



## Contents

1.	Hybrid model fitted to Eysenck Extraversion Scale data .....	1
1.1	The 2PL Model.....	2
1.2	The 1PL Model.....	8
1.3	A Special Model with Equality Constraints.....	13
1.4	Trace Lines and information curves for the 2PL model .....	20

## 1. 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.

- Would you do almost anything for a dare?
- Do you stop and think things over before doing anything?
- Do you generally do and say things quickly without stopping to think?
- When people shout at you, do you shout back?
- Do you like doing things in which you have to act quickly?
- Do you often do things on the spur of the moment?
- Are you usually carefree?
- Do you often long for excitement?
- Are you slow and unhurried in the way you move?

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 to illustrate the models of interest here.

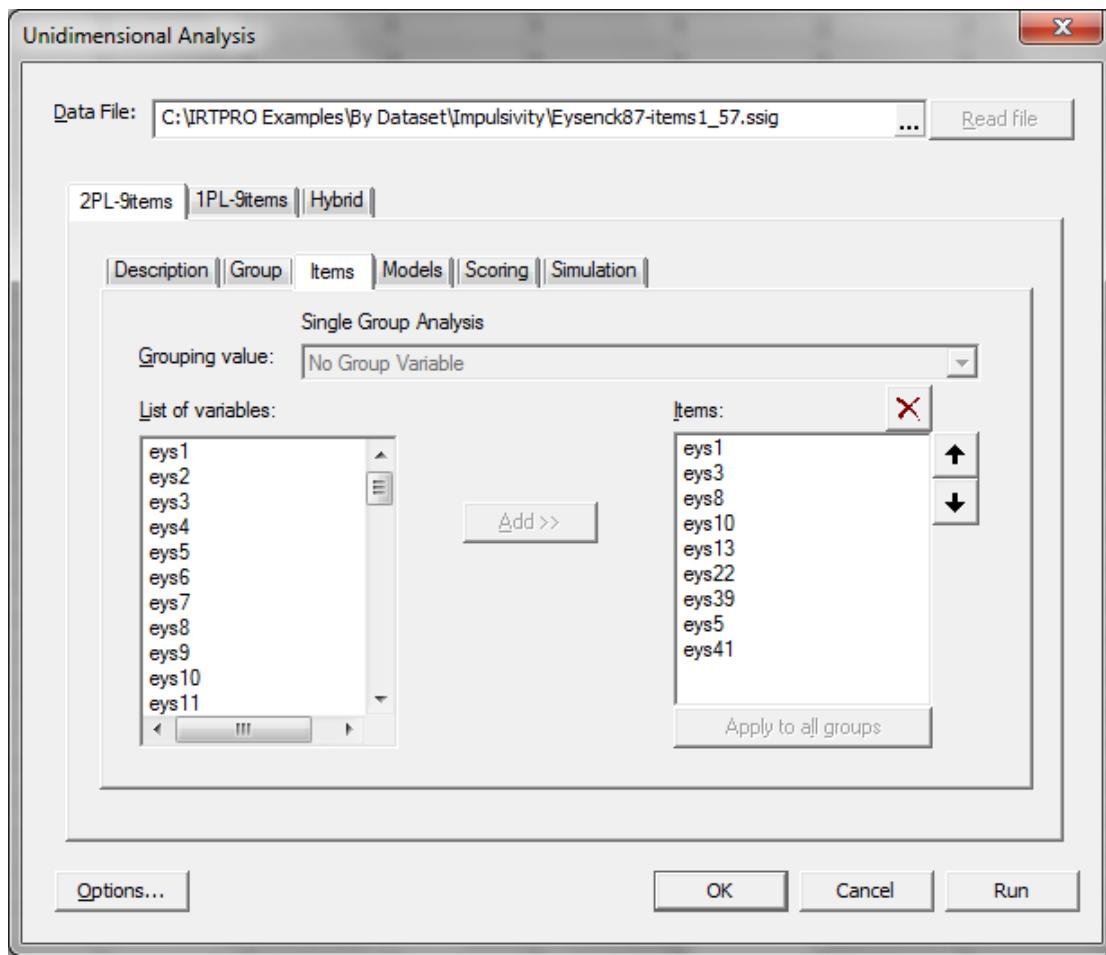
Below is an IRTPRO spreadsheet presentation of the data (**By Dataset\Impulsivity\Eysenck87-items1\_57.ssig**) showing the first 15 cases for the items eys1 to eys10.

