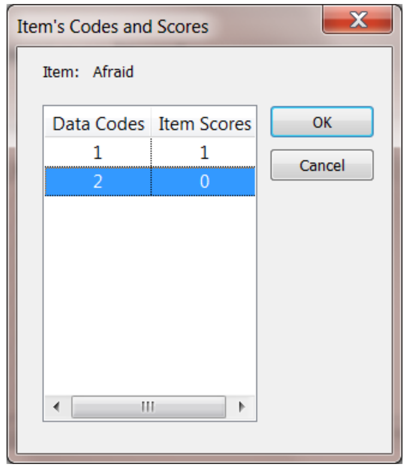
There are several ways to select all of the items: (1) click on any item, and then press <**Ctrl**>**A** to select all, and then either (a) drag the set from the **List of variables** box to the **Items** box, or (2) double-click each item in turn, or (3) select each item and press the **Add >>** button.

Because the **2PL** model is the default for dichotomous items, the entry of information for the analysis would have been complete, if the items were properly scored. However, in this case we would like to recode the item scores of the first 11 items (the Anxiety-plus adjectives) from (0,1) to (1,0) so that Anxiety-plus words are also scored 1 if checked, and scored 0 if not checked. In the data file, 1 is equal to checked and 2 is not checked.

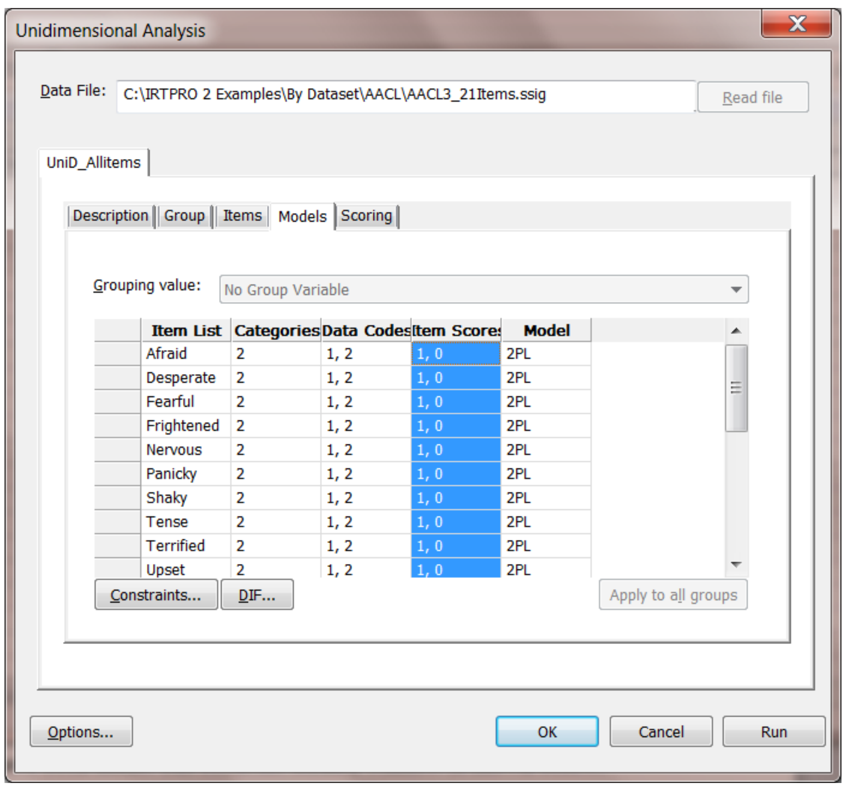
To recode the item scores, use the **Shift** key to select the first eleven cells in the **Item Scores** column then right-click to display the **Recode Item Scores …** option.



Selecting of the **Recode Item Scores…** option invokes the **Item's Codes and Scores** dialog and allows one to change values in the **Item Scores** text boxes.

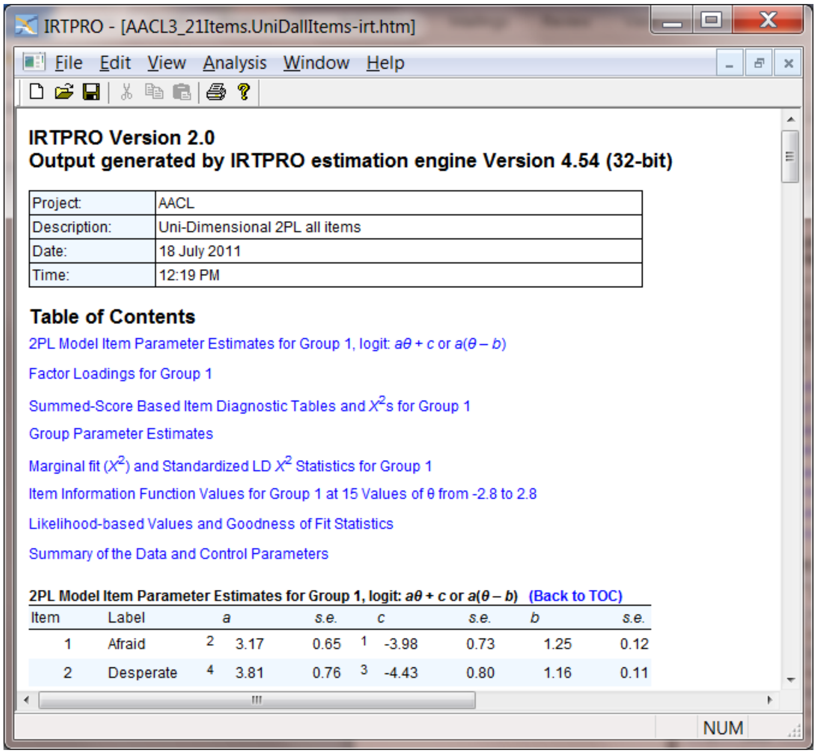


Click **OK** when the changes have been made. The resulting **Models** window is shown below.



Clicking the **Run** button in the lower left of the **Unidimensional Analysis** dialog produces the results (after a wait of only a few seconds in this case, for this small problem; other problems may require more time). The results are displayed in web-page format. A portion of the output containing the table of contents and the estimated parameters is shown below.

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.



IRTPRO Version 2.0
Output generated by IRTPRO estimation engine Version 4.54 (32-bit)

Project:	AACL
Description:	Uni-Dimensional 2PL all items
Date:	18 July 2011
Time:	12:19 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	0.65	1	-3.98	0.73	1.25	0.12
2	Desperate	4	3.81	0.76	3	-4.43	0.80	1.16	0.11

The table of parameter estimates is shown on the next 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). To be rendered comparable to normal ogive discrimination parameters, the IRTPRO estimates of the a parameters could be divided by 1.7.

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	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 & 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 χ^2 statistics based on the local dependence statistic proposed by Chen & 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 may 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 (X2) and Standardized LD X2 Statistics for Group 1   (Back to TOC)

Item	Marginal 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

Item	Marginal 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

6. Multiple groups analysis and DIF

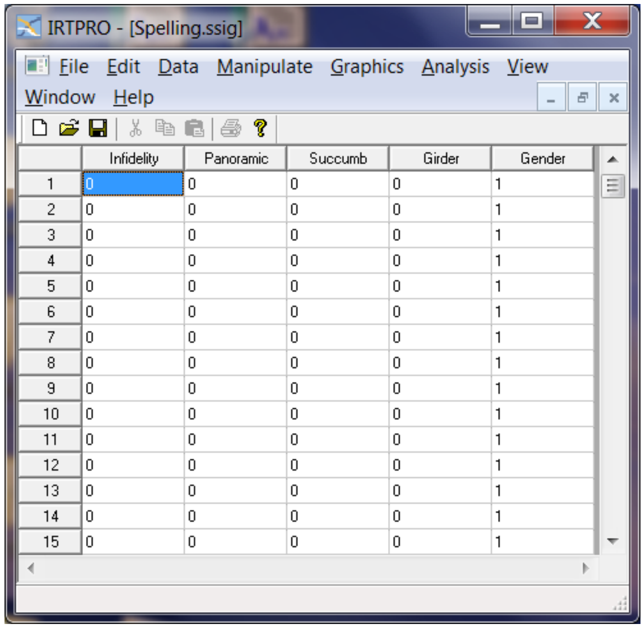
6.1 Detection of differential item functioning

The trace line model has many uses, but Lord (1977, 1980) observed that the trace line is ideally suited to defining differential item functioning (DIF), or lack thereof.

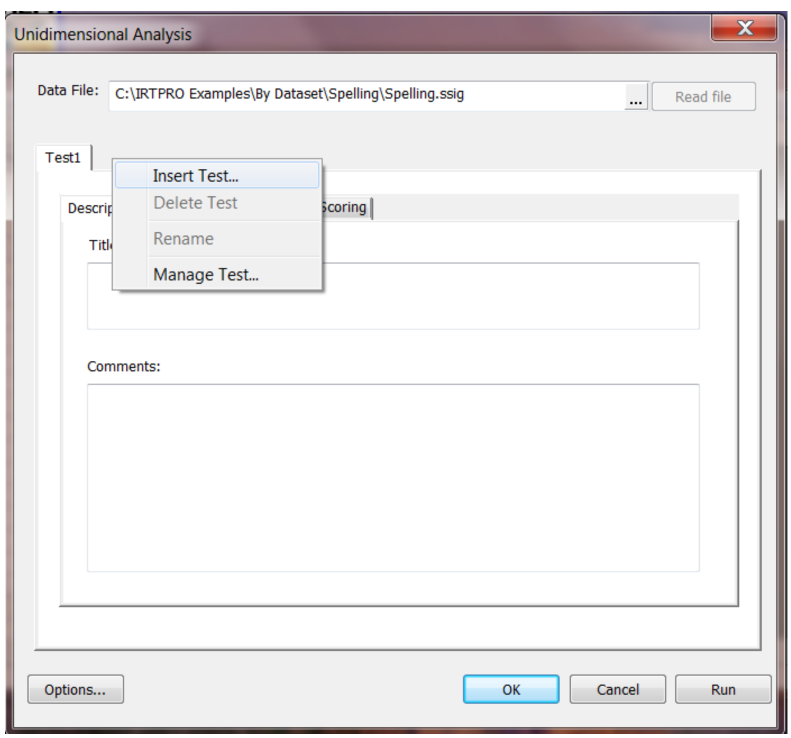
The value of the trace line at each level of θ is the conditional probability of a correct response given that level of ability of proficiency. If we are considering the possibility that an item may function differently (exhibit DIF) for some focal group relative to some reference group, then in the context of IRT we are considering whether the trace lines differ for the two groups. If the trace lines are the same, there is no DIF. If the trace lines differ, there is DIF. Because the trace line for an item is determined by the item parameters, Lord (1977, 1980) noted that the question of DIF detection could be approached by comparing estimates of the item parameters between groups.

We illustrate DIF with data obtained from 659 undergraduates at the University of Kansas. The data obtained are based on a conventional orally administered spelling test. The reference group for this analysis includes the male students (N = 285) and the focal group is made up of the female students (N = 374). The original test comprised 100 words, but only 4 have been selected for use here. The four words to be spelled are Infidelity, Panoramic, Succumb, and Girder. These four items were selected because preliminary analysis suggested that they have very nearly equal discrimination parameters; this is convenient for purposes of illustration. The data were free-response, so there is no guessing to be considered. The words Infidelity, Panoramic and Succumb were selected to comprise an "anchor" (a set of items believed to involved no DIF) with information over a range of the θ–continuum. The word Girder is the studied item; it was selected because it shows substantial differential difficulty for the two groups in these data. Thissen, Steinberg, and Wainer (1993) describe several DIF analyses of these data.

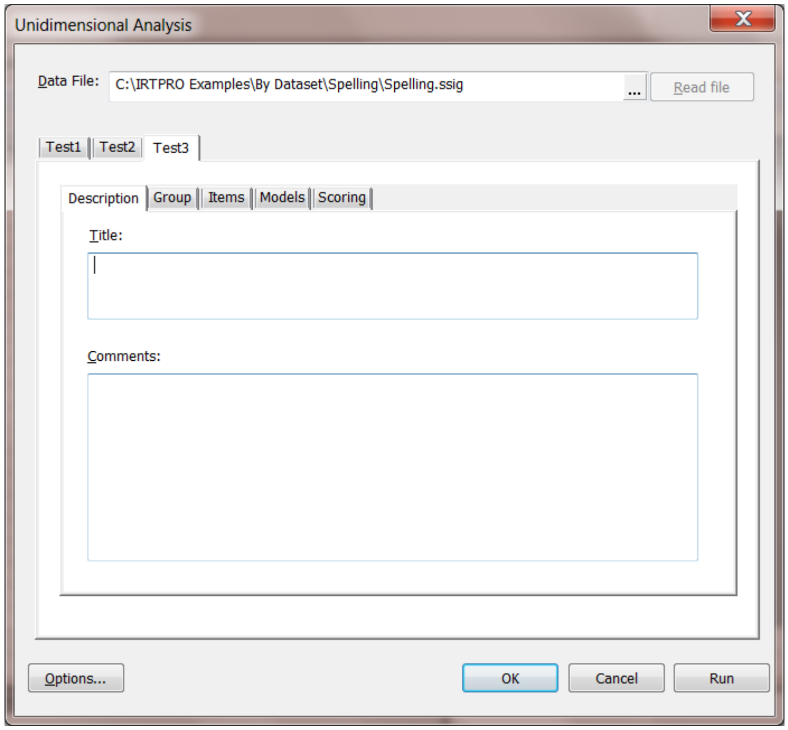
To see the data, use the **Open** file dialog under the **File** menu of IRTPRO, navigate to the **C:\IRTPRO Examples\By Dataset\Spelling** folder, select **Files of type: IRTPRO Data File (*.ssig)** in the **Open File** dialog, and open the file **Spelling.ssig**. The first 15 cases of the data are as follows:



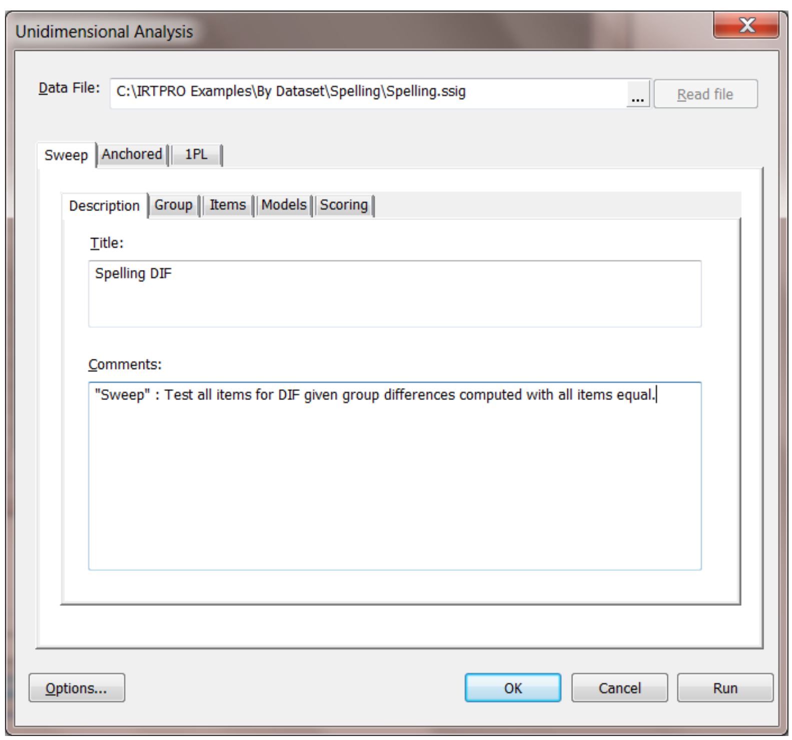
To initiate the unidimensional IRT analysis, select **Unidimensional IRT …** under the **Analysis** menu. We start by adding two additional tests in order to fit a total of three models to the same data. By right-clicking on the right side of the **Test1** tab, a pop-up menu is obtained that enable the user to enter a new test, delete a test or rename an existing test, the default test names being **Test1, Test2, …** Once the **Test2** tab is displayed, the procedure is repeated to obtain **Test1, Test2,** and **Test3.**



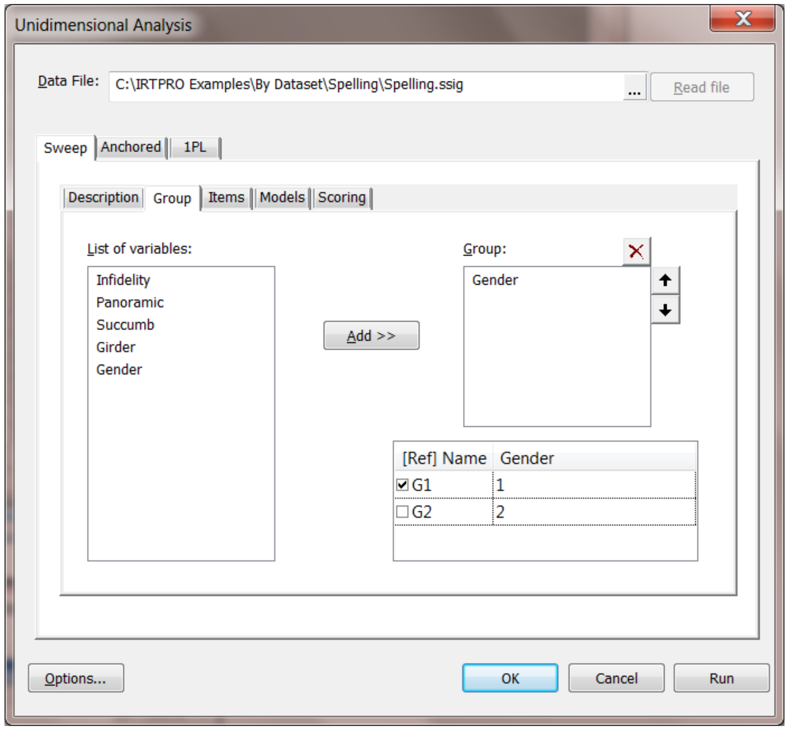
Right-click on each of the test tabs to invoke the **Insert Test…**, **Delete Test**, etc. drop-down menu. Click the **Rename** button and replace Test1 with Sweep, Test2 with Anchored, and Test3 with 1PL, the latter being more descriptive of the type of models to be fitted to the item response data.



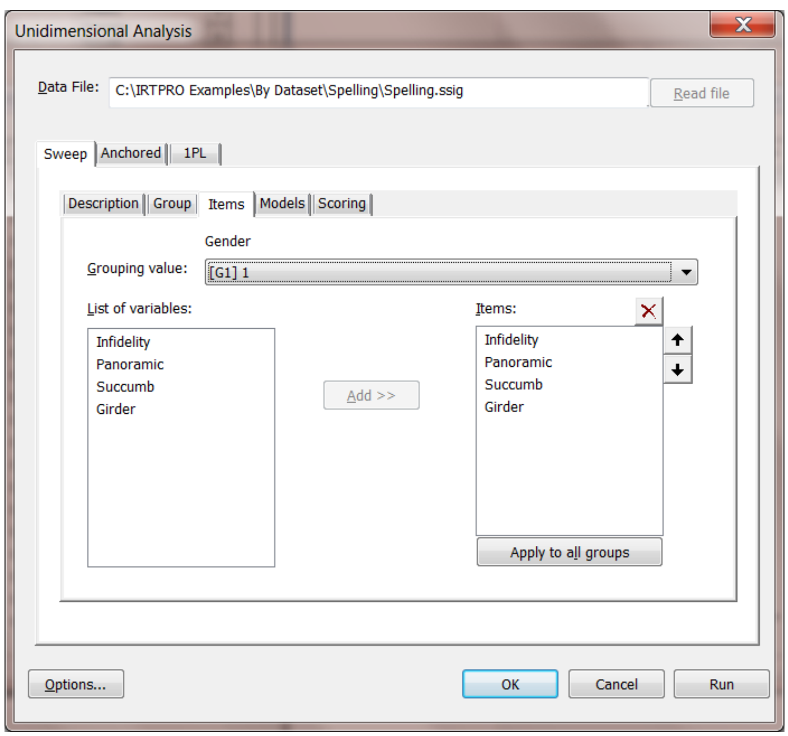
Start by clicking the **Sweep** tab to create the first test. Click the **Description** tab (the default tab for a new analysis) to add a title and comments.



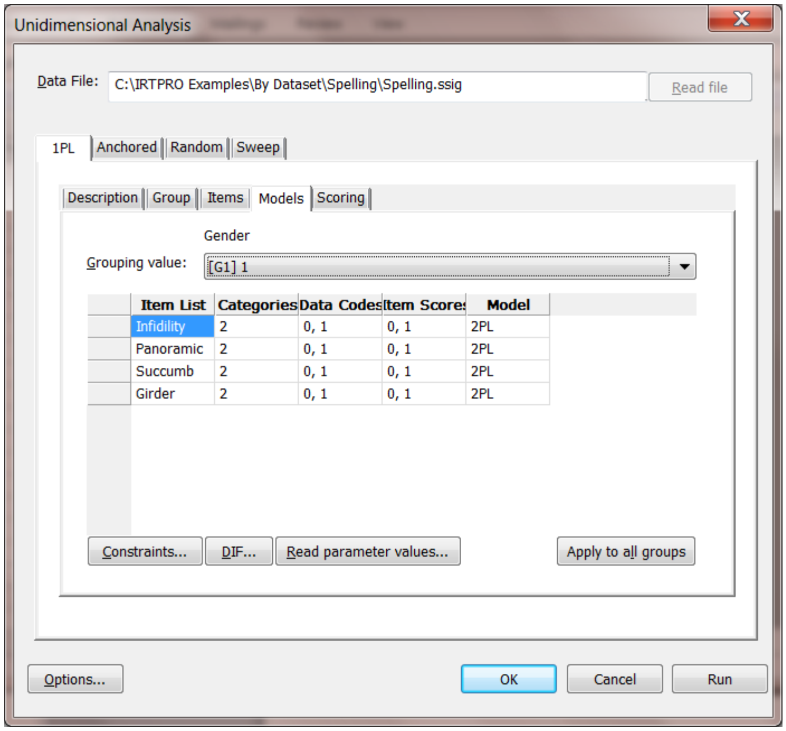
Next, click the **Groups** tab and select Gender as the grouping variable. The default reference group is the first group as shown below.



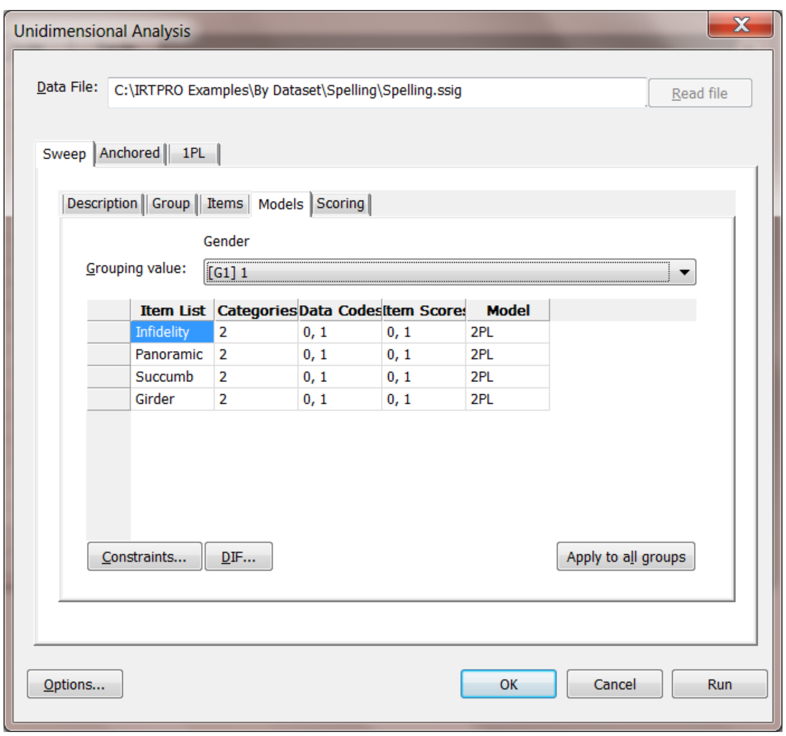
Use the **Items** tab in the **Sweep** tab in the **Unidimensional Analysis** dialog to select the four items to be analyzed, and click on the **Apply to all groups** button to select those items for both groups:



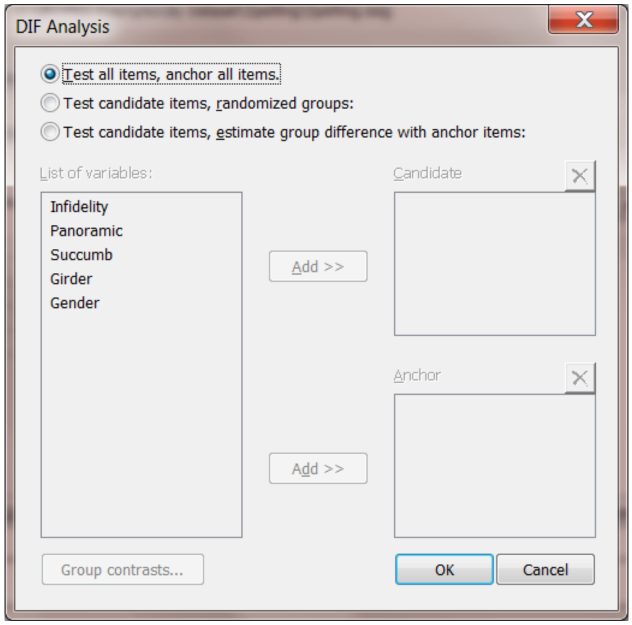
Proceeding to the **Models** tab, it is seen that all four items have two categories and that the 2PL model is displayed as the default for each item.



In the **Models** window, click the **DIF…** button:



This procedure brings up the DIF dialog shown next.



In this analysis we accept the default **Test all items, anchor all items**, so we click the **OK** button to return to the **Models** window and click the **Run** button to initiate the analysis. Portions of the output are as follows.

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	Infidelity	2	0.88	0.26	1	1.30	0.18	-1.48	0.37
2	Panoramic	4	1.48	0.35	3	0.78	0.19	-0.52	0.14
3	Succumb	6	1.86	0.54	5	-1.19	0.27	0.64	0.13
4	Girder	8	1.44	0.36	7	0.72	0.19	-0.50	0.14

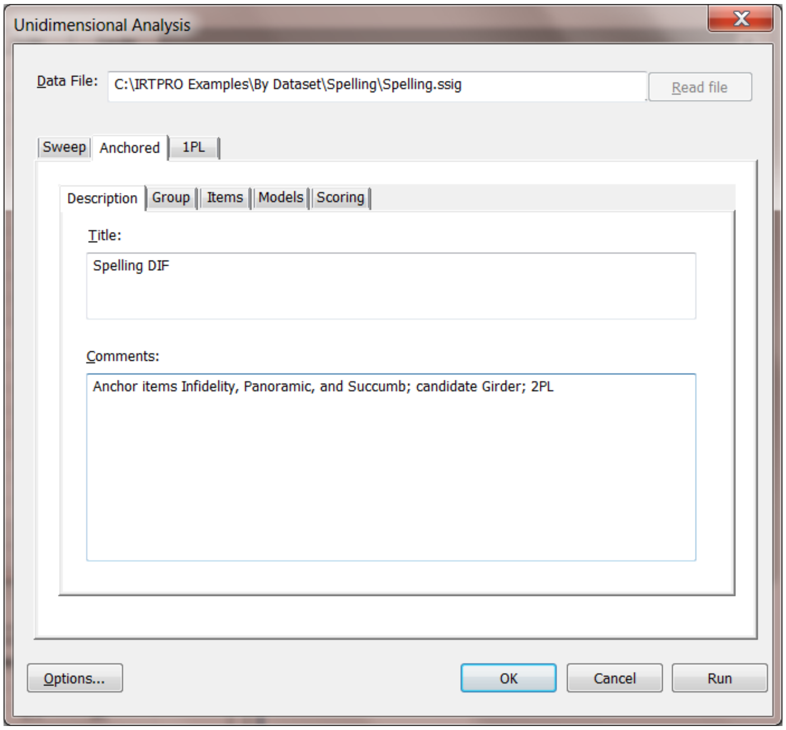
2PL Model Item Parameter Estimates for Group 2, logit: $a\theta + c$ or $a(\theta - b)$ **(Back to TOC)**

Item	Label	a		s.e.	c		s.e.	b	s.e.
1	Infidelity	10	1.62	0.43	9	2.29	0.34	-1.41	0.22
2	Panoramic	12	1.51	0.37	11	0.94	0.19	-0.62	0.11
3	Succumb	14	1.14	0.29	13	-0.78	0.13	0.69	0.21
4	Girder	16	1.64	0.43	15	0.14	0.16	-0.09	0.09

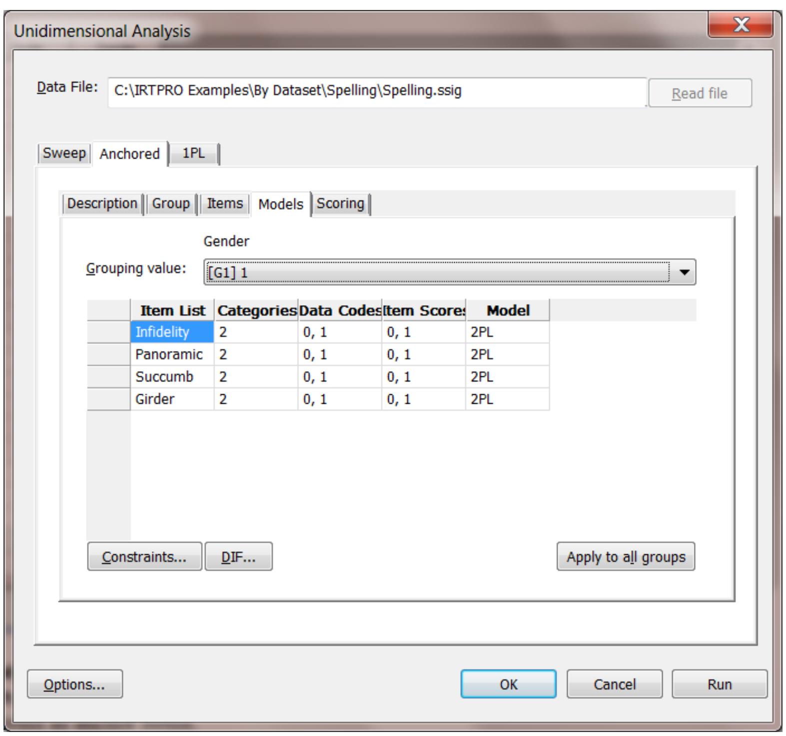
DIF Statistics for Graded Items **(Back to TOC)**

Item numbers in:										
Group 1	Group 2	Total X^2	d.f.	p	X^2_a	d.f.	p	$X^2_{c\|a}$	d.f.	p
1	1	7.5	2	0.0235	2.2	1	0.1348	5.3	1	0.0219
2	2	0.5	2	0.7765	0.0	1	0.9516	0.5	1	0.4787
3	3	2.1	2	0.3544	1.4	1	0.2393	0.7	1	0.4069
4	4	8.2	2	0.0167	0.1	1	0.7164	8.1	1	0.0045

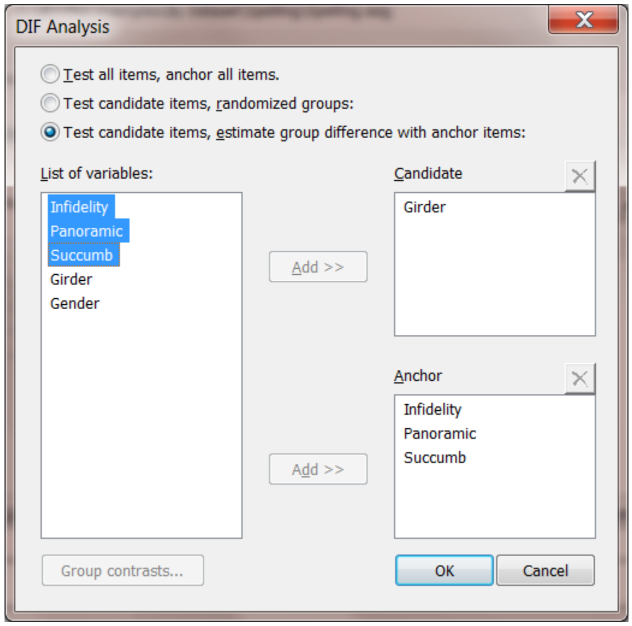
To return to the **Unidimensional Analysis** environment, close the output window and select this option from the **Analysis** option on the main menu bar. To proceed to setting up the second test, click the **Anchored** Tab and select **Description** to provide a title and comments. From the **Group** tab, Gender is again selected as the grouping variable and using the **Items** tab all four items are selected.



Proceeding to the **Models** tab, it is seen that each item is associated with the 2PL model.



Next, we click the **DIF** button to display the **DIF Analysis** window. Once this window is displayed, the **Test candidate items, estimate group difference with anchor items:** option is selected. Next Girder is selected as candidate item and the remaining three items as anchor items.



Once done, click the **OK** button to return to the **Models** window and click the **Run** button to initiate the analysis. Portions of the output are as follows.

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	Infidelity	2	1.18	0.24	1	1.59	0.16	-1.34	0.24
2	Panoramic	4	1.47	0.28	3	0.67	0.16	-0.46	0.11
3	Succumb	6	1.51	0.30	5	-1.09	0.17	0.72	0.13
4	Girder	8	1.47	0.37	7	0.73	0.19	-0.50	0.13

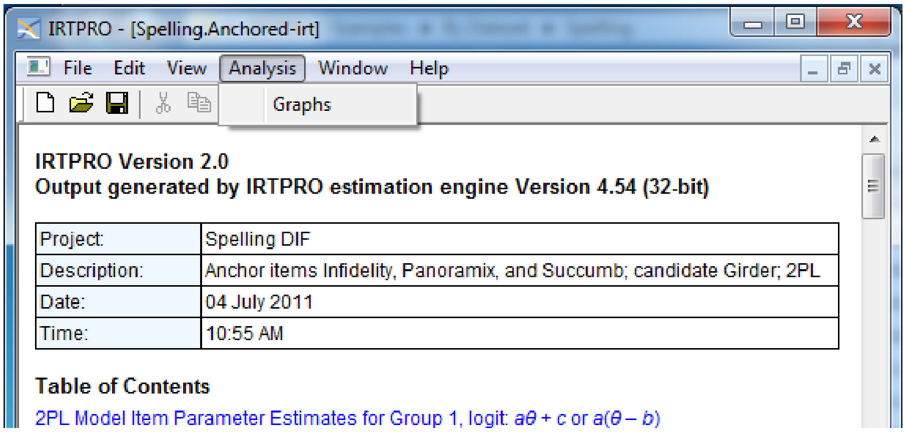
2PL Model Item Parameter Estimates for Group 2, logit: $a\theta + c$ or $a(\theta - b)$ **(Back to TOC)**

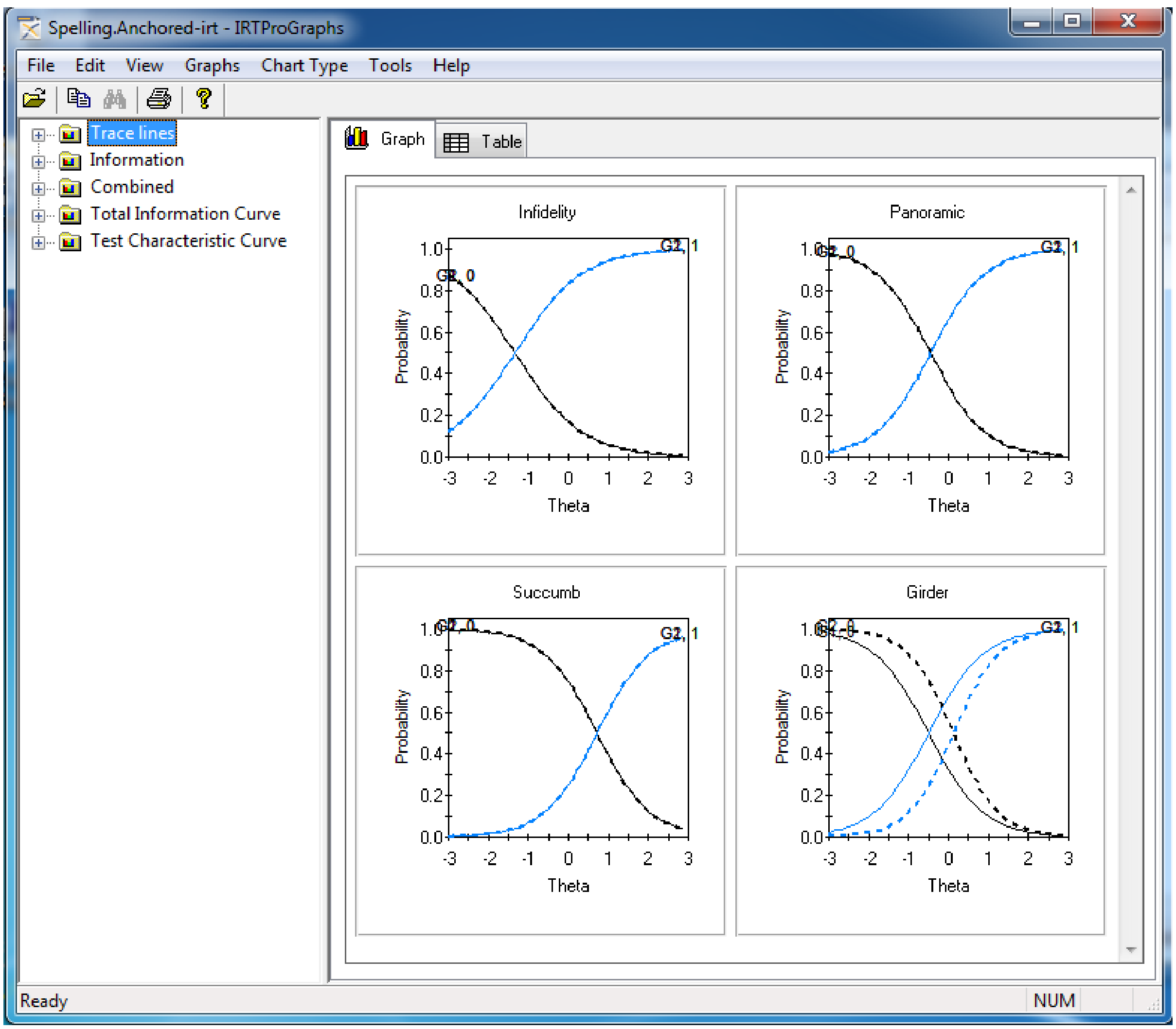
Item	Label	a		s.e.	c		s.e.	b	s.e.
1	Infidelity	2	1.18	0.24	1	1.59	0.16	-1.34	0.24
2	Panoramic	4	1.47	0.28	3	0.67	0.16	-0.46	0.11
3	Succumb	6	1.51	0.30	5	-1.09	0.17	0.72	0.13
4	Girder	10	1.77	0.54	9	-0.21	0.22	0.12	0.12

DIF Statistics for Graded Items **(Back to TOC)**

Item numbers in:										
Group 1	Group 2	Total X^2	d.f.	p	X^2_a	d.f.	p	$X^2_{c\|a}$	d.f.	p
4	4	11.2	2	0.0036	0.2	1	0.6413	11.0	1	0.0009

While the output is displayed, the **Graphs** option is available under the **Analysis** item on the main menu bar.

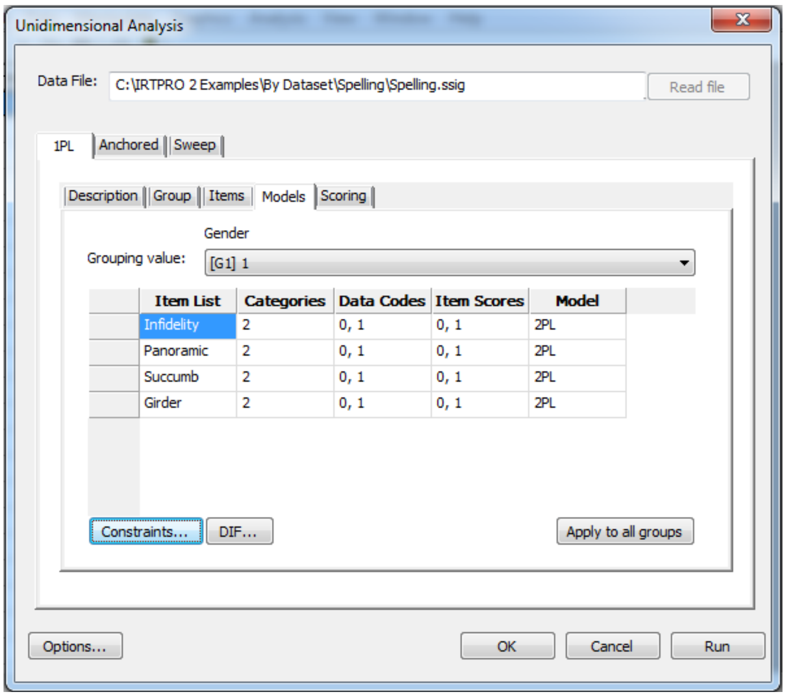




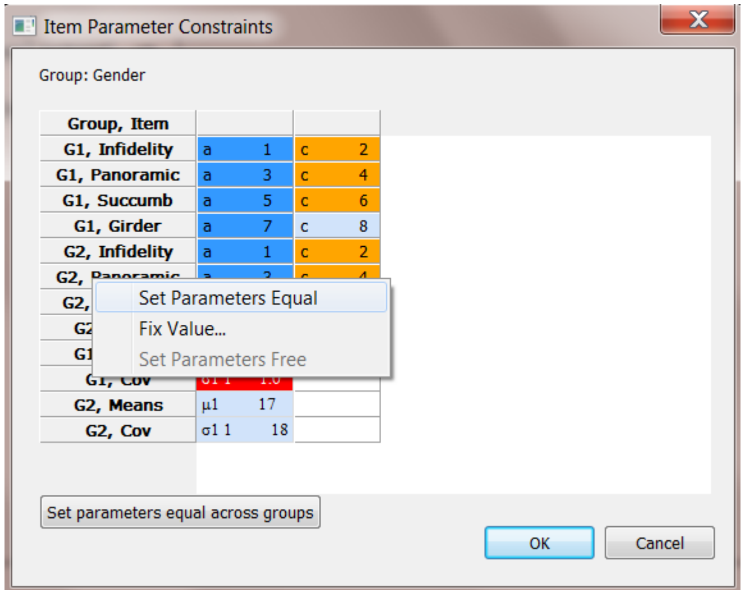
To return to the **Unidimensional Analysis** environment, close the graphics window (if it is open) and the output window, and select this option from the **Analysis** option on the main menu bar. To proceed to setting up the third test, click the **1PL** Tab.

Use the **Description** tab to provide a title and comments. From the **Group** tab, Gender is again selected as the grouping variable and using the **Items** tab all four items are selected.

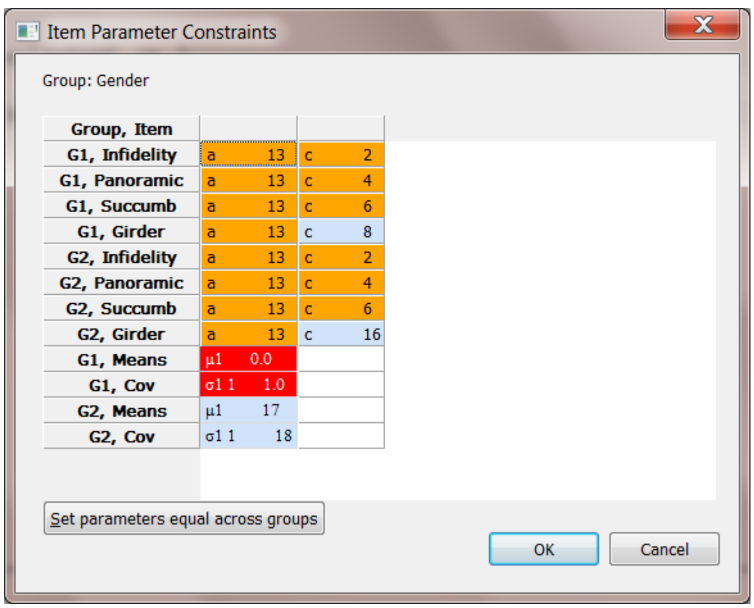
Next, we click the **DIF** button to display the **DIF Analysis** window. Once this window is displayed, just as in the Anchored example, the **Test candidate items, estimate group difference with anchor items:** option is selected. Next Girder is selected as candidate item and the remaining three items as anchor items. Click **OK** to return to the **Models** tab.



To use the 1PL model for the analysis, emulating Thissen, Steinberg, and Wainer (1993), we next click the **Constraints** button in the **Models** tab to bring up the **Constraints** dialog, and set all the slope (a) parameters for both groups equal:



Note that it is important that such constraints are added after the DIF dialog, so that the constraints are in addition to a set of constraints automatically imposed by the anchored- DIF selection:



Once done, click the **OK** button to return to the **Models** window and click the **Run** button to initiate the analysis. Portions of the output are as follows.

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	Infidelity	6	1.41	0.15	1	1.70	0.16	-1.20	0.14
2	Panoramic	6	1.41	0.15	2	0.66	0.15	-0.47	0.10
3	Succumb	6	1.41	0.15	3	-1.08	0.14	0.76	0.13
4	Girder	6	1.41	0.15	4	0.72	0.17	-0.51	0.12

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

Item	Label	a	s.e.	c	s.e.	b	s.e.
1	Infidelity	[6] 1.41	0.15	[1] 1.70	0.16	-1.20	0.14
2	Panoramic	[6] 1.41	0.15	[2] 0.66	0.15	-0.47	0.10
3	Succumb	[6] 1.41	0.15	[3] -1.08	0.14	0.76	0.13
4	Girder	[6] 1.41	0.15	[5] -0.21	0.18	0.15	0.13

DIF Statistics for Graded Items (Back to TOC)

Item numbers in:										
Group 1	Group 2	Total X^2	d.f.	p	X^2_a	d.f.	p	$X^2_{c\|a}$	d.f.	p
4	4	14.4	2	0.0007	0.0	1	1.0000	14.4	1	0.0001

These DIF statistics are slightly deceptive; they list a two degrees of freedom total X^2, and a X^2 test between the *a* parameters even though the *a* parameters are constrained equal (so the latter test must be 0.0), and there is only one degree of freedom for the test of the *b* parameter, which is the same as the overall test in this case. The software is not smart enough to count arbitrary constraints when doing DIF tests, and the user has to correct these things.

6.2 Analysis of the 2000 Program for International Student Assessment (PISA) data

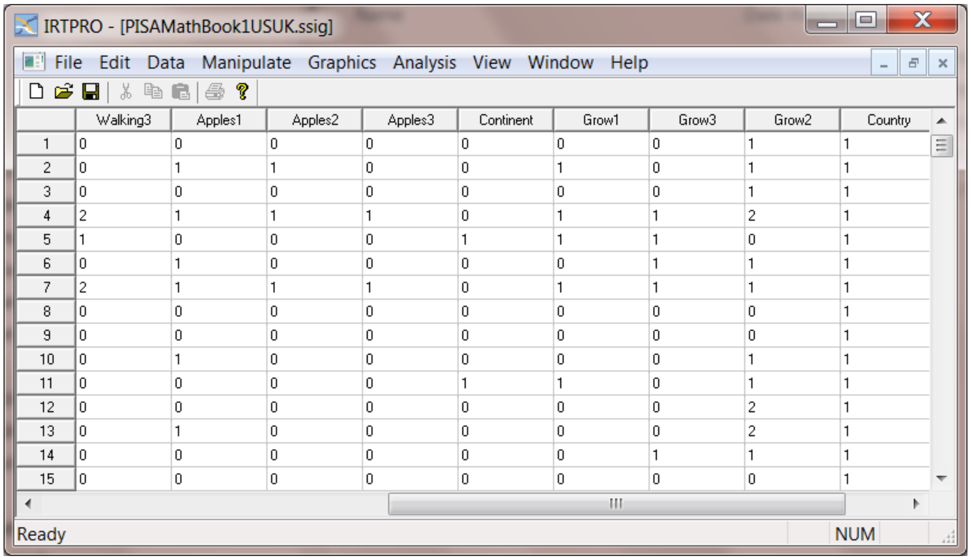
The Program for International Student Assessment (PISA) is a worldwide evaluation of 15-year-old school pupils' scholastic performance, performed first in 2000 and repeated every three years. It is coordinated by the Organization for Economic Co-operation and Development (OECD), with a view to improving educational policies and outcomes.

In this section, three analyses based on the 2000 PISA (see Adams & Wu, 2002) data are discussed: traditional statistics, unidimensional IRT and unidimensional Rasch. Note that our "Rasch" model is Thissen's (1982) Rasch model, which differs from the traditional Rasch model where all slopes are assumed equal to 1.0. For the unidimensional analysis, 2PL (items with two categories) and GPCredit (items with more than two categories) models were fitted. Furthermore, the item parameters across groups were set equal and the mean and variance of the UK group relative to the US reference group were freely estimated. In the Rasch analysis, we imposed additional equality constraints across items (all items have the same slope). This turns the General Partial Credit (GPC) model into Partial Credit (PC) model. Fit indices and log-likelihood all point to the less constrained IRT model as a better fitting model when compared to the Rasch model.

The dataset contains responses by a subset of students to 14 items from math booklet 1. The grouping variable is Country, defined as follows: group 1 is the United States (US) and Group 2 is the United Kingdom (UK). There are 358 students in the US group and 889 in the UK group.

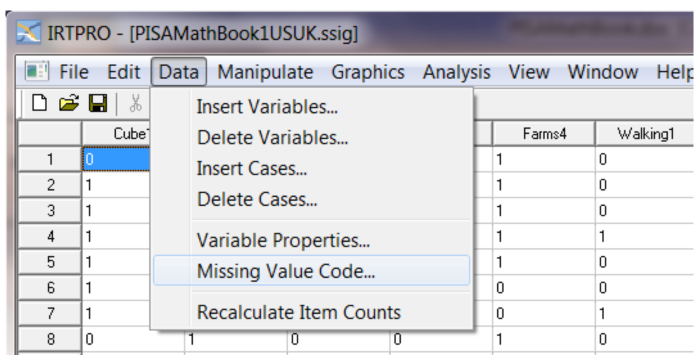
A fourth analysis, testlet response theory (TRT), based on this dataset is given in Section 7.2. For the TRT analysis, a multidimensional model for more than one group is fitted. The mean/variance of the primary math dimension is freely estimated in the UK group and the additional dimensions are there to account for local dependence among items in the same testlet.

The dataset **PISAMathBook1USUK.ssig** is located in the folder **IRTPRO Examples\By Data Set\PISA MathBook1** and when opened is displayed as a spreadsheet. Below we show the first 15 cases for items Walking3 to Grow2 and the grouping variable Country.