	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

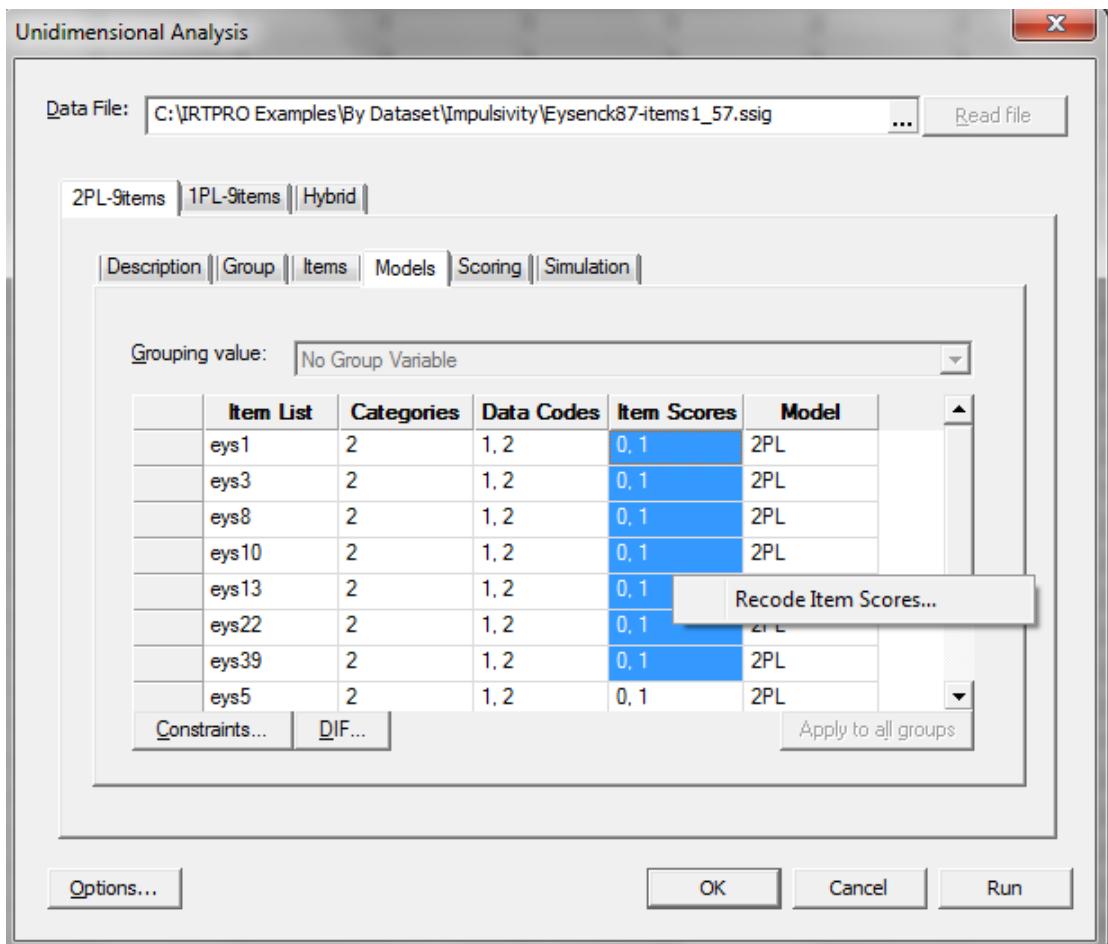
## 1.1 The 2PL Model

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**.