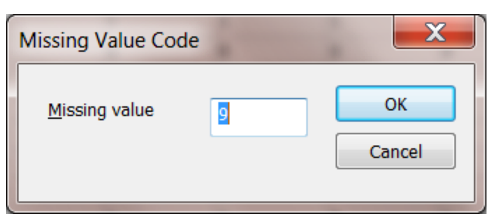


	Walking3	Apples1	Apples2	Apples3	Continent	Grow1	Grow3	Grow2	Country
1	0	0	0	0	0	0	0	1	1
2	0	1	1	0	0	1	0	1	1
3	0	0	0	0	0	0	0	1	1
4	2	1	1	1	0	1	1	2	1
5	1	0	0	0	1	1	1	0	1
6	0	1	0	0	0	0	1	1	1
7	2	1	1	1	0	1	1	1	1
8	0	0	0	0	0	0	0	0	1
9	0	0	0	0	0	0	0	0	1
10	0	1	0	0	0	0	0	1	1
11	0	0	0	0	1	1	0	1	1
12	0	0	0	0	0	0	0	2	1
13	0	1	0	0	0	0	0	2	1
14	0	0	0	0	0	0	1	1	1
15	0	0	0	0	0	0	0	0	1

Missing values in the data are coded as 9. To define the missing code, select the **Missing Value Code…** from the **Data** drop-down menu.

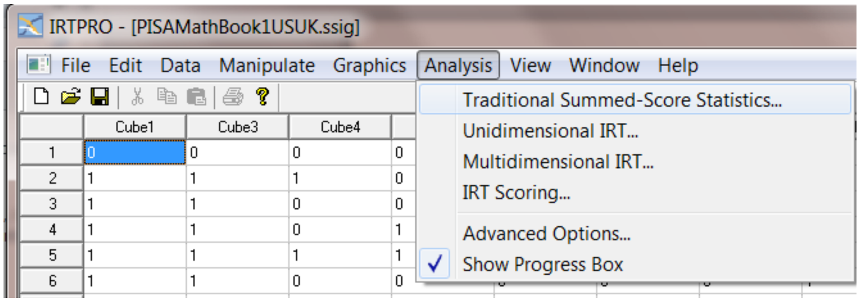


Enter the missing value and click **OK** when done. Save the **.ssig** file to make this change permanent.

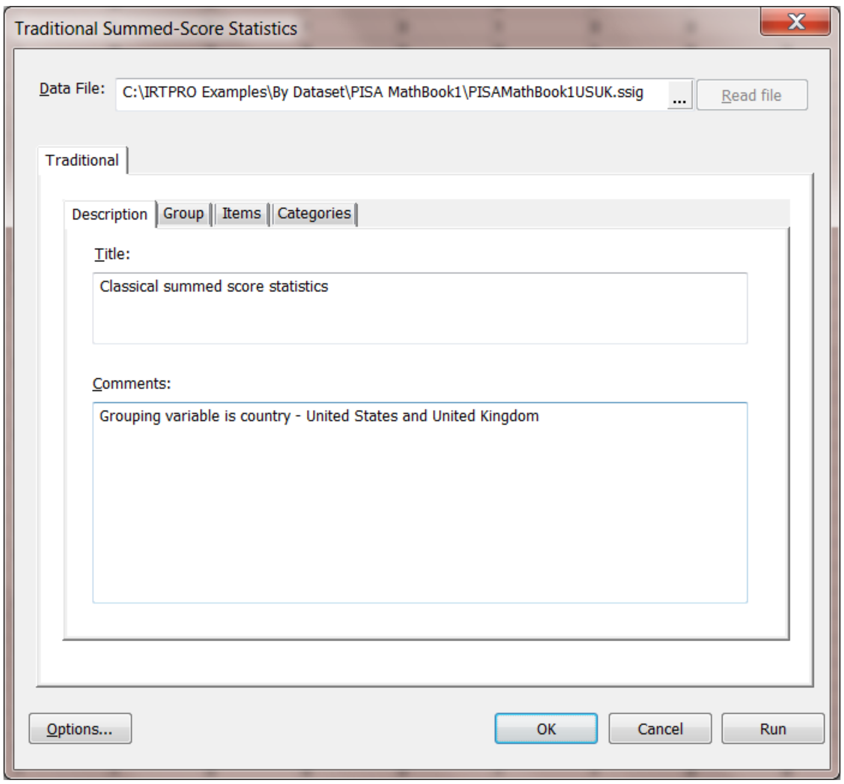


6.2.1 Traditional Statistics

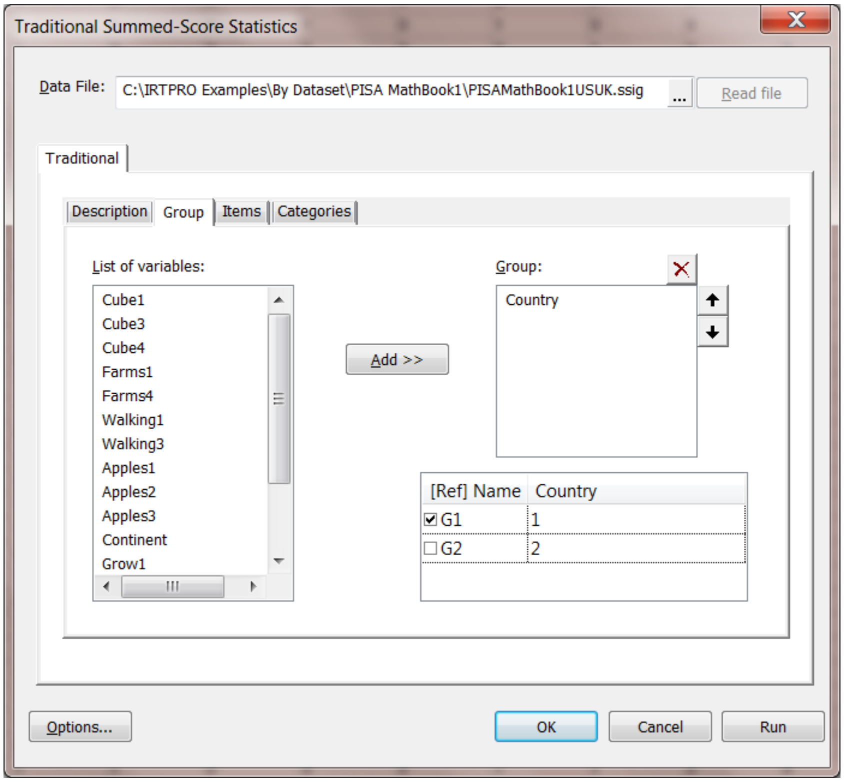
To view the statistics for these data, select **Traditional Summed-Score Statistics …** from the **Analysis** menu.



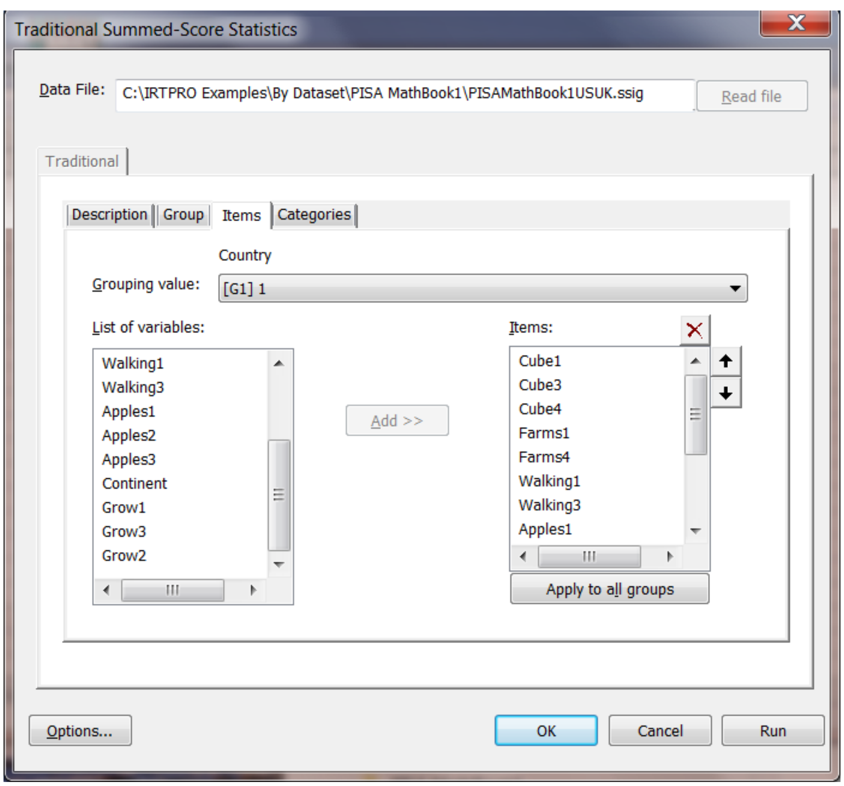
Right click on the **Test1** tab and rename **Test1** to **Traditional**. The **Traditional Summed-Score Statistics** dialog appears. Enter the title and comments in the **Description** tab as shown below.



Proceed to the **Group** tab and select Country form the list of variables. The reference group (by default) is the United States (Country = 1).

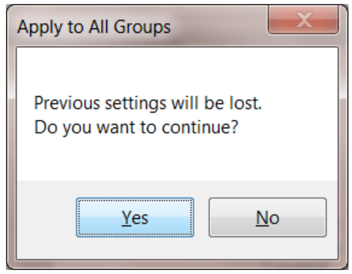


Next we proceed to the **Items** tab and select all 14 items for the first group.

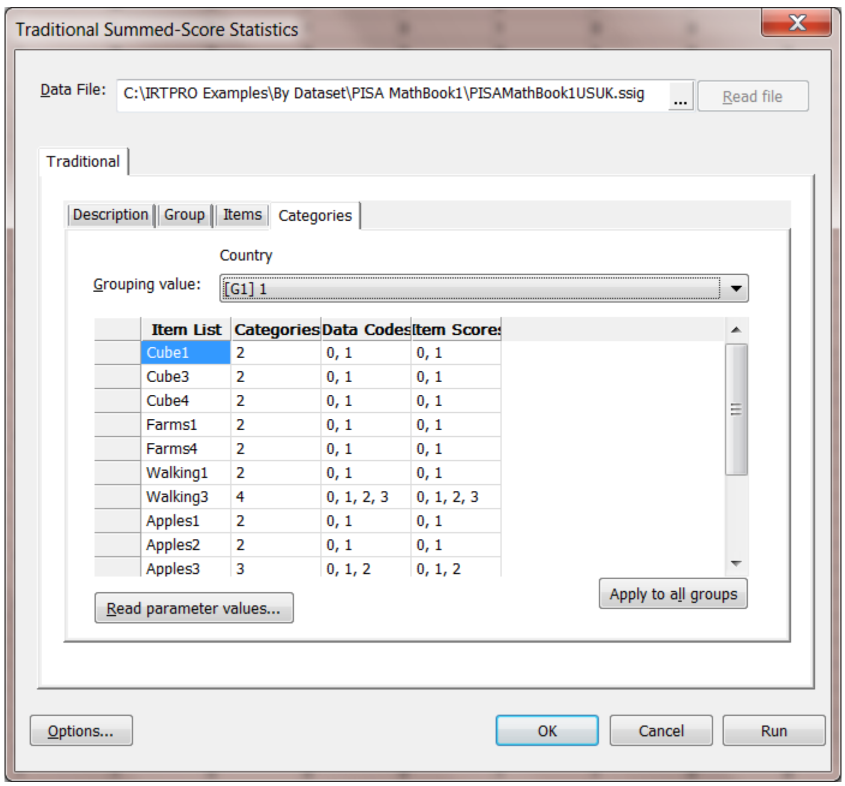


Since the responses from the second group (UK) are based on the same 14 items, click the **Apply to all groups** button to automatically select these items. Select the **Yes** option from the **Apply to All**

Groups pop-up message box.



IRTPRO computes the number of categories and associated values for each item. By clicking the **Categories** tab, these values are displayed as shown next for the first group. To see the corresponding values for the second group, this group may be selected from the **Grouping value:** drop down list.



When the **Run** button is clicked, the output appears, excerpts of which are listed below.

IRTPRO Version 2.0
Output generated by IRTPRO estimation engine Version 4.54 (32-bit)

Project:	Classical summed score statistics
Description:	Grouping variable is country - United States and United Kingdom
Date:	19 May 2011
Time:	12:33 PM

Table of Contents

Coefficient Alpha, calculated using listwise deletion is 0.8317 for the US group.

Item and (Weighted) Summed-Score Statistics for Group 1 **(Back to TOC)**
Coefficient alpha: 0.8317
Complete data N: 357

The table below is a summary of the coefficient Alpha, if each item in turn is deleted. For example, if item 7 is deleted, the reliability coefficient based on the remaining 13 items equals 0.8151.

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	0.557	0.497	0.4262	0.8234
2	0.754	0.432	0.3913	0.8255
3	0.333	0.472	0.4115	0.8243
4	0.462	0.499	0.5342	0.8166
5	0.594	0.492	0.4546	0.8217
6	0.238	0.427	0.5609	0.8165
7	0.524	0.713	0.5510	0.8151
8	0.499	0.501	0.4629	0.8211
9	0.227	0.419	0.5859	0.8153
10	0.232	0.589	0.5044	0.8181
11	0.345	0.582	0.4890	0.8193
12	0.443	0.497	0.4930	0.8192
13	0.507	0.501	0.4216	0.8237
14	1.168	0.757	0.3577	0.8341

The tables for Item 1 and Item 7 below give the frequency count for each category of an item as well as the number of missing values for the item in question. Similar results are produced for the remaining 12 items, but are not shown here.

Item	Cube1			(Back)
1	Category:	0	1	Missing
Frequencies:		159	199	0
For listwise-complete data:				
Frequencies:		158	199	
Average (wtd) Score:		4.43	8.83	
Std. Dev. (wtd) Score:		3.10	3.94	

Results for item numbers 2 to 6 and 8 to 14 appear in the output, but are not shown here.

Item	Walking3					(Back)
7	Category:	0	1	2	3	Missing
Frequencies:		207	125	16	10	0
For listwise-complete data:						
Frequencies:		206	125	16	10	
Average (wtd) Score:		4.74	9.01	12.25	15.90	
Std. Dev. (wtd) Score:		3.05	3.43	2.82	1.52	

The corresponding results for the UK group are given below.

Item and (Weighted) Summed-Score Statistics for Group 2 **(Back to TOC)**
Coefficient alpha: 0.8175
Complete data N: 887

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

Item	Response Average	Std. Dev.	With Item Deleted Item-Total Correlation	Coefficient α
1	0.717	0.451	0.3210	0.8139
2	0.840	0.367	0.3981	0.8103
3	0.391	0.488	0.3858	0.8100
4	0.676	0.468	0.5266	0.8013
5	0.609	0.488	0.2446	0.8188
6	0.398	0.490	0.6016	0.7959
7	0.611	0.880	0.6007	0.7947
8	0.689	0.463	0.3988	0.8092
9	0.278	0.448	0.5831	0.7983
10	0.360	0.657	0.5692	0.7955
11	0.583	0.695	0.5731	0.7951
12	0.630	0.483	0.3485	0.8123
13	0.696	0.460	0.4328	0.8072
14	1.374	0.675	0.3202	0.8178

The distribution of values over categories for the second group (UK) is listed next. Results for item numbers 2 to 6 and 8 to 14 appear in the output, but are not shown here.

Item	Cube1			(Back)
1	Category:	0	1	Missing
Frequencies:		252	637	0
For listwise-complete data:				
Frequencies:		251	636	
Average (wtd) Score:		6.07	9.95	
Std. Dev. (wtd) Score:		3.46	3.97	

Item	Walking3					(Back)
7	Category:	0	1	2	3	Missing
Frequencies:		534	219	85	51	0
For listwise-complete data:						
Frequencies:		532	219	85	51	
Average (wtd) Score:		6.49	11.03	13.85	15.78	
Std. Dev. (wtd) Score:		2.96	2.86	2.40	2.20	

The final part of the output is a summary of sample sizes and number of items per group.

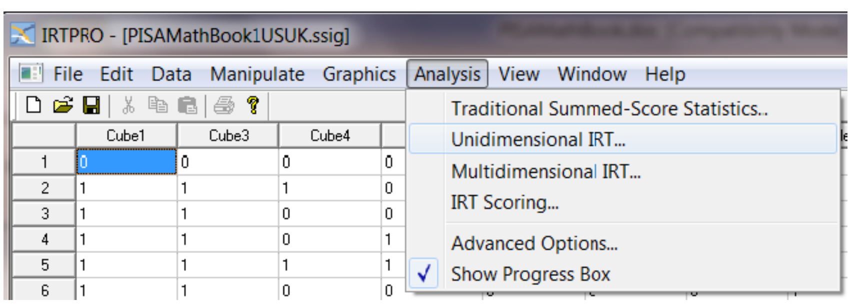
Summary of the Data and Control Parameters (Back to TOC)

Group:	Group 1	Group 2
Sample Size	358	889
Number of Items	14	14

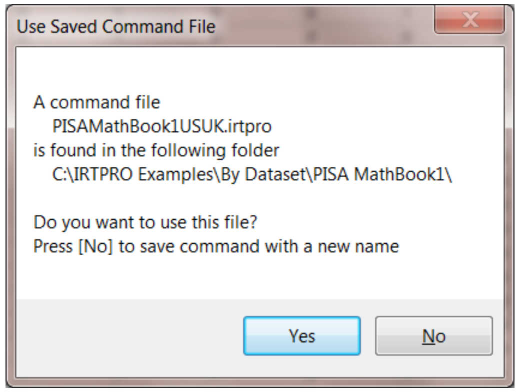
6.2.2 Unidimensional IRT

In this example a mixture of 2PL and general partial credit models are fitted to the data. Since the previous example dealt with traditional summed-score statistics, the following steps need to be followed to fit unidimensional models to the data:

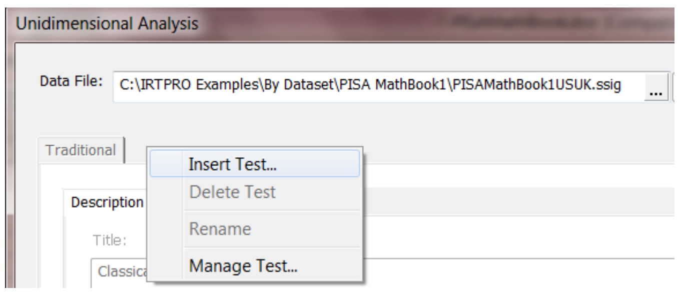
- Close the spreadsheet and then re-open **PISAMathbook1USUK.ssig**
- Select the **Analysis, Unidimensional IRT…** option from the main menu bar



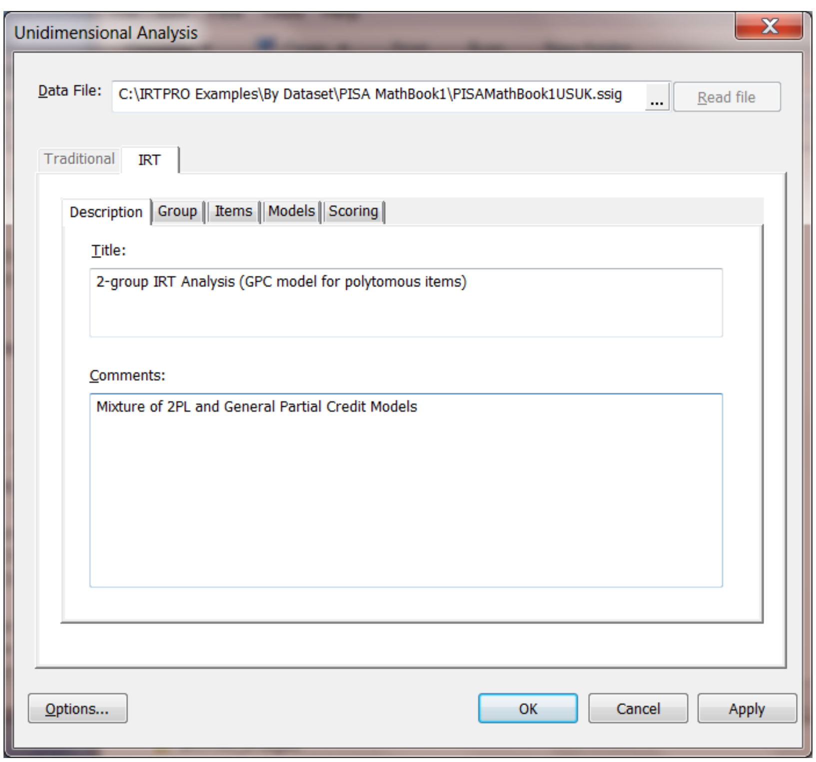
Click **Yes** to use the same command file.



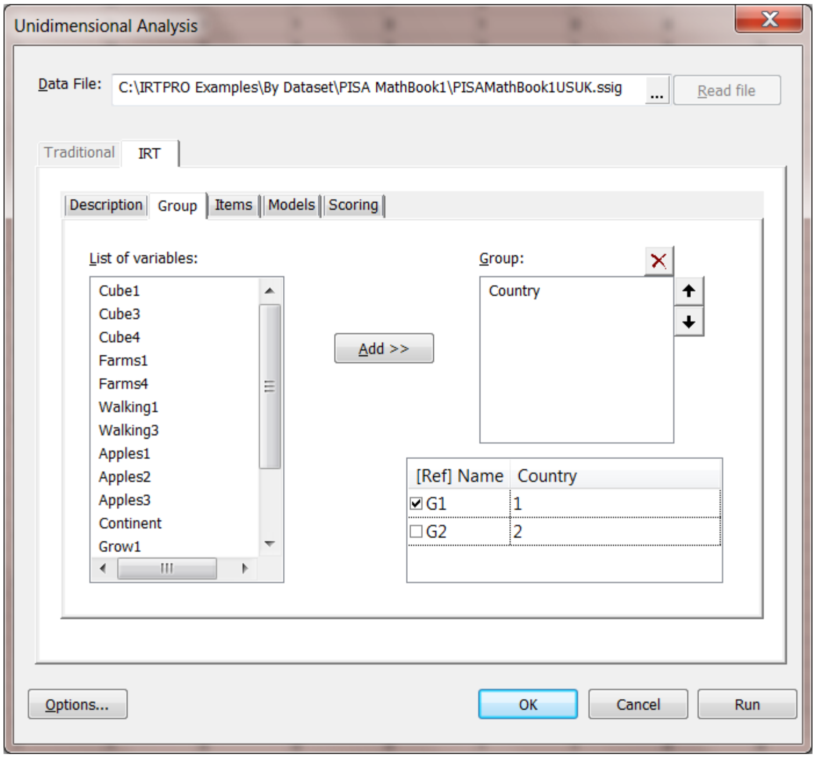
This action produces the **Unidimensional Analysis** window. To proceed, right-click next to the **Traditional** test tab (right-hand side) to insert a second test. The default tab is **Test2**. Rename to **IRT** by right-clicking this tab and then selecting the **Rename** option.



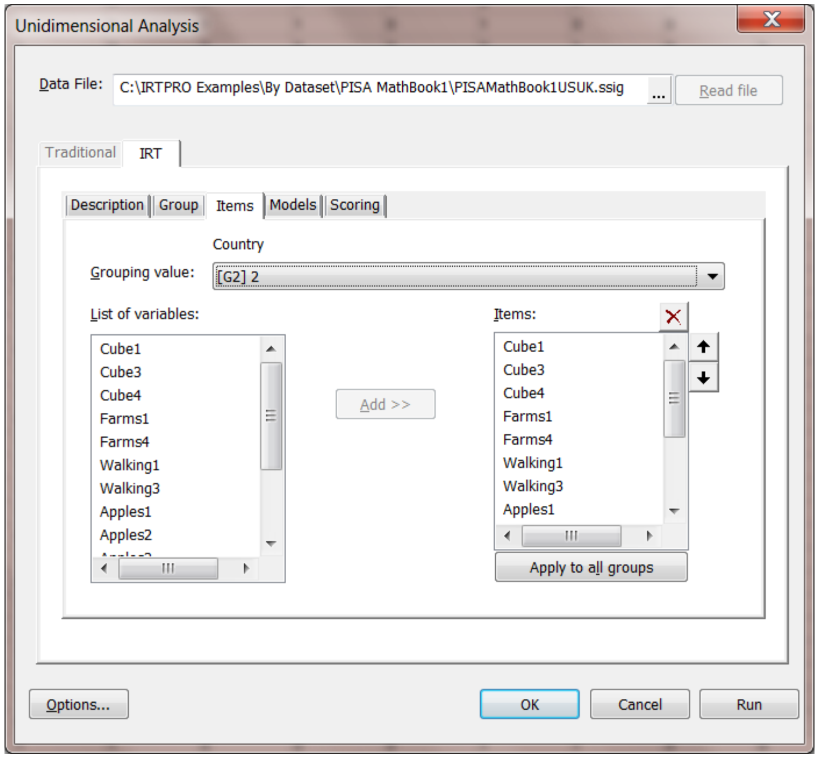
Once this is action is completed a title and comments may be added in the **Description** window as shown below.



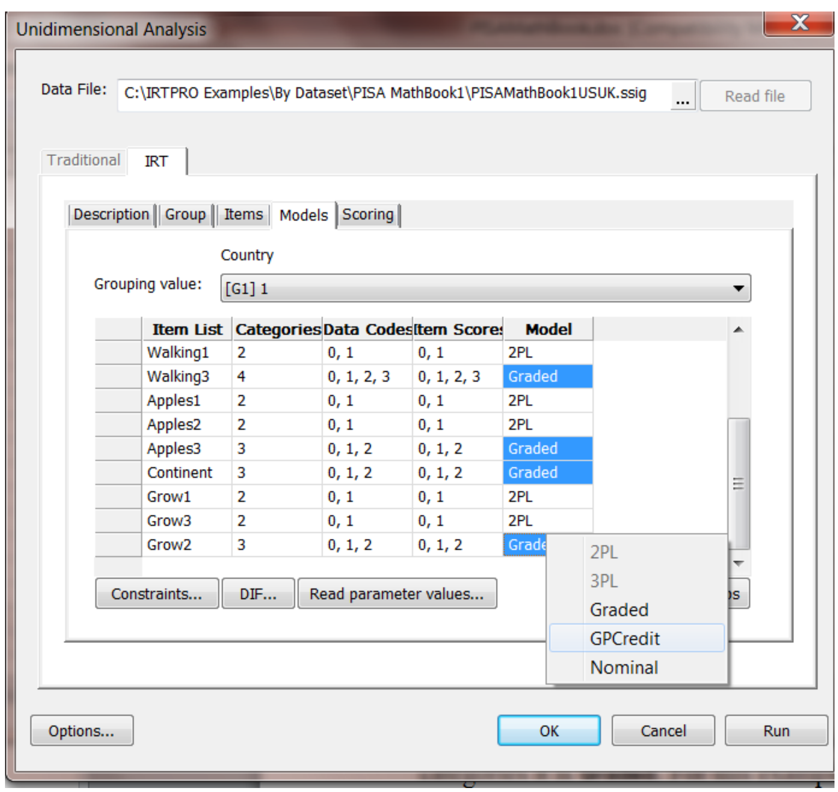
Proceed to the **Group** tab and select Country as the group variable.



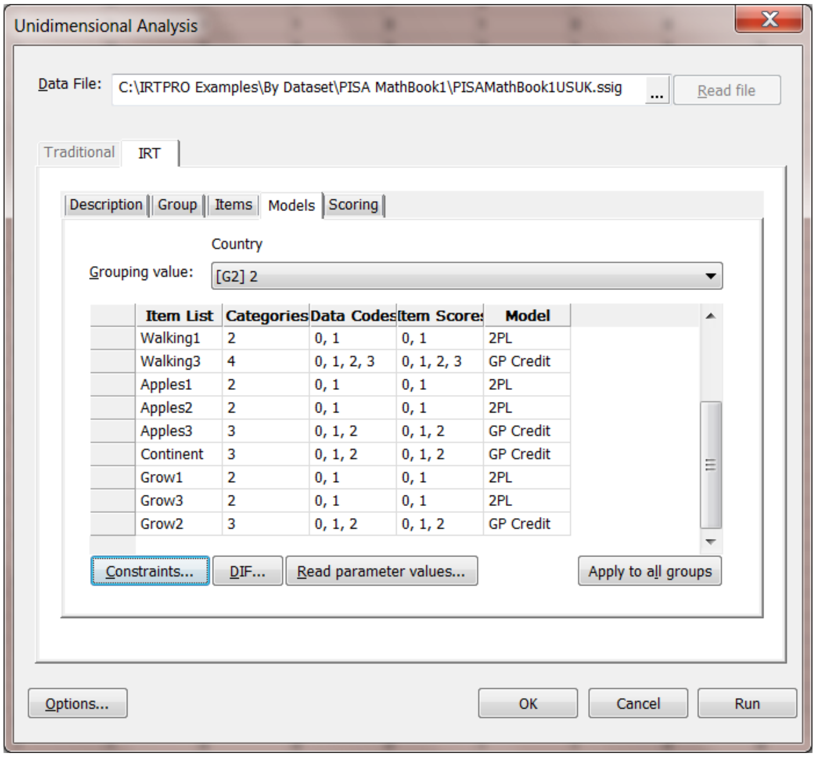
Using the **Items** tab, select all 14 items and click the **Apply to all groups** button. To demonstrate that these items were also selected for the second group, we change the value of the group variable from the first group (US) to the second group via the **Grouping variable:** drop-down list.



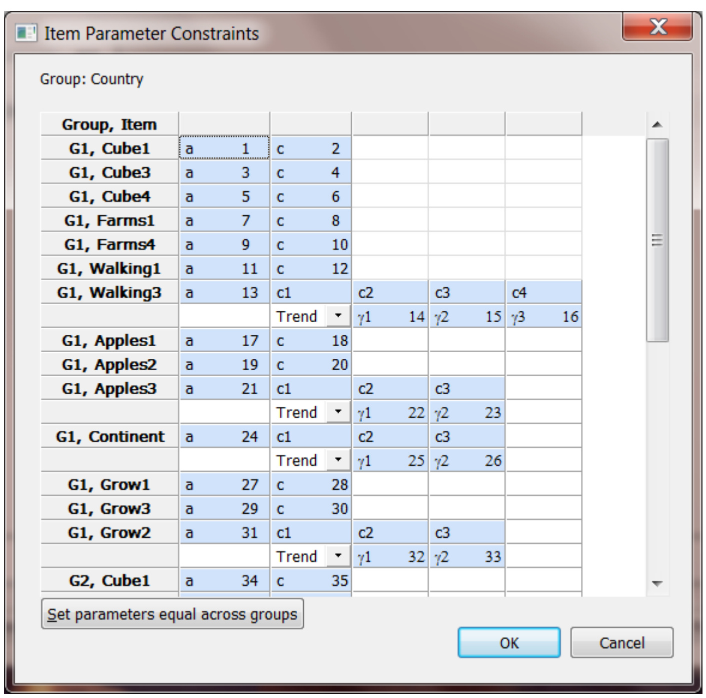
The default model for all items with 2 categories is 2PL and for those items with more than two categories it is Graded. For this example, we replace all the graded models with general partial credit models by selecting all the items in question. Next right-click on any of the selected cells and choose the GPCredit option.



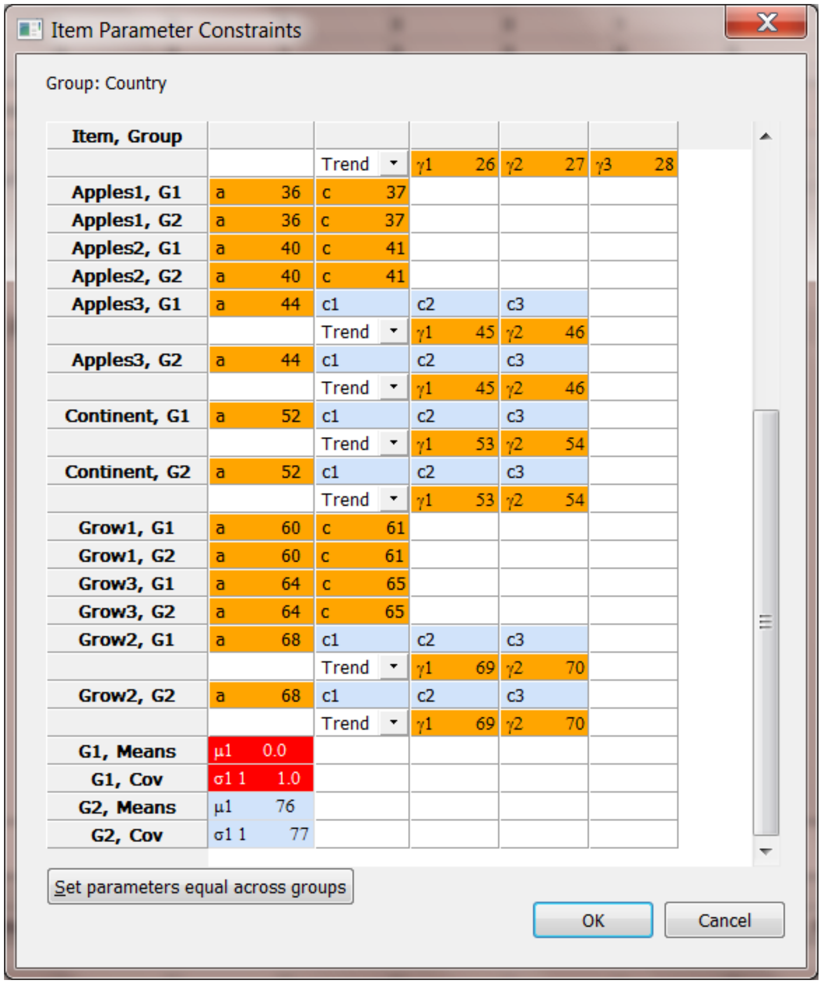
This action produces a revised **Models** window. Click **Apply to all groups**.



In this example, the item parameters across groups were set equal and the mean and variance of the UK group relative to the US reference group were freely estimated. To set parameters equal across groups, click the **Constraints…** button (see the **Models** window above.)



This action produces the **Item Parameter Constraints** window. By clicking the **Set parameters equal across groups** button IRTPRO sets corresponding parameters equal across groups. Note that this action is only performed for items that are present in all groups and have the same number of categories for each group. To get a clearer picture of the imposed contraints, double-click on the **Group, Item** tab to change the sorting to **Item, Group** as shown below. The **Item Parameter Constraints** window also shows that the mean and variance parameters of the second group (UK) are estimated freely.



Click **OK** to return to the **Models** window, then click the **Run** button to start the analysis. Portions of the output file is given below.

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	Cube1	[2] 1.11	0.17	[1] 0.47	0.12	-0.42	0.12
2	Cube3	[4] 1.57	0.22	[3] 1.48	0.17	-0.94	0.14
3	Cube4	[6] 1.12	0.18	[5] -1.04	0.13	0.93	0.20
4	Farms1	[8] 2.14	0.33	[7] -0.02	0.22	0.01	0.10
5	Farms4	[10] 0.79	0.14	[9] 0.19	0.10	-0.24	0.13
6	Walking1	[12] 2.68	0.44	[11] -2.18	0.30	0.81	0.18
8	Apples1	[18] 1.46	0.22	[17] 0.21	0.15	-0.14	0.10
9	Apples2	[20] 2.79	0.50	[19] -3.05	0.33	1.09	0.22
12	Grow1	[28] 1.24	0.20	[27] -0.06	0.13	0.05	0.11
13	Grow3	[30] 1.46	0.22	[29] 0.26	0.15	-0.18	0.10

GPC Model Item Parameter Estimates, logit: a[k(θ - b) + Σdk]

Item	Label	a	s.e.	b	s.e.	d_1	d_2	s.e.	d_3	s.e.	d_4	s.e.
7	Walking3	[13] 1.65	0.25	1.52	0.27	0.00	0.69	0.13	-0.30	0.09	-0.39	0.12
10	Apples3	[21] 2.24	0.38	1.50	0.27	0.00	0.07	0.07	-0.07	0.07		
11	Continent	[24] 1.64	0.32	1.32	0.25	0.00	0.55	0.08	-0.55	0.08		
14	Grow2	[31] 0.69	0.10	-0.66	0.14	0.00	0.94	0.19	-0.94	0.19		

Nominal Model Slopes and Scoring Function Contrasts for Group 1, logit: (a skθ + ck); s = Tα (Back to TOC)

Item	Label	a	s.e.	Contrasts	α_1	s.e.	α_2	s.e.	α_3	s.e.
7	Walking3	[13] 1.65	0.25	Trend	1.00	-----	0.00	-----	0.00	-----
10	Apples3	[21] 2.24	0.38	Trend	1.00	-----	0.00	-----		
11	Continent	[24] 1.64	0.32	Trend	1.00	-----	0.00	-----		
14	Grow2	[31] 0.69	0.10	Trend	1.00	-----	0.00	-----		

Nominal Model Scoring Function Values for Group 1, logit: (a skθ + ck); s = Tα (Back to TOC)

Item	Category	s_1	s_2	s_3	s_4
7	Walking3	0.00	1.00	2.00	3.00
10	Apples3	0.00	1.00	2.00	
11	Continent	0.00	1.00	2.00	
14	Grow2	0.00	1.00	2.00	

Nominal Model Intercept Contrasts for Group 1, logit: (a skθ + ck); c = Tγ (Back to TOC)

Item	Label	Contrasts	γ_1	s.e.	γ_2	s.e.	γ_3	s.e.
7	Walking3	Trend	[14] -2.51	0.26	[15] 1.03	0.16	[16] 0.29	0.07
10	Apples3	Trend	[22] -3.36	0.44	[23] 0.15	0.16		
11	Continent	Trend	[25] -2.17	0.17	[26] 0.89	0.12		
14	Grow2	Trend	[32] 0.46	0.07	[33] 0.65	0.08		

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

Item	Label	a	s.e.	c	s.e.	b	s.e.
1	Cube1	[2] 1.11	0.17	[1] 0.47	0.12	-0.42	0.12
2	Cube3	[4] 1.57	0.22	[3] 1.48	0.17	-0.94	0.14
3	Cube4	[6] 1.12	0.18	[5] -1.04	0.13	0.93	0.20
4	Farms1	[8] 2.14	0.33	[7] -0.02	0.22	0.01	0.10
5	Farms4	[10] 0.79	0.14	[9] 0.19	0.10	-0.24	0.13
6	Walking1	[12] 2.68	0.44	[11] -2.18	0.30	0.81	0.18
8	Apples1	[18] 1.46	0.22	[17] 0.21	0.15	-0.14	0.10
9	Apples2	[20] 2.79	0.50	[19] -3.05	0.33	1.09	0.22
12	Grow1	[28] 1.24	0.20	[27] -0.06	0.13	0.05	0.11
13	Grow3	[30] 1.46	0.22	[29] 0.26	0.15	-0.18	0.10

GPC Model Item Parameter Estimates, logit: a[k(θ - b) + Σdk]

Item	Label	a	s.e.	b	s.e.	d_1	d_2	s.e.	d_3	s.e.	d_4	s.e.
7	Walking3	[13] 1.65	0.25	1.52	0.27	0.00	0.69	0.13	-0.30	0.09	-0.39	0.12
10	Apples3	[21] 2.24	0.38	1.50	0.27	0.00	0.07	0.07	-0.07	0.07		
11	Continent	[24] 1.64	0.32	1.32	0.25	0.00	0.55	0.08	-0.55	0.08		
14	Grow2	[31] 0.69	0.10	-0.66	0.14	0.00	0.94	0.19	-0.94	0.19		

Nominal Model Slopes and Scoring Function Contrasts for Group 2, logit: (a skθ + ck); s = Tα (Back to TOC)

Item	Label	a	s.e.	Contrasts	α_1	s.e.	α_2	s.e.	α_3	s.e.
7	Walking3	[13] 1.65	0.25	Trend	1.00	-----	0.00	-----	0.00	-----
10	Apples3	[21] 2.24	0.38	Trend	1.00	-----	0.00	-----		
11	Continent	[24] 1.64	0.32	Trend	1.00	-----	0.00	-----		
14	Grow2	[31] 0.69	0.10	Trend	1.00	-----	0.00	-----		

Nominal Model Scoring Function Values for Group 2, logit: (a skθ + ck); s = Tα (Back to TOC)

Item	Category	s_1	s_2	s_3	s_4
7	Walking3	0.00	1.00	2.00	3.00
10	Apples3	0.00	1.00	2.00	
11	Continent	0.00	1.00	2.00	
14	Grow2	0.00	1.00	2.00	

Nominal Model Intercept Contrasts for Group 2, logit: (a skθ + ck); c = Tγ (Back to TOC)

Item	Label	Contrasts	γ_1	s.e.	γ_2	s.e.	γ_3	s.e.
7	Walking3	Trend	[14] -2.51	0.26	[15] 1.03	0.16	[16] 0.29	0.07
10	Apples3	Trend	[22] -3.36	0.44	[23] 0.15	0.16		
11	Continent	Trend	[25] -2.17	0.17	[26] 0.89	0.12		
14	Grow2	Trend	[32] 0.46	0.07	[33] 0.65	0.08		

Likelihood-based Values and Goodness of Fit Statistics (Back to TOC)

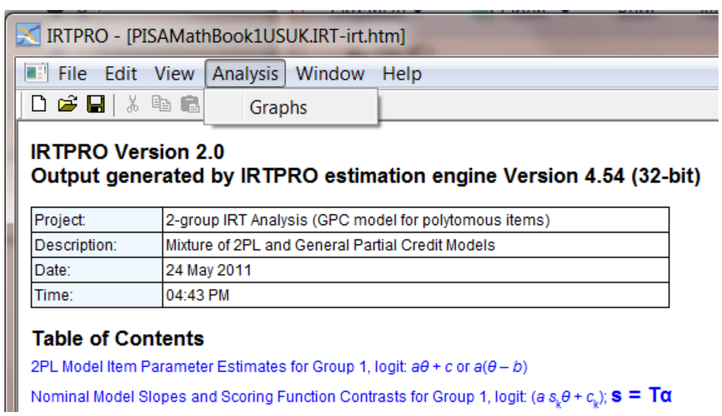
Statistics based on the loglikelihood	
-2loglikelihood:	21233.40
Akaike Information Criterion (AIC):	21303.40
Bayesian Information Criterion (BIC):	21482.90

Statistics based on one- and two-way marginal tables			
M_2	Degrees of freedom	Probability	RMSEA
785.98	333	0.0001	0.03
Note: M_2 is based on full marginal tables.			
Note: Model-based weight matrix is used.			

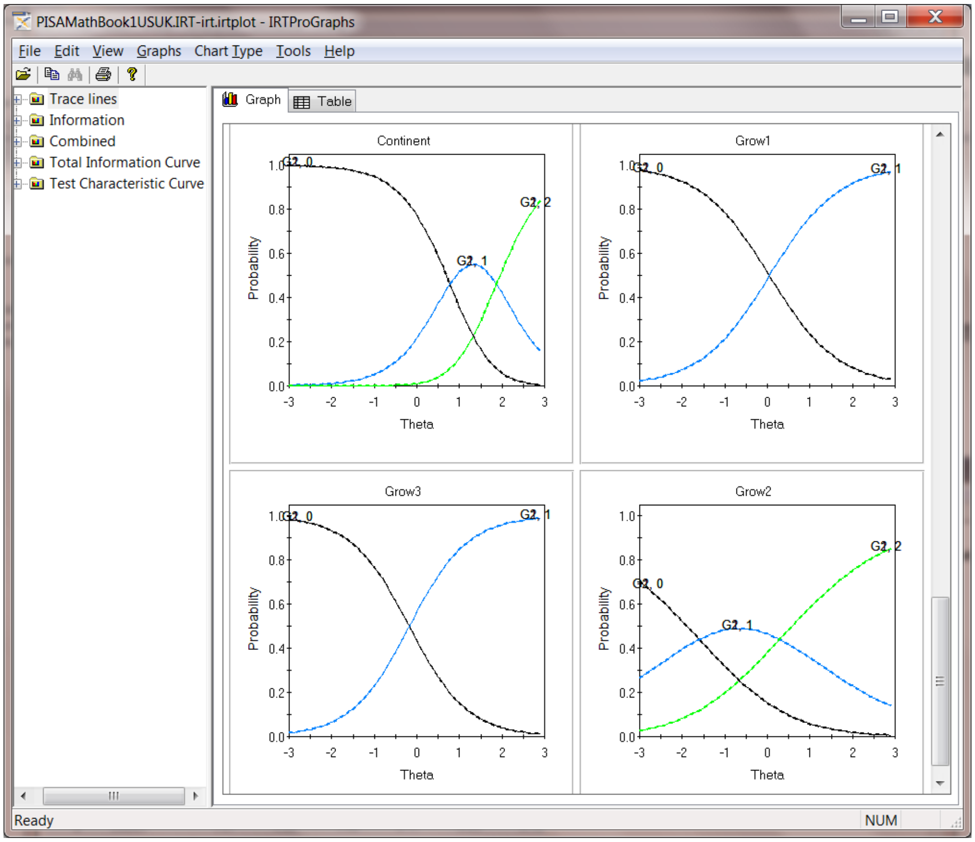
Group Parameter Estimates (Back to TOC)

Group	Label	μ	s.e.	σ^2	s.e.	σ	s.e.
1	G1	0.00	-----	1.00	-----	1.00	-----
2	G2	[67] -0.13	0.09	[68] 0.85	0.18	[68] 0.92	0.10

The option to display trace lines, information curves and test characteristic curves are available for all types of unidimensional analyses. While the output file is displayed, select the **Analysis, Graphs** option.



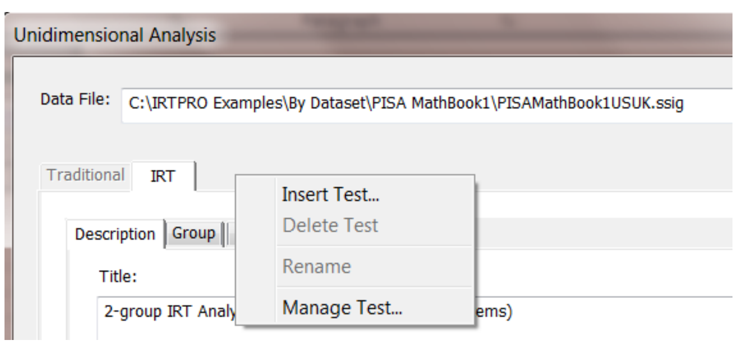
The default display shows the trace lines for the items in each group. To illustrate, the trace lines for the items Continent, Grow1, Grow3 and Grow2 are shown for the second (UK) group.



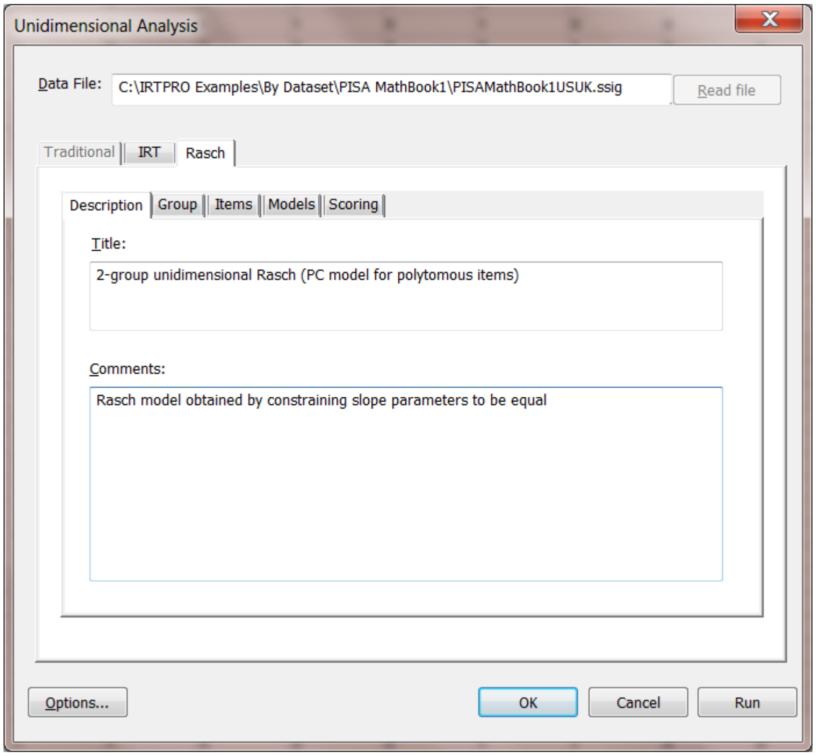
6.2.3 Unidimensional Rasch

In the Rasch analysis to be considered next, we imposed additional equality constraints across items (all items have the same slope). This turns the general partial credit (GPC) model into the partial credit (PC) model. Close the output file generated in the previous section to display the data **PISAMathbook1USUK.ssig**. Select the **Unidimensional IRT…** option from the **Analysis** menu.

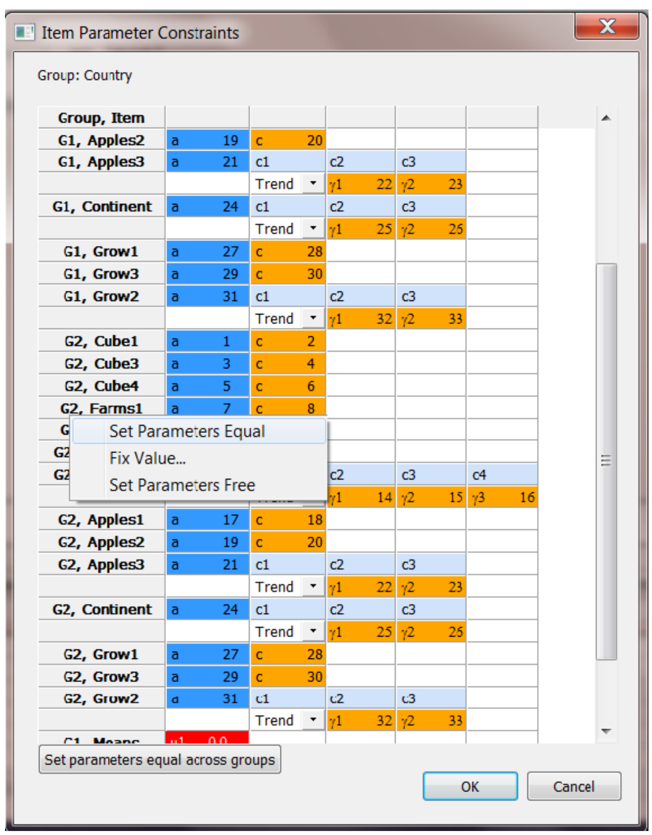
Right click on the righthand-side of the **IRT** tab to insert **Test3**. Once the **Test3** tab is displayed, right click on this tab and rename it to **Rasch**.



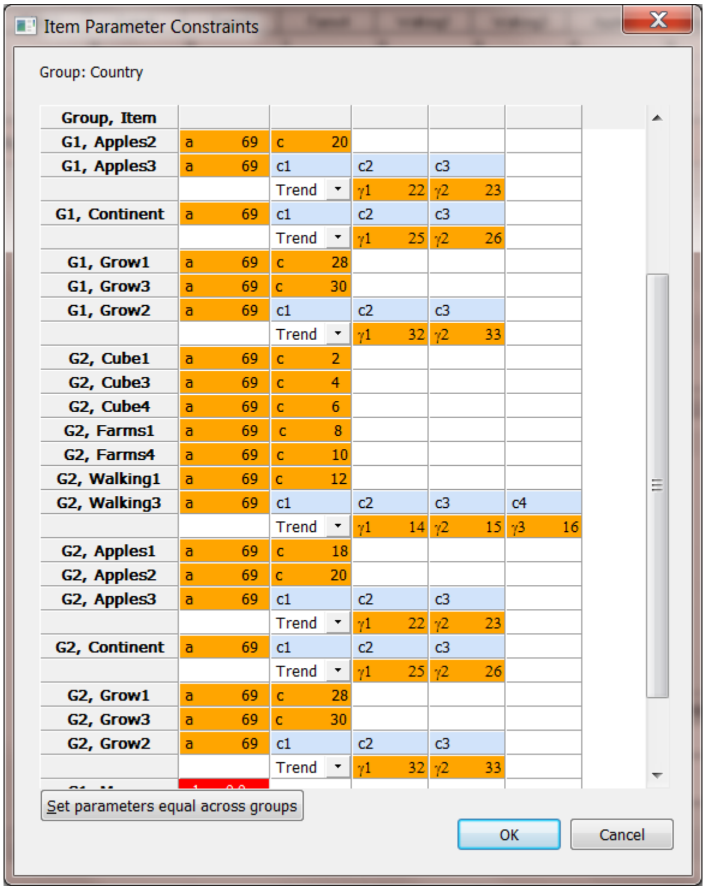
Once this is done, enter a title and comments.



Follow the steps described in the previous section to select the group variable and items. Next click the **Constraints** button on the **Models** window. Click the **Set parameters equal across groups** button. Then, for both groups, select all the cells containing the slope parameters (denoted as "a").



Right click on any one of the selected cells and select the **Set Parameters Equal** option from the drop-down menu. This action results in all the slope parameter numbers being set to a single number as shown in the **Constraints** window below. Click **OK** to return to the **Models** window.



To start the analysis, click the **Run** button. Portions of the output is given next.

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	Cube1	[20] 1.39	0.11	[1] 0.45	0.10	-0.32	0.09
2	Cube3	[20] 1.39	0.11	[2] 1.45	0.11	-1.04	0.14
3	Cube4	[20] 1.39	0.11	[3] -1.19	0.11	0.86	0.06
4	Farms1	[20] 1.39	0.11	[4] 0.12	0.11	-0.09	0.08
5	Farms4	[20] 1.39	0.11	[5] 0.06	0.11	-0.05	0.08
6	Walking1	[20] 1.39	0.11	[6] -1.33	0.11	0.96	0.07
8	Apples1	[20] 1.39	0.11	[10] 0.23	0.10	-0.16	0.08
9	Apples2	[20] 1.39	0.11	[11] -1.88	0.12	1.35	0.08
12	Grow1	[20] 1.39	0.11	[16] -0.09	0.10	0.06	0.07
13	Grow3	[20] 1.39	0.11	[17] 0.28	0.10	-0.20	0.08

GPC Model Item Parameter Estimates, logit: a[k(θ - b) + Σdk]

Item	Label	a	s.e.	b	s.e.	d_1	d_2	s.e.	d_3	s.e.	d_4	s.e.
7	Walking3	[20] 1.39	0.11	1.62	0.08	0.00	0.73	0.07	-0.33	0.09	-0.39	0.10
10	Apples3	[20] 1.39	0.11	1.70	0.09	0.00	-0.13	0.07	0.13	0.07		
11	Continent	[20] 1.39	0.11	1.41	0.08	0.00	0.58	0.06	-0.58	0.06		
14	Grow2	[20] 1.39	0.11	-0.32	0.08	0.00	0.74	0.06	-0.74	0.06		

Group Parameter Estimates (Back to TOC)

Group	Label	μ	s.e.	σ^2	s.e.	σ	s.e.
1	G1	0.00	-----	1.00	-----	1.00	-----
2	G2	[21] 0.51	0.07	[22] 0.84	0.09	[22] 0.92	0.05

Marginal Reliability for Response Pattern Scores: 0.82

Likelihood-based Values and Goodness of Fit Statistics (Back to TOC)

Statistics based on the loglikelihood	
-2loglikelihood:	21597.38
Akaike Information Criterion (AIC):	21641.38
Bayesian Information Criterion (BIC):	21754.20

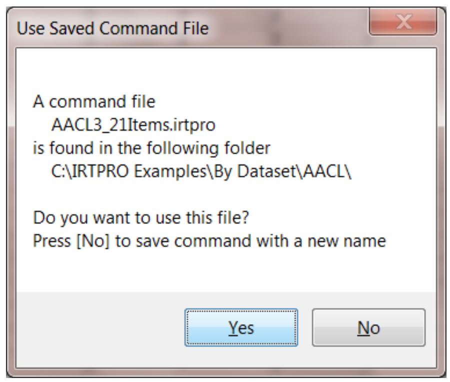
Statistics based on one- and two-way marginal tables			
M_2	Degrees of freedom	Probability	RMSEA
1214.58	346	0.0001	0.04
Note: M_2 is based on full marginal tables.			
Note: Model-based weight matrix is used.			

The deviance statistic (-2 log likelihood) for the Rasch model is reported above as 21597.38. The corresponding value for the IRT model (see previous section) is 21233.40. The chi-squared difference test therefore yields a value of 21597.38 – 21233.40 = 363.98. The degrees of freedom for testing between the IRT and Rasch models is 13 (14 slope parameters were estimated in the case of the IRT model versus one for the Rasch model). Since the chi-squared difference test is highly significant, we conclude that the IRT model provides a better fit to the item responses when compared to the Rasch model. Information-theoretic indices of fit (AIC and BIC) also point to the IRT model as better fitting.

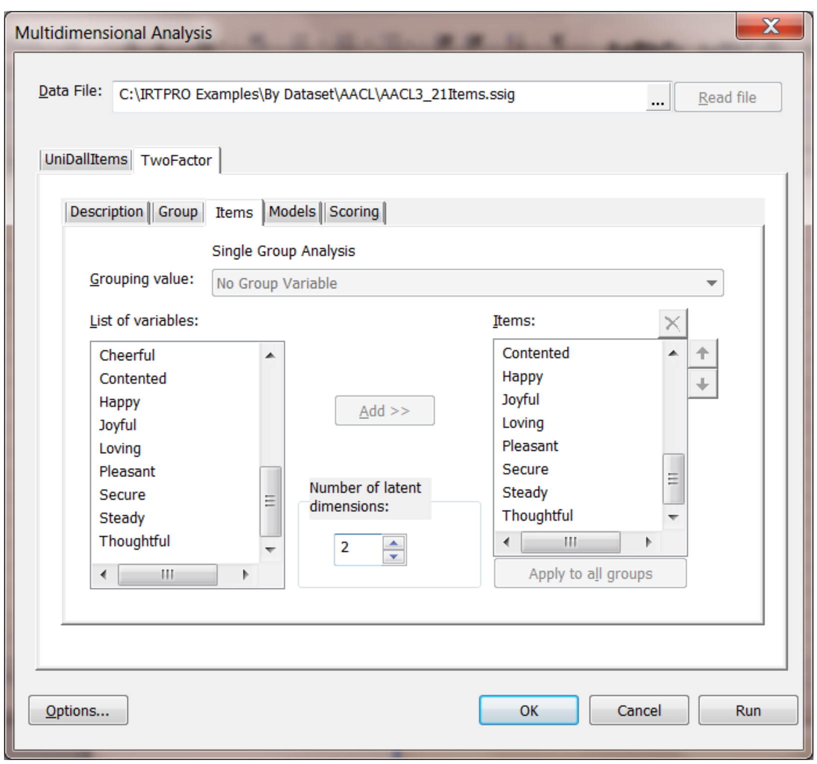
7. Multidimensional analysis

7.1 Multidimensional analysis of the Affect Adjective Check List (AACL)

To obtain a better fit than was obtained with a unidimensional model in Section 5.4, 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. See Section 5.4 for a description of the data and the recoding of item scores. Open the file **AACL_21Items.ssig** stored in the folder **C:\IRTPRO Examples\by DataSet\AACL**, and then select **Multidimensional IRT …** under the **Analysis** menu. A **Use Saved Command File** dialog appears. Click the **Yes** button and insert a second test by right-clicking on the right-hand side of the **UniDallItems** tab.



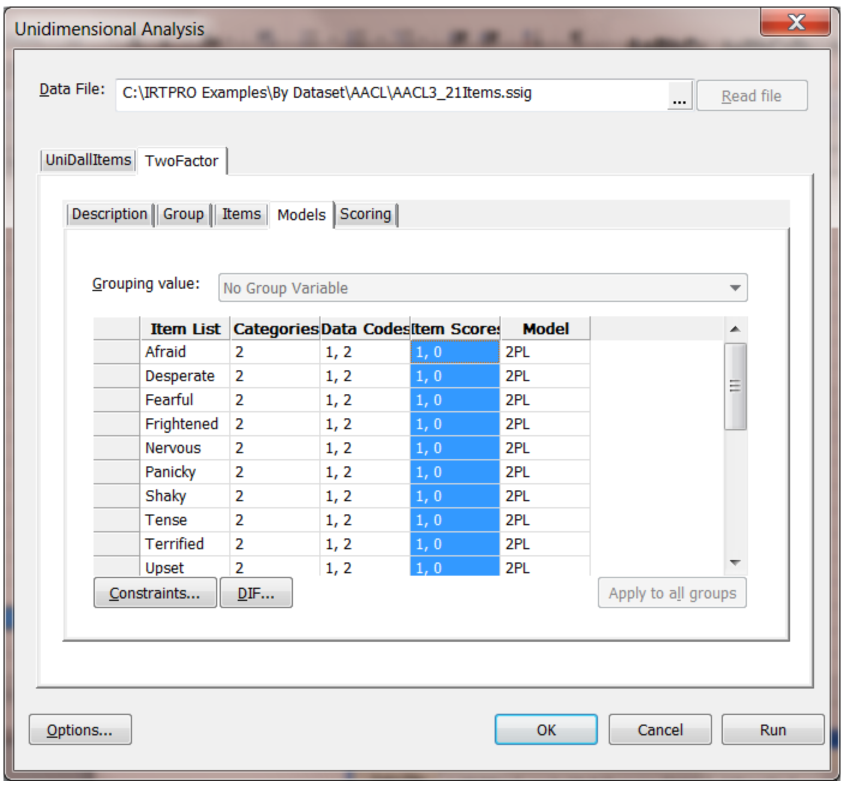
Rename the **Test2** tab to **TwoFactor**, enter a title and comments and select all 21 items. Change the number of dimensions to two as shown below.



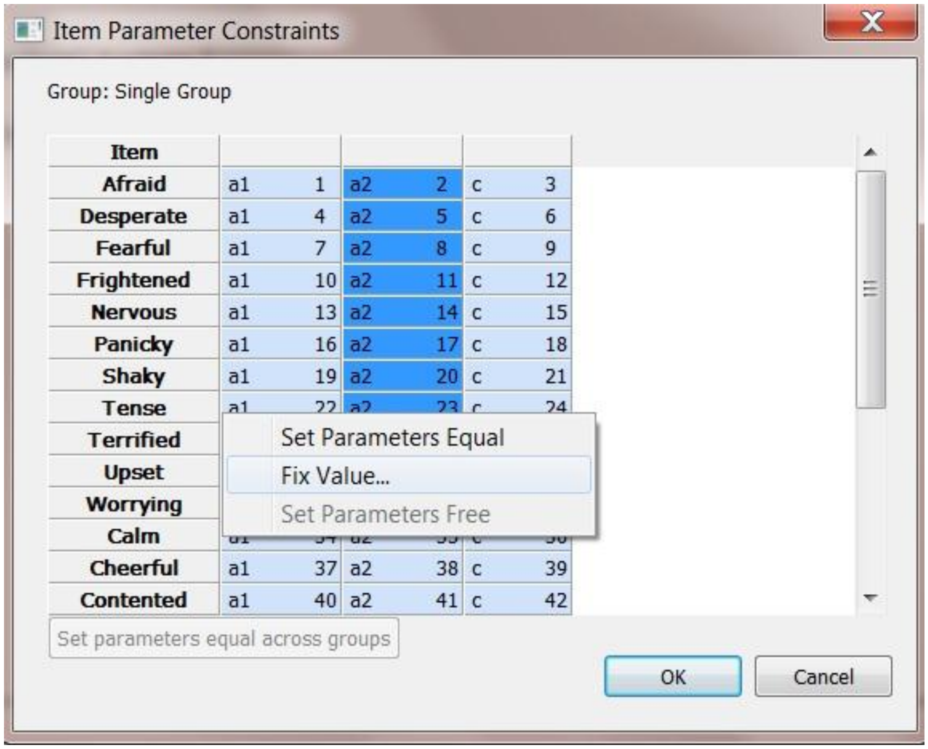
The particular two-dimensional model to be fitted to these data is a "simple structure" confirmatory factor analysis (CFA) model, which has non-zero slopes (or factor loadings) for the first eleven "anxiety-plus" adjectives on the first factor, non-zero slopes (or factor loadings) for the final ten "anxiety-minus" 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 set up the constraints in the IRTPRO graphical user interface (GUI), we click on the **Models** tab. Recode item scores of the first eleven items as explained in Section 5.4.

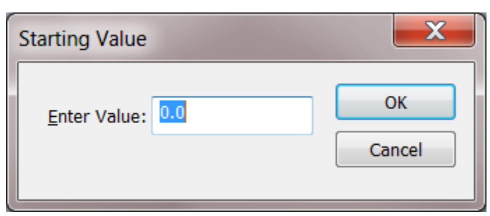


Once done, click on the **Constraints** button beneath the item list, and that shows the **Item Parameter Constraints** window:



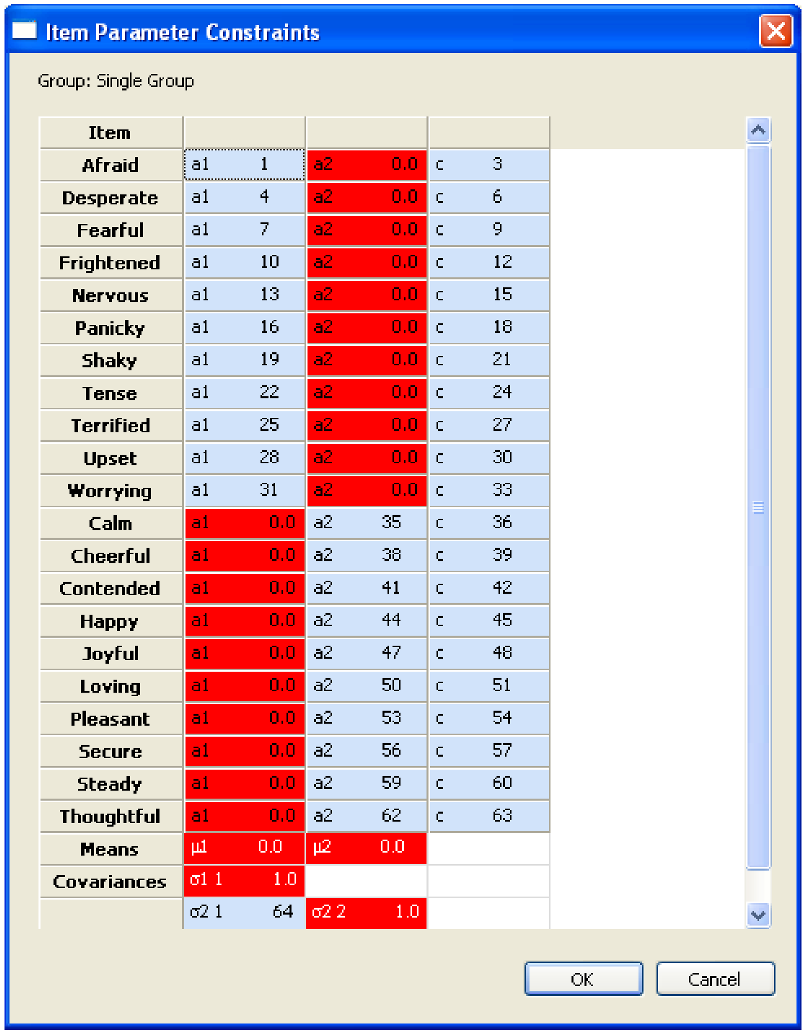
The **Item Parameter Constraints** window lists the items in the leftmost column, and then, for each item, the model parameters are indicated symbolically (*a* for slopes, and *c* for intercepts).

Select the a_2 cells for items Afraid to Worrying and right click to invoke the dialog for setting parameters equal, to free parameters or to fix parameter values. Select the **Fix Value…** option and since the default value is 0.0, click the **OK** button. 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.



Next, select the a_1 cells for items Calm to Thoughtful and right click to invoke the dialog for setting parameters equal, to free parameters or to fix parameter values. Select the **Fix Value…** option and click the **OK**.

Finally, select the covariance cell (σ_{21}) and right-click to select the **Set Parameters Free** option. 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 as shown below.



Item Parameter Constraints

Group: Single Group

Item			
Afraid	a1 1	a2 0.0	c 3
Desperate	a1 4	a2 0.0	c 6
Fearful	a1 7	a2 0.0	c 9
Frightened	a1 10	a2 0.0	c 12
Nervous	a1 13	a2 0.0	c 15
Panicky	a1 16	a2 0.0	c 18
Shaky	a1 19	a2 0.0	c 21
Tense	a1 22	a2 0.0	c 24
Terrified	a1 25	a2 0.0	c 27
Upset	a1 28	a2 0.0	c 30
Worrying	a1 31	a2 0.0	c 33
Calm	a1 0.0	a2 35	c 36
Cheerful	a1 0.0	a2 38	c 39
Contended	a1 0.0	a2 41	c 42
Happy	a1 0.0	a2 44	c 45
Joyful	a1 0.0	a2 47	c 48
Loving	a1 0.0	a2 50	c 51
Pleasant	a1 0.0	a2 53	c 54
Secure	a1 0.0	a2 56	c 57
Steady	a1 0.0	a2 59	c 60
Thoughtful	a1 0.0	a2 62	c 63
Means	μ1 0.0	μ2 0.0	
Covariances	σ1 1 1.0		
	σ2 1 64	σ2 2 1.0	

OK Cancel

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 (σ_{21}) is estimated – that is the correlation between the two latent variables.

When the parameters are as desired, we click **OK** in the **Item Parameter Constraints** window, then **Run** in the **Multidimensional Analysis** dialog box, and wait… longer than for the unidimensional run, because multidimensional runs take more time.

When the parameters have been estimated, the output appears. We note that 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. We also note that 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, we 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	-----		
4 3	0.55	0.06	1.00	-----

We observe that the correlation between the latent variables that account 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 we fitted in Section 5.4 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 -2 loglikelihood for the unidimensional 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 unidimensional 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 incredibly significant. So there is strong evidence that these data need a two dimensional model to be fitted.

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 (X2) and Standardized LD X2 Statistics for Group 1 (Back to TOC)

Item	Marginal 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

Item	Marginal 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. That *could* be modeled as well, but we leave that exercise to the reader.

7.2 Analysis of Quality of Life data

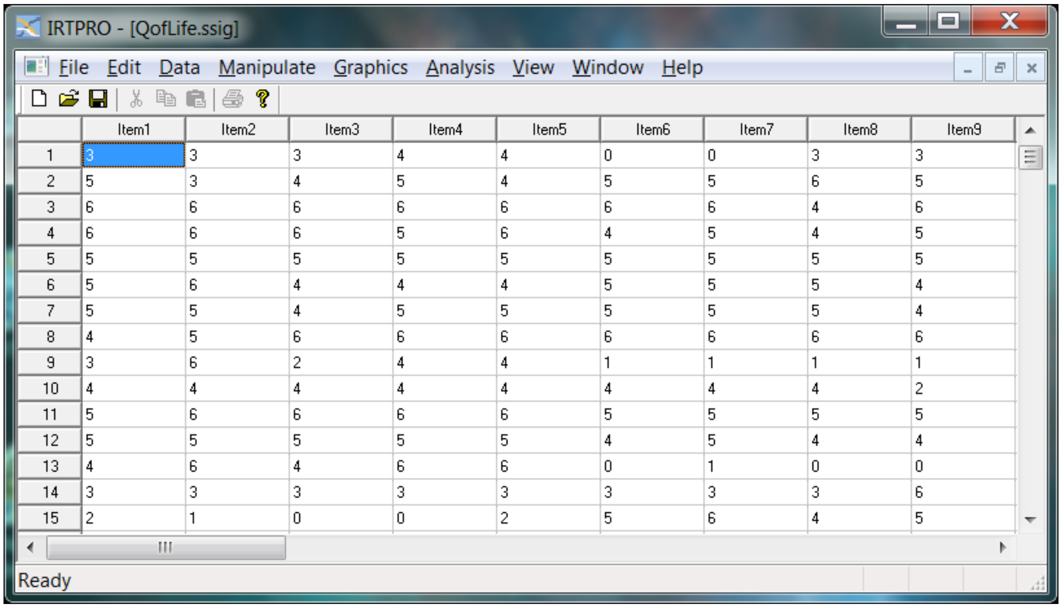
7.2.1 Exploratory Factor Analysis (EFA)

To illustrate the implementation of exploratory factor analysis for graded response data, the "Quality of Life Interview for the Chronically Mentally Ill" (Lehman, 1988) was analyzed based on the item responses of 586 chronically mentally ill patients. The scale consists of seven subdomains (Family, Finance, Health, Leisure, Living, Safety, and Social), each with 4 to 6 items for a total of 34 items. In addition, there is one global life satisfaction item, yielding a total of 35 items. Each item is rated on a 7-point scale with the following response categories: 1 = terrible; 2 = unhappy; 3 = mostly dissatisfied; 4 = mixed, about equally satisfied and dissatisfied; 5 = mostly satisfied; 6 = pleased; and 7 = delighted. The table below provides a brief description of each of the 35 items.

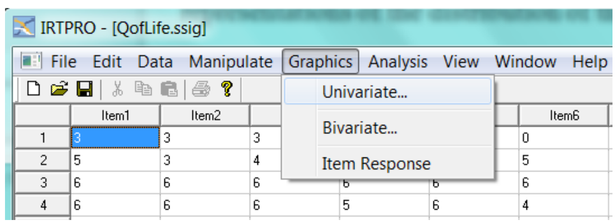
Table 7.1: Description of items in the Lehman (1988) Quality-of-Life Rating Scale Data (N =586)

Global Item1: Global life satisfaction as a whole Family Item2: Family Item3: Amount of family contact Item4: Family with interaction Item5: General family stuff Finance Item6: Total money you get Item7: Amount pay for basic needs Item8: Financial well-being Item9: Money for fun	Health Item10: Health in general Item11: Medical care Item12: How often see doctor Item13: Talk to therapist Item14: Physical condition Item15: Emotional well-being Leisure Item16: Way spend free time Item17: Amount of free time Item18: Chance to enjoy time Item19: Amount of fun Item20: Amount of relaxation Item21: Pleasure from TV	Living Item22: Living arrangements Item23: Food Item24: Privacy Item25: Amount of freedom Item26: Prospect of staying Safety Item27: Neighborhood safety Item28: Safe at home Item29: Police access Item30: Protect robbed/attack Item31: Personal safety Social Item32: Do things with others Item33: Time with others Item34: Social interactions Item35: People in general

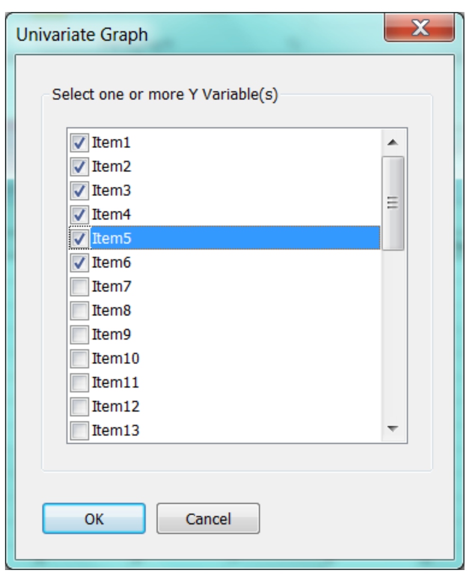
To open the data file, select the file **QolLife.ssig** from the **IRTPRO Examples\By Data Set\Quality of Life** folder. The first 15 cases for Item1 to Item9 are displayed below in spreadsheet format. Each of the items has seven categories and therefore the available models are graded, general partial credit and nominal.



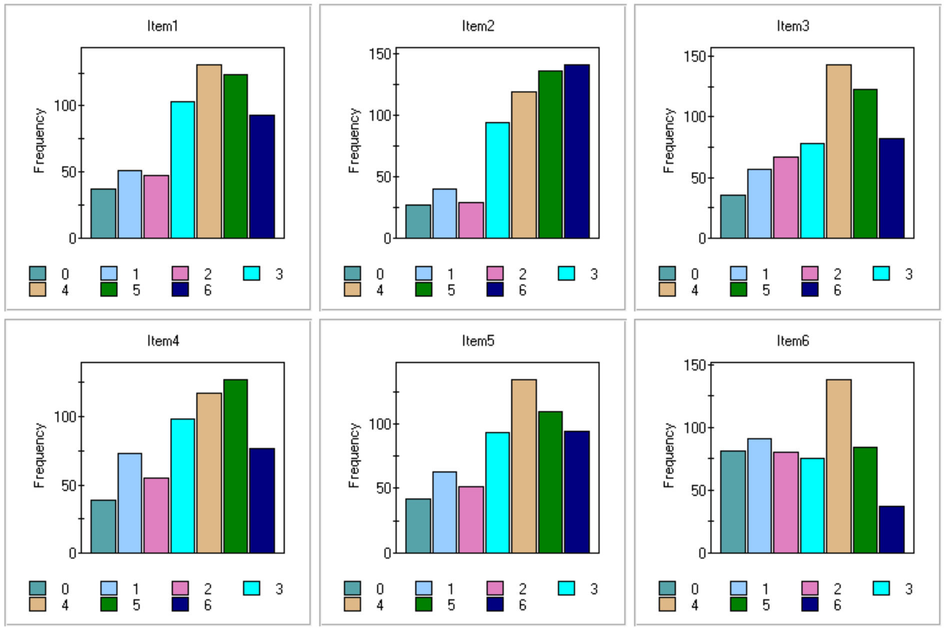
From the main menu bar, select the **Graphics**, **Univariate…** option to obtain bar chart representations of the distribution of the items over category values.



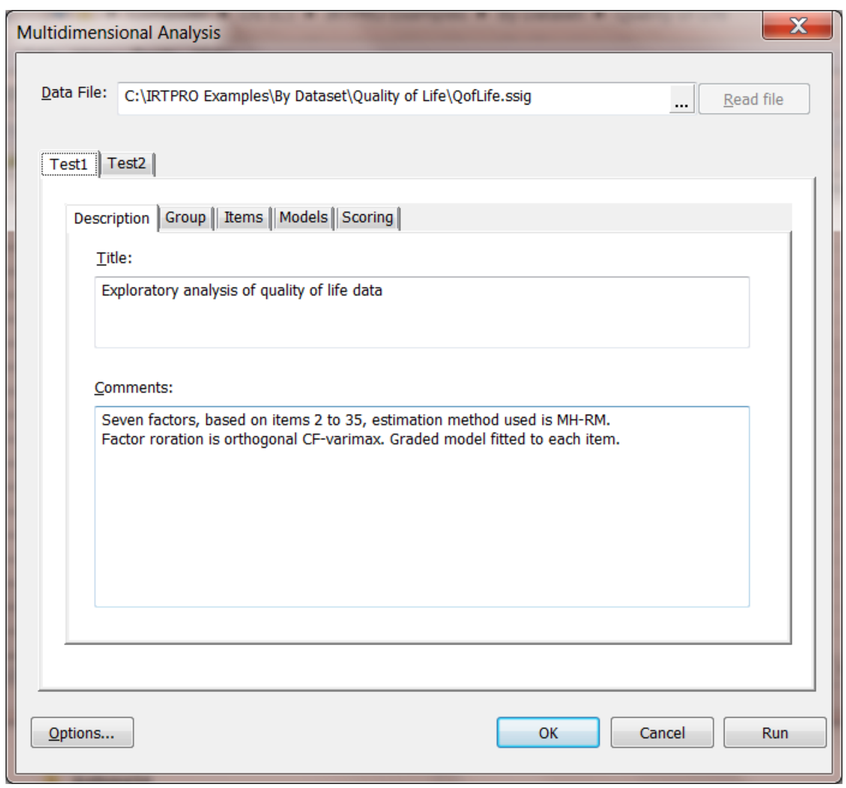
By selecting this option, the **Univariate Graph** window is displayed enabling one to select a list of items to be displayed graphically. To illustrate, the first six items are selected as shown.



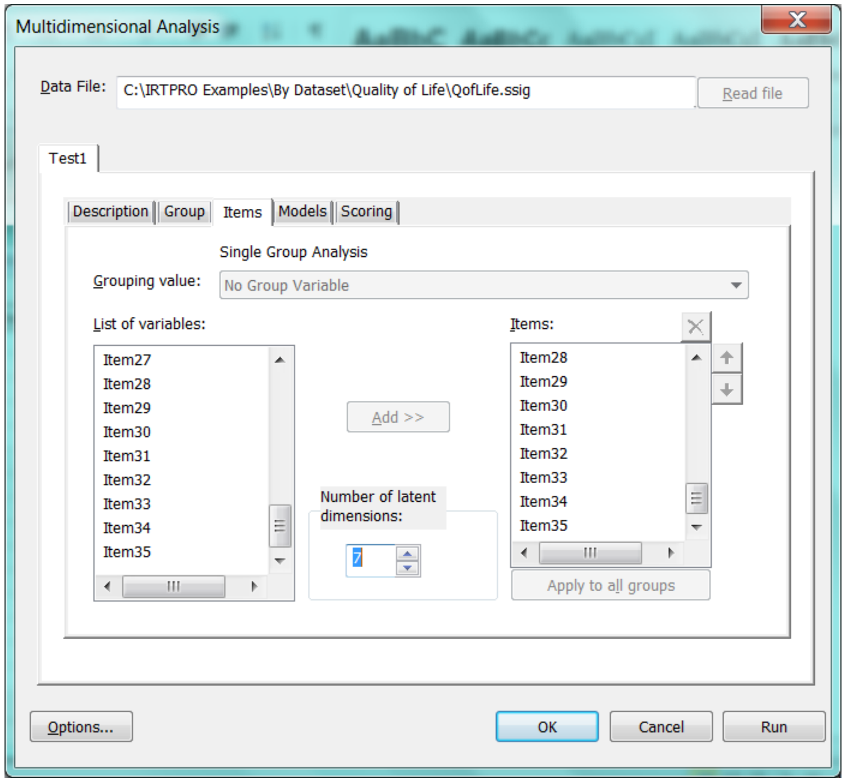
Next, click the **OK** button to obtain the bar charts. Each chart presents the distributions of responses over the seven categories.



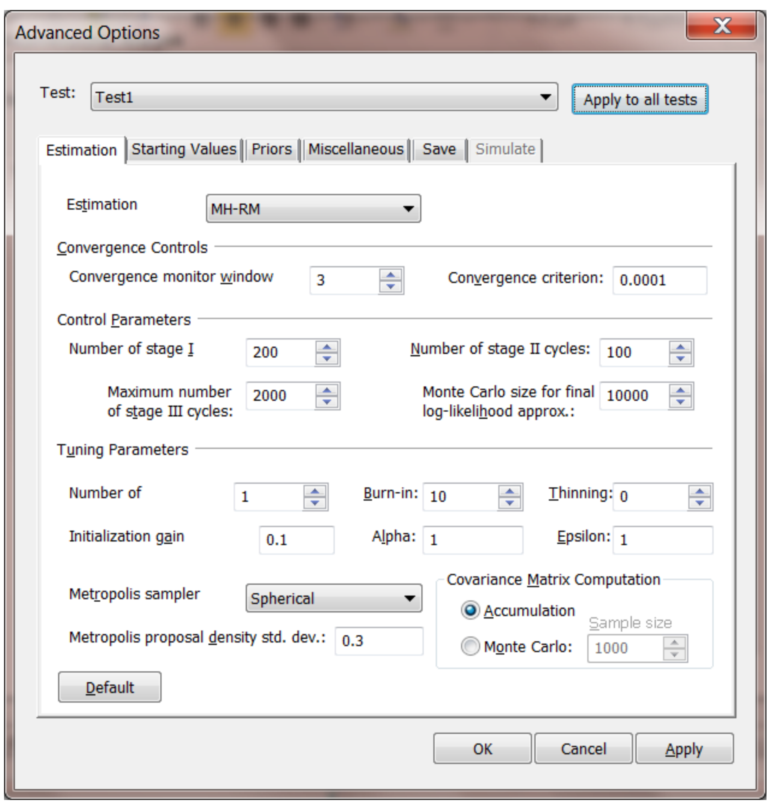
To start the exploratory factor analysis (EFA), select the **Analysis, Multidimensional…** option from the main menu bar and use the **Title** and **Comments** text boxes to describe the analysis.



Since the dataset consists of a single group, the **Group** tab is not used and we skip to the **Items** tab to select Item2 to Item35. In the **Number of latent dimensions:** field enter "7".

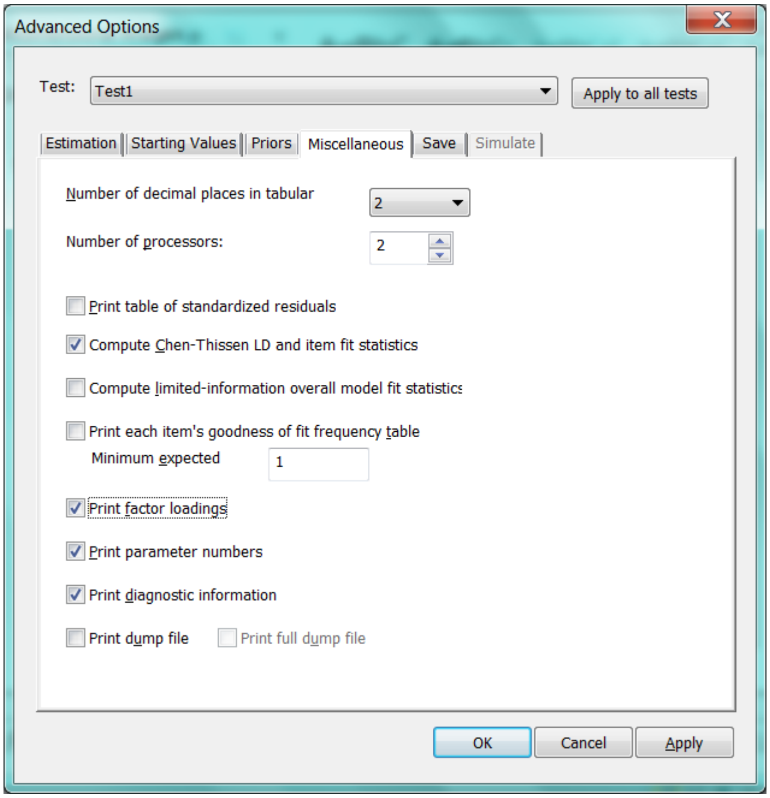


To change the estimation method from the default (Bock-Aitkin) to MH-RM, check the **Options** button (bottom-left on screen displayed above).

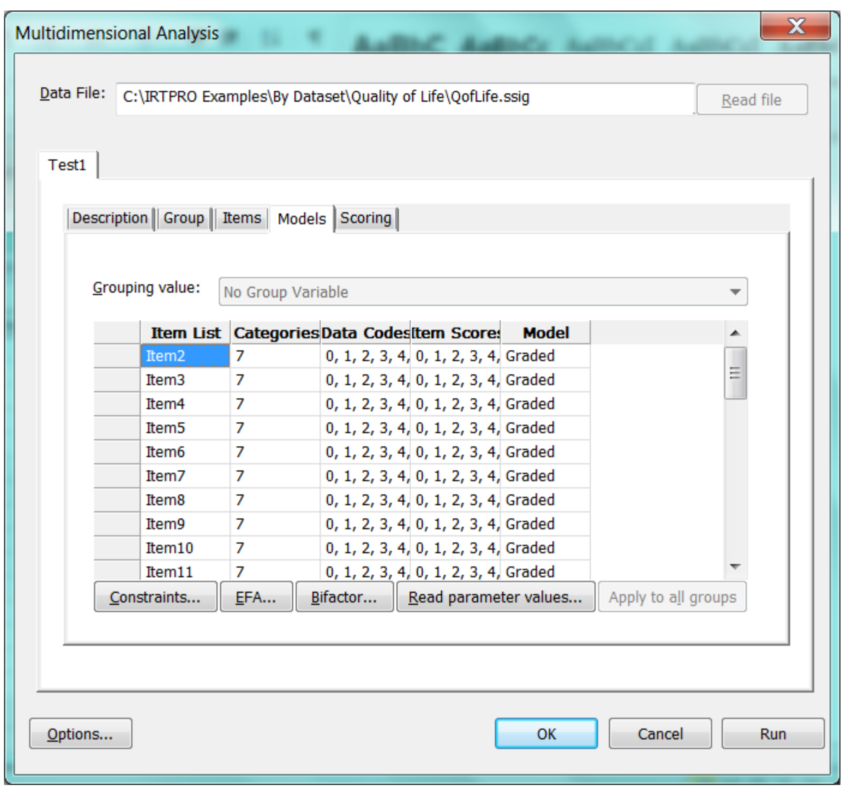


To obtain a listing of factor loading in the output file, select the **Miscellaneous** tab in the **Advanced**

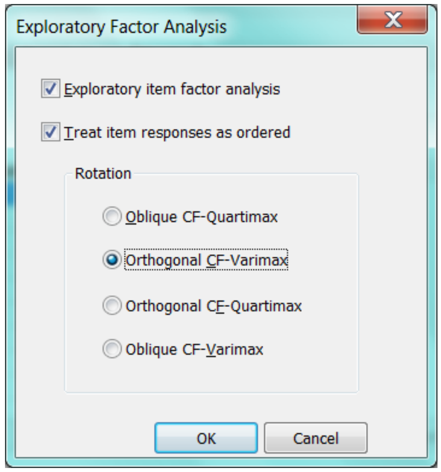
Options window and click the **Print Factor Loadings** check box.



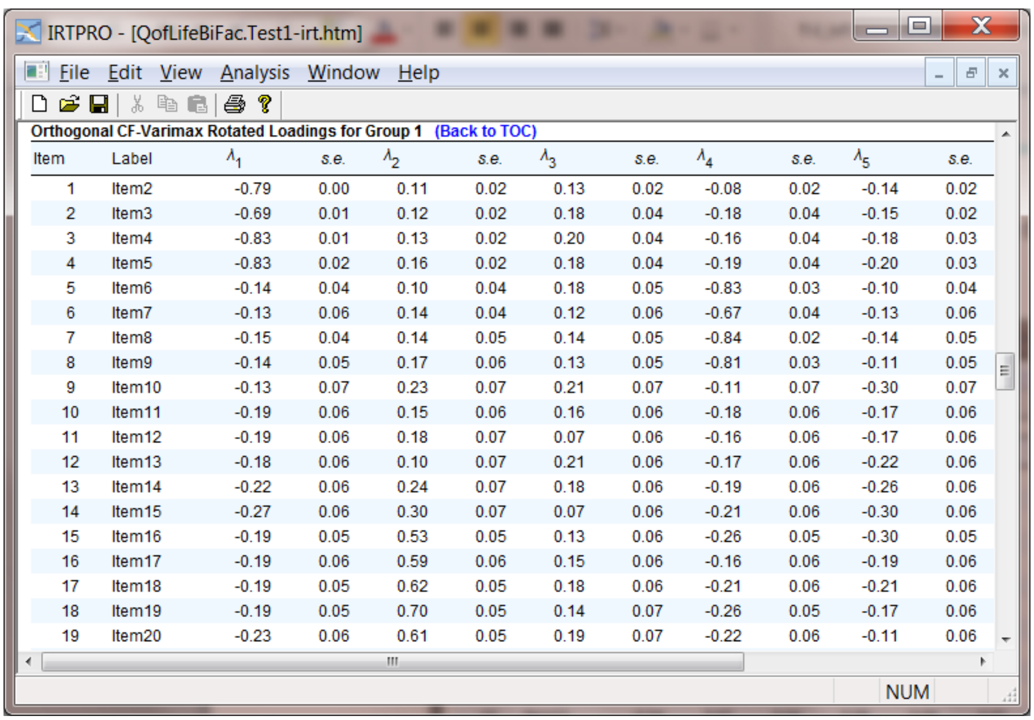
Next, proceed to the **Models** tab. Click the **EFA** button to continue.



To confirm that you intend to do an exploratory item factor analysis, click the appropriate checkbox. In doing so, IRTPRO automatically sets up the EFA parameter constraints and the **Constraints** button, **Models** tab, is disabled. There are four rotation types available as shown in the dialogue below. For this analysis, the **Orthogonal CF-Varimax** method is selected.



Portion of the output, listing the rotated factor loadings of the first five factors for item2 to item20, is displayed below.



Orthogonal CF-Varimax Rotated Loadings for Group 1 (Back to TOC)

Item	Label	λ_1	s.e.	λ_2	s.e.	λ_3	s.e.	λ_4	s.e.	λ_5	s.e.
1	Item2	-0.79	0.00	0.11	0.02	0.13	0.02	-0.08	0.02	-0.14	0.02
2	Item3	-0.69	0.01	0.12	0.02	0.18	0.04	-0.18	0.04	-0.15	0.02
3	Item4	-0.83	0.01	0.13	0.02	0.20	0.04	-0.16	0.04	-0.18	0.03
4	Item5	-0.83	0.02	0.16	0.02	0.18	0.04	-0.19	0.04	-0.20	0.03
5	Item6	-0.14	0.04	0.10	0.04	0.18	0.05	-0.83	0.03	-0.10	0.04
6	Item7	-0.13	0.06	0.14	0.04	0.12	0.06	-0.67	0.04	-0.13	0.06
7	Item8	-0.15	0.04	0.14	0.05	0.14	0.05	-0.84	0.02	-0.14	0.05
8	Item9	-0.14	0.05	0.17	0.06	0.13	0.05	-0.81	0.03	-0.11	0.05
9	Item10	-0.13	0.07	0.23	0.07	0.21	0.07	-0.11	0.07	-0.30	0.07
10	Item11	-0.19	0.06	0.15	0.06	0.16	0.06	-0.18	0.06	-0.17	0.06
11	Item12	-0.19	0.06	0.18	0.07	0.07	0.06	-0.16	0.06	-0.17	0.06
12	Item13	-0.18	0.06	0.10	0.07	0.21	0.06	-0.17	0.06	-0.22	0.06
13	Item14	-0.22	0.06	0.24	0.07	0.18	0.06	-0.19	0.06	-0.26	0.06
14	Item15	-0.27	0.06	0.30	0.07	0.07	0.06	-0.21	0.06	-0.30	0.06
15	Item16	-0.19	0.05	0.53	0.05	0.13	0.06	-0.26	0.05	-0.30	0.05
16	Item17	-0.19	0.06	0.59	0.06	0.15	0.06	-0.16	0.06	-0.19	0.06
17	Item18	-0.19	0.05	0.62	0.05	0.18	0.06	-0.21	0.06	-0.21	0.06
18	Item19	-0.19	0.05	0.70	0.05	0.14	0.07	-0.26	0.05	-0.17	0.06
19	Item20	-0.23	0.06	0.61	0.05	0.19	0.07	-0.22	0.06	-0.11	0.06

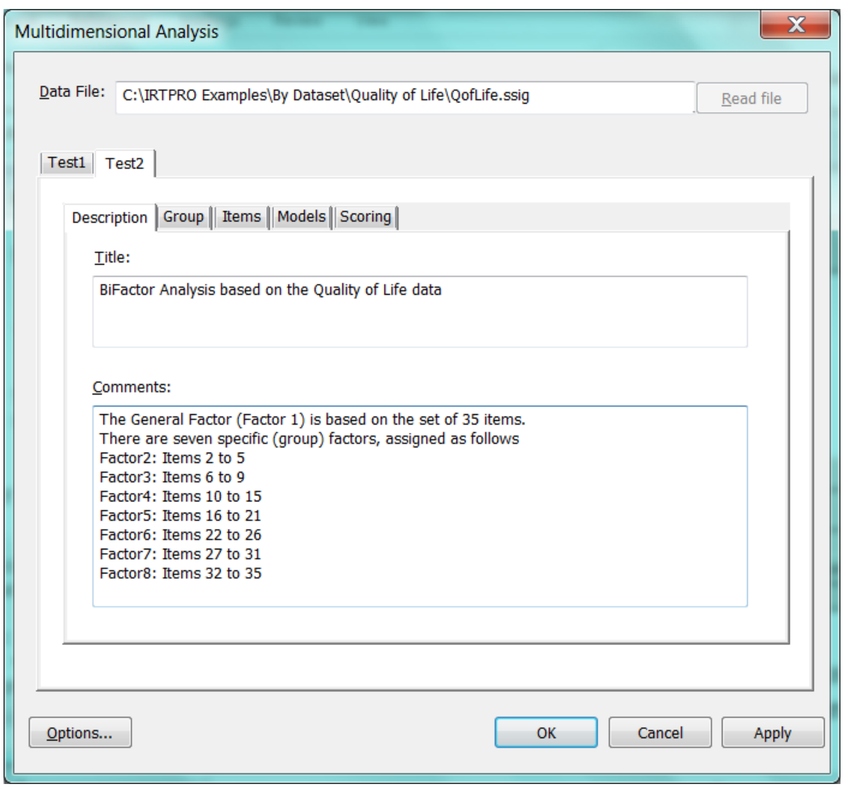
By looking at all factor loadings that exceed 0.40 in absolute value, the EFA confirms the presence of seven subdomains (F1 = Family, F2 = Leisure, F3 = Living, F4 = Finance, F5 = Safety, F6 = Social and F7 = Health) listed in the table at the beginning of this section.

7.2.2 Bifactor Analysis

A plausible factorial structure for many types of psychological and educational tests exhibits a general factor and one or more group or method factors. A bifactor model can represent this type of factorial structure. The bifactor structure results from the constraint that each item has a nonzero loading on the primary dimension and, at most, one of the group factors. Using maximum marginal likelihood estimation of item parameters, the bifactor restriction leads to a major simplification of the likelihood equations and (a) permits analysis of models with large numbers of group factors, (b) permits conditional dependence within identified subsets of items, and (c) provides more parsimonious factor solutions than an unrestricted full-information item factor analysis in some cases. Analysis of data obtained from 586 chronically mentally ill patients, described in the previous section, reveals a clear bifactor structure, partially demonstrated by fitting an EFA model to the data with orthogonal rotation of the factors.

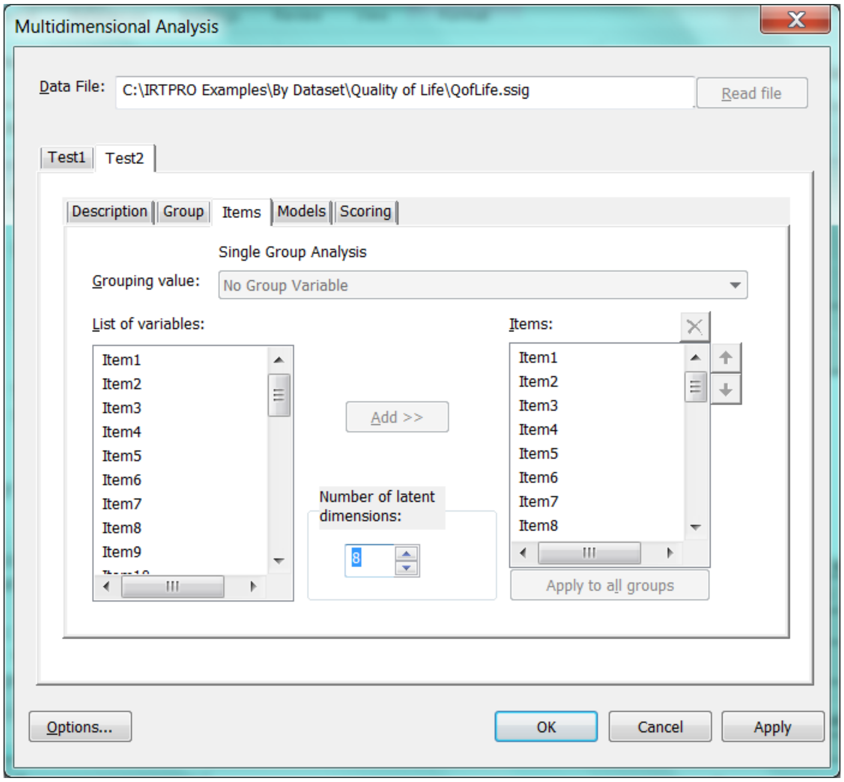
The bifactor model was originally introduced to extend the Spearman one-factor model for intelligence tests to include so-called "group" factors. Including these mutually uncorrelated factors enables the researcher to explain departures from the common (general) factor. The mutually uncorrelated factors assumption makes it possible to do numerical quadrature in two dimensions.

To define the analysis using the user's interface, select the **Analysis**, **Multidimensional IRT…** option from the main menu bar. For this analysis, we assume a total of eight factors, these being a general factor and seven additional mutually uncorrelated factors.

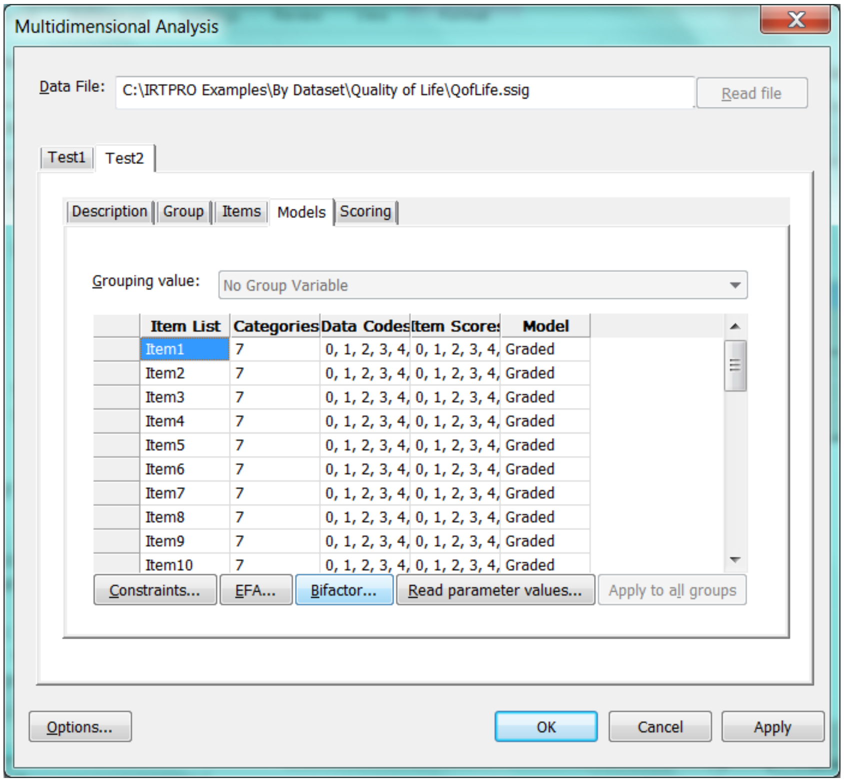


The **Description** tab shows the assignment of items to the group factors. Note that the mutually uncorrelated factors assumption implies that any given item can be assigned to only one group factor. Also, note that one or more of the 35 items do not have to be assigned to any group factor. In this illustration, item1 is not assigned to any one of the additional factors. Since the dataset consists of a single group, the **Groups** tab is skipped and under the **Items** tab all 35 items are selected.