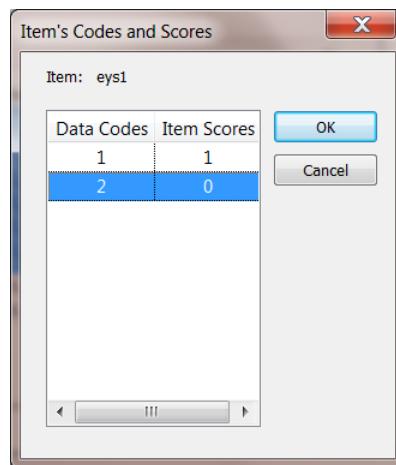
From the main menu bar, we select the **Analysis, Unidimensional IRT...** option. Click on the right-hand side of the **Test1** tab and insert a second test. Next, right-click on the **Test1** tab and rename it to **2PL-9Items**. Likewise, rename the **Test2** tab to **1PL-9Items**. Starting with the **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.



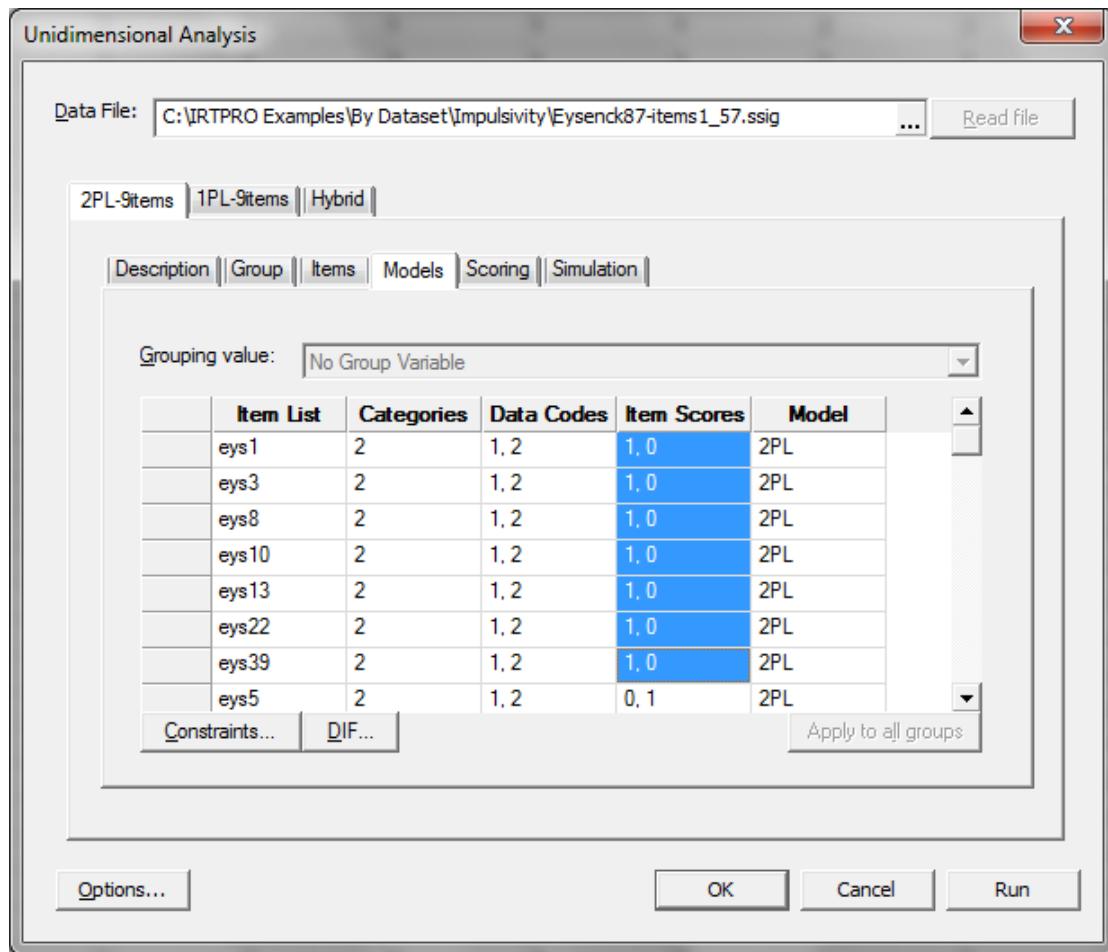
By clicking on the **Models** tab of the **Unidimensional Analysis** window, the list of items selected and model type is displayed. In the IRTPRO data, 1 = Yes and 2 = No. Therefore, (see Section 0) all of the items except Eys5 and Eys41 need to be reverse keyed on the **Models** dialog. The recoding 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.