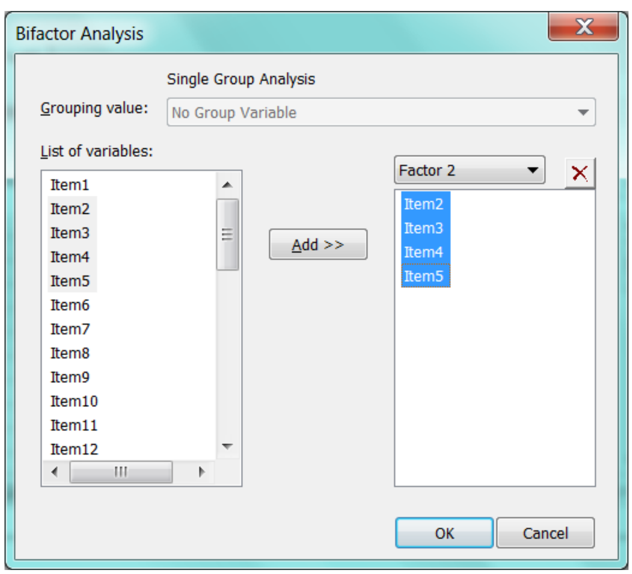
In the case of a multidimensional analysis, the **Models** tab contains a field for the number of latent dimensions, the number to be entered being 8 in the present example.



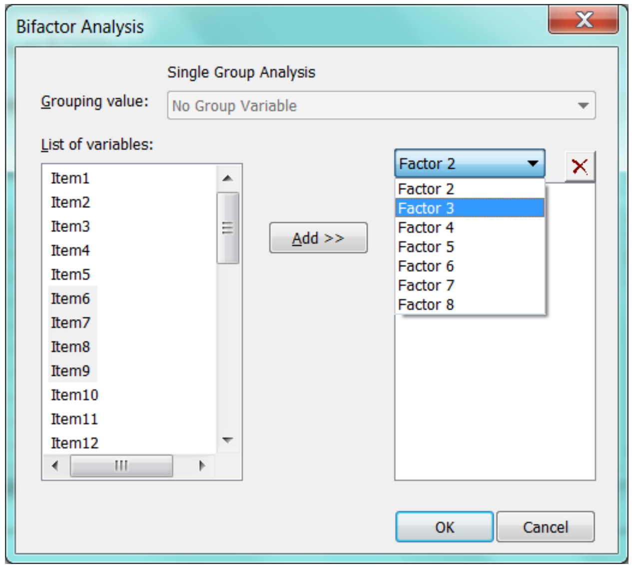
The graded model (the default model when the number of categories for an item is greater than two) is used for each item and hence we click the **Bifactor** button to assign items to the additional factors.



Selection of the **Bifactor** option activates a **Bifactor Analysis** window, allowing one to select items from the **List of variables** for each additional factor. Below, Factor 2 is selected and Item2 to Item5 are assigned to this factor.

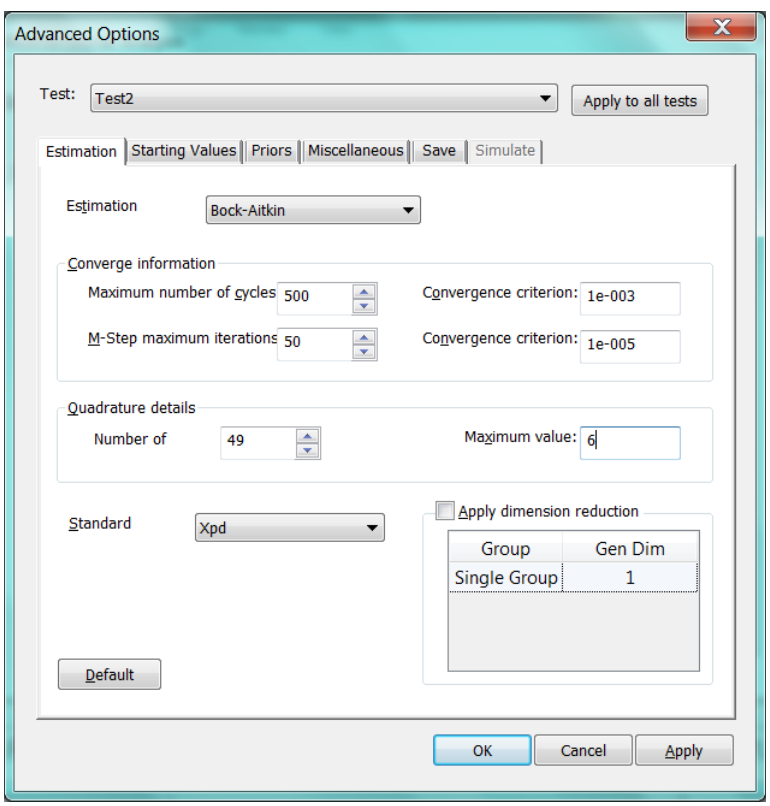


The next dialog illustrates the selection of Item6 to Item9 and of Factor 3 using the drop-down list.

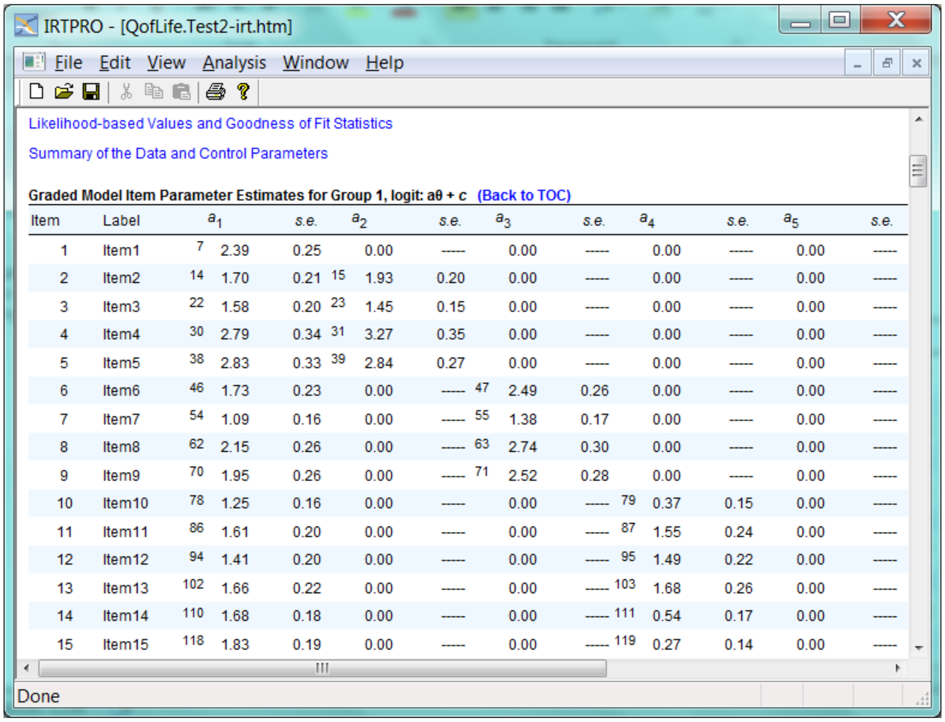


Note that when either an **EFA** or **Bifactor** analysis is specified, the **Constraints** option is no longer available since parameter constraints are, in these cases, generated by the user's interface.

Access to the **Advanced Options** window is obtained by clicking the **Options** button on the **Multidimensional Analysis** window. For the current analysis, the Bock-Aitkin estimation method is selected and the **Convergence information**, **Quadrature details**, and method to be used to calculate the standard errors are specified.



A portion of the output is shown below. Note that all the slope parameters (a_1) for the general factor are estimated. In the case of the additional factors, slopes are only estimated for the list of items assigned to a group factor.



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

Item	Label	a_1	s.e.	a_2	s.e.	a_3	s.e.	a_4	s.e.	a_5	s.e.
1	Item1	[7] 2.39	0.25	0.00	-----	0.00	-----	0.00	-----	0.00	-----
2	Item2	[14] 1.70	0.21	[15] 1.93	0.20	0.00	-----	0.00	-----	0.00	-----
3	Item3	[22] 1.58	0.20	[23] 1.45	0.15	0.00	-----	0.00	-----	0.00	-----
4	Item4	[30] 2.79	0.34	[31] 3.27	0.35	0.00	-----	0.00	-----	0.00	-----
5	Item5	[38] 2.83	0.33	[39] 2.84	0.27	0.00	-----	0.00	-----	0.00	-----
6	Item6	[46] 1.73	0.23	0.00	-----	[47] 2.49	0.26	0.00	-----	0.00	-----
7	Item7	[54] 1.09	0.16	0.00	-----	[55] 1.38	0.17	0.00	-----	0.00	-----
8	Item8	[62] 2.15	0.26	0.00	-----	[63] 2.74	0.30	0.00	-----	0.00	-----
9	Item9	[70] 1.95	0.26	0.00	-----	[71] 2.52	0.28	0.00	-----	0.00	-----
10	Item10	[78] 1.25	0.16	0.00	-----	0.00	-----	[79] 0.37	0.15	0.00	-----
11	Item11	[86] 1.61	0.20	0.00	-----	0.00	-----	[87] 1.55	0.24	0.00	-----
12	Item12	[94] 1.41	0.20	0.00	-----	0.00	-----	[95] 1.49	0.22	0.00	-----
13	Item13	[102] 1.66	0.22	0.00	-----	0.00	-----	[103] 1.68	0.26	0.00	-----
14	Item14	[110] 1.68	0.18	0.00	-----	0.00	-----	[111] 0.54	0.17	0.00	-----
15	Item15	[118] 1.83	0.19	0.00	-----	0.00	-----	[119] 0.27	0.14	0.00	-----

In many practical applications, the bifactor model provides a natural alternative to the traditional conditionally independent unidimensional IRT model. When conditional dependence is likely, as in the case of paragraph comprehension tests, tests in which there are two or more methods of item presentation, or personality or other items that have a two-level structure with an underlying general factor, the item bifactor solution provides an excellent alternative. An attractive by-product of this model is that it requires only the evaluation of a two-dimensional integral, regardless of the number of subtests, paragraphs, or content areas.

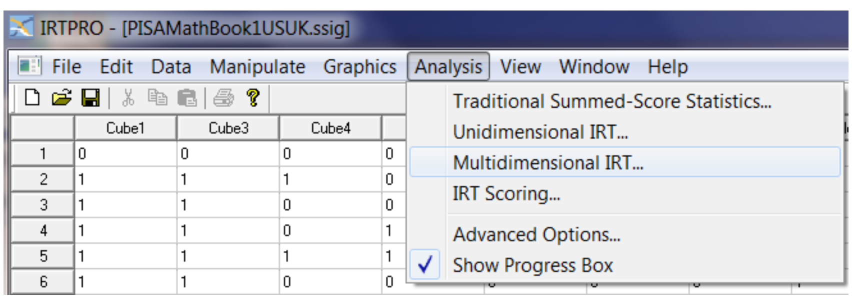
In the ordinal response case, the bifactor model provides a very general multidimensional model for graded response data. In mental health measurement, rating scales are typically constructed by sampling items from domains related to a single underlying construct, as in the quality-of-life scale analyzed in the illustration. In these cases, a priori knowledge of which item belongs to which subdomain is available, and the bifactor model is a natural choice. Similarly, in educational measurement problems, tests are often constructed by creating a series of subtests or so-called "testlets" (Wainer & Kiely, 1987) within which items have similar content or focus, and these testlets are then combined to form a test. In this case, item groupings are also known in advance, and the bifactor model applies. Regardless of the number of testlets, the relevant integrals in the full-information maximum marginal likelihood solution always reduce to 2 and can be approximated to any practical degree of accuracy.

7.3 Testlet Response Theory (TRT) analysis of the PISA data

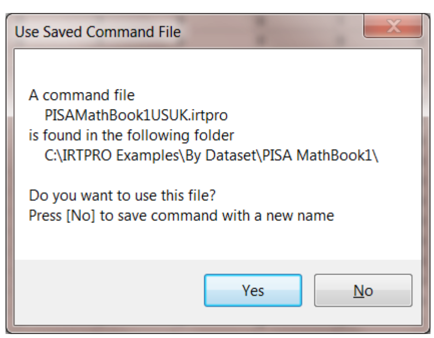
The Program for International Student Assessment (PISA), conducted triennially by the Organization for Economic Co-operation and Development (OECD) since year 2000, is an international educational assessment system that focuses on the 15-year olds' reading literacy, mathematics literacy, and science literacy. The format of the PISA can be best described as testlet-based. A testlet in this instance is a collection of test items organized around the same stimulus. For instance, it is standard practice in reading assessments to base several questions on one reading passage so that each question can measure a different aspect of the examinee's comprehension of the passage. PISA is noteworthy in that testlets are employed in all three sections namely reading, math, and science. For instance, a typical form of PISA mathematics assessment (in year 2000) consists of 14 items that can be divided into five testlets made up of nonoverlapping sets of items. Some testlets are longer, with more than two items, and some are shorter, with only 2 items. Critically, an item belongs to one and only one testlet.

Section 6.2 provides a more detailed description of the data. The analysis of a testlet response theory (TRT) model based on the PISA data is given in this section. For the TRT analysis, a multidimensional model for more than one group is fitted. The mean and variance of the primary math dimension is freely estimated in the UK group and the additional dimensions are there to account for local item dependence in the same testlet.

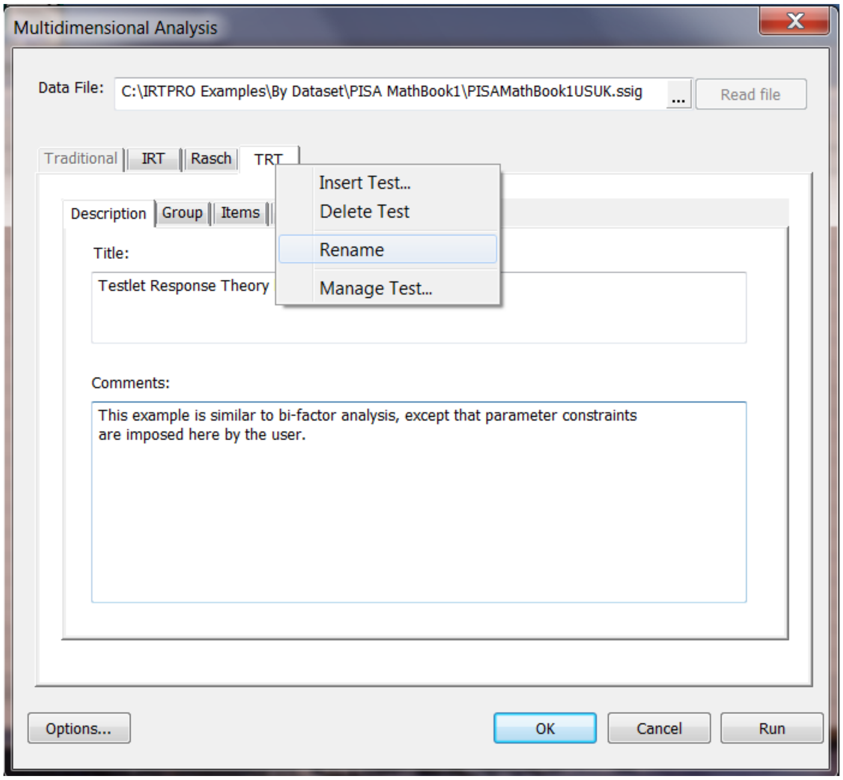
The dataset **PISAMathBook1USUK.ssig** is located in the folder **IRTPRO Examples\By Data Set\PISA MathBook1\.** To start the analysis open this file and from the main-menu bar select the **Analysis, Multidimensional IRT…** option.



Once this option is selected, the following message will be displayed if the analyses described in Sections 6.2.1 to 6.2.3 were performed.



Click **Yes** to use the same command file. This action produces the **Multidimensional Analysis** window. To proceed, right-click next to the **Rasch** test tab (right-hand side) to insert a fourth test. The default tab is **Test4**. Rename to **TRT** by right-clicking this tab and then selecting the **Rename** option. Once done, add a title and (optional) comments.

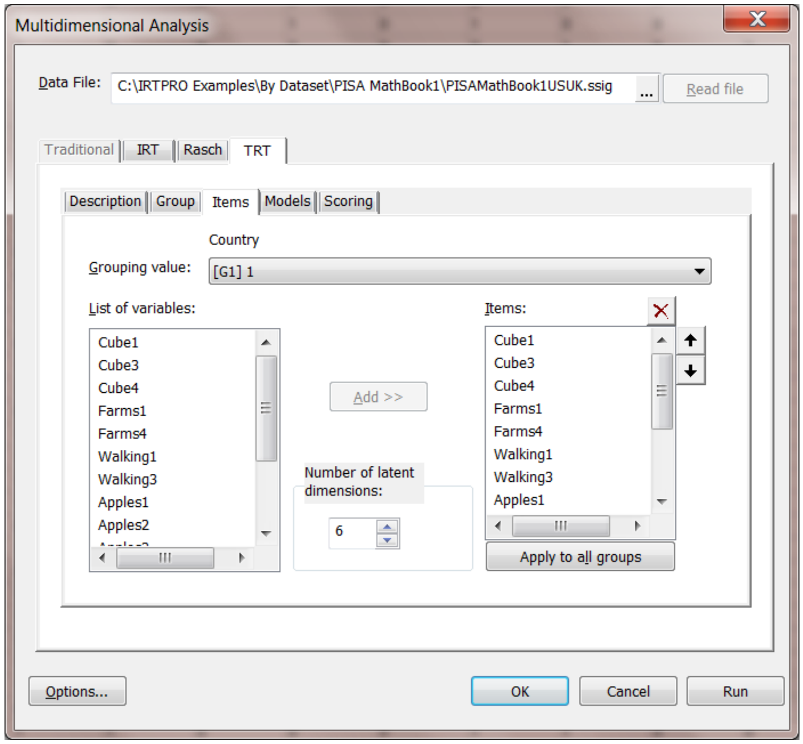


Once this action is completed, click the **Group** tab and select Country as the grouping variable (1 = US, 2 = UK). Next, use the **Items** tab to select the 14 items for the first group and the click the **Apply to all groups** button to select the same set of items for the UK group. As mentioned at the start, the 14 items can be regarded as consisting of five testlets. The five testlets are as follows:

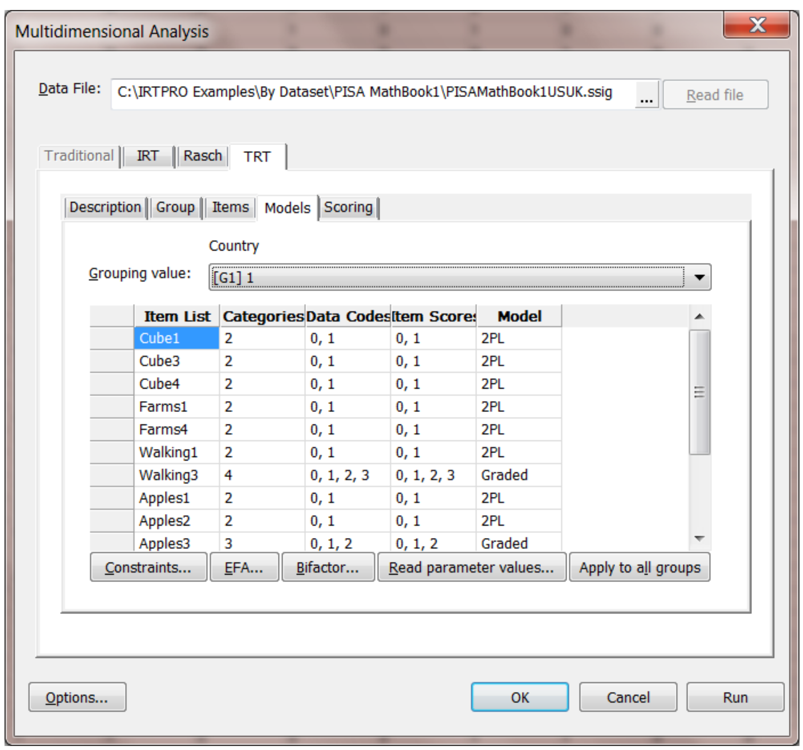
- Testlet 1: Cube1, Cube2 and Cube3
- Testlet 2: Farms1 and Farms4
- Testlet 3: Walking1 and Walking3
- Testlet 4: Apples1, Apples2 and Apples3
- Testlet 5: Grow1, Grow3 and Grow2

In what follows, we postulate that there are six factors, the first being a general mathematics achievement factor and the last five describing each testlet. It is further assumed that these testlets are mutually uncorrelated. The testlets are related to the general factor, but the testlet-specific factors in the TRT model are not correlated with the first/general dimension. This assumption allows one to solve the likelihood and derivatives equations using two-dimensional, rather than six-dimensional quadrature. Note that the item Continent is assigned to the general factor only.

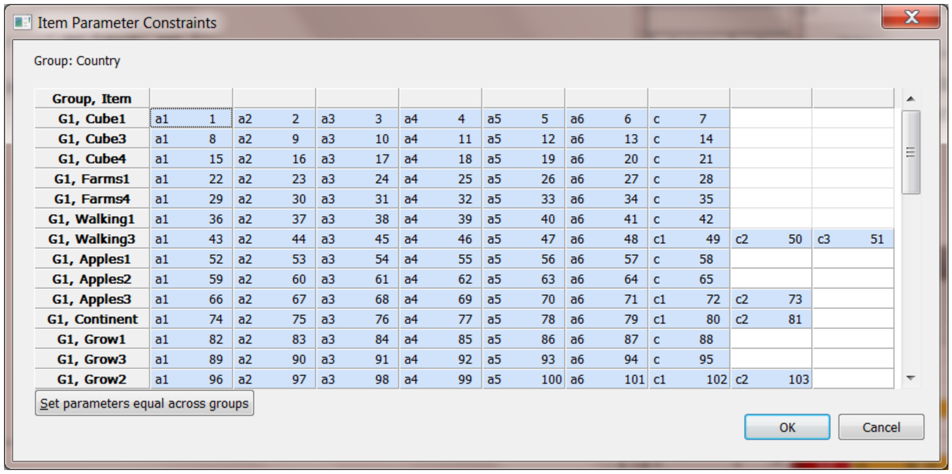
Before proceeding to the **Models** tab, change the **Number of latent dimensions** to 6 as shown.



For illustrative purposes, the default model types (2PL and Graded) are used.

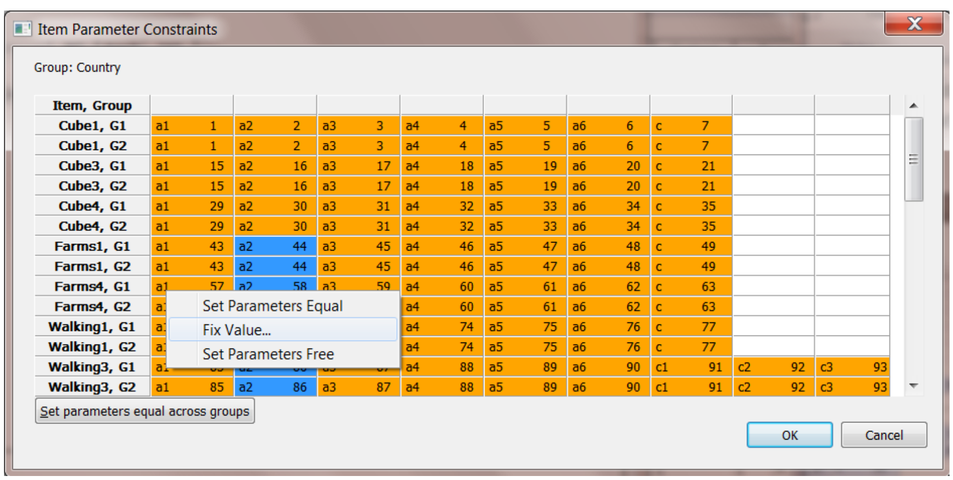


To specify that the five testlets are mutually uncorrelated, contraints are imposed on the slope parameters. Access to the **Item Parameter Constraints** window is obtained by clicking on the **Constraints…** button in the **Models** window.

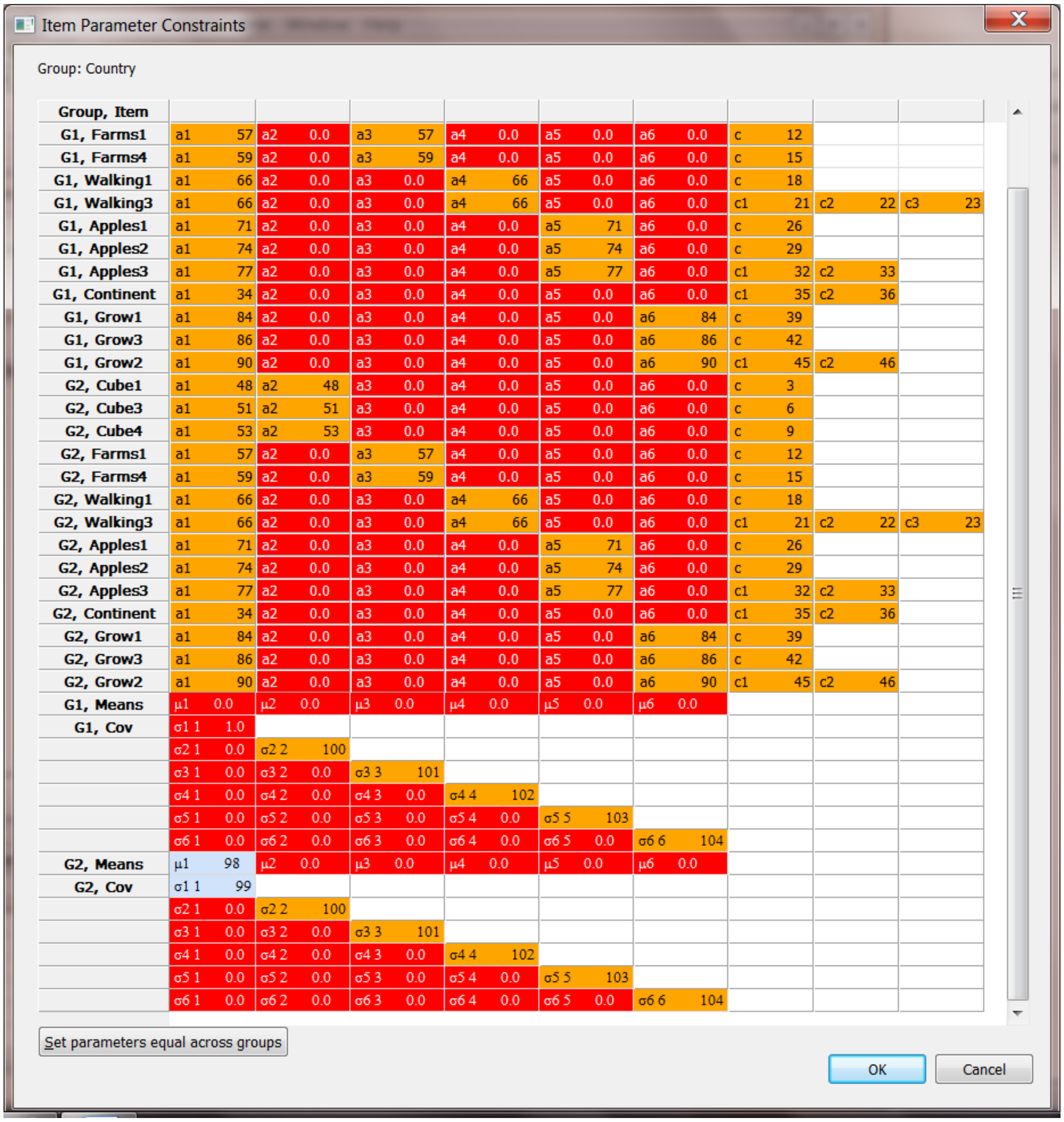


Double click the **Group, Item** button to change the sorting order to Item, Group. Once this is done, click the **Set parameters equal across groups** button and then start by selecting all the a2 cells below the Cube4, G2 cell. Right click and from the drop-down menu, select the **Fix Value…** option.

The default value is 0.0 and by clicking **OK**, all the selected cells will become red in color and show a value of 0.0.

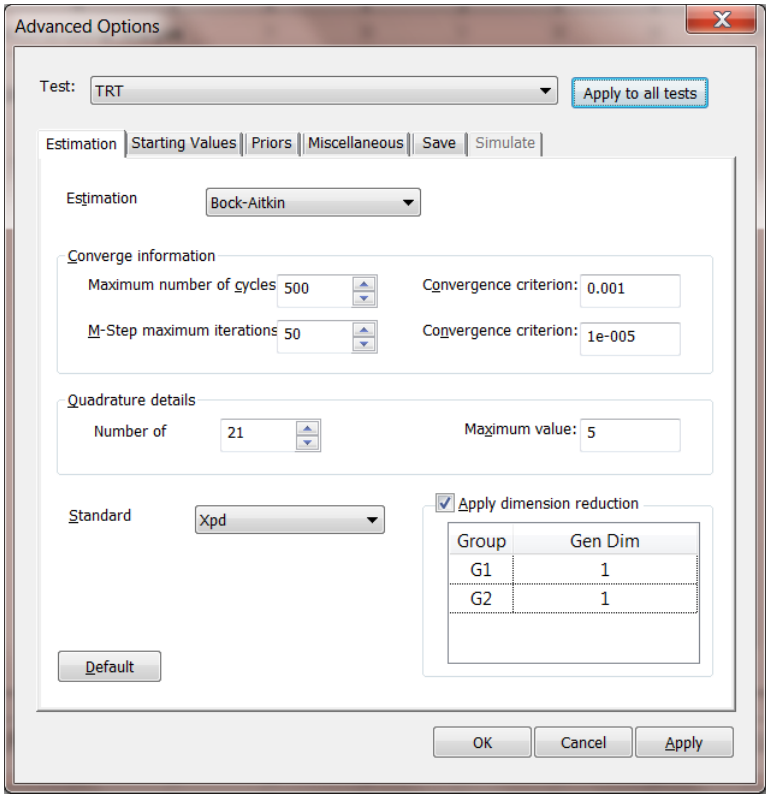


Repeat this procedure for testlets 2 to 5 by fixing all the cells, not belonging to a specific testlet, equal to zero as shown below.



A crucial step for the TRT model involves setting all the second tier slopes equal to each respective item's general–f actor slope, as shown above. Note also that the variances of factors two to six have been freed, and set equal across groups.

When done, click the **OK** button to return to the **Models** window, then click the **Options** button to obtain the **Advanced Options** window. Select the **Estimation** tab and set the number of quadrature points equal to 21 and the integration range from –5 to 5 (**Maximum value**: 5). Change the **Standard error estimation** method to **Xpd** and select the **Apply dimension reduction** option and set the number of general dimensions to 1 for both groups. Also change the convergence criteria values to those shown below.



Click **OK** and then the **Run** button to start the analysis. Portions of the output is listed below. First, the parameter estimates and estimated standard errors are given for all the items that have only two categories (2PL model) followed by the parameter estimates for the items associated with the Graded model.

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

Item	Label	a_1	s.e.	a_2	s.e.	a_3	s.e.	a_4	s.e.	a_5	s.e.	a_6	s.e.	c	s.e.
1	Cube1	[21] 0.99	0.12	[21] 0.99	0.12	0.00	-----	0.00	-----	0.00	-----	0.00	-----	[1] 0.62	0.10
2	Cube3	[22] 3.34	1.08	[22] 3.34	1.08	0.00	-----	0.00	-----	0.00	-----	0.00	-----	[2] 3.51	1.05
3	Cube4	[23] 1.24	0.15	[23] 1.24	0.15	0.00	-----	0.00	-----	0.00	-----	0.00	-----	[3] -1.26	0.15
4	Farms1	[24] 2.47	0.35	0.00	-----	[24] 2.47	0.35	0.00	-----	0.00	-----	0.00	-----	[4] -0.01	0.17
5	Farms4	[25] 0.75	0.09	0.00	-----	[25] 0.75	0.09	0.00	-----	0.00	-----	0.00	-----	[5] 0.20	0.08
6	Walking1	[26] 2.65	0.20	0.00	-----	0.00	-----	[26] 2.65	0.20	0.00	-----	0.00	-----	[6] -2.31	0.22
8	Apples1	[27] 1.51	0.15	0.00	-----	0.00	-----	0.00	-----	[27] 1.51	0.15	0.00	-----	[10] 0.25	0.12
9	Apples2	[28] 2.76	0.28	0.00	-----	0.00	-----	0.00	-----	[28] 2.76	0.28	0.00	-----	[11] -3.27	0.31
12	Grow1	[30] 1.25	0.13	0.00	-----	0.00	-----	0.00	-----	0.00	-----	[30] 1.25	0.13	[17] -0.04	0.11
13	Grow3	[31] 1.53	0.16	0.00	-----	0.00	-----	0.00	-----	0.00	-----	[31] 1.53	0.16	[18] 0.31	0.12

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

Item	Label	a_1	s.e.	a_2	s.e.	a_3	s.e.	a_4	s.e.	a_5	s.e.	a_6	s.e.
7	Walking3	[26] 2.65	0.20	0.00	-----	0.00	-----	[26] 2.65	0.20	0.00	-----	0.00	-----
10	Apples3	[29] 3.05	0.37	0.00	-----	0.00	-----	0.00	-----	[29] 3.05	0.37	0.00	-----
11	Continent	[16] 1.97	0.17	0.00	-----	0.00	-----	0.00	-----	0.00	-----	0.00	-----
14	Grow2	[32] 0.88	0.08	0.00	-----	0.00	-----	0.00	-----	0.00	-----	[32] 0.88	0.08

Graded Model Item Parameter Estimates for Group 1, logit: aθ + c

Item	Label	c_1	s.e.	c_2	s.e.	c_3	s.e.
7	Walking3	[7] -1.80	0.20	[8] -4.87	0.30	[9] -6.69	0.37
10	Apples3	[12] -4.04	0.44	[13] -6.14	0.57		
11	Continent	[14] -1.33	0.16	[15] -4.07	0.22		
14	Grow2	[19] 1.81	0.11	[20] -0.55	0.08		

Since the parameter estimates were constrained to be equal across groups, the corresponding results for the second group are not shown here. Note, however, that there are small differences between the factor loadings for the two groups. This can be attributed to the fact that the mean and variance associated with the general factor were estimated freely for the UK group. The factor loadings for the general factor and those associated with each testlet are larger or equal to 0.40 and highly significant (z-value = parameter estimate divided by standard error).

Factor Loadings for Group 1 (Back to TOC)

Item	Label	λ_1	s.e.	λ_2	s.e.	λ_3	s.e.	λ_4	s.e.	λ_5	s.e.	λ_6	s.e.
1	Cube1	0.44	0.06	0.44	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	Cube3	0.64	0.03	0.64	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	Cube4	0.50	0.05	0.50	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	Farms1	0.78	0.03	0.00	0.00	0.78	0.03	0.00	0.00	0.00	0.00	0.00	0.00
5	Farms4	0.40	0.07	0.00	0.00	0.40	0.07	0.00	0.00	0.00	0.00	0.00	0.00
6	Walking1	0.77	0.03	0.00	0.00	0.00	0.00	0.77	0.03	0.00	0.00	0.00	0.00
7	Walking3	0.77	0.03	0.00	0.00	0.00	0.00	0.77	0.03	0.00	0.00	0.00	0.00
8	Apples1	0.64	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.64	0.06	0.00	0.00
9	Apples2	0.79	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.79	0.04	0.00	0.00
10	Apples3	0.81	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.81	0.04	0.00	0.00
11	Continent	0.76	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	Grow1	0.56	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.56	0.06
13	Grow3	0.63	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.63	0.05
14	Grow2	0.44	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.44	0.06

Factor Loadings for Group 2 (Back to TOC)

Item	Label	λ_1	s.e.	λ_2	s.e.	λ_3	s.e.	λ_4	s.e.	λ_5	s.e.	λ_6	s.e.
1	Cube1	0.45	0.06	0.45	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	Cube3	0.65	0.04	0.65	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	Cube4	0.50	0.06	0.50	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	Farms1	0.81	0.05	0.00	0.00	0.81	0.05	0.00	0.00	0.00	0.00	0.00	0.00
5	Farms4	0.40	0.07	0.00	0.00	0.40	0.07	0.00	0.00	0.00	0.00	0.00	0.00
6	Walking1	0.80	0.05	0.00	0.00	0.00	0.00	0.80	0.05	0.00	0.00	0.00	0.00
7	Walking3	0.80	0.05	0.00	0.00	0.00	0.00	0.80	0.05	0.00	0.00	0.00	0.00
8	Apples1	0.65	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.65	0.07	0.00	0.00
9	Apples2	0.82	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.82	0.05	0.00	0.00
10	Apples3	0.84	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.84	0.05	0.00	0.00
11	Continent	0.78	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	Grow1	0.57	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.57	0.06
13	Grow3	0.64	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.64	0.06
14	Grow2	0.45	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.45	0.06

Group Parameter Estimates: (Back to TOC)

Group	Label	μ_1	s.e.	μ_2	s.e.	μ_3	s.e.	μ_4	s.e.	μ_5	s.e.	μ_6	s.e.
1	G1	0.00	-----	0.00	-----	0.00	-----	0.00	-----	0.00	-----	0.00	-----
2	G2	[33] 0.51	0.07	0.00	-----	0.00	-----	0.00	-----	0.00	-----	0.00	-----

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

θ_1	s.e.	θ_2	s.e.	θ_3	s.e.	θ_4	s.e.	θ_5	s.e.	θ_6	s.e.
1.00	-----										
0.00	-----	[35] 1.19	0.22								
0.00	-----	0.00	-----	[36] 0.15	0.12						
0.00	-----	0.00	-----	0.00	-----	[37] 0.26	0.06				
0.00	-----	0.00	-----	0.00	-----	0.00	-----	[38] 0.21	0.06		
0.00	-----	0.00	-----	0.00	-----	0.00	-----	0.00	-----	[39] 0.32	0.10

Latent Variable Variance-Covariance Matrix for Group 2 (Back)

θ_1	s.e.	θ_2	s.e.	θ_3	s.e.	θ_4	s.e.	θ_5	s.e.	θ_6	s.e.
[34] 0.91	0.12										
0.00	-----	[35] 1.19	0.22								
0.00	-----	0.00	-----	[36] 0.15	0.12						
0.00	-----	0.00	-----	0.00	-----	[37] 0.26	0.06				
0.00	-----	0.00	-----	0.00	-----	0.00	-----	[38] 0.21	0.06		
0.00	-----	0.00	-----	0.00	-----	0.00	-----	0.00	-----	[39] 0.32	0.10

Likelihood-based Values and Goodness of Fit Statistics (Back to TOC)

Statistics based on the loglikelihood	
-2loglikelihood:	21001.78
Akaike Information Criterion (AIC):	21079.78
Bayesian Information Criterion (BIC):	21279.79

The deviance statistic (-2 log likelihood) for the TRT model is reported above as 21001.78. The corresponding value for the IRT model (see Section 6.2.2) is 21233.40. The chi-squared difference test therefore yields a value of 21233.40 - 21001.78 = 231.62. The degrees of freedom for testing between which of the TRT or IRT models provide the better fit are 17 (39 parameters were estimated in the case of the TRT model versus 22 for the IRT model). Since the chi-squared difference test is highly significant, we conclude that the TRT model provides a better fit to the item responses when compared to the IRT model. Information-theoretic indices of fit (AIC and BIC) also point to the TRT model as better fitting.

7.4 Two-tier analysis of PISA Read and Math items

Cai (2010) proposed a two-tier item factor analysis model that subsumes standard multidimensional IRT models, bifactor IRT models, and testlet response theory (TRT) models as special cases. Features of the model lead to a reduction in the dimensionality of the latent variable space and consequently significant computational savings.

Similar to the success story of full-information item bifactor analysis (see *e.g.*, Gibbons *et al*., 2007, 2008), the existence of certain special features and restrictions can result in significant computational savings for maximum marginal likelihood estimation while keeping the model flexible enough to represent a variety of structures commonly found in educational and psychological measurement.

The two-tier model also generalizes the bifactor or testlet models in the types of observed variables that can be included, permitting an arbitrary mixture of dichotomous, ordinal, and nominal items. Extending the subdomain scoring strategies discussed by Gibbons *et al*. (2007) for item bifactor models, the two-tier model conveniently provides individual response pattern based IRT scale scores (as posterior expected values) for all latent variables in the model. Finally, the two-tier item factor analysis (IFA) model highlights the benefit of analytically reducing the dimensionality of latent variable space whenever possible.

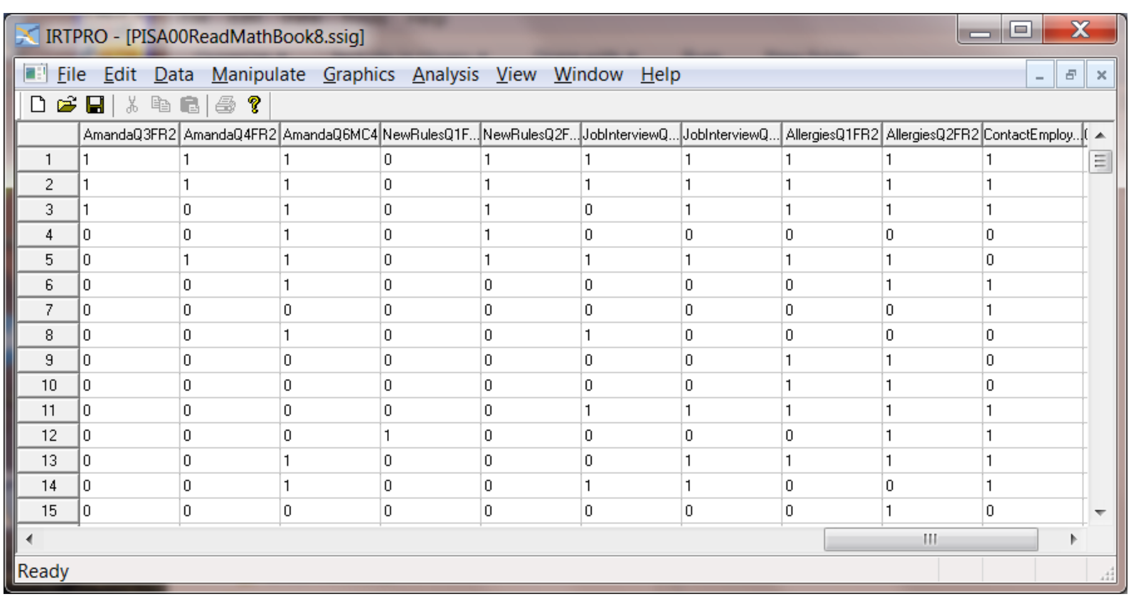
The key to the two-tier modeling framework rests on the recognition that the dimensions (latent variables, factors, latent traits, etc.) in an IFA model can be grouped into two tiers or classes: 1) primary dimensions, and 2) specific dimensions. The distinction is not based so much on the theoretical importance or breadth of the measured latent constructs as on the pattern of factor loadings and the factor inter-correlations. In the two-tier model, the primary dimensions and specific dimensions are uncorrelated. In addition, the specific dimensions are assumed mutually orthogonal and an item can load on at most one specific dimension, just as in a bifactor or testlet response model. On the other hand, the primary dimensions may be correlated among themselves, and the model imposes no further restrictions on the relation between items and primary dimensions beyond necessary conditions for identification.

As mentioned in Section 7.3, the format of the PISA can be best described as testlet-based. For instance, it is standard practice in reading assessments to base several questions on one reading passage so that each question can measure a different aspect of the examinee's comprehension of the passage. PISA is noteworthy in that testlets are employed in all three sections, namely reading, math, and science. For instance, a typical form of PISA reading assessment (in year 2000) consists of about 30 items that can be divided into 8 or 9 testlets made up of non-overlapping sets of items. Some testlets are longer, with 4 or 5 items, and some are shorter, with only 2 items. Critically, an item belongs to one and only one testlet.

As an illustration, consider only the reading and mathematics sections. By design, the reading items measure reading literacy (primary dimension 1) and the math items measure mathematics literacy (primary dimension 2). Test construction results in two dimensions that are strongly correlated (Adams & Wu, 2002), which is understandable because before solving a math problem one must be able to read the instructions first. However, a two-factor model does not entirely reflect the underlying structure of PISA. A dominating feature of testlet-based assessments is that the item responses from within the same testlet tend to be more correlated than across testlets. In the case of PISA, within-testlet residual dependence remains even after controlling for the influence of the two primary dimensions.

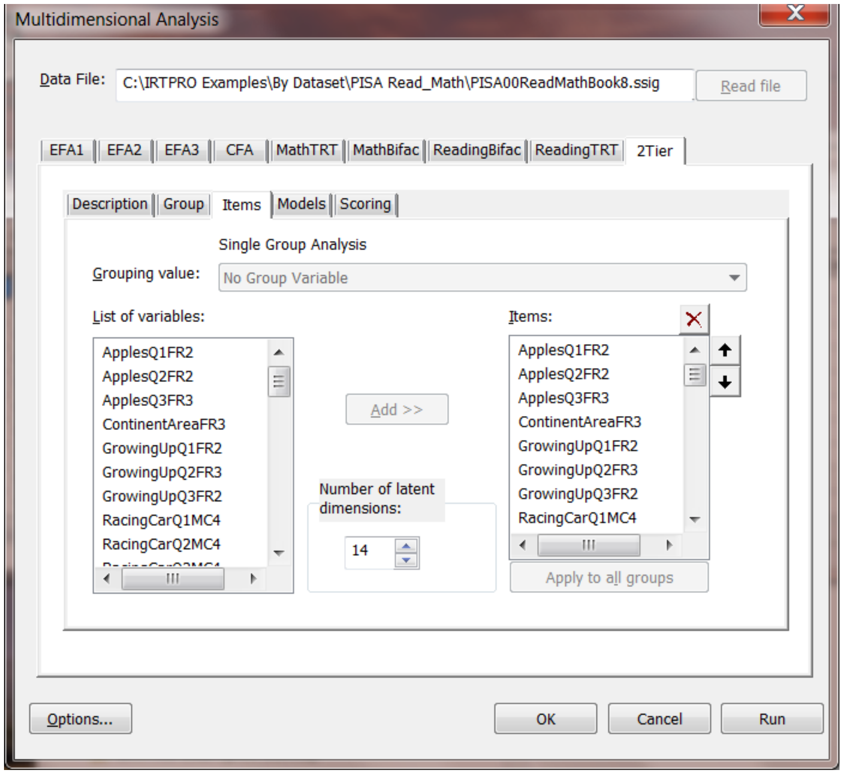
One approach to analyzing the data would be to break the analysis into two parts and fit standard item bifactor (or testlet) models to the first set of math items and the second set of reading items separately. However, if indeed the two primary dimensions are correlated, the two-tier model can utilize that correlation to produce more accurate scores. The ability to "borrow strength" from other parts of the model to enhance statistical prediction is an essential benefit of the two-tier model over separate bifactor analyses that would ignore the correlations among the primary factors.

A set of examples based on the PISA math and read items is contained in the command file **PISAReadMathBook8.irtpro** and is based on the IRTPRO dataset **PISAReadMathBook8.ssig**. These files are located in the folder **IRTPRO Examples\By Dataset\PISA Read_Math**. The first 15 cases for a number of reading items are shown below. The reader is referred to Cai (2011) for a detailed description of this two-tier analysis.

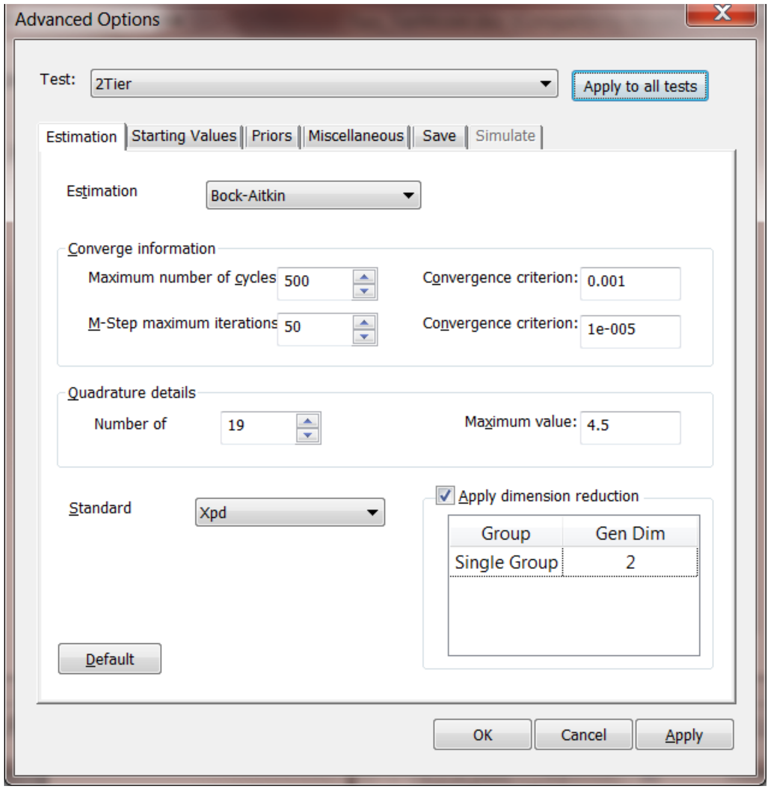


	AmandaQ3FR2	AmandaQ4FR2	AmandaQ6MC4	NewRulesQ1F...	NewRulesQ2F...	JobInterviewQ...	JobInterviewQ...	AllergiesQ1FR2	AllergiesQ2FR2	ContactEmploy...
1	1	1	1	0	1	1	1	1	1	1
2	1	1	1	0	1	1	1	1	1	1
3	1	0	1	0	1	0	1	1	1	1
4	0	0	1	0	1	0	0	0	0	0
5	0	1	1	0	1	1	1	1	1	0
6	0	0	1	0	0	0	0	0	1	1
7	0	0	0	0	0	0	0	0	0	1
8	0	0	1	0	0	1	0	0	0	0
9	0	0	0	0	0	0	0	1	1	0
10	0	0	0	0	0	0	0	1	1	0
11	0	0	0	0	0	1	1	1	1	1
12	0	0	0	1	0	0	0	0	1	1
13	0	0	1	0	0	0	1	1	1	1
14	0	0	1	0	0	1	1	0	0	1
15	0	0	0	0	0	0	0	0	1	0

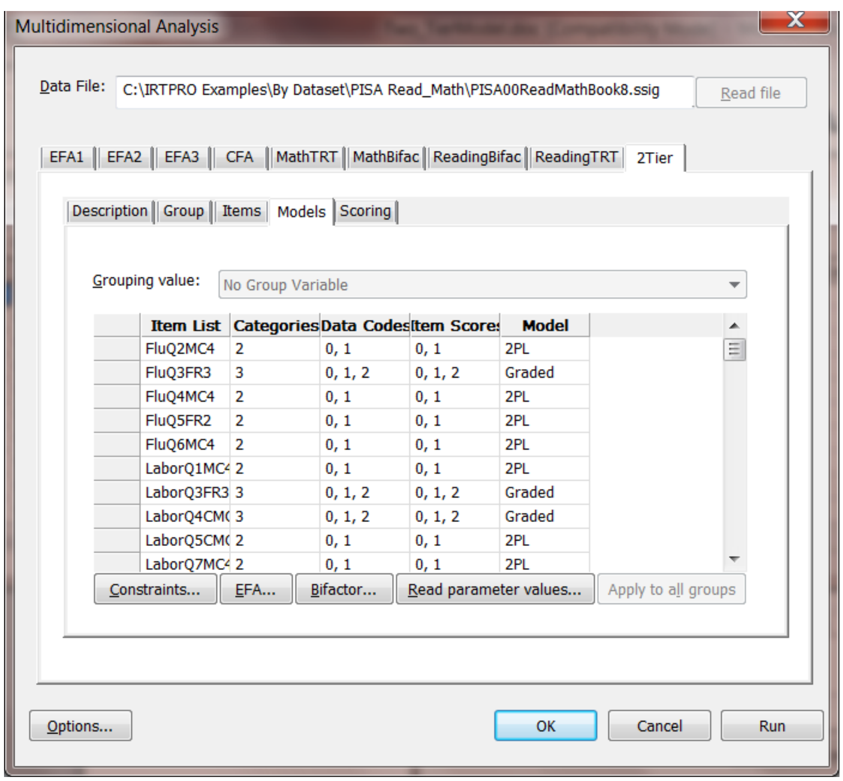
Select the **Analysis, Multidimensional** option and click **Yes** when prompted to use the existing command file. Select the **2Tier** test tab and then click the **Items** tab to obtain the display below. Note that there are two general dimensions and 12 testlets yielding a total of 14 dimensions.



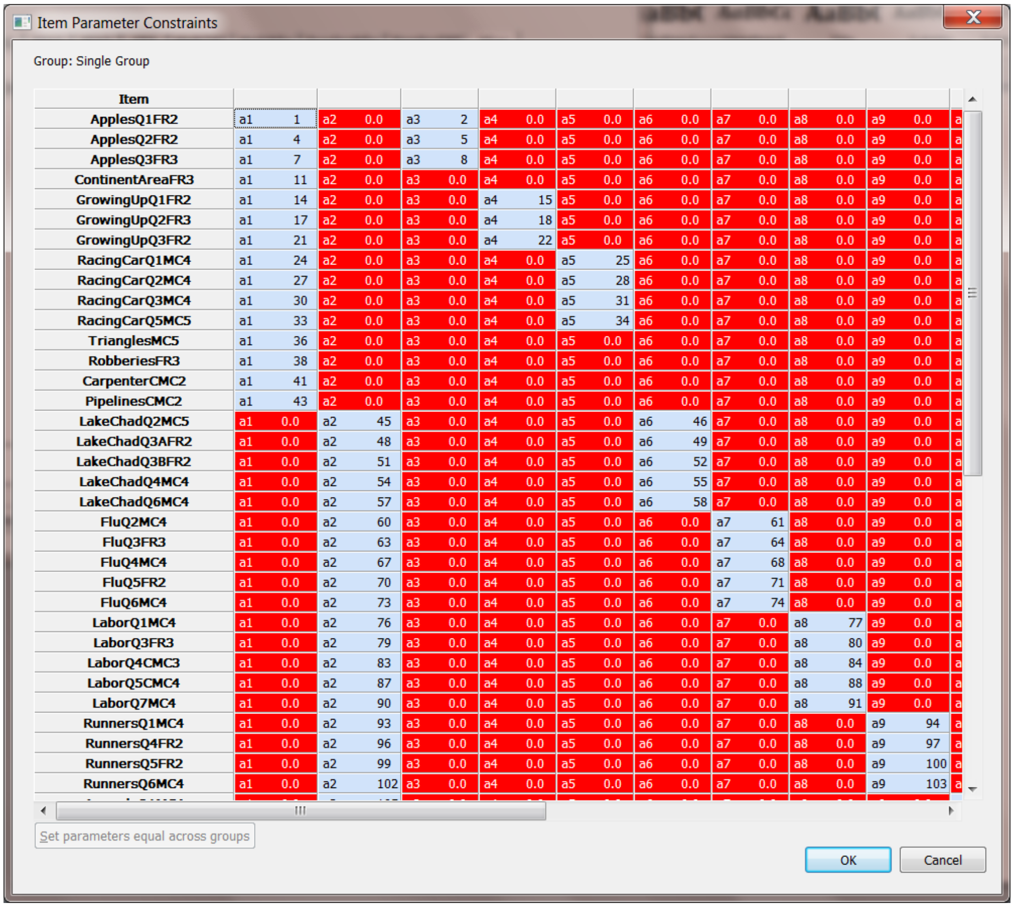
To view the estimation settings that were selected for the two-tier analysis, click the **Options...** button (see above). Note that the **Apply dimension reduction** option is selected and that the number of general dimensions is set to two. Also note that no grouping variable was selected for ths analysis thus assuming a single group.



Click the **OK** button to return to the **Multidimensional Analysis** window then click the **Models** tab to see the list of models that were selected.



Next click the **Constraints…** button to view the **Item Parameter Constraints** window. This window graphically illustrates that all the slope parameters are fixed at zero, except those belonging to the two main dimesions and to the various testlets.



Item									
ApplesQ1FR2	a1 1	a2 0.0	a3 2	a4 0.0	a5 0.0	a6 0.0	a7 0.0	a8 0.0	a9 0.0
ApplesQ2FR2	a1 4	a2 0.0	a3 5	a4 0.0	a5 0.0	a6 0.0	a7 0.0	a8 0.0	a9 0.0
ApplesQ3FR3	a1 7	a2 0.0	a3 8	a4 0.0	a5 0.0	a6 0.0	a7 0.0	a8 0.0	a9 0.0
ContinentAreaFR3	a1 11	a2 0.0	a3 0.0	a4 0.0	a5 0.0	a6 0.0	a7 0.0	a8 0.0	a9 0.0
GrowingUpQ1FR2	a1 14	a2 0.0	a3 0.0	a4 15	a5 0.0	a6 0.0	a7 0.0	a8 0.0	a9 0.0
GrowingUpQ2FR3	a1 17	a2 0.0	a3 0.0	a4 18	a5 0.0	a6 0.0	a7 0.0	a8 0.0	a9 0.0
GrowingUpQ3FR2	a1 21	a2 0.0	a3 0.0	a4 22	a5 0.0	a6 0.0	a7 0.0	a8 0.0	a9 0.0
RacingCarQ1MC4	a1 24	a2 0.0	a3 0.0	a4 0.0	a5 25	a6 0.0	a7 0.0	a8 0.0	a9 0.0
RacingCarQ2MC4	a1 27	a2 0.0	a3 0.0	a4 0.0	a5 28	a6 0.0	a7 0.0	a8 0.0	a9 0.0
RacingCarQ3MC4	a1 30	a2 0.0	a3 0.0	a4 0.0	a5 31	a6 0.0	a7 0.0	a8 0.0	a9 0.0
RacingCarQ5MC5	a1 33	a2 0.0	a3 0.0	a4 0.0	a5 34	a6 0.0	a7 0.0	a8 0.0	a9 0.0
TrianglesMC5	a1 36	a2 0.0	a3 0.0	a4 0.0	a5 0.0	a6 0.0	a7 0.0	a8 0.0	a9 0.0
RobberiesFR3	a1 38	a2 0.0	a3 0.0	a4 0.0	a5 0.0	a6 0.0	a7 0.0	a8 0.0	a9 0.0
CarpenterCMC2	a1 41	a2 0.0	a3 0.0	a4 0.0	a5 0.0	a6 0.0	a7 0.0	a8 0.0	a9 0.0
PipelinesCMC2	a1 43	a2 0.0	a3 0.0	a4 0.0	a5 0.0	a6 0.0	a7 0.0	a8 0.0	a9 0.0
LakeChadQ2MC5	a1 0.0	a2 45	a3 0.0	a4 0.0	a5 0.0	a6 46	a7 0.0	a8 0.0	a9 0.0
LakeChadQ3AFR2	a1 0.0	a2 48	a3 0.0	a4 0.0	a5 0.0	a6 49	a7 0.0	a8 0.0	a9 0.0
LakeChadQ3BFR2	a1 0.0	a2 51	a3 0.0	a4 0.0	a5 0.0	a6 52	a7 0.0	a8 0.0	a9 0.0
LakeChadQ4MC4	a1 0.0	a2 54	a3 0.0	a4 0.0	a5 0.0	a6 55	a7 0.0	a8 0.0	a9 0.0
LakeChadQ6MC4	a1 0.0	a2 57	a3 0.0	a4 0.0	a5 0.0	a6 58	a7 0.0	a8 0.0	a9 0.0
FluQ2MC4	a1 0.0	a2 60	a3 0.0	a4 0.0	a5 0.0	a6 0.0	a7 61	a8 0.0	a9 0.0
FluQ3FR3	a1 0.0	a2 63	a3 0.0	a4 0.0	a5 0.0	a6 0.0	a7 64	a8 0.0	a9 0.0
FluQ4MC4	a1 0.0	a2 67	a3 0.0	a4 0.0	a5 0.0	a6 0.0	a7 68	a8 0.0	a9 0.0
FluQ5FR2	a1 0.0	a2 70	a3 0.0	a4 0.0	a5 0.0	a6 0.0	a7 71	a8 0.0	a9 0.0
FluQ6MC4	a1 0.0	a2 73	a3 0.0	a4 0.0	a5 0.0	a6 0.0	a7 74	a8 0.0	a9 0.0
LaborQ1MC4	a1 0.0	a2 76	a3 0.0	a4 0.0	a5 0.0	a6 0.0	a7 0.0	a8 77	a9 0.0
LaborQ3FR3	a1 0.0	a2 79	a3 0.0	a4 0.0	a5 0.0	a6 0.0	a7 0.0	a8 80	a9 0.0
LaborQ4CMC3	a1 0.0	a2 83	a3 0.0	a4 0.0	a5 0.0	a6 0.0	a7 0.0	a8 84	a9 0.0
LaborQ5CMC4	a1 0.0	a2 87	a3 0.0	a4 0.0	a5 0.0	a6 0.0	a7 0.0	a8 88	a9 0.0
LaborQ7MC4	a1 0.0	a2 90	a3 0.0	a4 0.0	a5 0.0	a6 0.0	a7 0.0	a8 91	a9 0.0
RunnersQ1MC4	a1 0.0	a2 93	a3 0.0	a4 0.0	a5 0.0	a6 0.0	a7 0.0	a8 0.0	a9 94
RunnersQ4FR2	a1 0.0	a2 96	a3 0.0	a4 0.0	a5 0.0	a6 0.0	a7 0.0	a8 0.0	a9 97
RunnersQ5FR2	a1 0.0	a2 99	a3 0.0	a4 0.0	a5 0.0	a6 0.0	a7 0.0	a8 0.0	a9 100
RunnersQ6MC4	a1 0.0	a2 102	a3 0.0	a4 0.0	a5 0.0	a6 0.0	a7 0.0	a8 0.0	a9 103

8. IRT scoring

8.1 Introduction

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. There are two instances under which the IRT scale scores may be one-to-one related (in a nonlinear fashion) to summed scores. First, when the one-parameter logistic (or in general, Rasch) model is used, the summed scores are sufficient statistics for the latent ability variable. Second, when the scale scores are based on summed-score posteriors for any IRT model, the summed scores can be directly translated into scale scores.

The main advantages of scale scores are that they:

- Remain comparable when items are added to or deleted from the tests.
- Weight the individual items optimally according to their discriminating powers.
- Have more accurate standard errors.
- Provide more flexible and robust adjustments for guessing than the classical corrections.
- Are on the same continuum as the item locations.

There are three types of IRT scale score estimation methods that IRTPRO supports:

- Bayes estimation (EAP)
- Summed Score EAP (SSEAP)
- Bayes modal estimation (MAP)

The three types of IRT scale score estimation methods are discussed in the sections to follow.

8.1.1 Bayes estimation (EAP)

The Bayes estimate is the mean of the posterior distribution of θ, given the observed response pattern x_i (Bock & Mislevy, 1982). It can be approximated as accurately as required by the Gaussian quadrature (see the section on MML estimation):

$$\overline{\theta}_i \cong \frac{\sum_{k=1}^{q} X_k P(\mathbf{x}_i \mid X_k) A(X_k)}{\sum_{k=1}^{q} P(\mathbf{x}_i \mid X_k) A(X_k)}.$$

This function of the response pattern $\mathbf{x}_i$ has also been called the expected a posteriori (EAP) estimator. A measure of its precision is the posterior standard deviation (PSD) approximated by

$$PSD(\overline{\theta}_i) \cong \frac{\sum_{k=1}^{q} (X_k - \overline{\theta}_i)^2 P(\mathbf{x}_i \mid X_k) A(X_k)}{\sum_{k=1}^{q} P(\mathbf{x}_i \mid X_k) A(X_k)}.$$

The weights, $A(X_k)$, in these formulas depend on the assumed distribution of θ. Theoretical weights, empirical weights, $A^*(X_k)$, or subjective weights are possibilities.

The EAP estimator exists for any answer pattern and has a smaller average error in the population than any other estimator, including the ML estimator. It is in general biased toward the population mean, but the bias is small within $\pm 3\sigma$ of the mean when the PSD is small (*e.g.*, less than 0.2σ). Although the sample mean of EAP estimates is an unbiased estimator of the mean of the latent population, the sample standard deviation is in general smaller than that of the latent population. This is not a serious problem if all the respondents are measured within the same PSD. But it could be a problem if respondents are compared using alternative test forms that have much different PSDs. The same problem occurs, of course, when number-right scores from alternative test forms with differing reliabilities are used to compare respondents. Tests administrators should avoid making comparisons between respondents who have taken alternative forms that differed appreciably in their psychometric properties. A further implication is that, if EAP estimates are used in computerized adaptive testing, the trials should not terminate after a fixed number of items, but should continue until a prespecified PSD is reached.

8.1.2 Summed Score EAP (SSEAP)

IRT models also imply posteriors for the summed scores, even if the IRT model used is not among the Rasch family of models. Without loss of generality, consider the dichotomous case first. For any IRT model with dichotomous item scores $(u_i = 0,1)$, the likelihood for any summed score $x = \sum u_i$ is

$$L_x(\theta) = \sum_{\sum u_i = x} L(\mathrm{u} \mid \theta)$$

where

$$L(\mathrm{u} \mid \theta) = \prod_i T(u_i \mid \theta)$$

and $T(u_i \mid \theta)$ is the traceline for response u to item i. The first summation is over all such response patterns that the summed score equals x. The probability of each score is

$$P_x = \int L_x(\theta)\, g(\theta)$$

and the expected θ associated with each summed score x is

$$E(\theta \mid x) = \frac{\int \theta L_x(\theta)\, g(\theta)}{P_x}$$

with posterior standard deviation given by

$$PSD\left(\theta \mid x = \sum u_i\right) = \left(\frac{\int \left[\theta - E\left(\theta \mid x\right)\right]^2 L_x\left(\theta\right) g\left(\theta\right)}{P_x} \right)^{\frac{1}{2}}.$$

8.1.3 Bayes modal estimation (MAP)

Similar to the Bayes estimator, but with a somewhat larger average error, is the Bayes modal or so-called maximum a posteriori (MAP) estimator. It is the value of θ that maximizes

$$P(\theta \mid x_i) = \sum_{j=1}^{n} \{x_{ij} \log_e P_i(\theta) + (1 - x_{ij}) \log_e [1 - P_i(\theta)]\} + \log_e g(\theta),$$

where $g(\theta)$ is the density function of a continuous population distribution of θ. The likelihood equation is

$$\sum_{j=1}^{n} \frac{x_{ij} - P_j(\theta)}{P_j(\theta)[1 - P_j(\theta)]} \cdot \frac{\partial P_j(\theta)}{\partial \theta} + \frac{\partial \log_e g(\theta)}{\partial \theta} = 0.$$

Analogous to the maximum likelihood estimate, the MAP estimate is calculated by Fisher scoring, employing the *posterior information*,

$$J(\theta) = I(\theta) + \frac{\partial^2 \log_e g(\theta)}{\partial \theta^2},$$

where the right-most term is the second derivative of the population log density of θ.

In the case of the 2PL model and a normal distribution of θ with variance σ^2, the posterior information is

$$I(\theta) = \sum_{j=1}^{n} a_j^2 P_j(\theta)[1 - P_j(\theta)] + \frac{1}{\sigma^2}.$$

The PSD of the MAP estimate, θ, is approximated by

$$\text{PSD}(\theta) = \sqrt{1 / I(\hat{\theta})}.$$

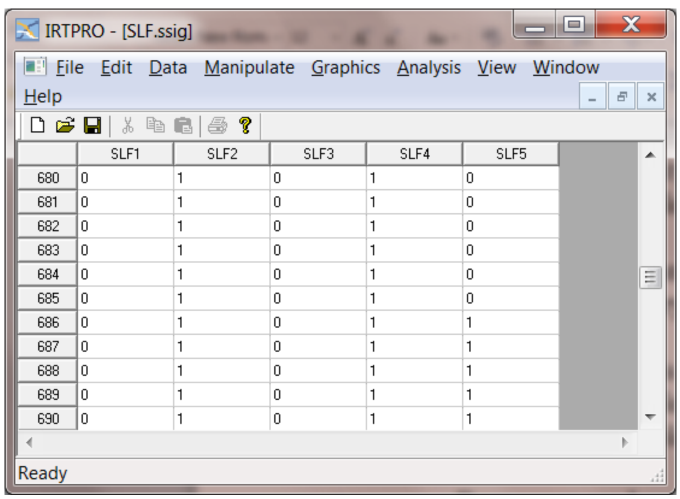
Like the EAP estimator, the MAP estimator exists for all response patterns, but is generally biased toward the population mean.

8.2 Scoring using a social life feelings (SLF) dataset

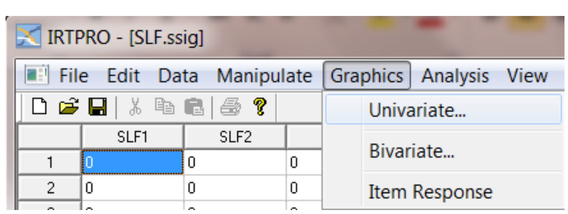
The dataset used in this section is taken from an extensive study of social life feelings reported in Schuessler (1982) and Krebs and Schuessler (1987). The aim was to establish scales for use in social research. According to Bartholomew (1998) the aim of the study was to establish scales for use in social research that were comparable in quality with those used in ability testing. For illustration purposes, the data used in this section is from the German sample consisting of the following five items:

1. Anyone can raise his standard of living if he is willing to work at it (SLS1).
2. Our country has too many poor people who can do little to raise their standard of living (SLS2).
3. Individuals are poor because of the lack of effort on their part (SLS3).
4. Poor people could improve their lot if they tried (SLS4).
5. Most people have a good deal of freedom in deciding how to live (SLS5).

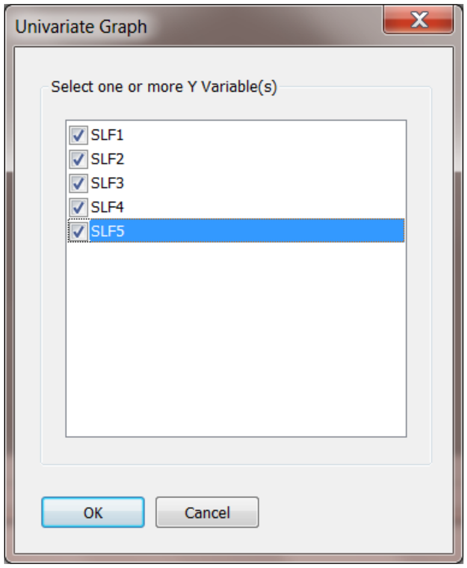
Responses are based on a sample size of 1490 individuals. The spreadsheet below displays item values for cases 680 to 690. The name of the dataset is **SLF.ssig** and is stored in the folder **IRTPRO Examples\By Dataset\Social Life Feelings**



In order to display the frequency distribution of the five items visually, the **Graphics**, **Univariate…** option is selected from the main menu bar.

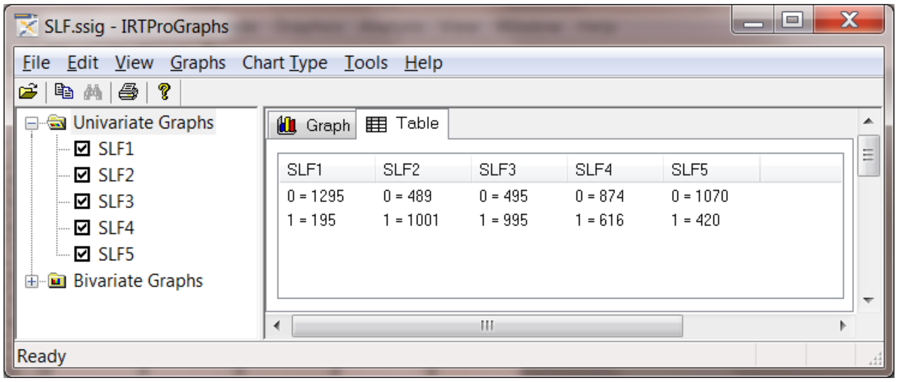


By making this selection, a **Univariate Graph** dialog is displayed.



After selecting the items (see above), click **OK** to obtain a bar chart presentation for each item. From this display, it can be concluded, for example, that a relatively small proportion of individuals have selected the category corresponding to "1" for the item SLF1.

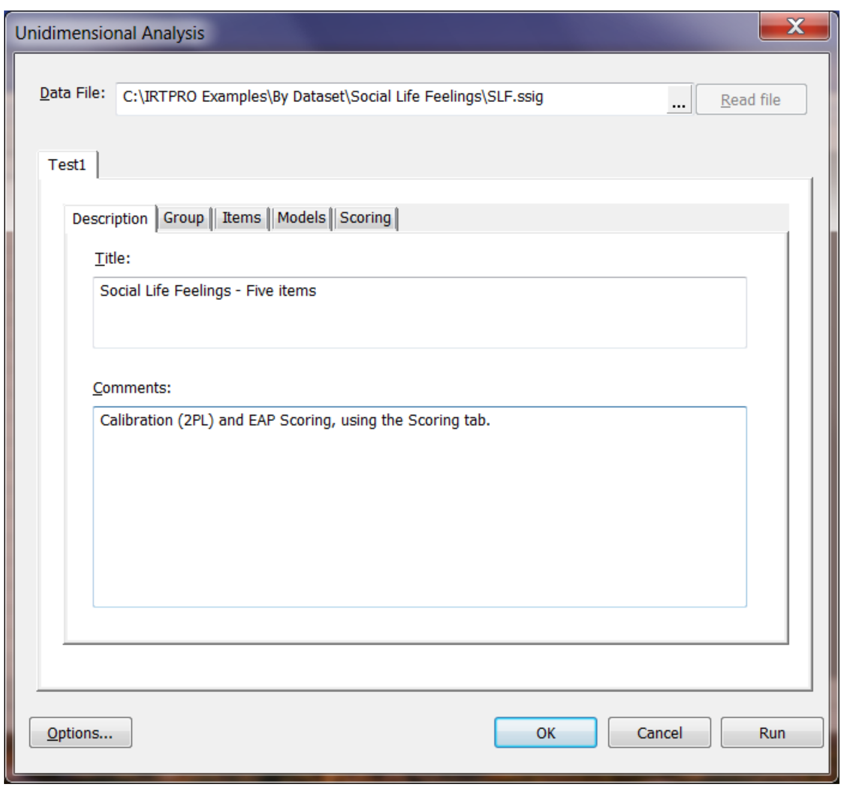
The frequency counts for each item can be displayed by clicking the **Table** icon in the **Graph** window. From this display it follows that all items are binary. We further conclude that there are no missing values present since the total count of the datavalues 0 and 1 equals 1490 for each item.



In Sections 8.2.1 and 8.2.2, examples that illustrate item scoring are presented. In the example presented in Section 8.2.1, 2PL models are fitted to the five items, labeled SLS1 to SLS5 respectively and EAP scores are computed. Use is made of the **Advanced Options, Miscellaneous; Save** selection to specify that the estimated parameters must be saved in a file with extension **-prm.txt**. The example in Section 8.2.2 demonstrates item scoring using a **-prm.txt** parameter file obtained from a previously executed IRT (calibration) run.

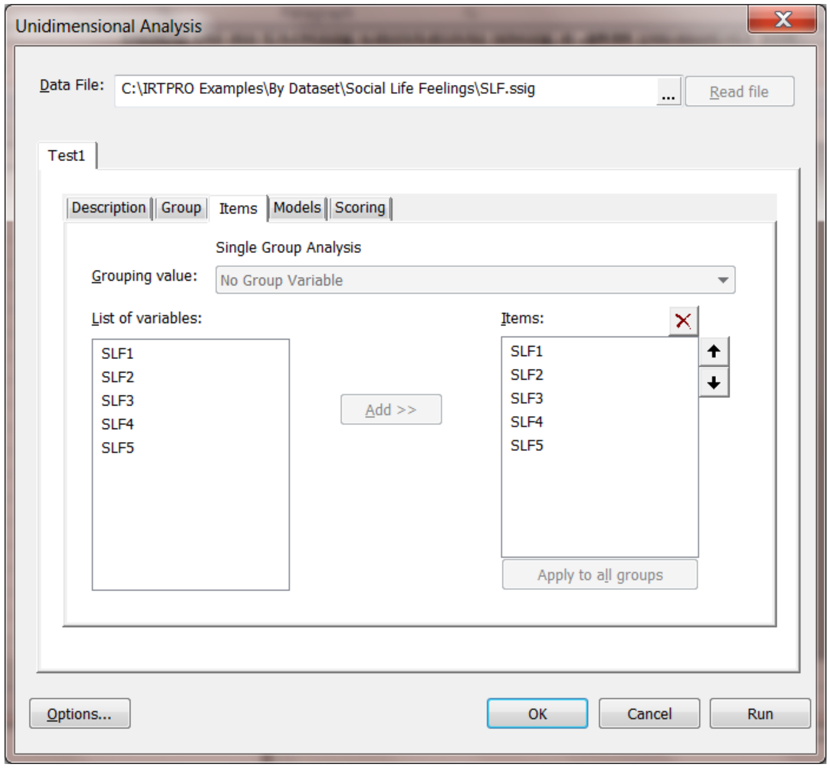
8.2.1 Calibration and Scoring

In this section 2PL models are fitted to the five items, labeled SLS1 to SLS5 respectively, and the estimated parameters are saved to a text file with extension **-prm.txt**. EAP scores (See Section 8.1.1) are also computed.

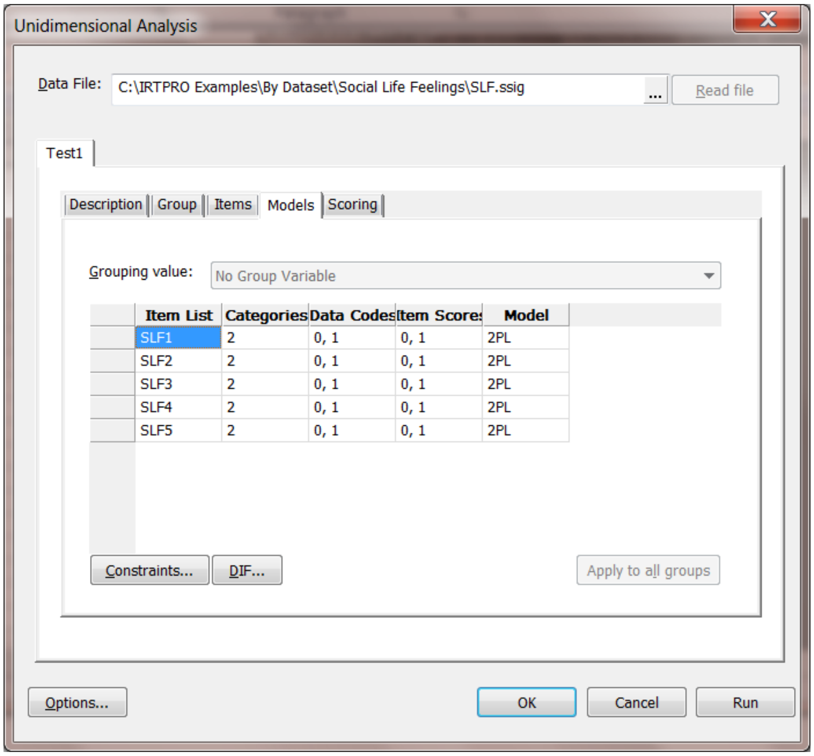


From the main menu bar, select the **Analysis, Unidimensional IRT…** option to obtain the **Unidimensional Analysis** window shown below. Use the **Description** tab to enter a title and comments.

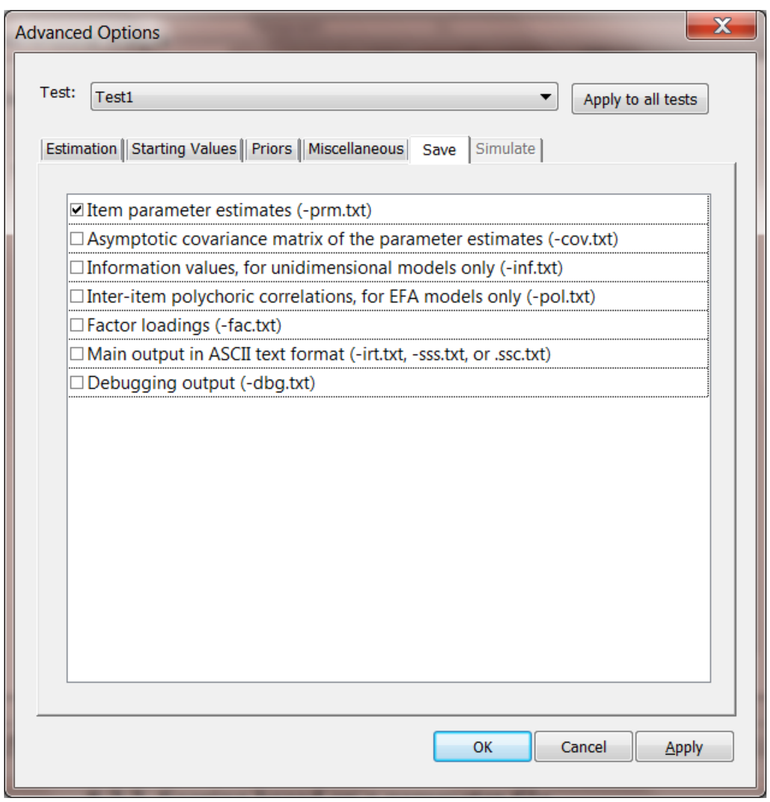
Since the dataset **SLF.ssig** is based on a single group (Germany, 1490 individuals), the **Group** tab is skipped and we proceed to the **Items** tab to select all five items.



The **Models** dialog displays the model-type to be fitted to each item. Since all items are binary, the default model is 2PL.



To ensure that the estimated parameters are save to a **–prm.txt** file, click the **Options…** button (bottom right-hand corner of the previous display). This action invokes the **Advance Options** window. Click the **Save** tab and make sure that **Item parameters estimates (–prm.txt)** is selected.

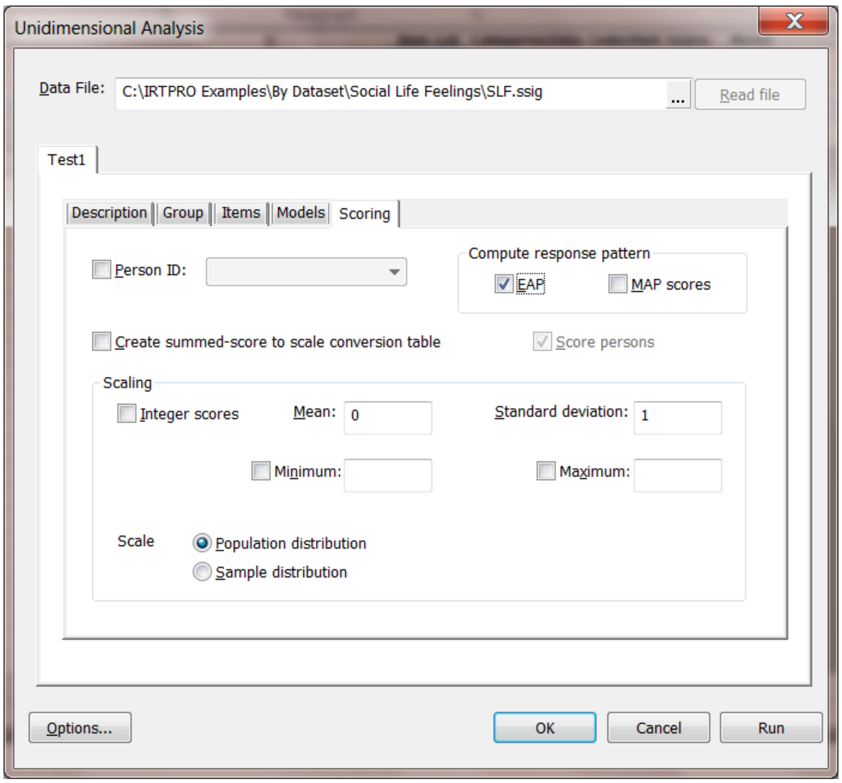


Click **OK** to return to the **Unidimensional Analysis** window and click on the **Scoring** tab to display the **Scoring** dialog.

Using this dialog, make the following selections:

- Scoring method: EAP
- Scaling: Mean = 0; Standard deviation = 1 (the defaults)
- Scale: Population distribution

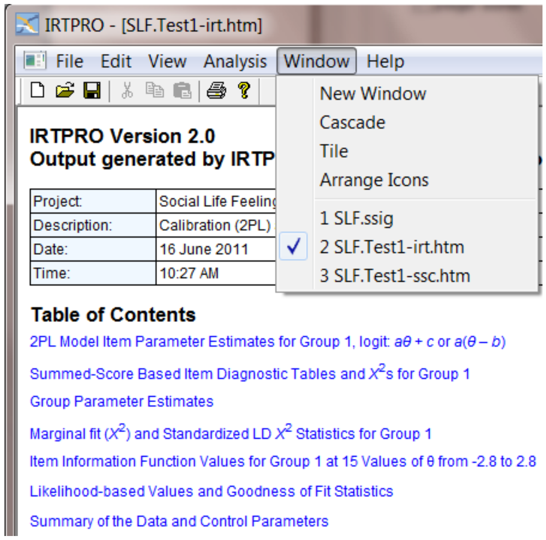
When done, click the **Run** button to start calibration and scoring.



If the analysis completes successfully, two output files are created with extensions:

- –**irt.htm** (calibration), and
- –**ssc.htm** (Scoring).

The Window menu (below) shows the selection of the IRT analysis (calibration) output.



Portions of the output are listed next. The first table gives the parameter estimates and standard error estimates for the five 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	SLF1	2	1.20	0.15	1	-2.35	0.13	1.97	0.18
2	SLF2	4	0.71	0.09	3	0.80	0.06	-1.11	0.14
3	SLF3	6	1.53	0.17	5	0.99	0.09	-0.65	0.06
4	SLF4	8	2.55	0.39	7	-0.67	0.12	0.26	0.04
5	SLF5	10	0.92	0.10	9	-1.10	0.07	1.19	0.12

Likelihood based statistics and fit statistics are given in the output shown below. The statistic: – 2loglikelihood (also called the deviance statistic) is used to compare nested models. Both the AIC and BIC statistics are used as a model selection tool.

Likelihood-based Values and Goodness of Fit Statistics (Back to TOC)

Statistics based on the loglikelihood	
-2loglikelihood:	8258.37
Akaike Information Criterion (AIC):	8278.37
Bayesian Information Criterion (BIC):	8331.43

The RMSEA value of 0.02 indicates a relatively good fit using the 2PL model.

Statistics based on the full item x item x ... classification			
G^2	Degrees of freedom	Probability	RMSEA
39.09	21	0.0095	0.02
X^2	Degrees of freedom	Probability	RMSEA
38.92	21	0.0100	0.02

Next, we select the output file generated for the scoring part of the analysis (**-ssc.htm**). Selected output is shown below. The first portion is a table containing the parameter estimates obtained in the calibration phase.

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

Item	Label	a	c	b
1	SLF1	1.20	-2.35	1.97
2	SLF2	0.71	0.80	-1.11
3	SLF3	1.53	0.99	-0.65
4	SLF4	2.55	-0.67	0.26
5	SLF5	0.92	-1.10	1.19

The next portion of the output shows that the item scores are saved to the text file **SLF.Test-sco.txt**. Text files can be opened with any text editor such as Notepad.

Summary of the Data and Control Parameters (Back to TOC)

Sample Size	1490
Number of Items	5
Number of Dimensions	1

Scoring Control Values

Response pattern EAPs are computed

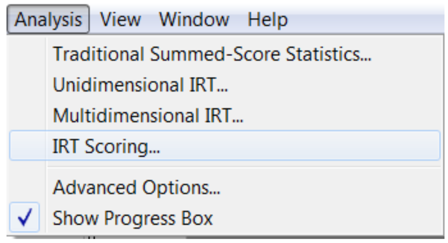
Output Files

HTML results and control parameters:	SLF.Test1-ssc.htm
Text scaled score file:	SLF.Test1-sco.txt

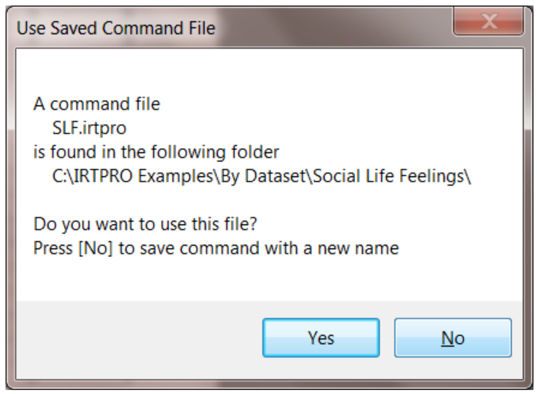
8.2.2 Scoring based on a parameter file

In this section, the summed-score EAP (SSEAP) and maximum a posteriori (MAP) scoring methods are considered. Use is made of the IRTPRO dataset **SLS.ssig** and the parameter estimates, obtained as described in the previous section, are read from a **–prm.txt** parameter file. Scoring is accomplished by selecting the **Analysis, IRT Scoring…** option from the main menu bar.

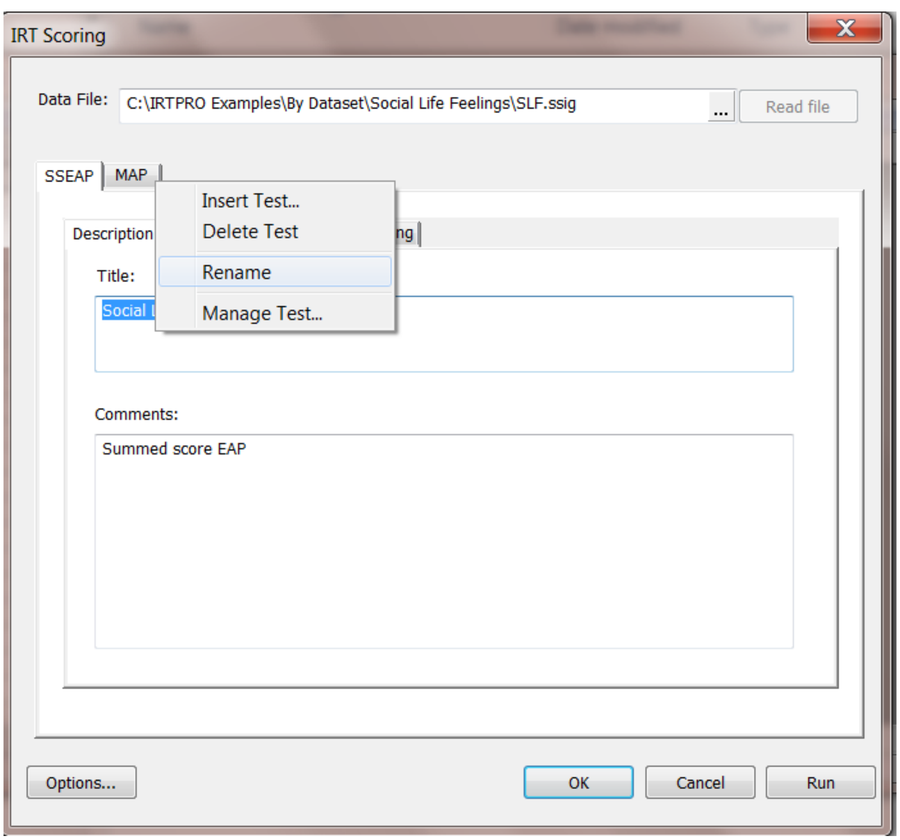
Start by opening the IRTPRO data file **SLS.ssig** located in the folder **IRTPRO Examples\By Dataset\Social Life Feelings\.** If this file is still open from a previous session, close it first and then re-open it, otherwise the **IRT Scoring…** option might be disabled.



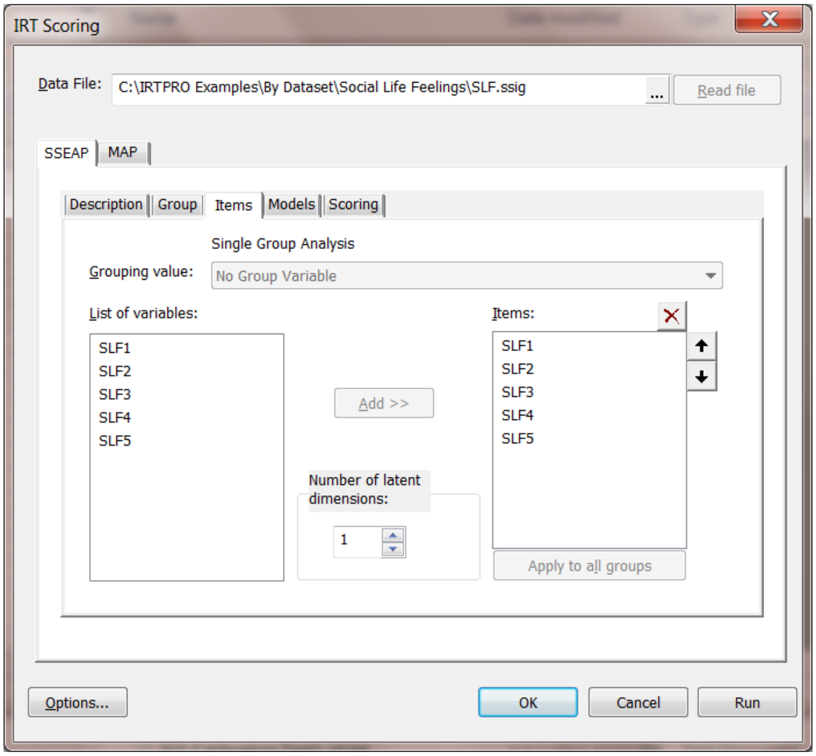
By selecting **IRT Scoring…**, a **Use Save Command File** message box is displayed. Since we do not want to overwrite the existing command file (generated in Section 8.2.1), the **No** button is clicked.



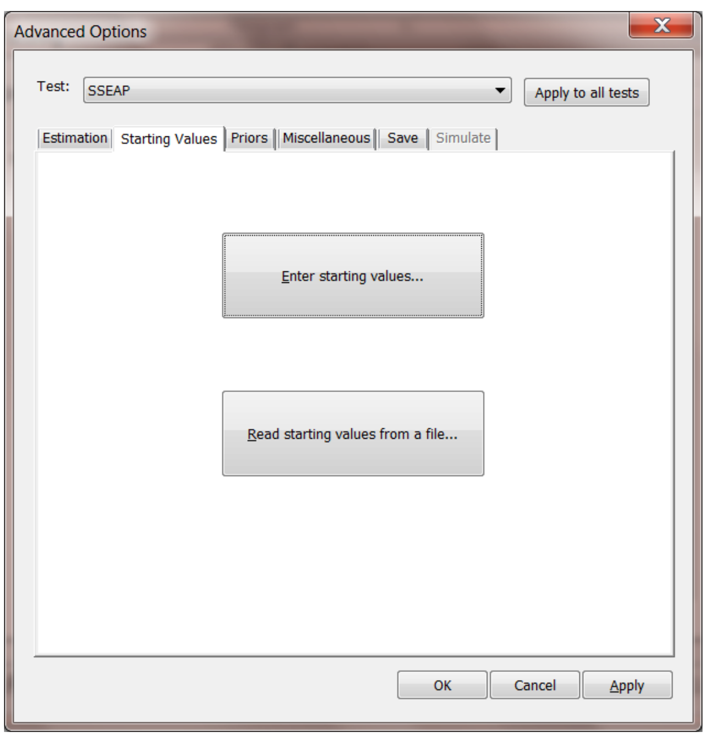
Use the **Insert Test…** and **Rename** options (obtained by right-clicking next to an existing test tab to insert a new test or on a tab to rename a test) to insert a second test and to rename the **Test1** and **Test2** tabs to **SSEAP** and **MAP** respectively as shown below.



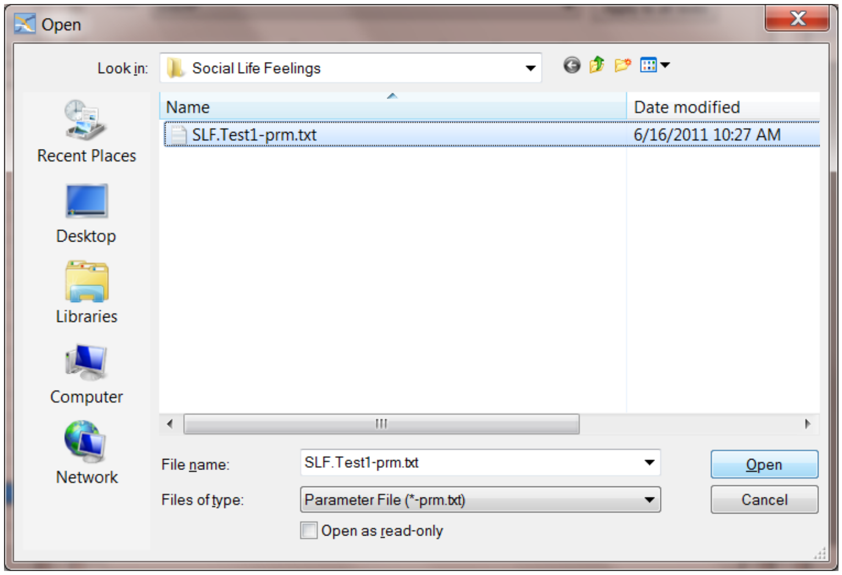
Starting with the **SSEAP** tab, enter a title and (optionally) comments as illustrated. Proceed to the **Items** tab and select the items SLF1 to SLF5.



Click the **Options...** button (lower right-hand corner in display above) to activate the **Advanced Options** window and click the **Starting Values** tab to obtain the dialog shown below.

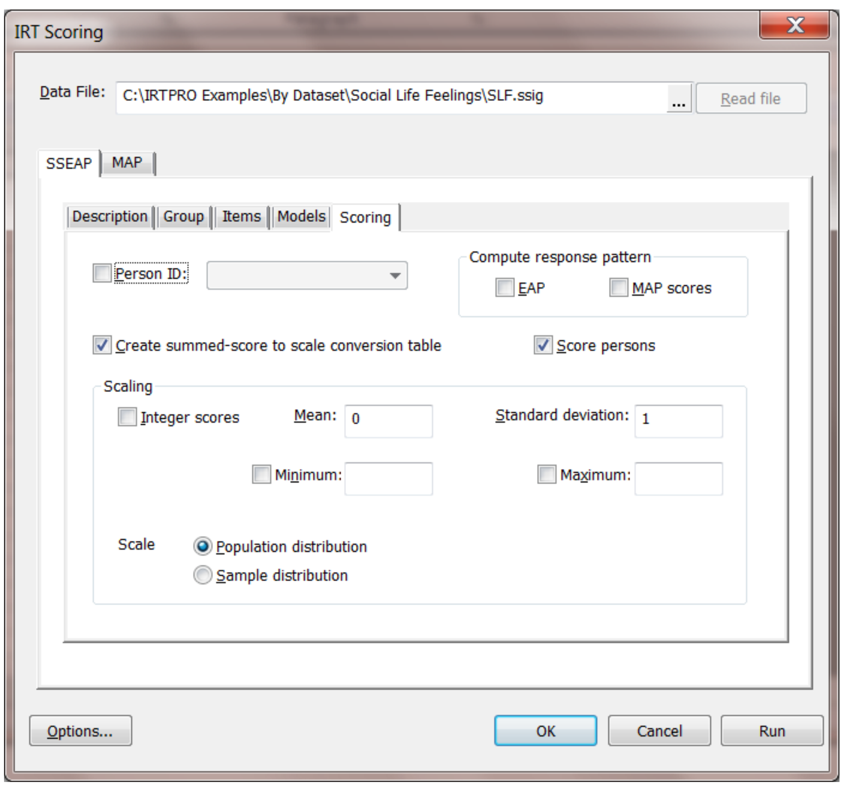


Next click the **Read starting values from a file…** button to display the **Open** dialog, then select **SLF.Test1-prm.txt**. Click the **Open** button to return to the IRT scoring menu.



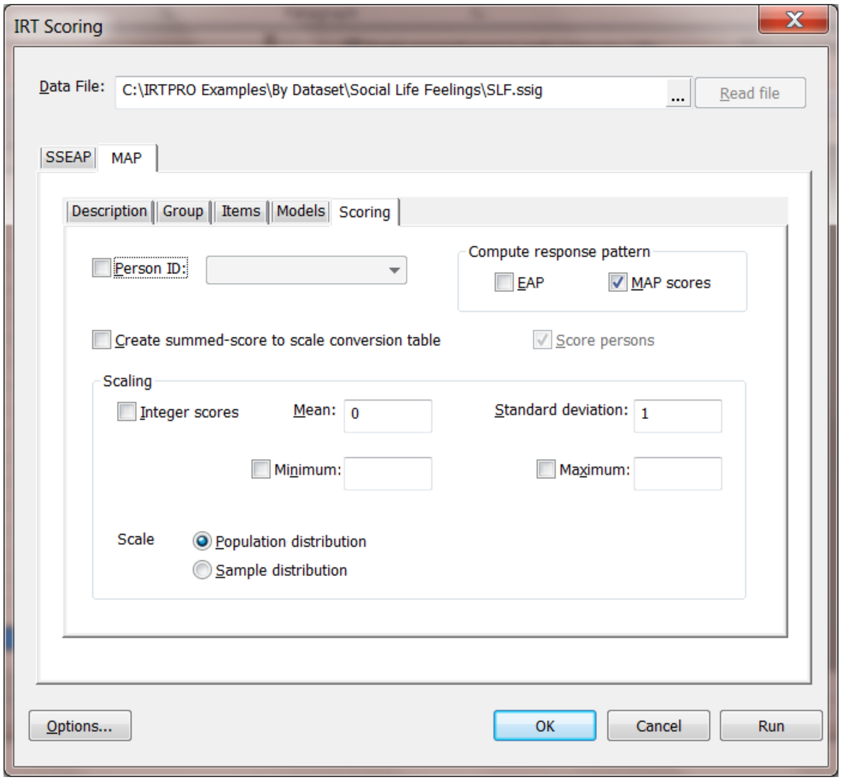
Click the **Scoring** tab and make the following selections:

- Check the **Create summed-score to scale conversion table** option.
- Check the **Score persons** option.
- Select **Scale: Population distribution**.