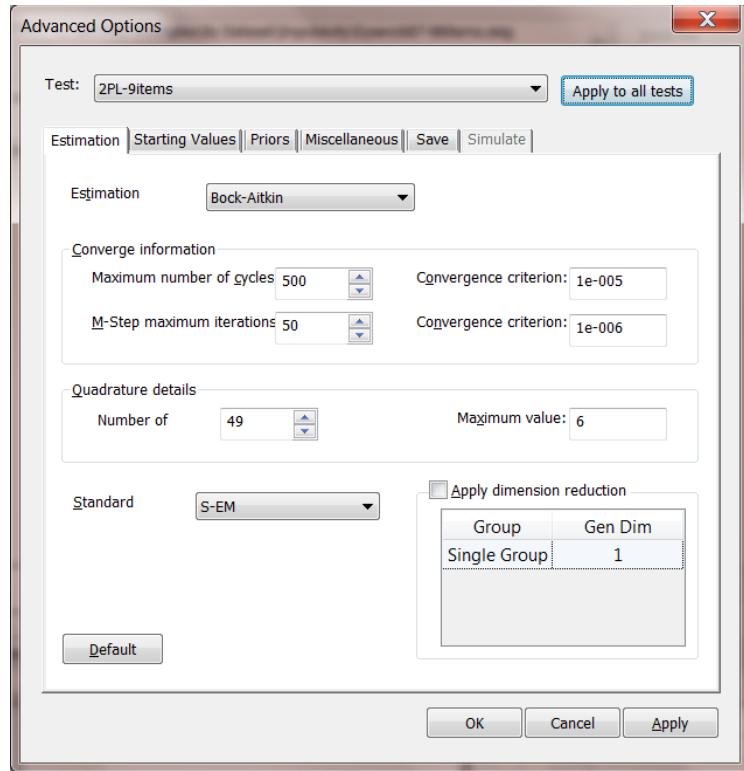
Selection of the **Recode Item Scores ...** option results in the display of the **Item's Codes and Scores** dialog where the **Item Scores** are changed in accordance with the answer key.



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



In this example, 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 bottom left 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:  $a\theta + c$  or  $a(\theta - b)$**  ([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
2	eys3	4	0.75	0.41	3	1.24	0.21
3	eys8	6	1.07	0.36	5	-0.43	0.21
4	eys10	8	0.83	0.42	7	-1.68	0.32
5	eys13	10	3.70	7.13	9	2.11	3.81
6	eys22	12	0.52	0.24	11	0.19	0.18
7	eys39	14	0.76	0.27	13	0.60	0.21
8	eys5	16	0.82	0.39	15	-1.60	0.37
9	eys41	18	0.13	0.19	17	0.56	0.16
						-4.16	6.04

**Summed-Score Based Item Diagnostic Tables and  $\chi^2$ 's for Group 1** ([Back to TOC](#))

S-X2 Item Level Diagnostic Statistics

Item	Label	$\chi^2$	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 <sub>2</sub>	Degrees of freedom	Probability	RMSEA
47.45	27	0.0088	0.06