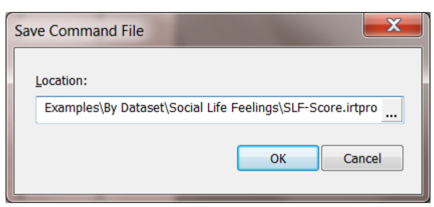


Finally, select the MAP test tab and repeat all the steps (**Description**, **Items**, **Starting values**) described above. However, the **Scoring** dialog should now contain the following selections:

- Check the **MAP scores** option.
- Select Scale: Population distribution.



The scoring procedure is started by clicking the **Run** button. At this stage, the user will have the opportunity to save the command file under a new name. In this case, the default name **SLF.irtpro** is changed to **SLF-Score.irtpro** to ensure that the command file generated in Section 8.2.1 is not overwritten. Click the **OK** button to start the analysis.



Selections of the output for the SSEAP scoring procedure are shown below:

Project:	Social Life Feelings
Description:	Summed score EAP
Date:	16 June 2011
Time:	11:03 AM

Table of Contents

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

Item	Label	a	c	b
1	SLF1	1.20	-2.35	1.97
2	SLF2	0.71	0.80	-1.11
3	SLF3	1.53	0.99	-0.65
4	SLF4	2.55	-0.67	0.26
5	SLF5	0.92	-1.10	1.19

Summed Score to Scale Score Conversion Table (Back to TOC)

Summed Score	EAP[θ|x]	SD[θ|x]	Modeled Proportion
0	-1.191	0.717	0.1087594
1	-0.679	0.682	0.2286010
2	-0.110	0.652	0.2617982
3	0.511	0.649	0.2268359
4	1.022	0.653	0.1359492
5	1.544	0.701	0.0380563

Marginal reliability of the scaled scores for summed scores = 0.55444

Scoring Control Values

Scale scores for summed scores are tabulated and computed	
Summed score equivalence threshold:	0.000010

Note that the file containing the scores is saved as **SL-Score.SSEAP-sco.txt** and can be opened with any text editor.

Output Files

HTML results and control parameters:	SLF-Score.SSEAP-ssc.htm
Text scaled score file:	SLF-Score.SSEAP-sco.txt

The results shown next were obtained for the MAP scoring procedure. The parameter estimates are those obtained from the parameter file created as described in Section 8.2.1.

Project:	Social Life Feelings
Description:	Response Pattern MAP
Date:	16 June 2011
Time:	11:03 AM

Table of Contents

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

Item	Label	a	c	b
1	SLF1	1.20	-2.35	1.97
2	SLF2	0.71	0.80	-1.11
3	SLF3	1.53	0.99	-0.65
4	SLF4	2.55	-0.67	0.26
5	SLF5	0.92	-1.10	1.19

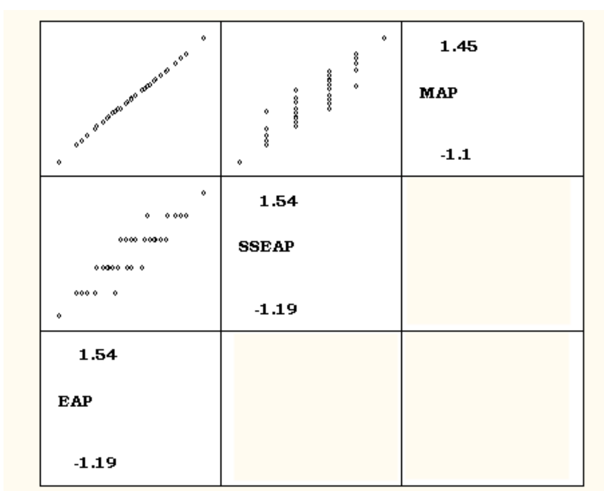
Scoring Control Values

Response pattern MAPs are computed

Output Files

HTML results and control parameters:	SLF-Score.MAP-ssc.htm
Text scaled score file:	SLF-Score.MAP-sco.txt

A matrix plot of the three sets of scores reveals an almost perfect correlation between the scores obtained with EAP and MAP. A plot of the SSEAP scores against the EAP and against the MAP scores shows strong positive correlation.



Conclusions drawn from the matrix scatter plot, are substantiated by calculating the sample statistics of the three sets of scores, the results being reported below:

Descriptive Statistics for three scoring methods

Correlation Matrix

	EAP	SSEAP	MAP
EAP	1.000		
SSEAP	0.962	1.000	
MAP	1.000	0.964	1.000

Means

EAP	SSEAP	MAP
0.000	-0.001	0.016

Standard Deviations

EAP	SSEAP	MAP
0.773	0.746	0.704

The correlations between the three scoring methods are shown below, followed by the means and standard errors of EAP, SSEAP and MAP. There is almost perfect correlation between EAP and MAP (1.000 to three decimal places).

In applied testing situations, the larger standard errors (about 10%) associated with the SSEAPs may be considered a reasonable penalty offset by the ease of summed score based IRT scoring.

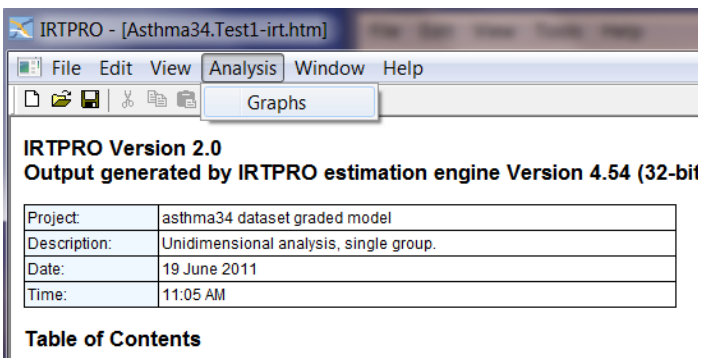
9. Model-based graphics

9.1 Introduction

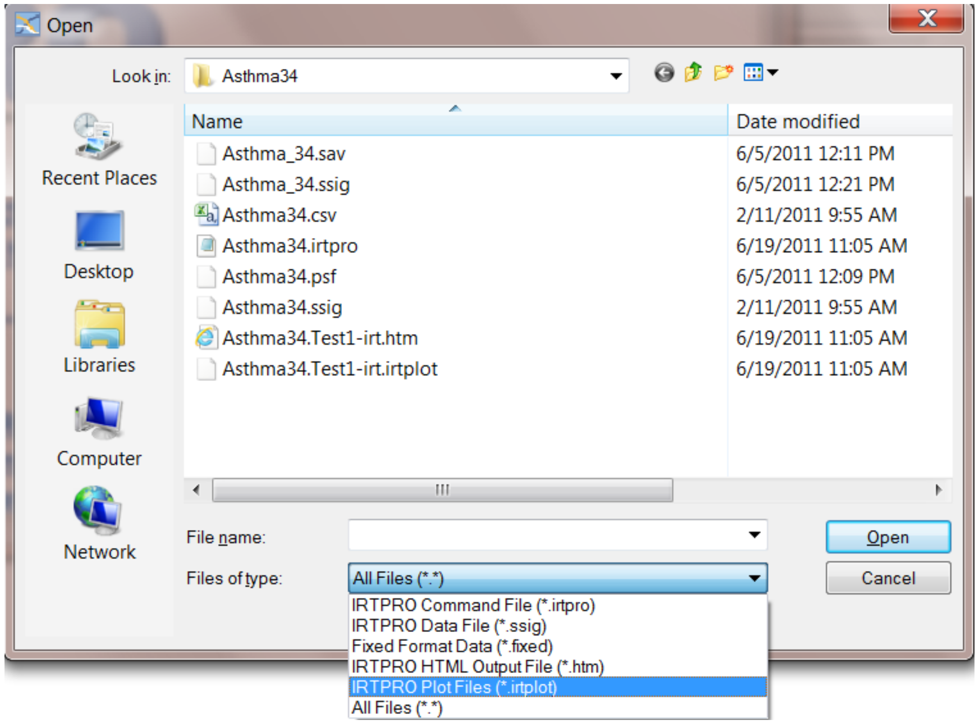
The model-based graphics option in IRTPRO is only available for unidimensional IRT models. There are five different display types available, these being:

- Trace lines (Section 9.2).
- Item information curves (Section 9.3).
- Combined display of trace lines and item information curves (Section 9.4).
- Total information curve (Section 9.5).
- Test characteristic curve (Section 9.6).

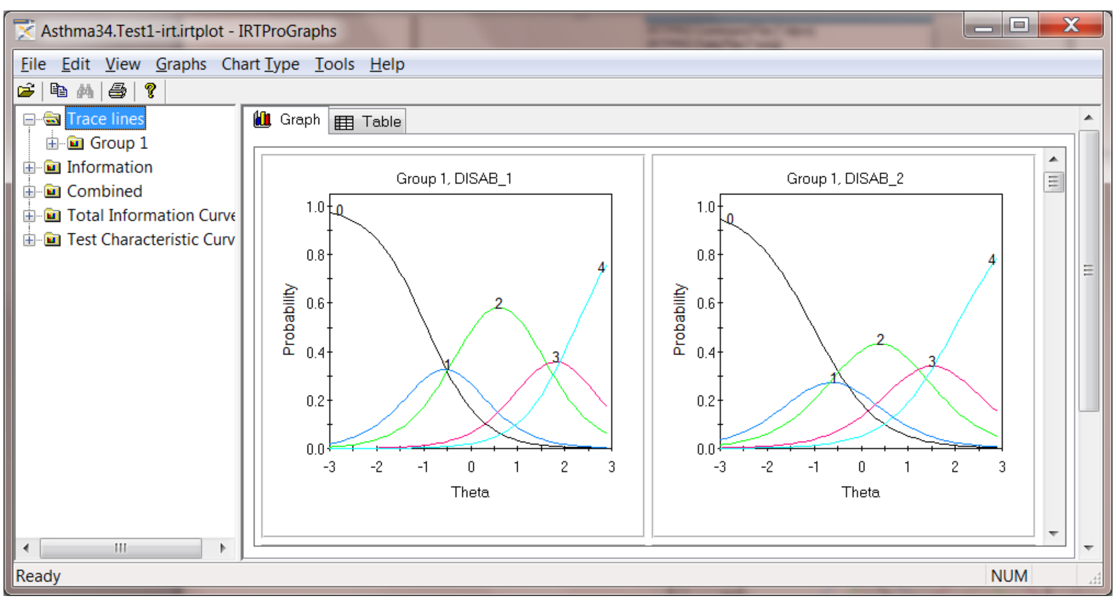
There are two ways to display IRT graphics. The first method is to run a unidimensional analysis. On successful completion of the analysis, an output file with extension **-irt.htm** is produced. With the **-irt.htm** content displayed, select **Analysis**, **Graphs** to obtain trace lines, information and test characteristic curves.



Alternatively, on successful completion of a run, IRTPRO produces a plot file with the extension **-irt.irtplot**. One can use the **File**, **Open** option to locate this file.



The **Graphics** window is displayed by either selecting the **Analysis, Graphs** option or by opening the plot file. By default, trace lines are displayed when this window is opened.

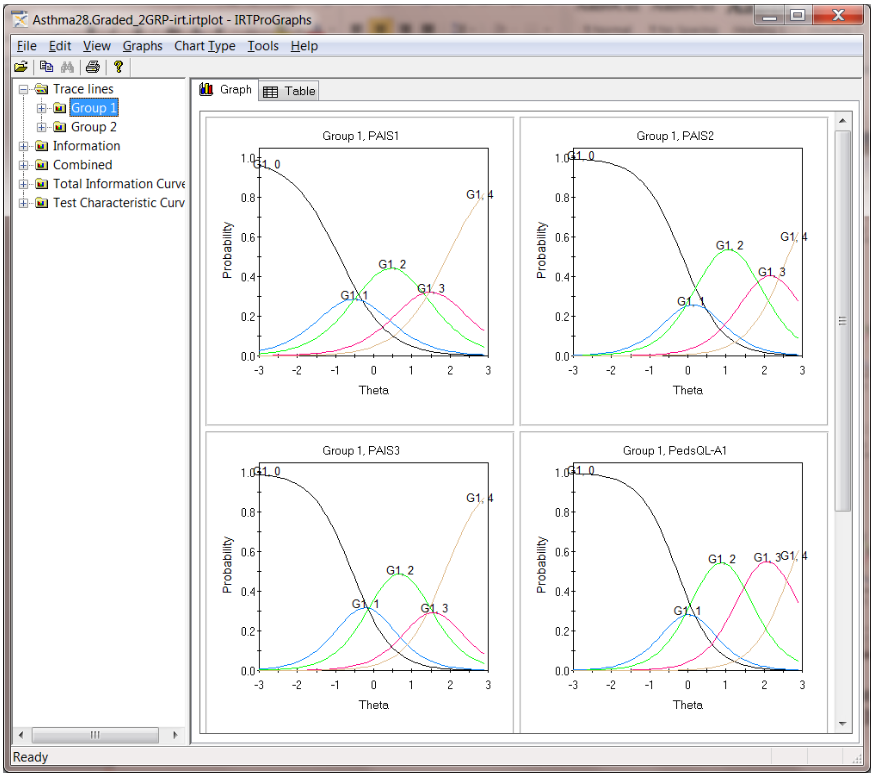


9.2 Trace lines

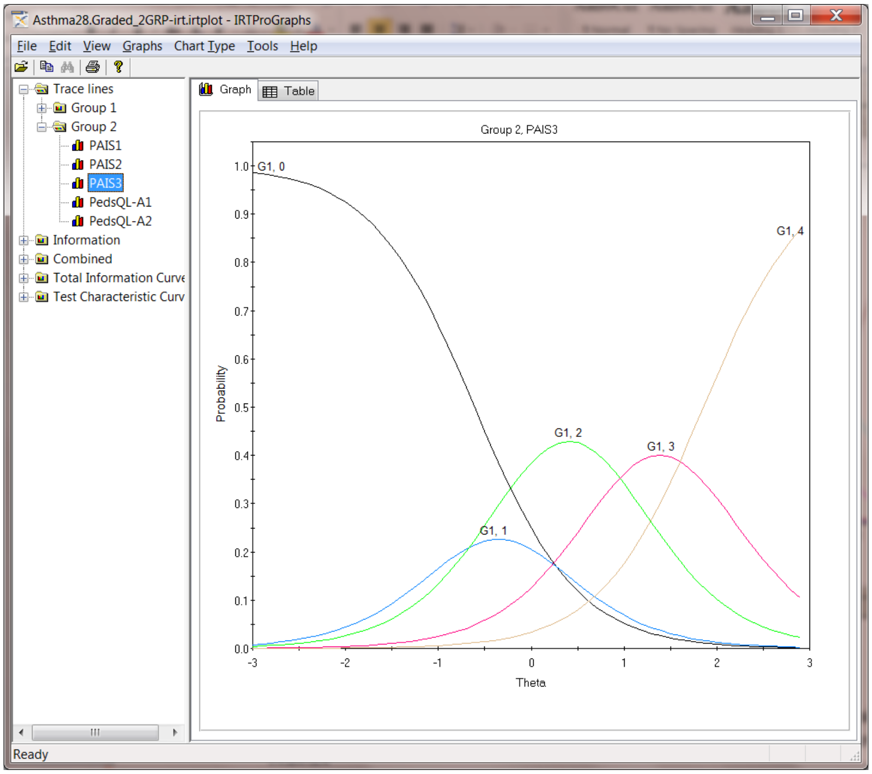
A trace line (item characteristic curve) is a nonlinear function that portrays the regression of the item score on the trait or ability measured in a test.

In the case of binary data, two trace lines are shown, one for the positive response and the other it's complement. In polytomous models such as the graded response model and nominal response models, trace lines for each response category are plotted. Each curve shows the selection probability of a category of the item as a function of θ.

By default, trace lines for all the items selected under the **Models** tab are displayed simultaneously for each group. Trace lines for the first four items of Group number 1 of a two-group analysis are shown below.



To view the trace lines for any specific item, double click on the relevant graphics box, or, if for example the item PAIS3 of the second group is of interest, click on the **Group 2, PAIS3** link using the trace line tree expansion shown below.



9.3 Item Information

Item information is a function of θ; it provides valuable insight about the precision of measurement provided by the item. It is of particular use in test construction, where these curves can be used to ensure the inclusion of different items that maximize the precision of measurement at different levels of θ in the test.

In the case of the 2PL model, for example, the item information function is given by (Hambleton & Swaminathan, 1985, Table 6-1)

$$a_i^2\left\{1+exp\left[-\left(\theta-b_i\right)\right]\right\}^{-1}\left\{1-\left(1+exp\left[-\left(\theta-b_i\right)\right]\right)^{-1}\right\}$$

with the maximum value directly proportional to the square of the item discrimination parameter, *a*. A larger value of *a* is associated with greater information. The maximum information is obtained at b_i.

For the three-parameter model, the information function is (Hambleton & Swaminathan, 1985, Table 6-1)

$$\frac{a_i^2\left\{1-\left(1+exp\left[-\left(\theta-b_i\right)\right]\right)^{-1}\right\}\left\{1+exp\left[-\left(\theta-b_i\right)\right]^{-1}-c_i\right\}^2}{\left\{1-c_i\right\}^2}$$

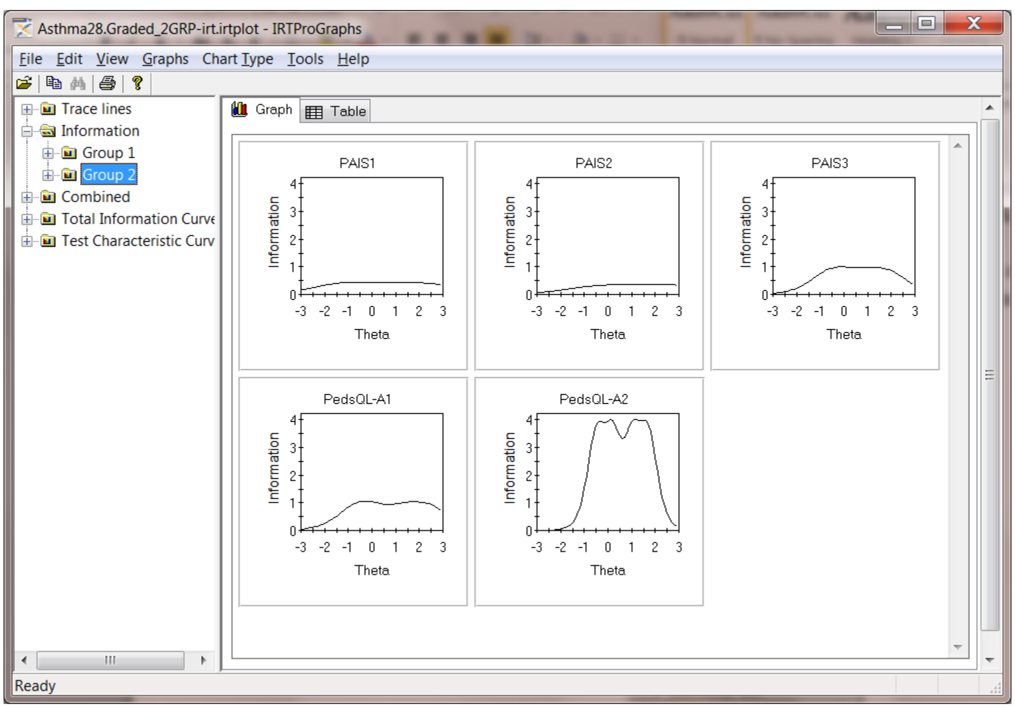
The maximum information is reached at

$$b_i + \frac{1}{a_i} ln\left[1/2 + 1/2\sqrt{1+8c_i}\right]$$

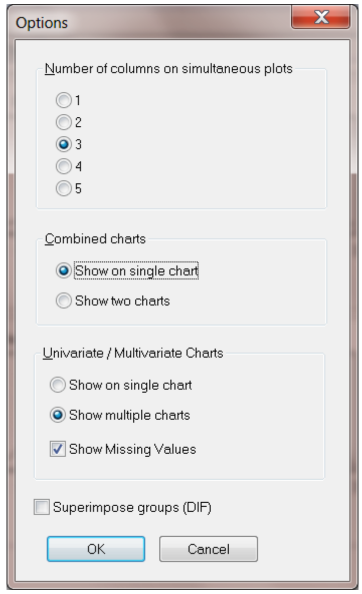
An increase in information is associated with a decrease in c_i. The maximum information is obtained when c_i = 0. Baker & Kim (2004) contains information function equations for most of the IRT models available in IRTPRO.

The slope of the trace line plays an important role in the information provided by an item. An increase in the slope means the item provides more information. The use of items with more information leads to smaller standard errors of measurement. By assessing these curves, items that contribute little information, and therefore contribute little to precision, may be identified and discarded.

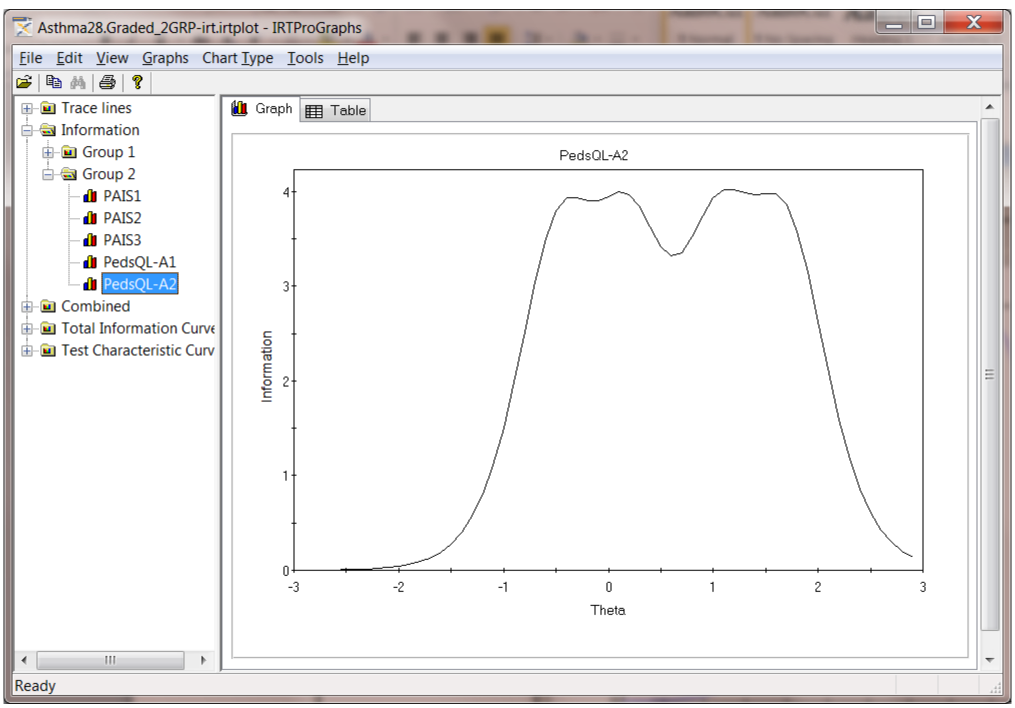
To obtain the item information curves, click on the **Information** link. As before, simultaneous displays of all the item information curves are obtained. By scrolling up or down, all items can be viewed if there are many items in a test.



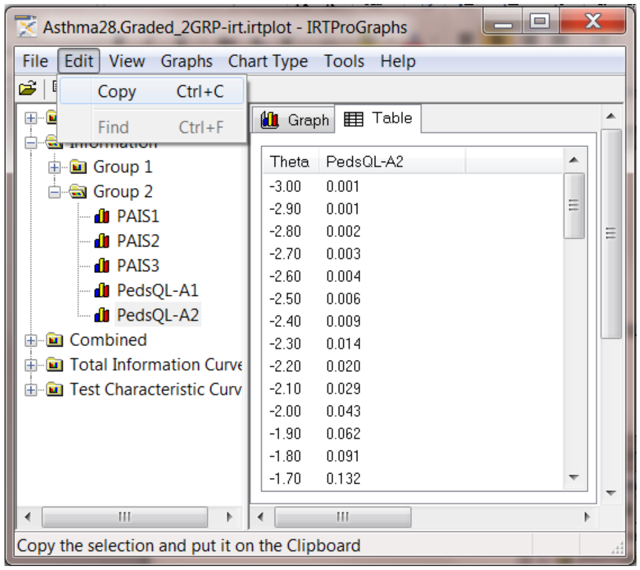
One can also change the number of columns of plots from three (see above) to less or more columns. This is accomplished by clicking **Tools** on the main graphics menu bar to select the **Options** dialog shown below.



Just as in the case of trace lines one can view one information curve at a time by expanding the Information tree and by clicking on the item to be displayed. This is illustrated below for PedsQL-A2 (item 5 of the second group).

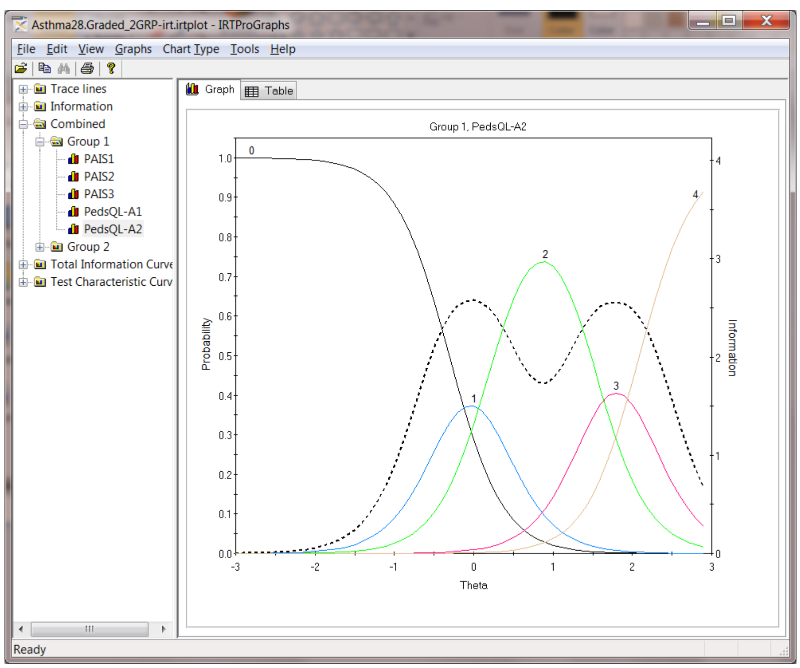


To view the co-ordinates used to plot a graph, the **Table** icon next to the **Graph** icon in the graphics window can be selected. Use **Edit**, **Copy** (**Ctrl**+**C**) to copy the contents of the table to the clipboard if you need to paste these values into another document.

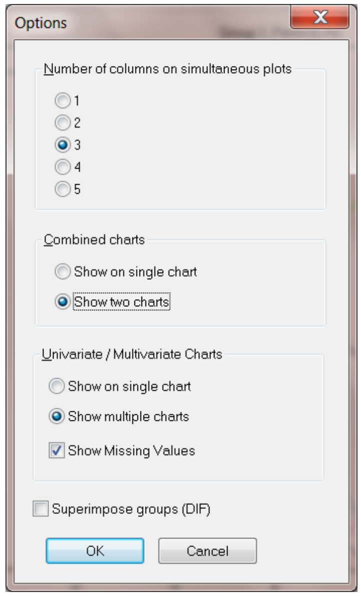


9.4 Combined trace lines-information curves

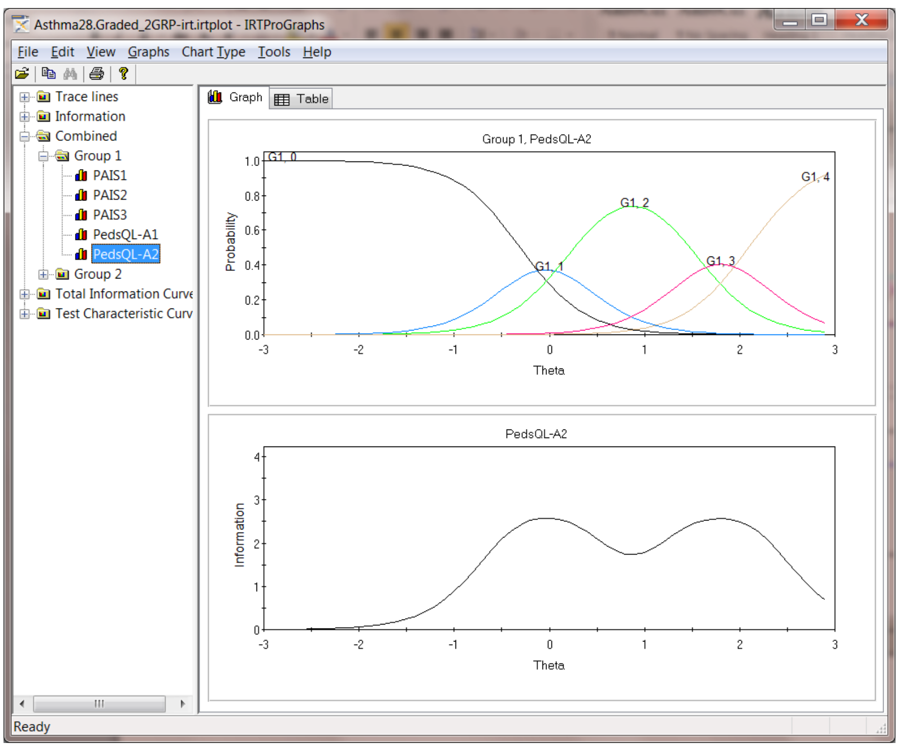
The user additionally has the option to obtain a combined trace lines-information curve presentation. This is illustrated below for Group1, item 5 (PedsQL-A2). Just as in the previous cases, the default selection is the simultaneous display of all the items that are obtained by clicking on **Combined**. Note that the information curve is displayed as a dashed line and that the information scale is shown on the right side of the graph.



For reporting purposes, a researcher may prefer to display the trace lines and information curves as two separate vertically stacked plots. This is achieved by selecting the **Tools, Option** dialog and by making the appropriate choice (**Show two charts)** in the **Combined charts** pane.



The **Show two charts** choice results in a trace lines graph stacked above the information graph for the item selected.



9.5 Total Information

The total information (or test information) function summarizes the information function for a set of items or test. The contribution of each item in the test to the total information is additive.

The slope of the trace line plays an important role in the information provided by an item. An increase in the slope means the item provides more information. The use of items with more information leads to smaller standard errors of measurement. By assessing these curves, items that contribute little information, and therefore contribute little to precision, may be identified and discarded.

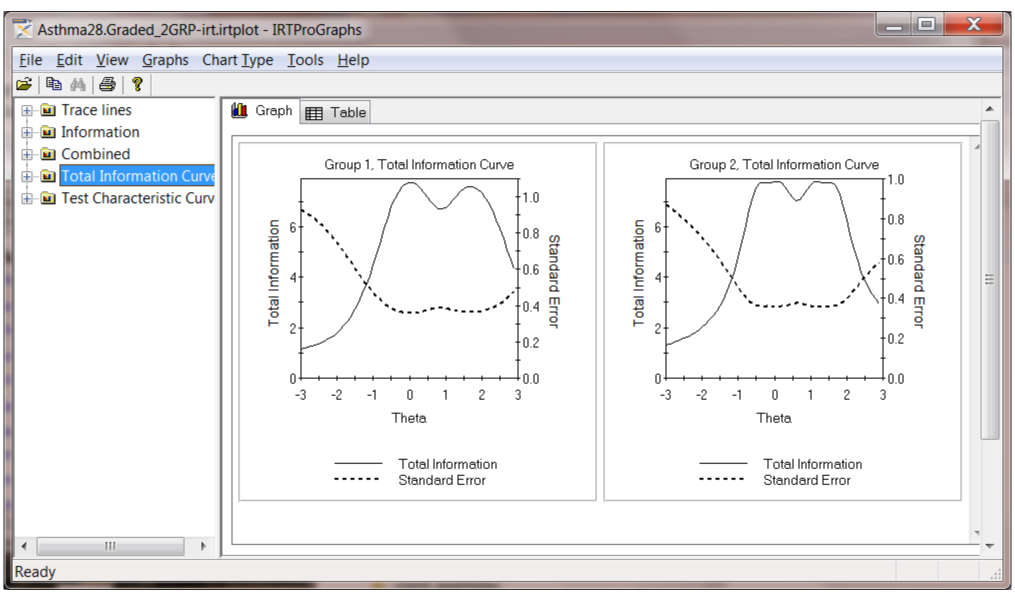
The measurement error variance of the MAP scale score is (on average) inversely related to the amount of information provided by a set of test items at any level of θ, so on average, the standard error of MAP estimates at ability level θ can be written as

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}.$$

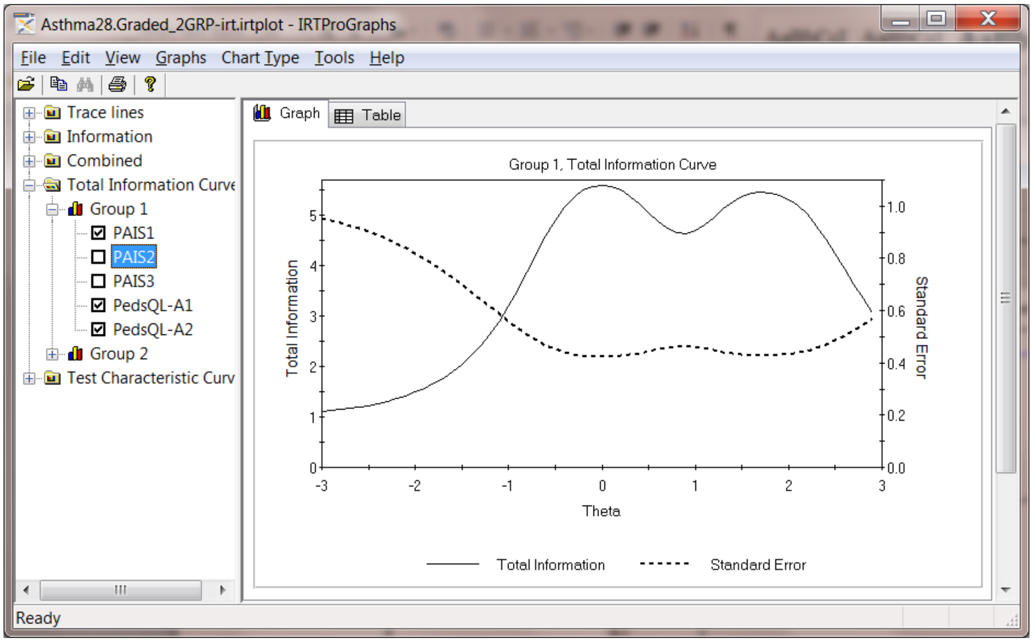
The contribution of both item and test information curves are summarized by Hambleton & Swaminathan (1985) as follows:

"The item and test information functions provide viable alternatives to the classical concepts of reliability and standard error. The information functions are defined independently of any specific group of examinees and represent the standard error of measurement at any chose ability level. Thus, the precision of measurement can be determined at any level of ability that is of interest. Furthermore, through the information function, the test constructor can precisely assess the contribution of each item to the precision of the total test and hence choose items in a manner that is not contradictory with other aspects of test construction."

The graph shown below is obtained by clicking on the **Total Information Curve** link. The solid line shows the total information curve at various value of theta. The corresponding standard errors are presented by the dashed line. The standard error scale is given on the right hand vertical axis. The curves below are based on all the items that are included in the model.



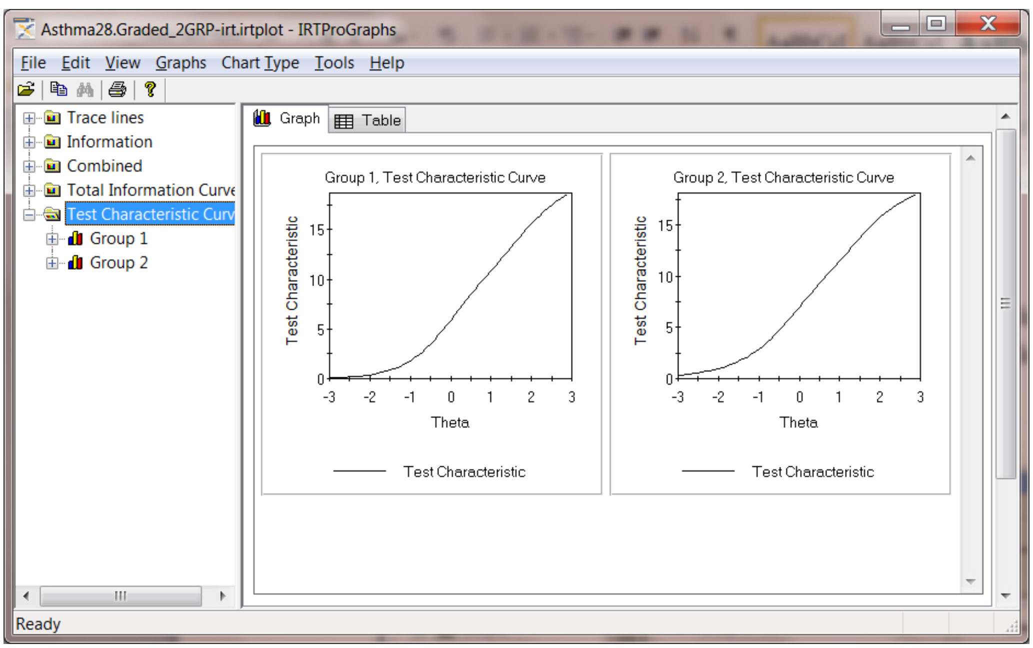
The next graphical presentation shows the total information and standard errors when Items 2 and 3 (PAIS2 and PAIS3) are omitted.



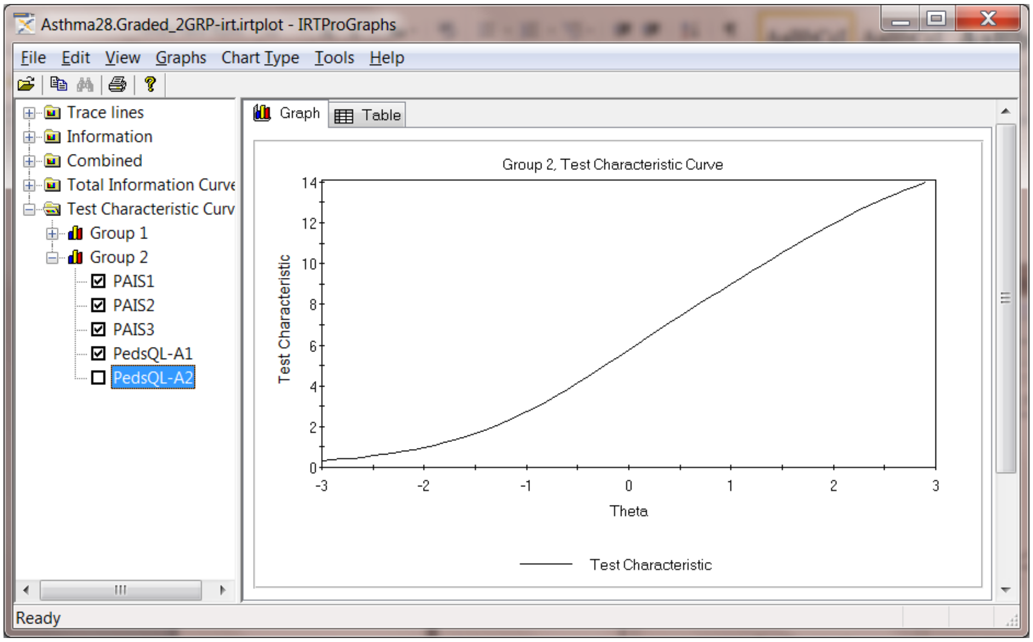
9.6 Test Characteristic Curves

The test characteristic curve is the expected value of the summed score for a test, or a set of items, as a function of θ.

In the image below, the test characteristic curves for Groups 1 and 2 are displayed.



The next graphical presentation shows the test characteristic curve for group 2 when the item PedsQL-A2 is omitted.

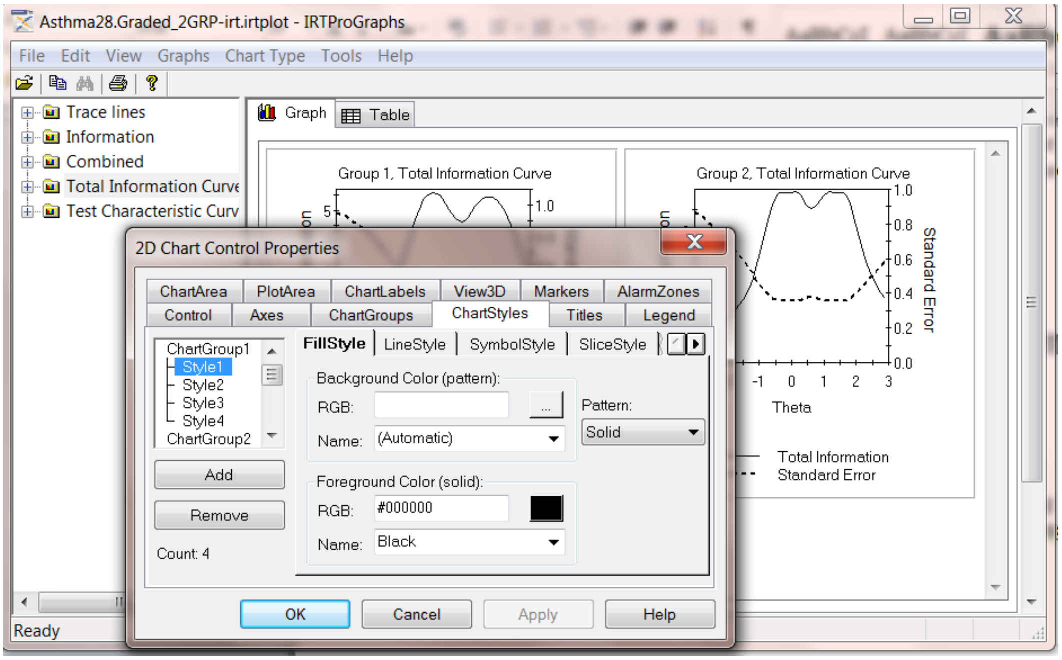


Polytomous items each has expected score curves, which show the expected item score as a function of θ, computed by taking literally the numerically values of the item scores $u = 0, 1, \ldots, m-1$ where m is the number of response categories. Expected score curves are simpler than trace line plots for, say, five response categories, because the latter have five curves and the expected score is only curve. While it is difficult to compare two items' sets of five trace lines, it is easier to compare two items' expected score curves.

In **IRTPROGraphs**, expected score curves for individual items can be graphed by selecting *only one item* with the check boxes in the list at the left side of the graphics window.

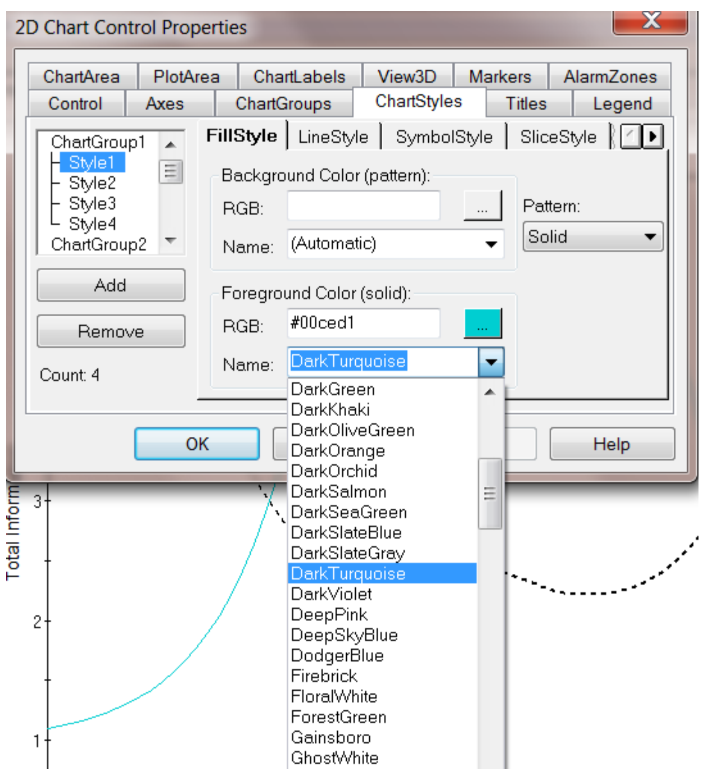
9.7 Controlling the appearance of a graph

The visual appearance of a graphical display can be changed before the graph is copied to another document. This is accomplished by right clicking in the plot area of the display that is to be modified. This action activates the **2D Chart Control Properties** dialog that provides the user with several options (options available depends on the graph type). In the illustration below, we wish to change the colors of the total information and standard error curves of the first group.

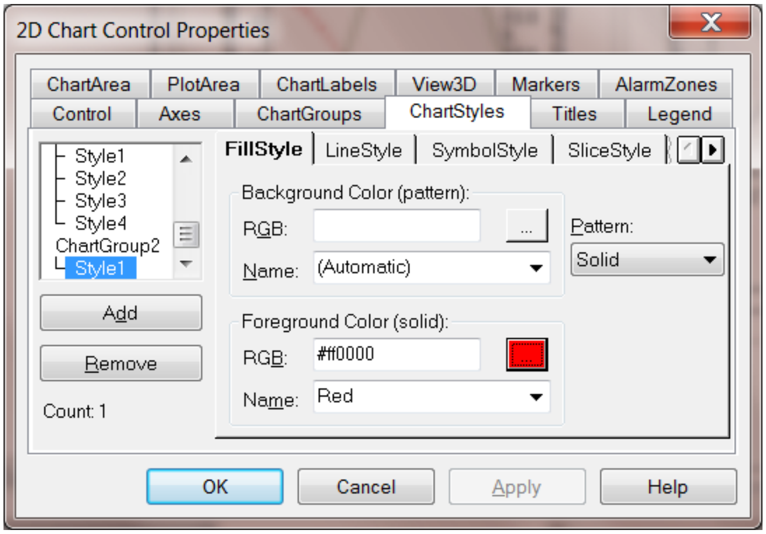


This is accomplished by selecting the **ChartStyles** tab and then by clicking on the **Name:** (of the color) arrow.

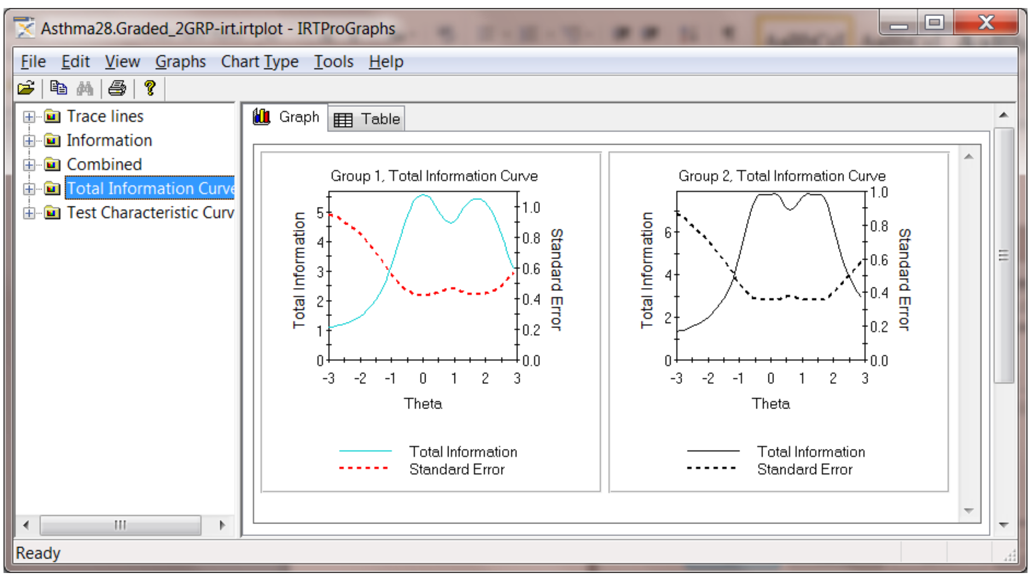
This action results in the display of a drop-down menu with a list of available colors. In the image shown below, the color Dark Turquoise was selected.



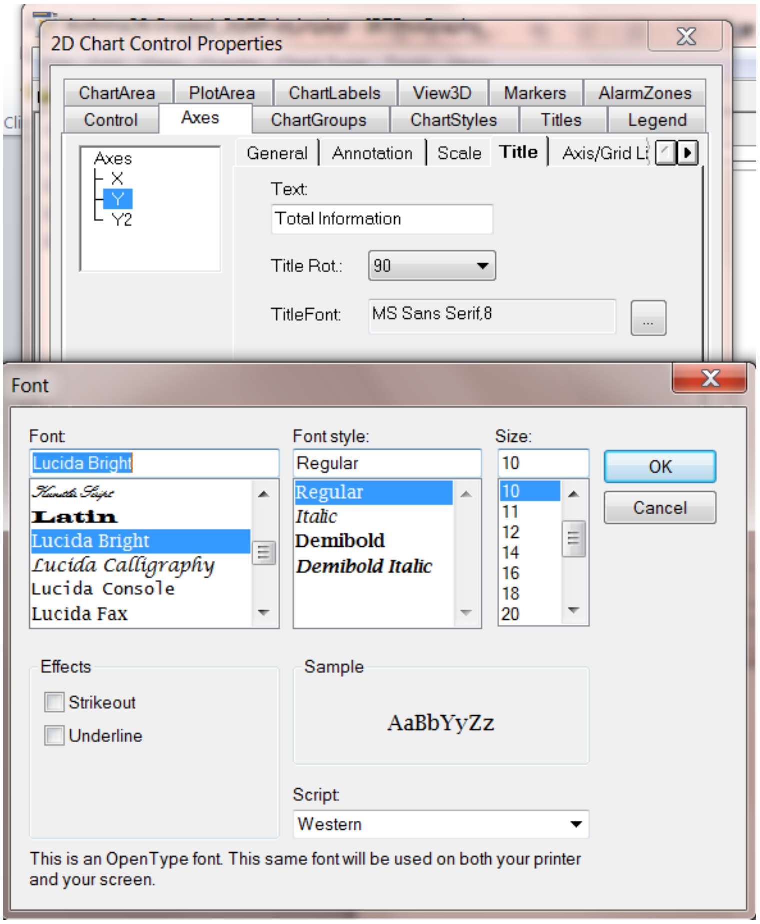
To change the color of the standard error curve from Black to Red, select **ChartGroup2, Style 1** and repeat the procedure described above.



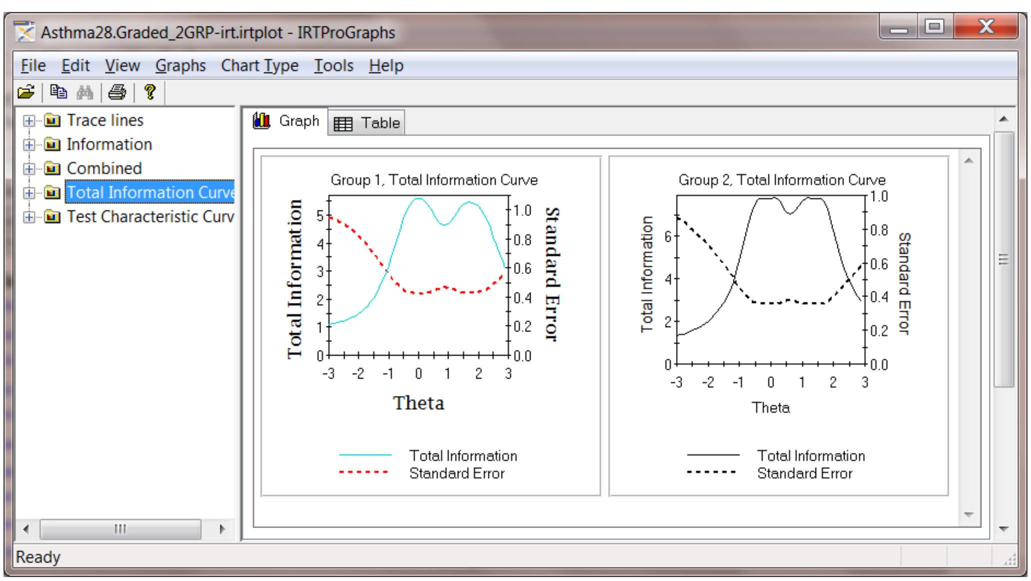
Click **OK** when done to obtain the revised graphical display shown next.



To change the format of the axes labels, select the **Axes** tab and click on the **TitleFont** selector button to display the **Font** dialog. To illustrate, we selected the Lucida Bright font, size 10.



Click **OK** when done to obtain the revised graphical display shown below. Chapter 10 contains additional examples illustrating the functionality of the **2D Chart Control Properties** dialog.

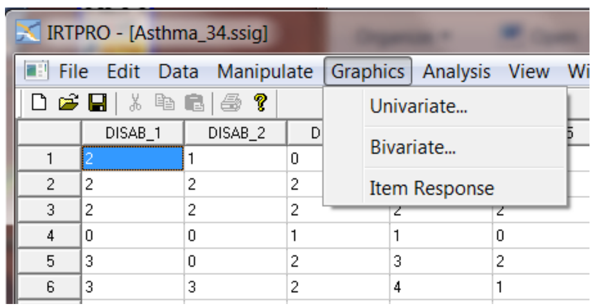


10. Data-based graphics

10.1 Introduction

Graphics are often useful for data exploration. Relationships and trends may be conveyed in an informal and simplified visual form via graphical displays. IRTPRO offers both data-based and model-based graphs (see Chapter 9). In the case of data-based graphs, IRTPRO distinguishes between univariate and bivariate graphs. Univariate graphs (see Section 10.2) 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 that purpose, bivariate graphs (see Section 10.3) are more appropriate.

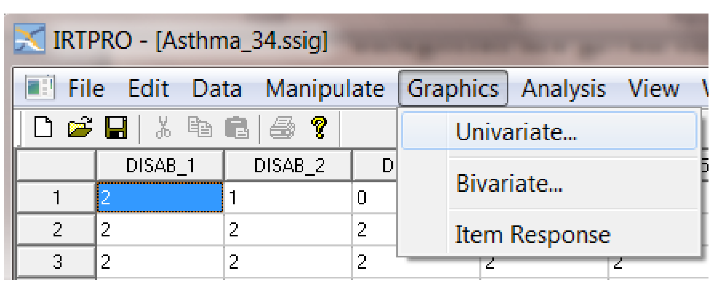
To make univariate or bivariate graphs, the IRTPRO dataset of interest must be the currently opened window. Click the **Graphics** button on the main menu-bar and make a selection between the **Univariate…** and **Bivariate…** options.



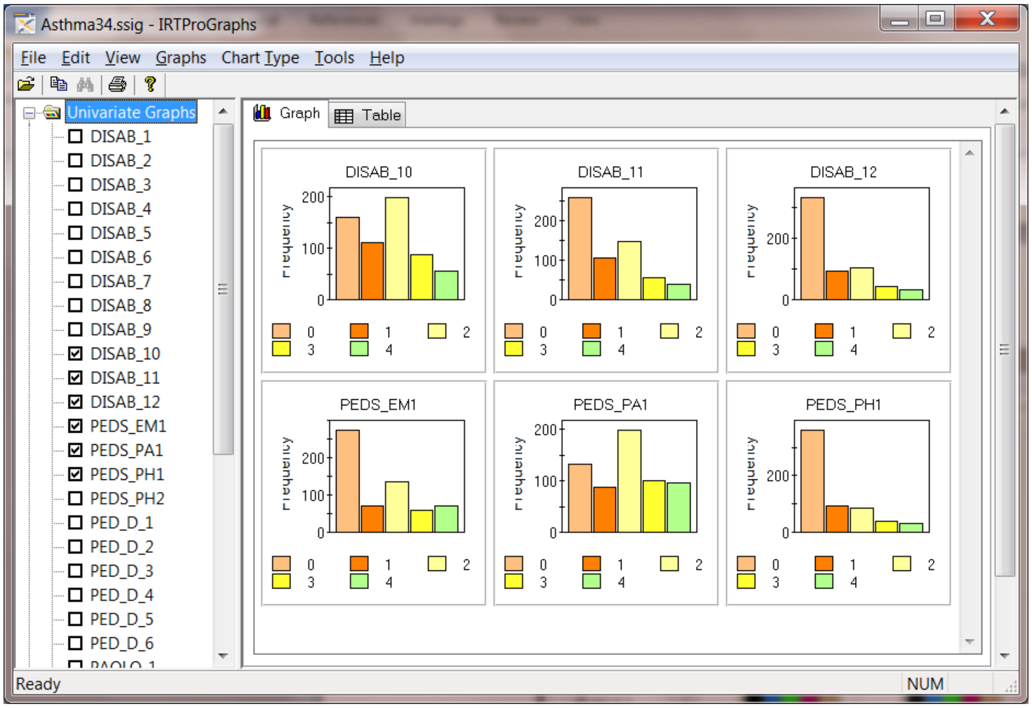
10.2 Univariate Graphs

The default graph-type is a bar chart for each item selected. A bar chart is a graphic representation of the frequency distribution of discrete or categorical data in which the values or categories are given on the horizontal axis and the frequencies are given on the vertical axis.

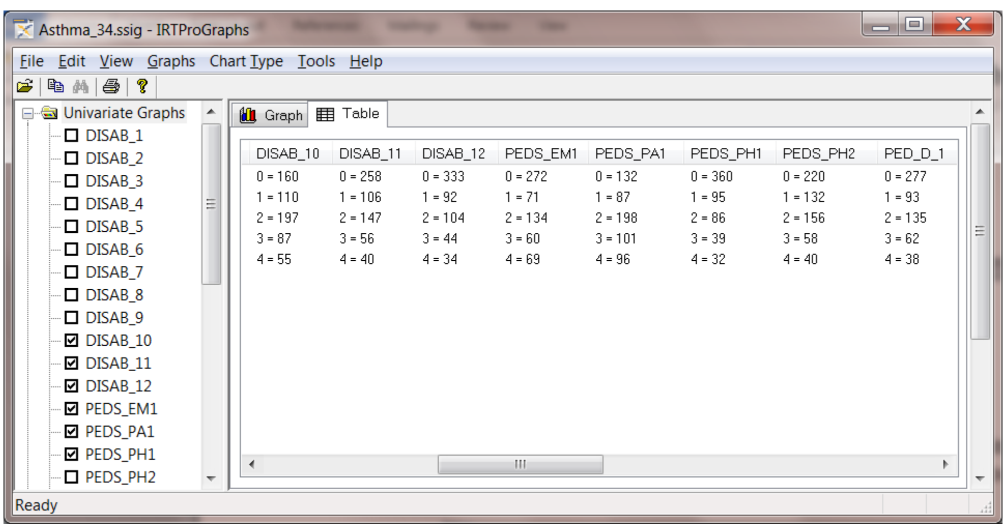
The image below shows the selection of the **Graphics, Univariate…** option.



Selection of this option opens a **Univariate Graph** dialog that enables one to select one or more of the variables in the data set. By clicking on the **OK** button, a simultaneous display of bar charts is obtained. The default display is to show the bar charts after removal of missing values.

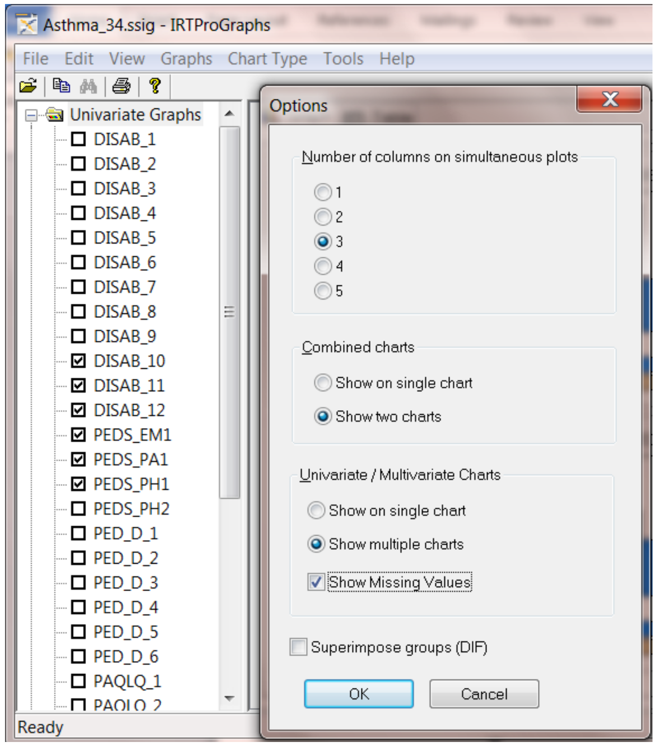


The data used to create the graphs can be viewed by selecting the **Table** "mode" as shown below. For example, for the DISAB_10 item, there are 160 values equal to 0, 110 values equal to 1, 197 values equal to 2, 87 values equal to 3, and 55 values equal to 4.

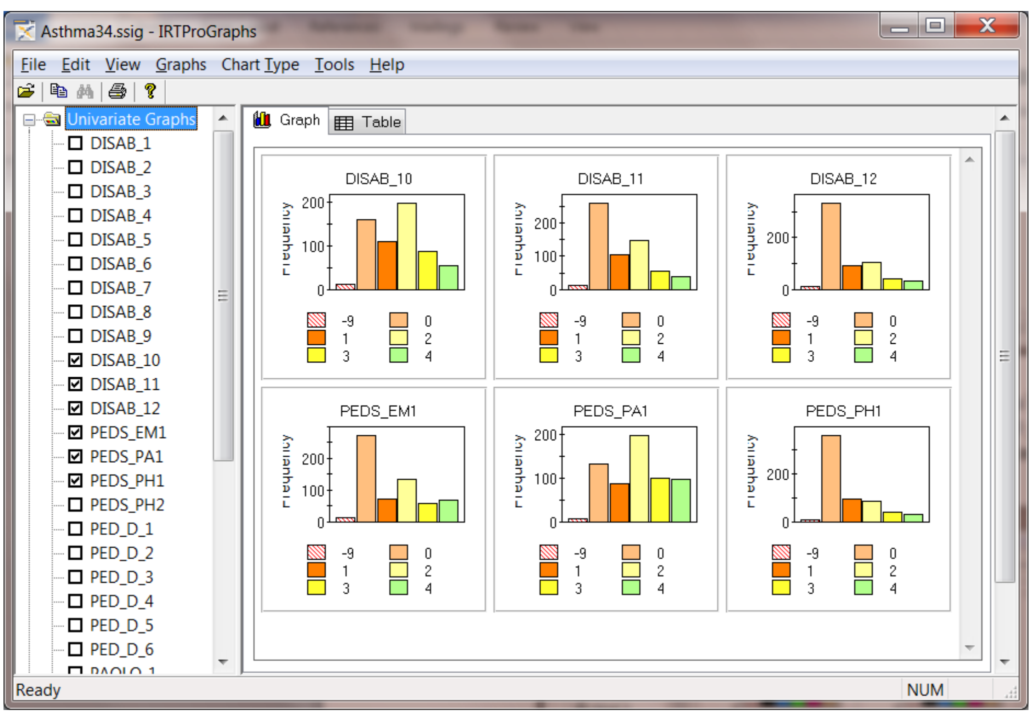


DISAB_10	DISAB_11	DISAB_12	PEDS_EM1	PEDS_PA1	PEDS_PH1	PEDS_PH2	PED_D_1
0 = 160	0 = 258	0 = 333	0 = 272	0 = 132	0 = 360	0 = 220	0 = 277
1 = 110	1 = 106	1 = 92	1 = 71	1 = 87	1 = 95	1 = 132	1 = 93
2 = 197	2 = 147	2 = 104	2 = 134	2 = 198	2 = 86	2 = 156	2 = 135
3 = 87	3 = 56	3 = 44	3 = 60	3 = 101	3 = 39	3 = 58	3 = 62
4 = 55	4 = 40	4 = 34	4 = 69	4 = 96	4 = 32	4 = 40	4 = 38

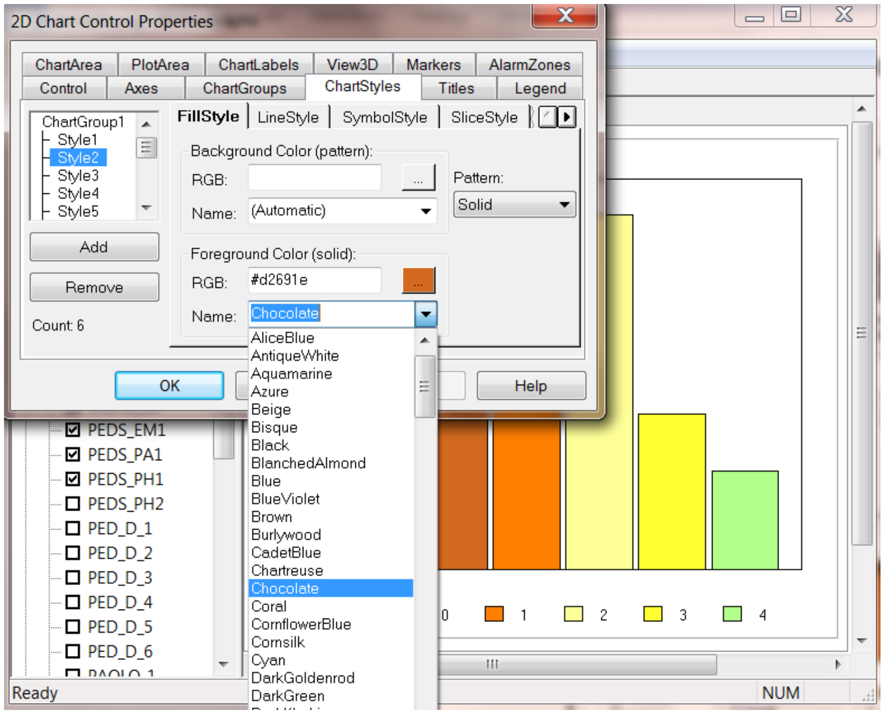
To display the missing values, click on the **Tools** button and make sure that the **Show Missing Values** check box is selected. The **Options** dialog also enables the user to select the number of columns on the simultaneous plots.



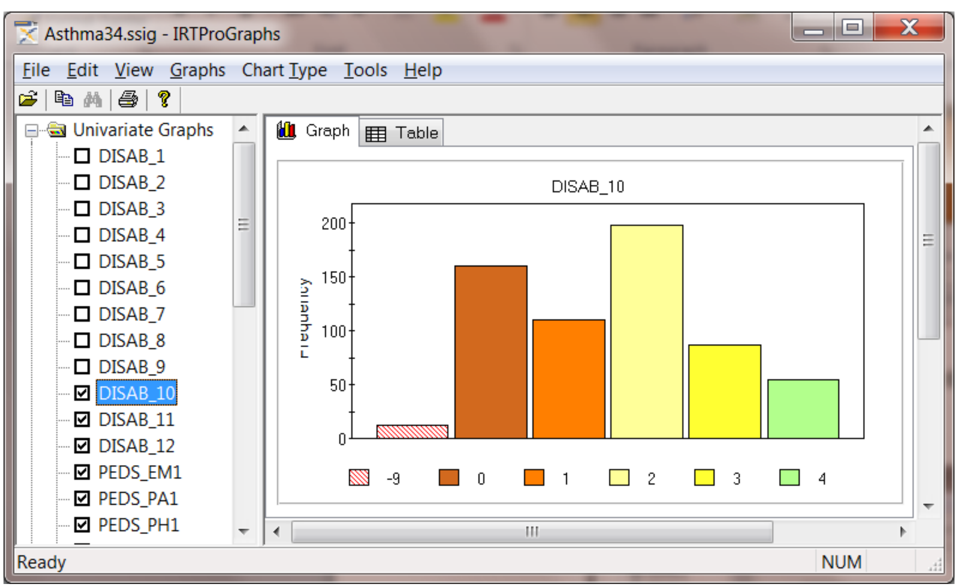
Below we show the bar charts for the six items selected. As can be seen, all the items selected have missing values (coded –9). By right clicking in the DISAB_10 plot area (for example), the **Chart Properties** dialog is obtained that can be used to change the color of each bar.



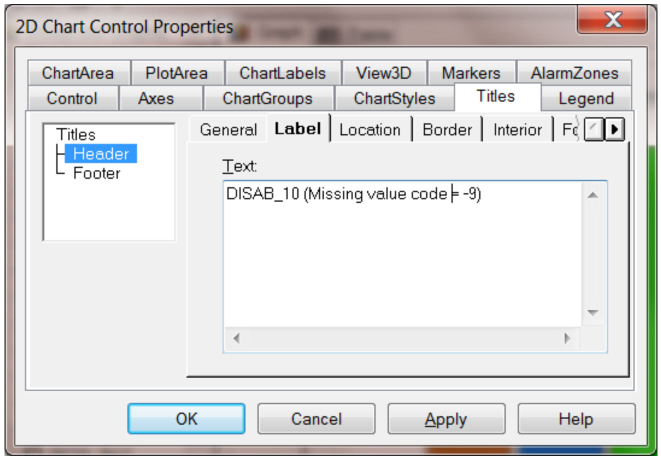
In the illustration below, use is made of the **ChartStyles** tab to change the color of the bar corresponding to the value "0" to chocolate.



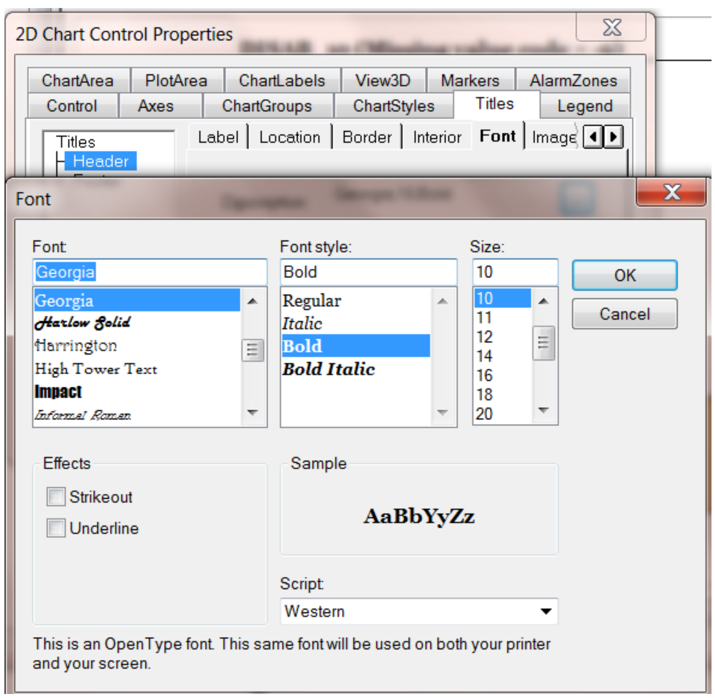
After making the desired color changes, click the **OK** button to view the modified display shown below.



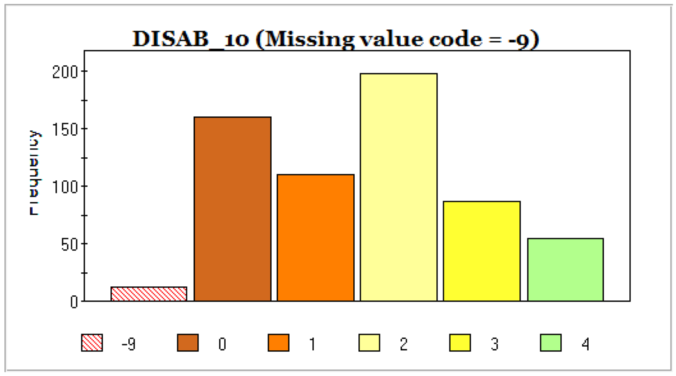
As an illustration, suppose that we want to change the text and font of the current title (DISAB_10). Right-click in the DISAB_10 plot area to display the **Chart Properties** dialog and select the **Titles** tab. Use the **Titles** window to first select the **Label** tab and then the **Font** tab. Change the **Header**, **Text** as shown below.



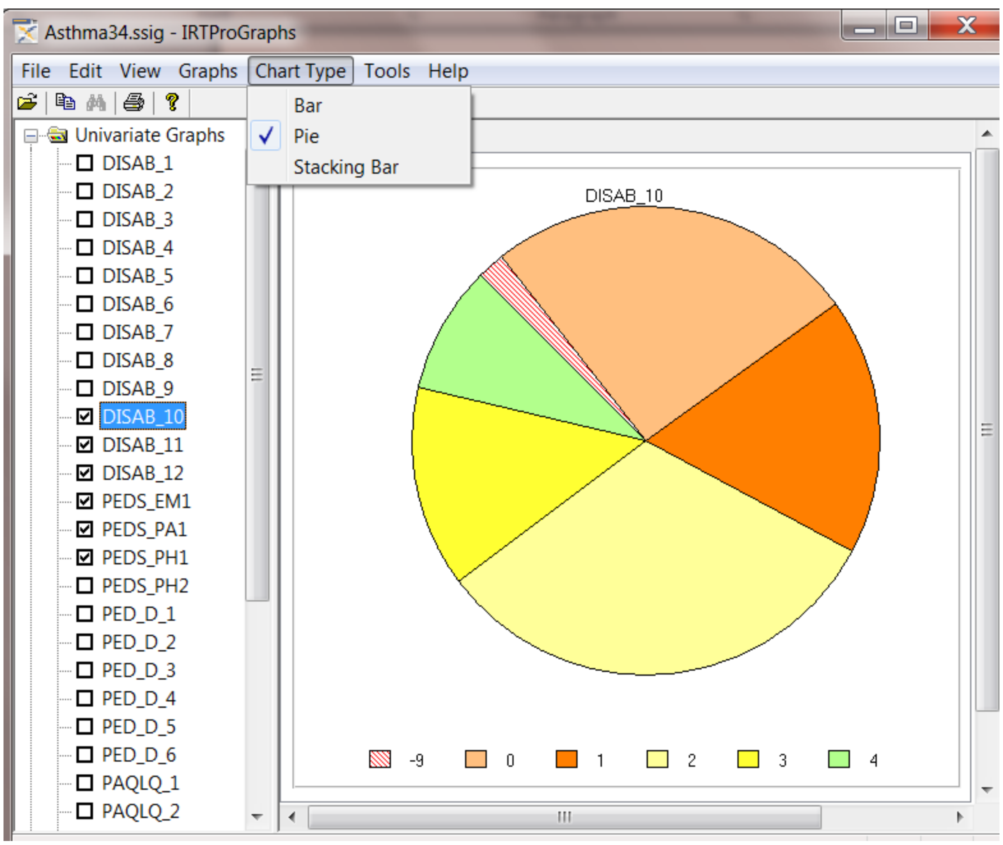
Next, use the **Font** tab to change to default font, font style and font size. In this case, it was changed to Georgia, Bold, 10.



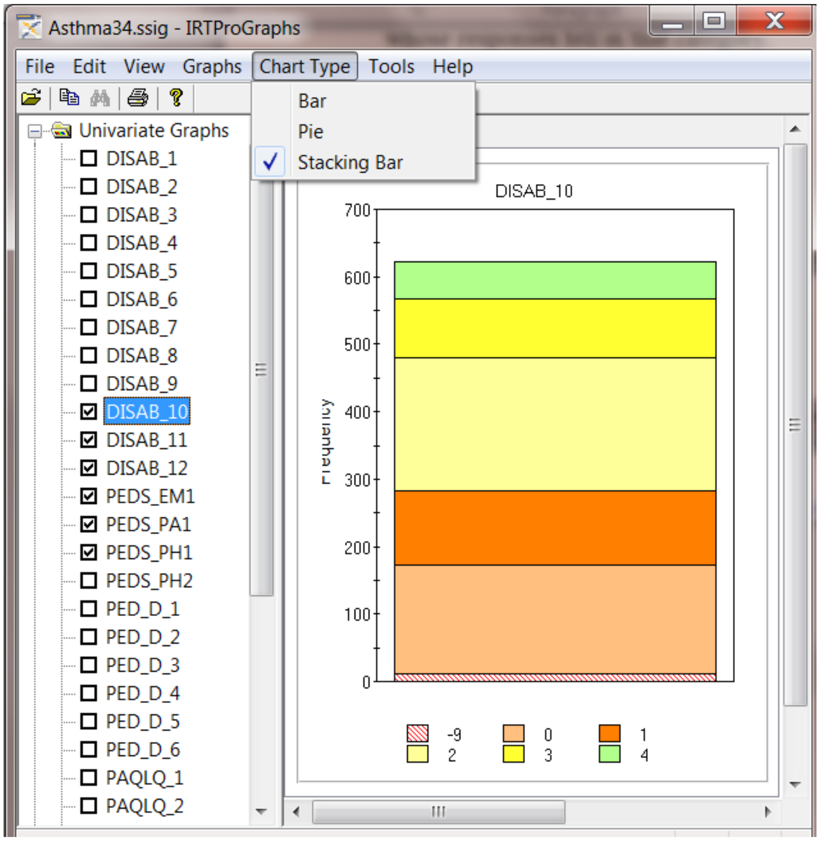
Click the **OK** button of the **Font** dialog to return to the **2D Chart Control Properties** dialog, then click **OK** to view the edited graph.



A pie chart display of the percentage distribution of a variable may be obtained by selecting the **Chart Type, Pie** option. A pie chart is a graphic representation of percentages or frequencies by means of a circle that is subdivided into slices in such a way that the areas of these slices are proportional to the percentages or frequencies. Pie charts may be customized by using the graph editing dialog boxes obtained by right-clicking in the plot area of the pie chart.

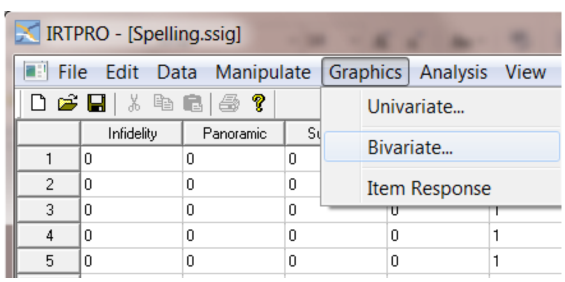


The distribution of frequencies over the categories of an item can also be displayed in the form of a stacked bar chart by selecting the **Chart Type, Stacking Bar** option. The bars in a stacked bar graph are divided into the categories of the item displayed. Each bar represents the number of examinees whose responses fell in that category.

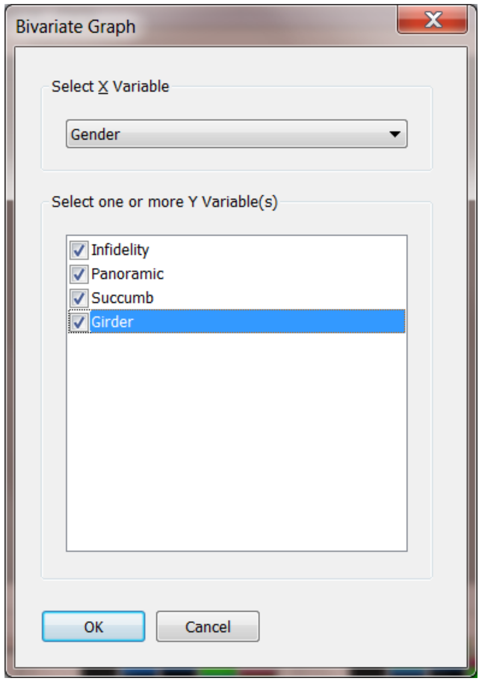


10.3 Bivariate Graphs

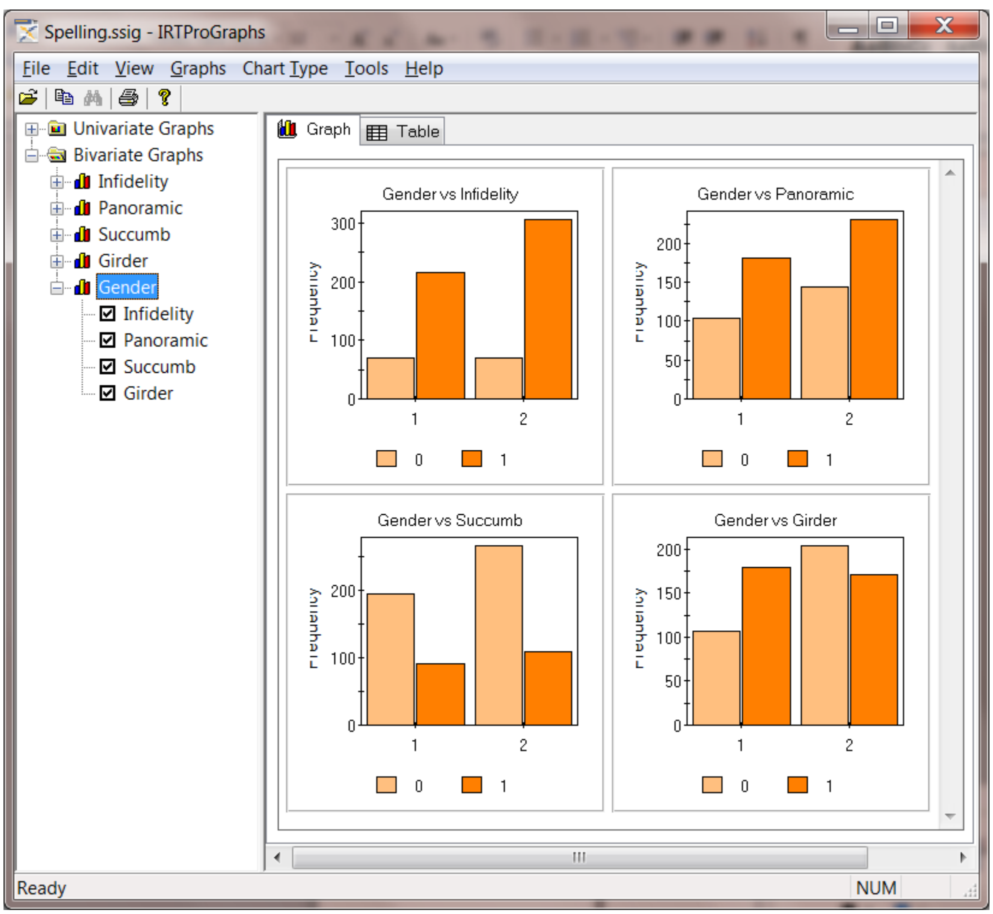
The **Graphics, Bivariate…** option allows us to graphically display a two-way frequency table.



Selection of this option results in the display of a **Bivariate Graph** dialog allowing the user to select an **X-variable** and one or more **Y-variables** to obtain a set of bivariate plots. In the following example, Gender is selected as the **X-variable** and the items Infidelity, Panoramic, Succumb and Girder as the **Y-variables.**

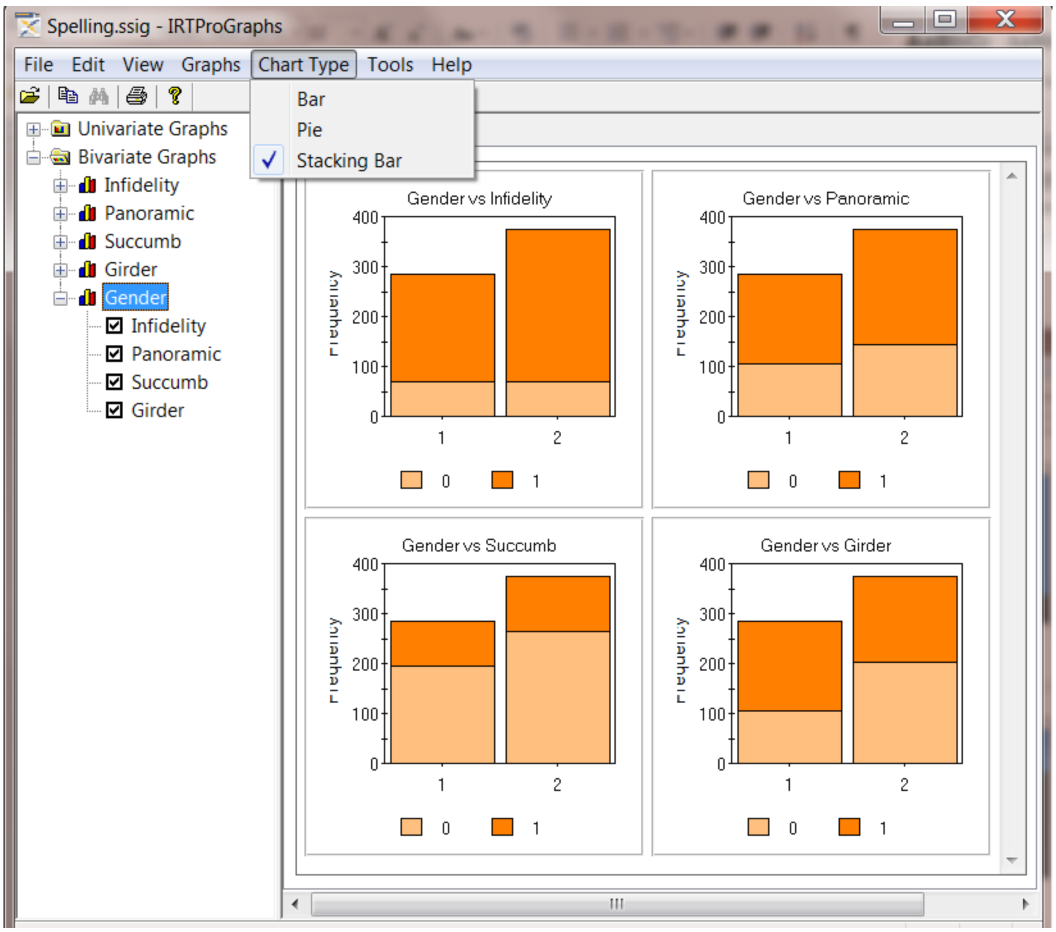


Clicking the **OK** button results in the following graphical display. Note that the categories of Gender (the **X-variable**) are displayed below the horizontal axis. Each category of a selected **Y-variable** corresponds to a color and the color legend is given at the bottom of the graph.

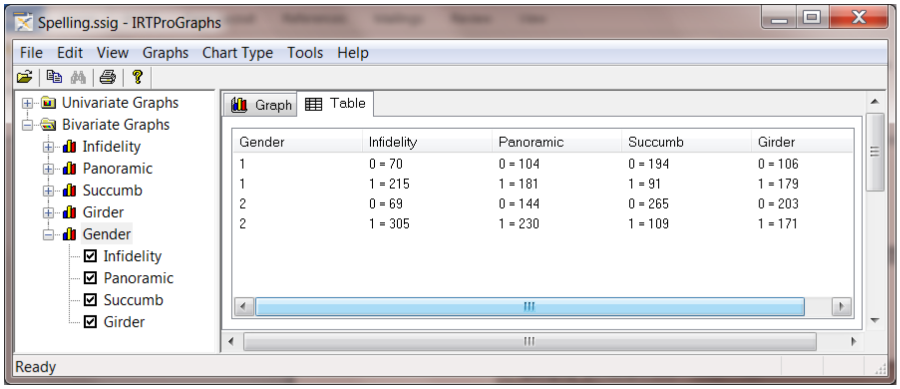


Note that in the left pane (above) all the items included in the model are listed. If a specific item is clicked, an expanded list of all the items (excluding the one selected) is displayed and any of these items may be selected as **Y-variables**.

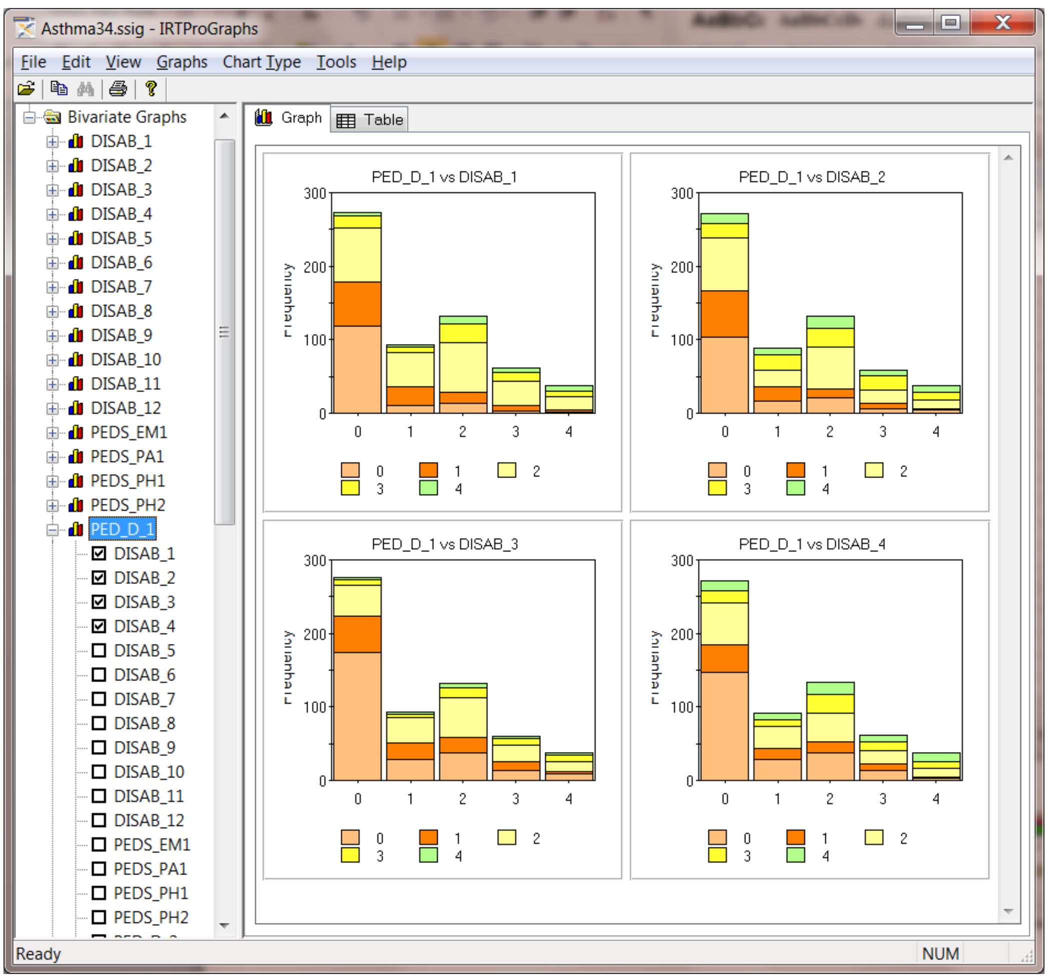
A more informative display of the relationship between two variables might be stacked bar-charts, obtained by selecting the **Chart Type, Stacking Bar** option. In the display below, it can be observed that there are more Gender = 2 subjects compared to the number off Gender = 1 subjects. Furthermore, for the item Girder a larger percentage Gender = 2 individuals chose the "0" category than is the case for Gender = 1.



As mentioned earlier, one can switch to the **Table** mode to view the data generated for the plots requested. The frequencies listed in the table below, substantiate the conclusions drawn from the stacked bar-charts display.

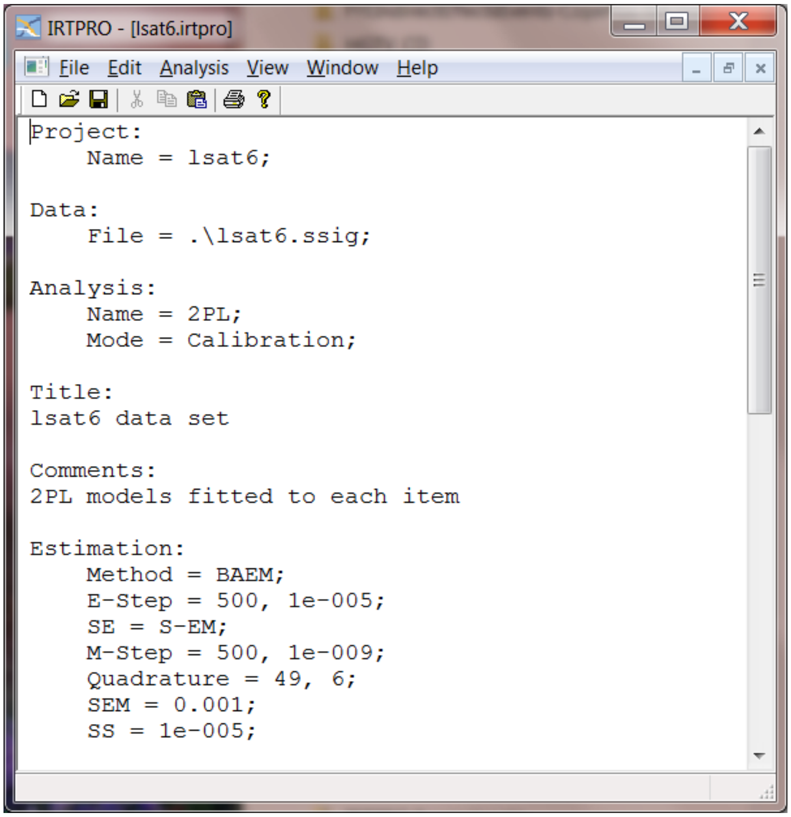


In the next illustration, bivariate charts are requested for the item PED_D_1 versus the items DISAB_1, DISAB_2, DISAB_3 and DISAB_4, from the IRTPRO dataset **Asthma34.ssig**. This dataset were selected since each item has more than two categories. In this case, the stacked bars representation is less cluttered than the side-by-side bar charts representation and usually easier to interpret visually.

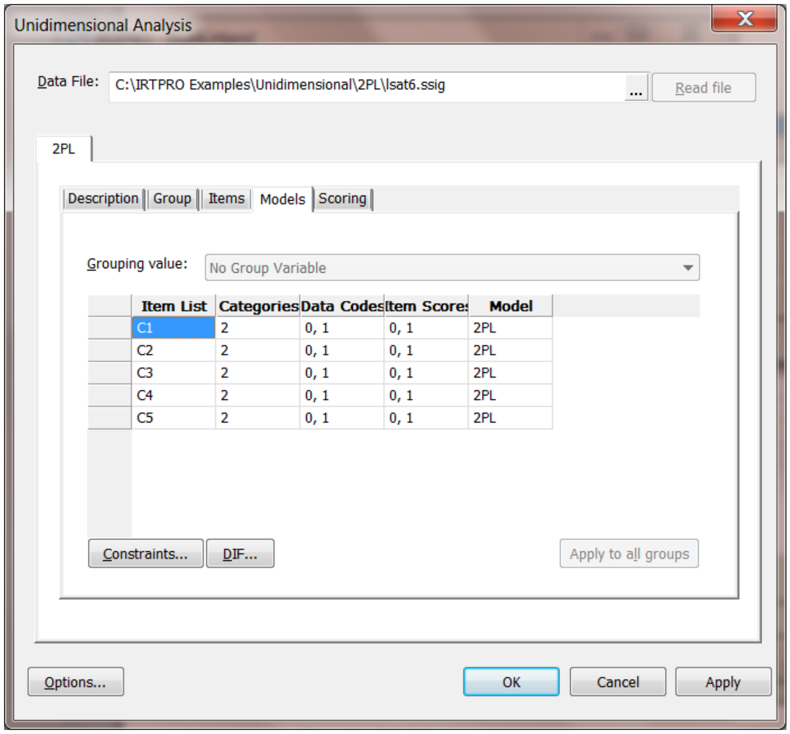


11. Estimation methods and settings

A brief discussion of the IRTPRO's estimation methods and their control parameters is provided in this chapter. To see the dialog boxes that permit access to the control parameters, open the syntax file (see Chapter 12) **lsat6.irtpro** from the **IRTPRO Examples\Unidimensional\2PL** folder, a portion of which is displayed below.

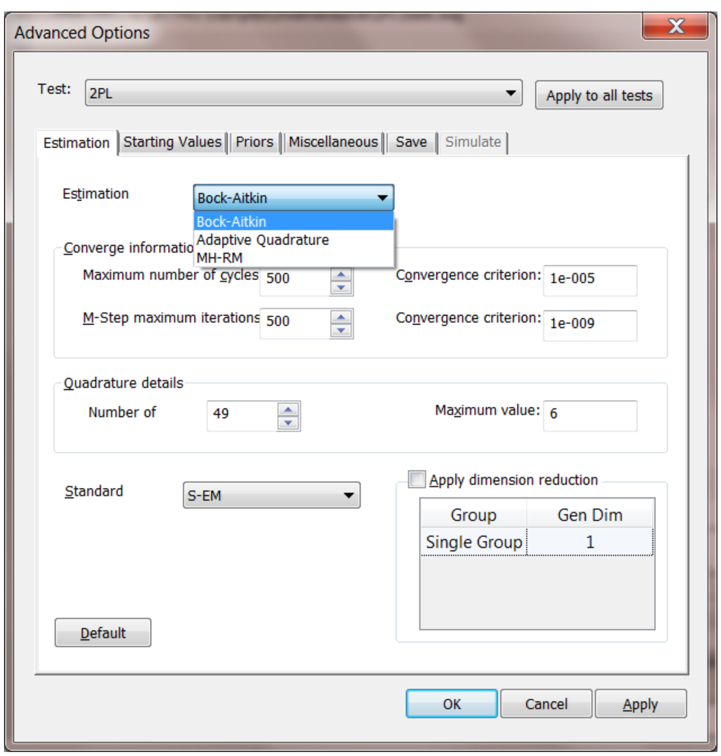


Next, select the **Analysis, Unidimensional** option from the main menu bar and click the **Options** button shown at the bottom of the **Unidimensional Analysis** window on the left.

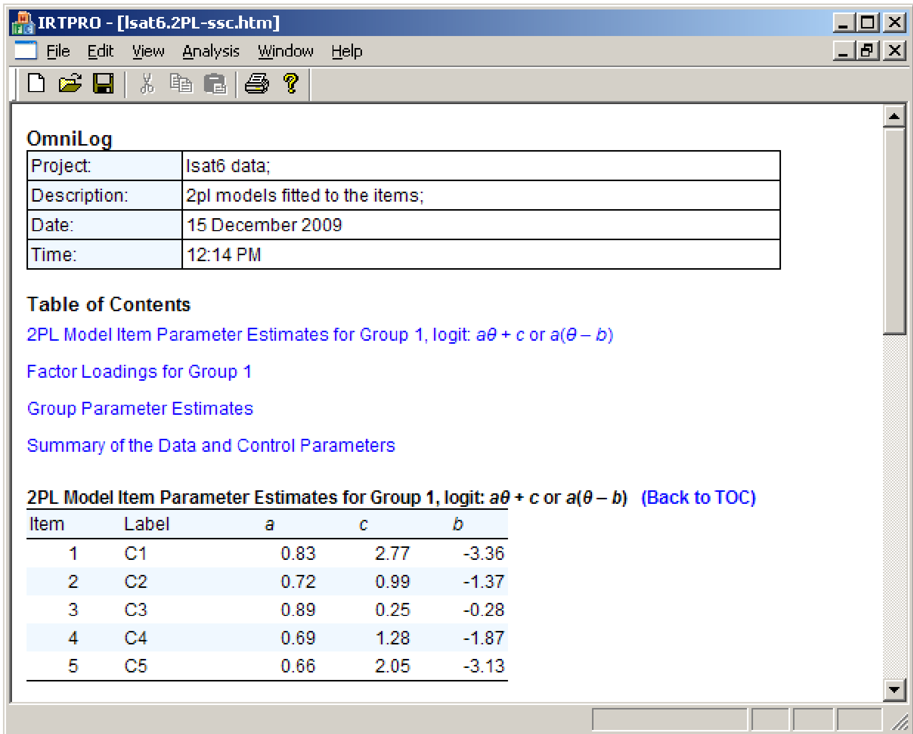


11.1 Bock-Aitkin EM

The default estimation method for a unidimensional analysis is the Bock-Aitkin method and the **Advanced Options** window shown below shows the default estimation settings. A researcher has the option to change these settings, for example, the number of quadrature points, the range over which these points are spread, the maximum number of cycles (E-step) and the maximum number of iterations (M-step).



A portion of the output, listing the parameter estimates for the Bock-Aitkin estimation method, is shown below.