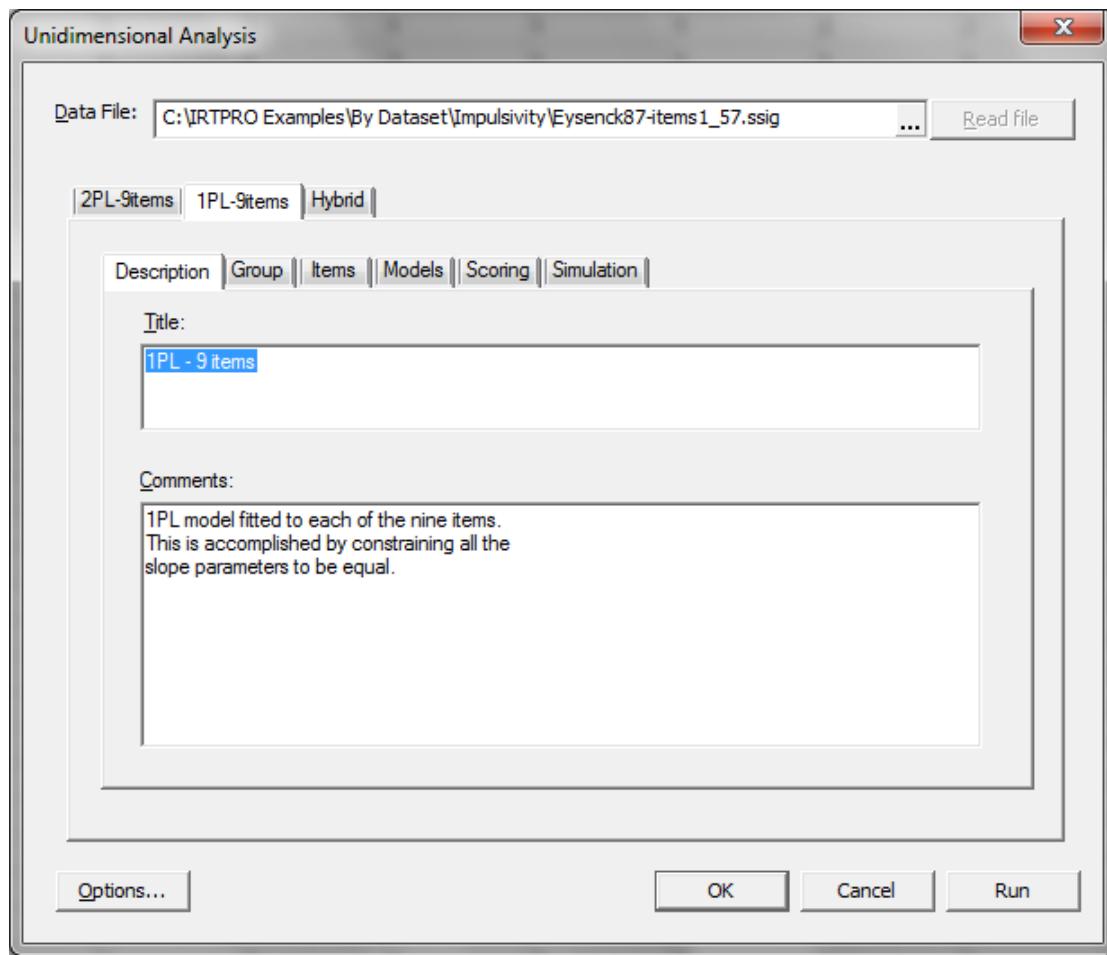
Note: M<sub>2</sub> is based on full marginal tables.

Note: Model-based weight matrix is used.