IRTPRO - [lsat6.2PL-ssc.htm]

File Edit View Analysis Window Help

OmniLog

Project:	lsat6 data;
Description:	2pl models fitted to the items;
Date:	15 December 2009
Time:	12:14 PM

Table of Contents

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

Factor Loadings for Group 1

Group Parameter Estimates

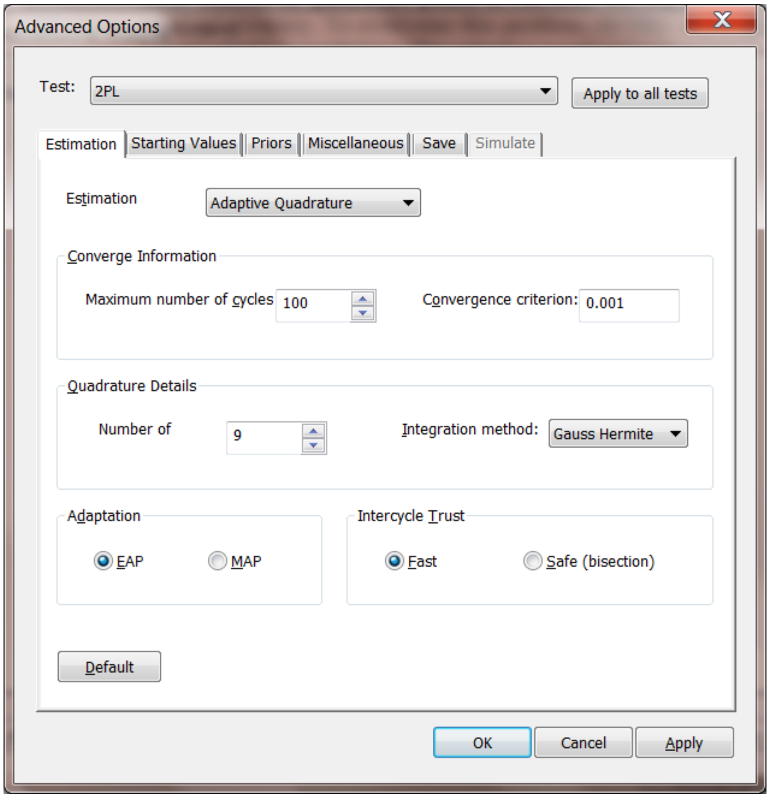
Summary of the Data and Control Parameters

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

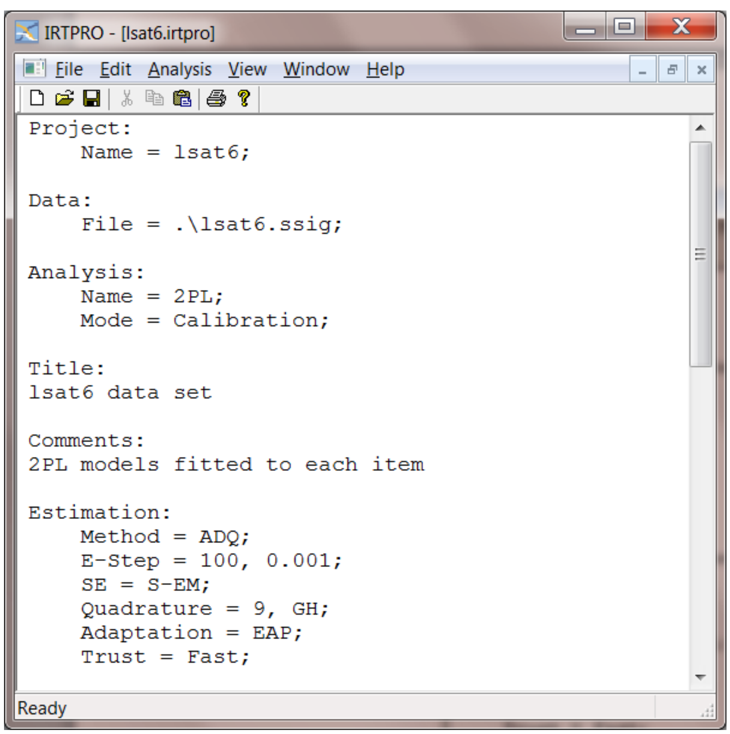
Item	Label	*a*	*c*	*b*
1	C1	0.83	2.77	-3.36
2	C2	0.72	0.99	-1.37
3	C3	0.89	0.25	-0.28
4	C4	0.69	1.28	-1.87
5	C5	0.66	2.05	-3.13

11.2 Adaptive Quadrature

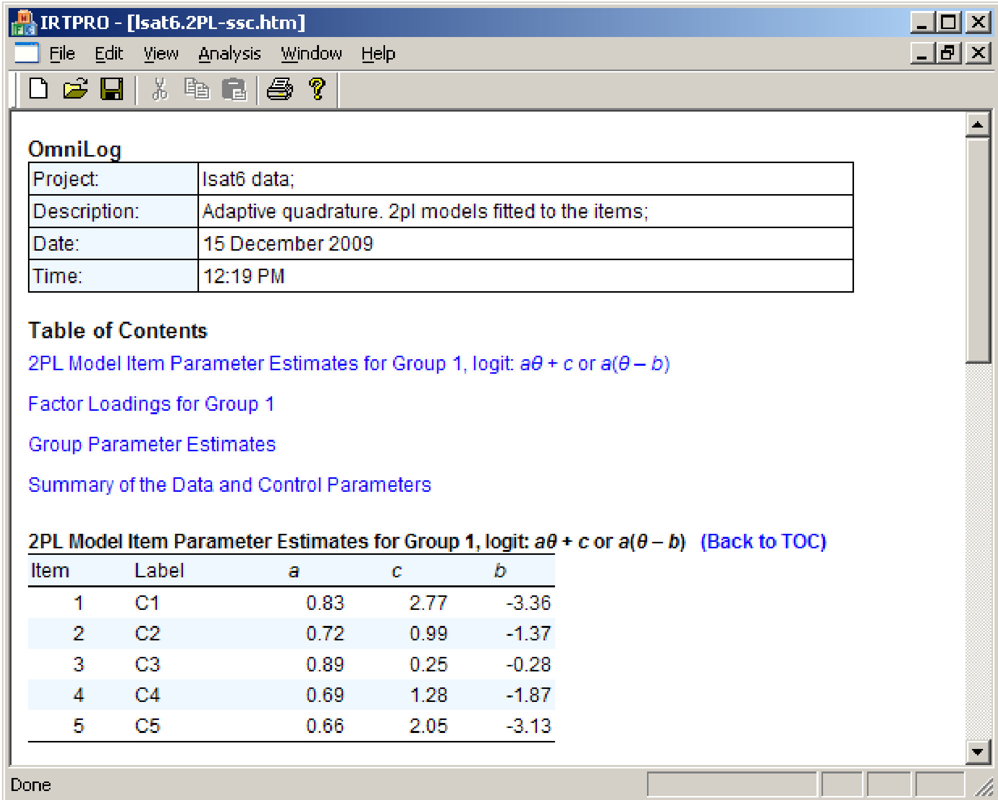
A problem with standard numerical quadrature as employed in the Bock-Aitkin procedure is that it has a fixed set of quadrature nodes for the posterior distribution of all persons. This often requires the use of a large number of quadrature points to calculate the log-likelihood and derivatives to an acceptable level of accuracy. To overcome this problem, IRTPRO also offers a numeric integration procedure called adaptive quadrature. The adaptive quadrature procedure uses the empirical Bayes means and covariances, updated at each iteration to essentially shift and scale the quadrature locations of each case (person) in order to place them under the peak of the corresponding integral.



The default adaptive quadrature settings are shown above. These settings can be changed or reset to the default values by clicking the **Default** button. Select the **OK** button when done. A portion of the updated syntax file is shown below.



Select the **Analysis**, **Run** option. The parameter estimates are shown below.



OmniLog

Project:	lsat6 data;
Description:	Adaptive quadrature. 2pl models fitted to the items;
Date:	15 December 2009
Time:	12:19 PM

Table of Contents

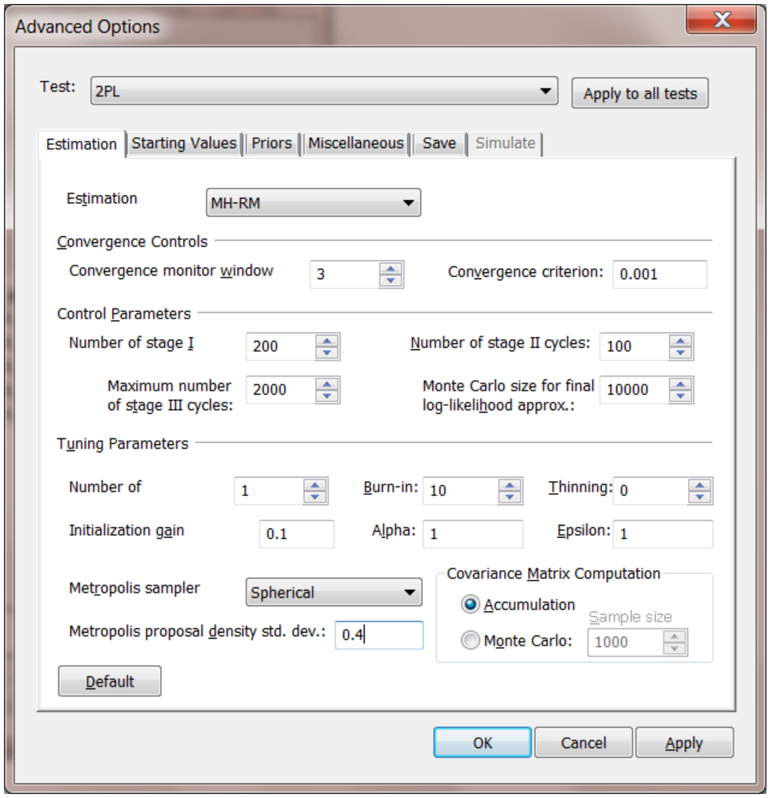
2PL Model Item Parameter Estimates for Group 1, logit: $a\theta + c$ or $a(\theta - b)$ (Back to TOC)

Item	Label	*a*	*c*	*b*
1	C1	0.83	2.77	-3.36
2	C2	0.72	0.99	-1.37
3	C3	0.89	0.25	-0.28
4	C4	0.69	1.28	-1.87
5	C5	0.66	2.05	-3.13

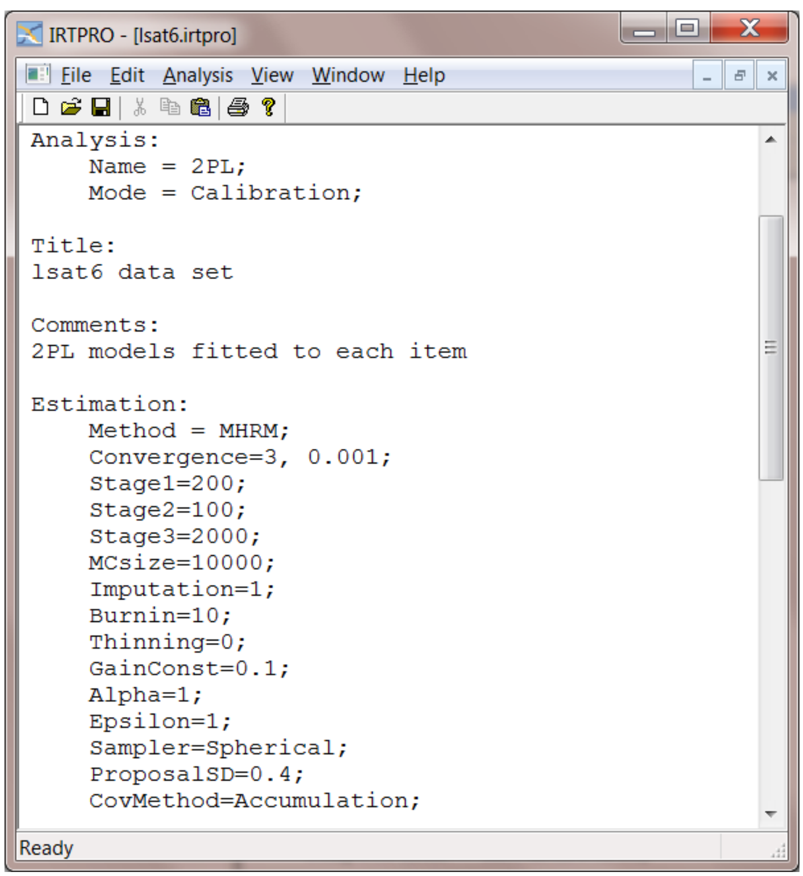
11.3 MH-RM

Li Cai (2010-b and 2010-c) proposed a Metropolis-Hastings Robbins-Monro (MH-RM) algorithm to address the "curse of dimensionality" that has plagued multidimensional IRT and high-dimensional latent structural equation modeling. The MH-RM algorithm performs favorably in comparative studies against established gold-standard methods such as Gaussian quadrature. The MH-RM algorithm is much more efficient than the MCEM algorithm in the use of Monte Carlo because the simulation size in MH-RM is fixed and usually small throughout the iterations. In addition, MH-RM produces an estimate of the parameter information matrix as a by-product that can be used subsequently for standard error estimation and goodness-of-fit testing.

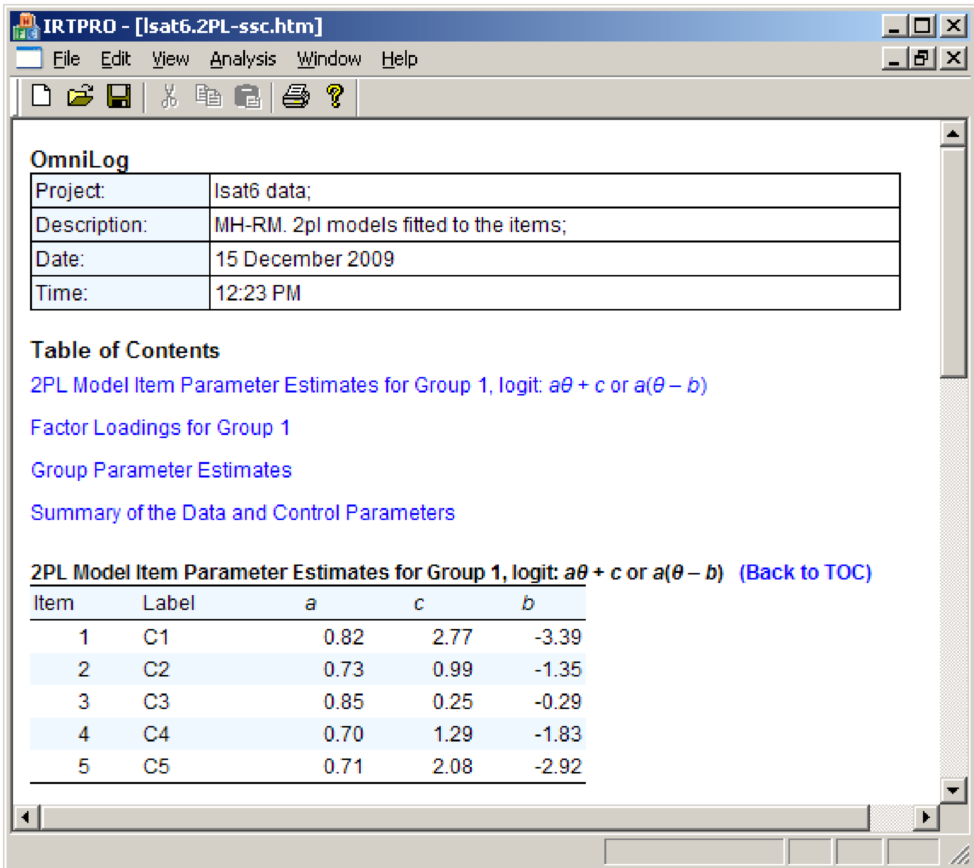
For practical data analysis, one can often achieve efficiency gains of several orders of magnitude over existing methods such as numerical quadrature if one uses MH-RM to estimate the parameters of the model. The MH-RM method is ideally suited for multi-dimensional analyses where the number of dimensions exceeds two or three.



A portion of the revised syntax file is shown below.



To run the 2PL model using the MH-RM method, select the **Analysis, Run** option. Some of the parameter estimates shown below differs a small amount from those obtained using the previously described estimation methods.


IRTPRO - [lsat6.2PL-ssc.htm]

File Edit View Analysis Window Help

OmniLog

Project:	lsat6 data;
Description:	MH-RM. 2pl models fitted to the items;
Date:	15 December 2009
Time:	12:23 PM

Table of Contents

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

Factor Loadings for Group 1

Group Parameter Estimates

Summary of the Data and Control Parameters

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

Item	Label	*a*	*c*	*b*
1	C1	0.82	2.77	-3.39
2	C2	0.73	0.99	-1.35
3	C3	0.85	0.25	-0.29
4	C4	0.70	1.29	-1.83
5	C5	0.71	2.08	-2.92

12. Syntax

12.1 Introduction

Examples covering the range of models that the IRTPRO program handles are presented in Chapters 4 to 8 in the following sequence:

- Chapter 4: Traditional summed-scored statistics
- Chapter 5: Unidimensional IRT
- Chapter 6: Multiple groups and DIF
- Chapter 7: Multidimensional IRT
- Chapter 8: IRT Scoring

IRTPRO distinguishes between three analysis modes:

- Traditional (Chapter 4)
- Calibration (Chapters 5 – 7)
- Scoring (Chapter 8)

Each analysis created by the GUI produces a syntax file with extension **.irtpro**, essentially being a record of a user's selections from the sequence of dialogs. Using the same IRTPRO **.ssig** data file (referred to as the Project Name), more than one analysis can be specified by inserting additional test tabs in the **Analysis** window. In this case, the syntax associated with each analysis (test) are appended to the same **.irtpro** file. The advantage of having a syntax file is that it enables one to duplicate an analysis. Specifically, suppose that a person A sends a copy of the IRTPRO **.ssig** and **.irtpro** files to person B. The latter will be able to duplicate the results obtained by person A, without the need to recreate the analysis.

Note that if a syntax file is opened, IRTPRO automatically fills the relevant dialogs and these can be viewed and modified using the **Analysis** menu on the main menu-bar. The **IRTPRO Examples** folder contains many syntax files illustrating the capabilities of the program.

12.2 Structure of a syntax file

A syntax file consists of paragraphs (commands), each starting with a command keyword followed by a punctuation mark (:). For example, regardless of the mode of analysis, the first five commands of a syntax file have the following structure:

```
Project:
   Name = <Name of the IRTPRO data file without the extension .ssig>;
Data:
   File = <Name of the IRTPRO data file with the extension .ssig >;
Analysis:
   Name = <Testn>;
   Mode = <Analysis type>;
Title:
  <Description of the analysis>
```

Comments :
<Additional comments about the analysis>

12.2.1 Project: and Data: commands

The first two commands are generated when a **.ssig** file is opened and an analysis mode (Section 12.1) is selected from the **Analysis** menu.

12.2.2 Analysis: command

This command contains two keywords (Name and Mode). The first keyword corresponds with the n-th test tab in the analysis window, the default being Test1 for a new analysis, followed by Test2, Test3,… if additional tabs are inserted on the **Analysis** window. These test tabs can be renamed by the user.

The second keyword can be any one of the following choices:

1. Mode = Traditional;
2. Mode = Calibration;
3. Mode = Scoring;

12.2.3 Title: and Comments: commands

These commands are associated with the **Description** tab on the **Analysis** window and are optional. Note that text entered in the title and comments paragraphs does not end with a semi-colon (;). However, semi-colons are required when indicated in paragraphs.

A subset of the available IRTPRO commands has the same functionality, regardless of the analysis type. These commands will be elaborated upon in this section, whereas paragraphs specific to an analysis type will be discussed in Sections 12.3 to 12.5.

Below is a typical syntax file containing two tests. The commands that are common to each analysis type are marked in bold and will be discussed in this section.

Project:
    Name = Spelling;
Data:
    File = .\Spelling.ssig;
Analysis:
    Name = Test1;
    Mode = Calibration;
Title:
Four item spelling test

Comments:
All item parameters constrained
to be equal across groups

```
Estimation:
   Method = BAEM;
   E-Step = 500, 1e-005;
   SE = S-EM;
   M-Step = 50, 1e-009;
   Quadrature = 49, 6;
   SEM = 0.001;
   SS = 1e-005;
Save:
   PRM, INF
Scoring:
   Pattern = EAP;
   Score Persons;
   Mean = 0;
   SD = 1;
Miscellaneous:
    Decimal  = 3;
    Processor  = 2;
   Print M2, CTLD, Loadings, P-Nums, Diagnostic;
   Min Exp = 1;
Groups:
   Variable = gender;
Group G1:
    Value  = (1);
    Dimension  = 1;
    Items  = item1, item2, item3, item4;
    Codes(item1, item2, item3, item4) = 0(0), 1(1);
    Model(item1, item2, item3, item4) = 2PL;
    Referenced;
    Mean = 0.0;
    Covariance  = 1.0;
Group G2:
    Value  = (2);
    Dimension = 1;
    Items  = item1, item2, item3, item4;
    Codes(item1, item2, item3, item4) = 0(0), 1(1);
    Model(item1, item2, item3, item4) = 2PL;
   Mean = Free;
   Covariance = Free;
Constraints:
   Equal = (G2, item1, Slope[0]), (G1, item1, Slope[0]);
   Equal = (G2, item1, Intercept[0]), (G1, item1, Intercept[0]);
       :
       :
   Equal = (G2, item4, Slope[0]), (G1, item4, Slope[0]);
   Equal = (G2, item4, Intercept[0]), (G1, item4, Intercept[0]);
```

Only a portion of the syntax for the second test is shown below. The important point to note is that each subsequent test starts with an **Analysis:** command.

```
Analysis:
        Name = Test2;
        Mode = Calibration;
Title:
Spelling test 4 items
```

Comments:
Parameters of items 1-3 equal, item 4 different
:
:

Next, we describe the remaining commands, common to the three modes of analyses. The commands that are mode specific will be discussed in Sections 12.3 to 12.5.

12.2.4 Estimation: command

The first keyword in the **Estimation:** command is the keyword **Method =** which has the following form:

Method = <estimation method>;

The method of estimation (<estimation method>) is BAEM (Bock-Aitkin), ADQ (Adaptive Quadrature) or MHRM (Metropolis-Hastings Robbins-Monro).

The keywords controlling the iterative procedure of each estimation method are listed below. For a further discussion, the reader is referred to Chapter 11.

12.2.4.1 BAEM

```
Estimation:
   Method = BAEM;
   E-Step = 500, 1e-005;
   SE = S-EM;
   M-Step = 500, 1e-009;
   Quadrature = 49, 6;
   SEM = 0.001;
   SS = 1e-005;
```

12.2.4.2 ADQ

```
Estimation:
   Method = ADQ;
   E-Step = 100, 0.001;
   SE = S-EM;
   Quadrature = 9, GH;
   Adaptation = EAP;
   Trust = Fast;
```

12.2.4.3 MHRM

```
Estimation:
   Method = MHRM;
   Convergence=3, 0.001;
   Stage1=200;
   Stage2=100;
   Stage3=2000;
```

```
 MCsize=10000;
Imputation=1;
Burnin=10;
Thinning=0;
GainConst=0.1;
Alpha=1;
Epsilon=1;
Sampler=Spherical;
ProposalSD=1;
CovMethod=Accumulation;
```

12.2.5 Save: command

The structure of the **Save:** command is as follows.

```
Save:
   <List of files to be saved with .txt extension>
```

Each name in the selected list must be followed by a comma, except the last one. Valid names are:

- PRM (Item parameter estimates **–prm.txt**)
- COV (Asymptotic covariance matrix of the parameter estimates **–cov.txt**)
- INF (Information values, unidimensional models only **–inf.txt**)
- POL (Inter item polychoric correlations, unidimensional models only **–pol.txt**)
- FAC (Factor loadings **–fac.txt**)
- IRT (Main output in ASCII text format **–irt.txt**, **-sss.txt** and **–ssc.txt**)
- DBG (Debugging output **–dbg.txt**)

12.2.6 Miscellaneous: command

Keywords contained in this paragraph are:

```
Decimal = <Number of decimal places>;
Processor = <Number of processors>;
Print <List of Additional results in addition to the standard output>;
```

A further description of the function of these keywords follows.

<Number of decimal places>; Can be a value of 2, 3 or 4 and controls the number of decimals to be written to the output file.

<Number of processors>; The number of processors requested for an analysis. If syntax is generated by the GUI, the default is the total number of processors available on a computer.

<List of Additional results>;

Each item in the list is followed by a comma, except for the last one. Available values are:

- StdRes (Print table of standardized residuals)
- CTLD (Compute Chen-Thissen LD and item fit statistics)
- M2 (Compute limited-information overall model fit statistics)
- GOF (Print each item's goodness of fit frequency table)
- Loadings (Print factor loadings)
- P-Nums (Print parameter numbers)

- Diagnostic (Print diagnostic information)

Note that if the printout of each item's goodness of fit frequency table (GOF) is requested, the keyword

Min Exp = <value>;

follows the **Print** <list>; statement, where <value> denotes the minimum expected frequency to be used when a frequency table is computed.

12.2.7 Groups: command

This paragraph is usually empty for a single group analysis, but for a multiple group analysis it has the form

Groups:
Variable = <variable) that defines the groups>;

Example: **Variable = Language;**

In the case of an exploratory factor analysis (EFA), this paragraph has the form

Groups:
Variable = <variable that defines the groups>; (multiple groups only)
EFA = <Rotation Method>; (See Section 12.4.3)

12.2.8 Groups Gn: command

For a single group analysis, the command becomes **Groups:** and this statement appears only once in an analysis. For a multiple group analysis this command is followed by the keyword **Value** = (n) where n refers to the actual value assigned to a grouping variable.

The keywords **Dimension** =, **Items** =, Codes and **Model** (when the analysis mode is **Traditional** the **Model** keyword is not used) follows next and is repeated for each groups as shown below:

Group G1:
Value = (1);
Dimension = 1;
Items = item1, item2, item3, item4; *(List of items selected)*
Codes (item1, item2, item3, item4) = 0(0), 1(1); *(Data values followed by scores shown in parenthesis for a given subset of items)*
Model(item1, item2, item3, item4) = 2PL; *(Name of model selected for a given subset of items)*
Referenced;
Mean = 0.0;
Covariance = 1.0;

```
Group G2:
        Value = (2);
         Dimension = 1;
         Items  = item1, item2, item3, item4;
        Codes(item1, item2, item3, item4) = 0(0), 1(1);

        Model(item1, item2, item3, item4) = 2PL;
         Mean  = free;
        Covariance  = free;

   Group G3:
     Value = (3);
        :
        :
```

For the reference group the keyword **Referenced;** appears just below the **Model** keyword, followed by fixed values for the **Mean(s) and Covariance(s)**. Typically the **Mean(s) and Covariance(s)** are set free in the remaining groups.

12.2.9 Remarks

The number of categories may vary from item to item. Therefore there can be several lines starting with the keyword **Codes**. Likewise, different models may be fitted to items and therefore there can be several lines starting with the keyword **Model**. The following example serves as illustration:

```
Codes(Item1) = 1(0), 2(1), 3(2), 4(2);
Codes(Item2, Item3) = 0(1), 1(0);
Codes(Item4) = 0(0), 1(1), 2(2);

Models(Item1) = Nominal;
Models(Item2, Item3) = 2PL;
Models(Item4) = Graded;
```

From the above, we note that the third and fourth categories of Item1 have been combined. Note also that the smallest score value is always 0, but does not have to be the first in the list. For example, Item2 and Item3 are recoded so that a "0" in the data set is assigned a score of "1" and a "1" in the data set is assigned a score of "0".

12.3 Traditional summed-scored statistics

Below we show a typical syntax file for a traditional summed-scored statistics analysis. The main difference between this syntax and the syntax for the remaining analysis modes is that in this case, there are no **Models** keywords and no **Constraints:** command.

```
Project:
   Name = Anxiety14itemsV7;
Data:
   File = .\Anxiety14itemsV7.ssig;
Analysis:
   Name = Test1;
   Mode = Traditional;
Title:
Six anxiety items selected from the file Anxiety14itemsV7.ssig
Comments:
To illustrate the computation of traditional statistics
Estimation:
   Method = BAEM;
   E-Step = 500, 1e-005;
   SE = S-EM;
   M-Step = 50, 1e-006;
   Quadrature = 49, 6;
   SEM = 0.001;
   SS = 1e-005;
Miscellaneous:
   Decimal = 2;
   Processors = 4;
   Print CTLD, P-Nums, Diagnostic;
   Min Exp = 1;
Groups:
Group :
   Dimension = 1;
   Items = Calm, Tense, Regretful, AtEase, Anxious, Nervous;
   Codes(Calm) = 1(0), 2(1), 3(2), 4(3), 5(4);
   Codes(Tense) = 1(0), 2(1), 3(2), 4(3), 5(4);
   Codes(Regretful) = 1(0), 2(1), 3(2), 4(3), 5(4);
   Codes(AtEase) = 1(0), 2(1), 3(2), 4(3), 5(4);
   Codes(Anxious) = 1(0), 2(1), 3(2), 4(3), 5(4);
   Codes(Nervous) = 1(0), 2(1), 3(2), 4(3), 5(4);
   Mean = 0.0;
   Covariance = 1.0;
```

12.4 Calibration

Calibration entails estimating parameters and standard errors for a wide range of IRT models. Although the general form of a syntax file contains many common features as pointed out in Section 12.2, there are keywords that are uniquely associated with the specific type of analysis selected. These aspects are discussed in more detail in Sections 12.4.1 to 12.4.3.

12.4.1 Unidimensional IRT

A typical syntax file for a multiple group unidimensional IRT analysis is shown below. Note that a multiple group analysis usually contains a **Constraints:** paragraph to ensure that the model fitted to the data is estimable. In the syntax file shown, corresponding item parameters are constrained to be equal across groups, but the elements of the mean (vector) and covariance (matrix) are set free in the groups that do not serve as the reference group. In this case, it is the mean and covariance of the second group that are set free.

```
Project:
   Name = PISAMathBook1USUK;
Data:
   File = .\PISAMathBook1USUK.ssig;
Analysis:
   Name = IRT;
   Mode = Calibration;
Title:
2-group IRT Analysis (GPC model for polytomous items)
Comments:
Mixture of 2PL and General Partial Credit Models
Estimation:
   Method = BAEM;
   E-Step = 500, 1e-005;
   SE = S-EM;
   M-Step = 50, 1e-006;
   Quadrature = 49, 6;
   SEM = 0.001;
   SS = 1e-005;
Miscellaneous:
   Decimal = 2;
   Processors = 2;
   Print CTLD, P-Nums, Diagnostic;
   Min Exp = 1;
Groups:
   Variable = Country;
Group G1:
   Value = (1);
   Dimension = 1;
   Items = Cube1, Cube3, Cube4, Farms1, Farms4, Walking1,  Walking3, Apples1,Apples2, Apples3,
   Continent, Grow1, Grow3, Grow2;
   Codes(Cube1, Cube3, Cube4, Farms1, Farms4, Walking1) = 0(0), 1(1);
   Codes(Walking3) = 0(0), 1(1), 2(2), 3(3);
   Codes(Apples1, Apples2) = 0(0), 1(1);
   Codes(Apples3, Continent) = 0(0), 1(1), 2(2);
   Codes(Grow1, Grow3) = 0(0), 1(1);
   Codes(Grow2) = 0(0), 1(1), 2(2);
   Model(Cube1, Cube3, Cube4, Farms1, Farms4, Walking1) = 2PL;
   Model(Walking3) = GP Credit;
   GammaMatrix(Walking3) = Trend;
   Model(Apples1, Apples2) = 2PL;
   Model(Apples3, Continent) = GP Credit;
   GammaMatrix(Apples3 Continent) = Trend;
   Model(Grow1, Grow3) = 2PL;
   Model(Grow2) = GP Credit;
   GammaMatrix(Grow2) = Trend;
   Referenced;
   Mean = 0.0;
   Covariance = 1.0;
Group G2:
   Value = (2);
   Dimension = 1;
   Items = Cube1, Cube3, Cube4, Farms1, Farms4, Walking1, Walking3, Apples1,
   Apples2, Apples3, Continent, Grow1, Grow3, Grow2;
   Codes(Cube1, Cube3, Cube4, Farms1, Farms4, Walking1) = 0(0), 1(1);
   Codes(Walking3) = 0(0), 1(1), 2(2), 3(3);
```

```
   Codes(Apples1, Apples2) = 0(0), 1(1);
   Codes(Apples3, Continent) = 0(0), 1(1), 2(2);
   Codes(Grow1, Grow3) = 0(0), 1(1);
   Codes(Grow2) = 0(0), 1(1), 2(2);
   Model(Cube1, Cube3, Cube4, Farms1, Farms4, Walking1) = 2PL;
   Model(Walking3) = GP Credit;
   GammaMatrix(Walking3) = Trend;
   Model(Apples1, Apples2) = 2PL;
   Model(Apples3, Continent) = GP Credit;
   GammaMatrix(Apples3 Continent) = Trend;
   Model(Grow1, Grow3) = 2PL;
   Model(Grow2) = GP Credit;
   GammaMatrix(Grow2) = Trend;
   Mean = Free;
   Covariance = Free;
Constraints:
   Equal = (G1, Cube1, Slope[0]), (G2, Cube1, Slope[0]);
   Equal = (G1, Cube1, Intercept[0]), (G2, Cube1, Intercept[0]);
   Equal = (G1, Cube3, Slope[0]), (G2, Cube3, Slope[0]);
   Equal = (G1, Cube3, Intercept[0]), (G2, Cube3, Intercept[0]);
   Equal = (G1, Cube4, Slope[0]), (G2, Cube4, Slope[0]);
   Equal = (G1, Cube4, Intercept[0]), (G2, Cube4, Intercept[0]);
   Equal = (G1, Farms1, Slope[0]), (G2, Farms1, Slope[0]);
   Equal = (G1, Farms1, Intercept[0]), (G2, Farms1, Intercept[0]);
   Equal = (G1, Farms4, Slope[0]), (G2, Farms4, Slope[0]);
   Equal = (G1, Farms4, Intercept[0]), (G2, Farms4, Intercept[0]);
   Equal = (G1, Walking1, Slope[0]), (G2, Walking1, Slope[0]);
   Equal = (G1, Walking1, Intercept[0]), (G2, Walking1, Intercept[0]);
   Equal = (G1, Walking3, Slope[0]), (G2, Walking3, Slope[0]);
   Equal = (G1, Walking3, Alpha[0]), (G2, Walking3, Alpha[0]);
   Equal = (G1, Walking3, Alpha[1]), (G2, Walking3, Alpha[1]);
   Equal = (G1, Walking3, Alpha[2]), (G2, Walking3, Alpha[2]);
   Equal = (G1, Walking3, Gamma[0]), (G2, Walking3, Gamma[0]);
   Equal = (G1, Walking3, Gamma[1]), (G2, Walking3, Gamma[1]);
   Equal = (G1, Walking3, Gamma[2]), (G2, Walking3, Gamma[2]);
   Equal = (G1, Apples1, Slope[0]), (G2, Apples1, Slope[0]);
   Equal = (G1, Apples1, Intercept[0]), (G2, Apples1, Intercept[0]);
   Equal = (G1, Apples2, Slope[0]), (G2, Apples2, Slope[0]);
   Equal = (G1, Apples2, Intercept[0]), (G2, Apples2, Intercept[0]);
   Equal = (G1, Apples3, Slope[0]), (G2, Apples3, Slope[0]);
   Equal = (G1, Apples3, Alpha[0]), (G2, Apples3, Alpha[0]);
   Equal = (G1, Apples3, Alpha[1]), (G2, Apples3, Alpha[1]);
   Equal = (G1, Apples3, Gamma[0]), (G2, Apples3, Gamma[0]);
   Equal = (G1, Apples3, Gamma[1]), (G2, Apples3, Gamma[1]);
   Equal = (G1, Continent, Slope[0]), (G2, Continent, Slope[0]);
   Equal = (G1, Continent, Alpha[0]), (G2, Continent, Alpha[0]);
   Equal = (G1, Continent, Alpha[1]), (G2, Continent, Alpha[1]);
   Equal = (G1, Continent, Gamma[0]), (G2, Continent, Gamma[0]);
   Equal = (G1, Continent, Gamma[1]), (G2, Continent, Gamma[1]);
   Equal = (G1, Grow1, Slope[0]), (G2, Grow1, Slope[0]);
   Equal = (G1, Grow1, Intercept[0]), (G2, Grow1, Intercept[0]);
   Equal = (G1, Grow3, Slope[0]), (G2, Grow3, Slope[0]);
   Equal = (G1, Grow3, Intercept[0]), (G2, Grow3, Intercept[0]);
   Equal = (G1, Grow2, Slope[0]), (G2, Grow2, Slope[0]);
   Equal = (G1, Grow2, Alpha[0]), (G2, Grow2, Alpha[0]);
   Equal = (G1, Grow2, Alpha[1]), (G2, Grow2, Alpha[1]);
   Equal = (G1, Grow2, Gamma[0]), (G2, Grow2, Gamma[0]);
   Equal = (G1, Grow2, Gamma[1]), (G2, Grow2, Gamma[1]);
```

12.4.1.1 The constraints: command

In the syntax file listed above, a set of **Equal** = ( ), ( ); keywords follow the **constraints:** command, each having the following structure.

Equal = (G1, Item Name, Parameter), (G2, Item Name, Parameter);

Note that the for all 2PL models there are two parameters; the slope (a1) and the intercept (c = Intercept[0]). For a two-dimensional model the slope parameters are denoted as a1 (Slope[0]) and a2 (Slope[1]). Consider, in this case, the statement:

Equal = (G1, Cube1, Intercept[0]), (G2, Cube1, Intercept[0]);

This statement instructs IRTPRO to set the c parameter of Cube1 in the first group equal to the c parameter of Cube1 in the second group.

12.4.2 DIF (Differential item functioning)

A typical syntax file for a DIF analysis is shown below. Note that a DIF analysis should always contain a **DIF <type>;** command where <type> equals one of the words Anchor, Random or All. This command should be inserted just before the **Constraints:** command. In the syntax file shown, the syntax for each type of DIF analysis is shown in red. Note that in an analysis (test) only one type can be specified at a time.

```
Project:
   Name = Spelling;
Data:
   File = .\Spelling.ssig;
Analysis:
   Name = Anchored;
   Mode = Calibration;
Title:
Spelling DIF
Comments:
Anchor items 1-2-3 candidate 4, 2PL
Estimation:
   Method = BAEM;
   E-Step = 500, 1e-005;
   SE = S-EM;
   M-Step = 50, 1e-009;
   Quadrature = 49, 6;
   SEM = 0.001;
   SS = 1e-005;
Save:
   PRM, COV
Miscellaneous:
   Decimal = 2;
   Processor = 1;
   Print GOF, P-Nums;
   Min Exp = 1;
```

```
Groups:
   Variable = Gender;
Group G1:
   Value = (1);
   Dimension = 1;
   Items = Infidelity, Panoramic, Succumb, Girder;
   Codes(Infidelity, Panoramic, Succumb, Girder) = 0(0), 1(1);
   Model(Infidelity, Panoramic, Succumb, Girder) = 2PL;
   Referenced;
   Mean = 0.0;
   Covariance = 1.0;
Group G2:
   Value = (2);
   Dimension = 1;
   Items = Infidelity, Panoramic, Succumb, Girder;
   Codes(Infidelity, Panoramic, Succumb, Girder) = 0(0), 1(1);
   Model(Infidelity, Panoramic, Succumb, Girder) = 2PL;
   Model(Panoramic) = 2PL;
   Model(Succumb) = 2PL;
   Model(Girder) = 2PL;
   Mean = Free;
   Covariance = Free;
DIF Anchor:
   Candidate = Girder;
   Anchors = Infidelity, Panoramic, Succumb;
Constraints:
   Equal = (G1, Infidelity, Slope[0]), (G2, Infidelity, Slope[0]);
   Equal = (G1, Infidelity, Intercept[0]), (G2, Infidelity, Intercept[0]);
   Equal = (G1, Panoramic, Slope[0]), (G2, Panoramic, Slope[0]);
   Equal = (G1, Panoramic, Intercept[0]), (G2, Panoramic, Intercept[0]);
   Equal = (G1, Succumb, Slope[0]), (G2, Succumb, Slope[0]);
   Equal = (G1, Succumb, Intercept[0]), (G2, Succumb, Intercept[0]);
```

The example above is for DIF analysis with a specified set of anchor items. Alternatively, the DIF paragraph could be either a randomized group analysis, or, for the two-stage analysis in which all items are examined for DIF, conditional on between–group differences estimated with all item parameters constrained equal across groups.

```
DIF Random:
   Candidate = Girder;
   Anchors = Infidelity, Panoramic, Succumb;
Constraints:
DIF All:
Constraints:
```

All items are constrained equal to the corresponding parameters in each group in the GUI–generated syntax file, and in the first stage of the two–stage DIF analysis, to estimate the mean and variance of the focal group(s).

12.4.2.1 The DIF <type>: command

IRTPRO distinguishes between three types of DIF (see Chapter 1 for details):

DIF Anchor:
Candidate = <List of candidate items names>; (Separate the names with a comma)
Anchors = <List of anchor items>; (Separate the names with a comma)

DIF Random:
Candidate = <List of candidate items>;
Anchors = <List of anchor items>;

DIF All:
(No keywords in this paragraph)

When selecting the DIF option, the GUI automatically generates the required equality constraints listed above.

12.4.3 Multidimensional (EFA, BIFAC and CFA)

In the multidimensional case, there are three modeling methods available in IRTPRO; these being exploratory factor analysis (EFA), bifactor analysis (BIFAC) and confirmatory factor analysis (CFA). Additional keywords, to be inserted in the general syntax framework, are uniquely associated with the specific type of analysis method selected. These keywords are printed in red in what follows.

12.4.3.1 Exploratory Factor Analysis (EFA), single group analysis

When either an EFA or BIFAC analysis is requested, the **Constraints:** command is not used. However, the keyword **EFA** = <Rotation Method>; must be inserted in the **Groups:** paragraph.

Groups:
EFA = <Rotation Method>;
Group:
Dimension = n; *(n = 2, 3, 4 …)*
(*Additional keywords not shown*)

There are four rotation methods available in IRTPRO:

- EFA = ObDQ; (Oblique CF Quartimax)
- EFA = OrV; (Orthogonal CF Varimax)
- EFA = OrDQ; (Orthogonal CF Quartimax)
- EFA = ObV; (Oblique CF Varimax)

12.4.3.2 Exploratory Factor Analysis (EFA), multiple group analysis

Groups:
Variable = Country;
EFA = <Rotation Method>;
Group G1:
Value = (1);
Dimension = n;

12.4.3.3 Bifactor Analysis (BIFAC)

As noted previously, when either an EFA or BIFAC analysis is requested, the **Constraints:** command is not used. However, in the case of a bifactor analysis the keyword **GenDim = 1;** must be inserted in the **Groups:** paragraph. See Section 7.4 for an example where **GenDim = 2;** is used. A syntax file for doing a bifactor analysis (See Section 7.2) is shown below. Keywords that are unique to a bifactor analysis are printed in a bold red typeface.

```
Project:
   Name = QofLife;
Data:
   File = .\QofLife.ssig;
Analysis:
   Name = BiFAC;
   Mode = Calibration;
Title:
Bifactor analysis of the quality of life data (35 items)
Comments:
One general factor and seven specific factors
Estimation:
   Method = BAEM;
   E-Step = 500, 0.001;
   SE = Xpd;
   M-Step = 50, 0.001;
   Quadrature = 36, 6;
   SEM = 0.001;
   SS = 1e-005;
Miscellaneous:
   Decimal = 2;
   Processors = 4;
   Print CTLD, P-Nums;
 Groups:
 Group :
   Dimension = 8;
   GenDim = 1;
   Items = Item1, Item2, Item3, Item4, Item5, Item6, Item7, Item8,Item9, Item10, Item11, Item12, Item13,
   Item14, Item15, Item16, Item17, Item18, Item19, Item20, Item21, Item22, Item23, Item24, Item25,
   Item26, Item27, Item28, Item29, Item30, Item31, Item32, Item33, Item34, Item35;
   Codes(Item1, Item2, Item3, Item4, Item5, Item6, Item7, Item8,Item9, Item10, Item11, Item12, Item13,
   Item14, Item15, Item16, Item17, Item18, Item19, Item20, Item21, Item22, Item23, Item24, Item25,
   Item26, Item27, Item28, Item29, Item30, Item31, Item32, Item33, Item34, Item35) = 0(0), 1(1), 2(2), 3(3),
   4(4), 5(5), 6(6);
   Model(Item1, Item2, Item3, Item4, Item5, Item6, Item7, Item8,Item9, Item10, Item11, Item12, Item13,
   Item14, Item15, Item16, Item17, Item18, Item19, Item20, Item21, Item22, Item23, Item24, Item25,
   Item26, Item27, Item28, Item29, Item30, Item31, Item32, Item33, Item34, Item35) = Graded;
   BFA(Item2, Item3, Item4, Item5) = 2;
   BFA(Item6, Item7, Item8, Item9) = 3;
   BFA(Item10, Item11, Item12, Item13, Item14, Item15) = 4;
   BFA(Item16, Item17, Item18, Item19, Item20, Item21) = 5;
   BFA(Item22, Item23, Item24, Item25, Item26) = 6;
   BFA(Item27, Item28, Item29, Item30, Item31) = 7;
   BFA(Item32, Item33, Item34, Item35) = 8;
```

As shown above, the last part of the **Groups:** paragraph contains one or more lines of the form:

BFA(list of item names) = <number of the associated group factor>;

Note that some items may be associated only with the general factor. In this example, Item1 is not included with any of the group factors. It is also important to note that the items included as group factors must form mutually exclusive sets.

12.4.3.4 Confirmatory Factor Analysis (CFA or IRT)

A syntax file for doing a confirmatory factor analysis (See Chapter 7) is shown below. Key to doing a CFA is the use of the **Constraints:** command to set parameters equal to zero. This part of the **Constraints:** paragraph is printed in red typeface. Note that for a CFA, one can additionally impose equality constraints (**Equal =**). This is typically required when doing a multiple group CFA.

```
Project:
   Name = AACL3_21Items;
Data:
   File = .\AACL3_21Items.ssig;
Analysis:
   Name = 2d-CFA;
   Mode = Calibration;
Title:
AACL dataset, 21 items
Comments:
2-Dimensional simple structure CFA
Estimation:
   Method = BAEM;
   E-Step = 500, 1e-005;
   SE = S-EM;
   M-Step = 50, 1e-006;
   Quadrature = 49, 6;
   SEM = 0.001;
   SS = 1e-005;
Scoring:
   Mean = 0;
   SD = 1;
Miscellaneous:
   Decimal = 2;
   Processors = 2;
   Print CTLD, Loadings, P-Nums, Diagnostic;
   Min Exp = 1;
Groups:
Group :
   Dimension = 2;
   Items = Afraid, Desperate, Fearful, Frightened, Nervous, Panicky, Shaky, Tense, Terrified, Upset,
Worrying, Calm, Cheerful, Contented, Happy, Joyful, Loving, Pleasant, Secure, Steady, Thoughtful;
   Codes(Afraid, Desperate, Fearful, Frightened, Nervous, Panicky, Shaky, Tense, Terrified, Upset,
Worrying, Calm, Cheerful, Contented, Happy, Joyful, Loving, Pleasant, Secure, Steady, Thoughtful) = 1(1),
2(0);
   Model(Afraid, Desperate, Fearful, Frightened, Nervous, Panicky, Shaky, Tense, Terrified, Upset,
Worrying, Calm, Cheerful, Contented, Happy, Joyful, Loving, Pleasant, Secure, Steady, Thoughtful) = 2PL;
   Means = 0.0, 0.0;
   Covariances = 1.0,
              Free, 1.0;
```

```
Constraints:
   (Afraid, Slope[1]) = 0.0;
   (Desperate, Slope[1]) = 0.0;
   (Fearful, Slope[1]) = 0.0;
   (Frightened, Slope[1]) = 0.0;
   (Nervous, Slope[1]) = 0.0;
   (Panicky, Slope[1]) = 0.0;
   (Shaky, Slope[1]) = 0.0;
   (Tense, Slope[1]) = 0.0;
   (Terrified, Slope[1]) = 0.0;
   (Upset, Slope[1]) = 0.0;
   (Worrying, Slope[1]) = 0.0;
   (Calm, Slope[0]) = 0.0;
   (Cheerful, Slope[0]) = 0.0;
   (Contented, Slope[0]) = 0.0;
   (Happy, Slope[0]) = 0.0;
   (Joyful, Slope[0]) = 0.0;
   (Loving, Slope[0]) = 0.0;
   (Pleasant, Slope[0]) = 0.0;
   (Secure, Slope[0]) = 0.0;
   (Steady, Slope[0]) = 0.0;
   (Thoughtful, Slope[0]) = 0.0;
```

12.5 Scoring

A syntax file for scoring (See Chapter 8) is shown below. This syntax file contains the **Scoring:** command which follows the **Estimation:** paragraph. Key to an item scoring analysis is the use of the **Constraints:** command to assign values (obtained from a previous calibration) to the model parameters. This part of the **Constraints:** paragraph is printed in red.

```
Project:
   Name = SLF;
Data:
   File = .\SLF.ssig;
Analysis:
   Name = SSEAP;
   Mode = Scoring;
Title:
Social Life Feelings
Comments:
Summed score EAP
Estimation:
   Method = BAEM;
   E-Step = 500, 1e-005;
   SE = S-EM;
   M-Step = 50, 1e-006;
   Quadrature = 49, 6;
   SEM = 0.001;
   SS = 1e-005;
Scoring:
   Create SS to SC table;
   Score Persons;
   Mean = 0;
   SD = 1;
```

```
Miscellaneous:
   Decimal = 2;
   Processor = 1;
   Print CTLD, P-Nums, Diagnostic;
   Min Exp = 1;
Groups:
Group :
   Dimension = 1;
   Items = SLF1, SLF2, SLF3, SLF4, SLF5;
   Codes(SLF1, SLF2, SLF3, SLF4, SLF5) = 0(0), 1(1);
   Model(SLF1, SLF2, SLF3, SLF4, SLF5) = 2PL;
Constraints:
   (SLF1, Slope[0]) = 1.19684;
   (SLF1, Intercept[0]) = -2.35356;
   (SLF2, Slope[0]) = 0.71455;
   (SLF2, Intercept[0]) = 0.79647;
   (SLF3, Slope[0]) = 1.53051;
   (SLF3, Intercept[0]) = 0.99190;
   (SLF4, Slope[0]) = 2.54698;
   (SLF4, Intercept[0]) = -0.66874;
   (SLF5, Slope[0]) = 0.92269;
   (SLF5, Intercept[0]) = -1.09696;
```

12.5.1 Scoring: command

The **Scoring:** paragraph contains a number of optional keywords and has the following structure:

Scoring:
ID = <Variable name>; (*Optional*)

Followed by one of the following scoring methods

- **Create** SS to SC table;
- **Pattern** = EAP;
- **Pattern** = MAP;

Score Persons; *(Optional when summed-score to scale conversion table (SC to SC) is requested.)*
Integer Score; (*Optional*)
Mean = <Mean score value>; (*Default = 0.0*)
SD = <Standard deviation value>; (*Default = 1.0*)
Minimum = <Minimum score>; (*Optional*)
Maximum = <Maximum score>; (*Optional*)

13. References

Adams, R. & Wu, M. (2002). *PISA 2000 Technical Report*. Paris: OECD.

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

Baker, F.B. & Kim, S. H. (2004). *Item response theory: Parameter estimation techniques*. New York, NY: Marcel Dekker, Inc.

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-a). A two-tier full-information item factor analysis model with applications. Psychometrika, **75**, 581-612.

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

Cai, L. (2010-c). Metropolis-Hastings Robbins-Monro Algorithm for Confirmatory Item Factor Analysis. *Journal of Educational and Behavioral Statistics*, **35**, 307-335.

Cai, L., Yang, J. S. & Hansen, M. (in press). Generalized full-information item bifactor analysis. *Psychological Methods*.

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.

Eysenck, H.J. & Eysenck, S.B.G. (1969). *Personality Structure and Measurement*. London: Routledge.

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.

Hambleton, R.K. & Swaminathan, H. (1985). Item Response Theory. Principles and applications. Boston: Kluwer.

Krebs, D. & Schuessler, K.F. (1987). Soziale Empfindunge: ein interkultureller Skalenvergleich bei *Deutschen und Amerikanern*. Monographien: *Sozialwissenschaftliche Methoden*, Frankfurt/Main, New York: Campus Verlag.

Lehman, A. F. (1988). A quality of life interview for the chronically mentally ill. *Evaluation and Program Planning*, **11**, 51–62.

Lord, F. M. (1977). A broad-range tailored test of verbal ability. *Applied Psychological Measurement*, **1**, 95-100.

Lord, F.M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Erlbaum.

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, RJ. (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). NewYork: 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.

Revelle, W., Humphreys, M.S., Simon, L. & Gilliland, K. (1980). The Interactive Effect of Personality, Time of day, and Caffeine: A Test of the Arousal Model. *Journal of Experimental Psychology: General*, **109**, 1-31.

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.

Schuessler, K.F. (1982). *Measuring Social Life Feelings*, The Jossey-Bass Social and Behavioral Science Series, Jossey-Bass.

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

Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, **30**, 526-537.

Snyder, M., & Gangestad, S. (1986). On the nature of self-monitoring: Matters of assessment, matters of validity. *Journal of Personality and Social Psychology*, **51**, 125-139.

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 item response model. In M. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models: Developments and applications*.

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.

Wainer, H. & Kiely, G.L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, **24**, 185-201.

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