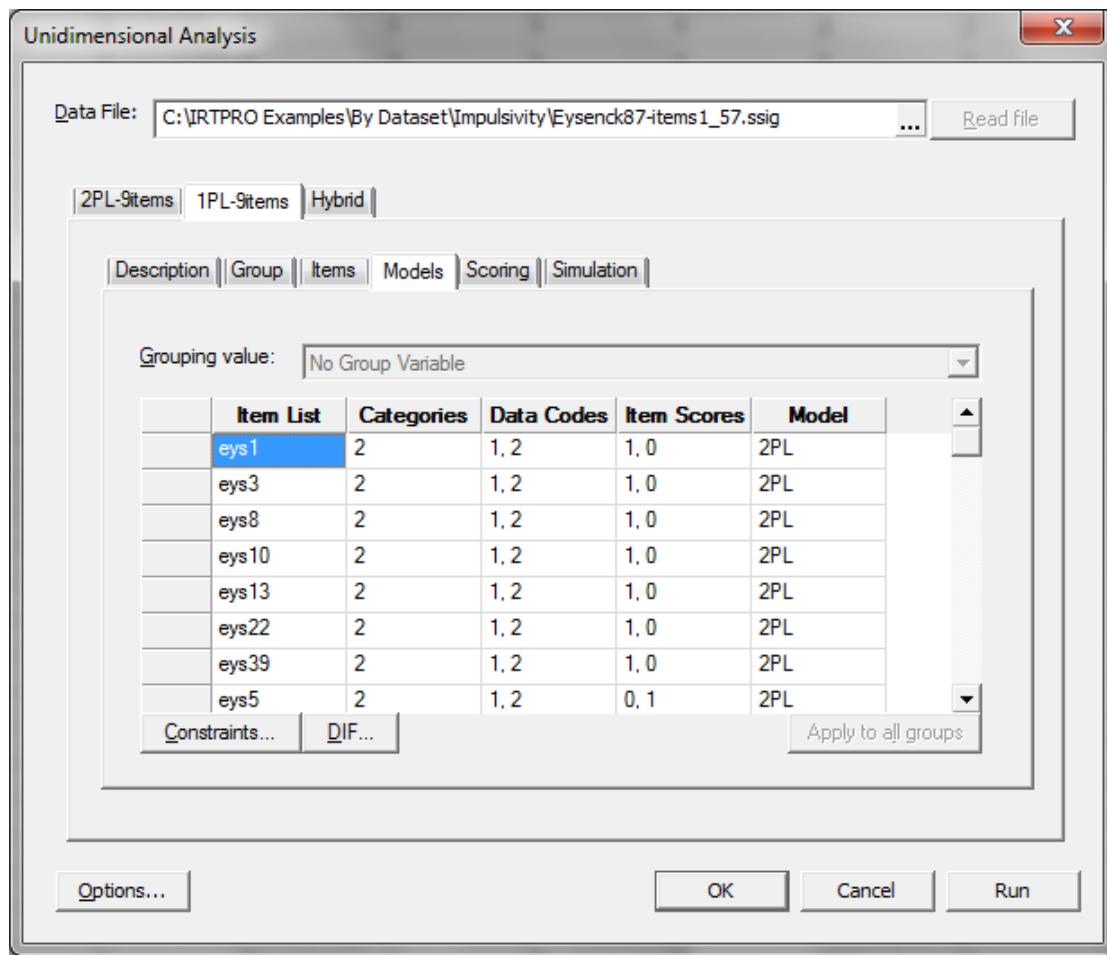
Before interpreting the variability among the slopes as indicative of reliable differences among the items' association with the latent variable being measured (impulsivity), we fit the 1PL model to the data to use the difference between the goodness of fit of the two models to compute a test of significance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

## 1.2 The 1PL Model

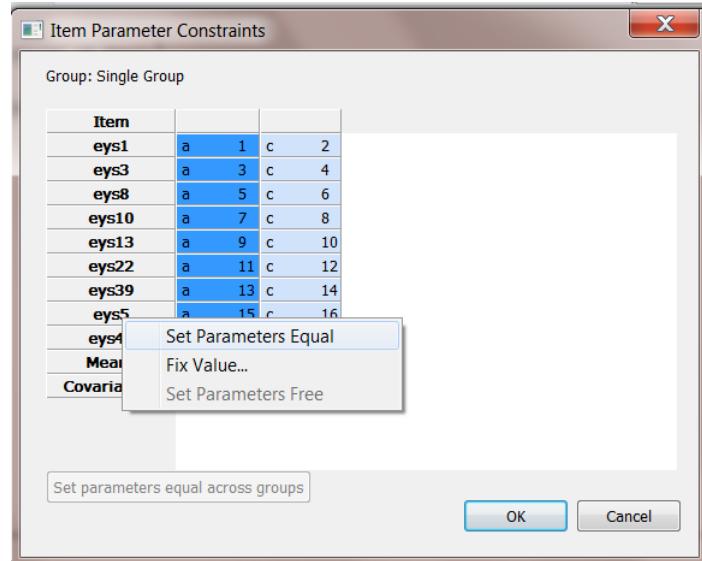
By closing the output file with extension .htm, we return to the **Analysis** option on the main menu bar, click on the **1PL-9items** tab. Start with the **Description** tab, and enter a title and comments.



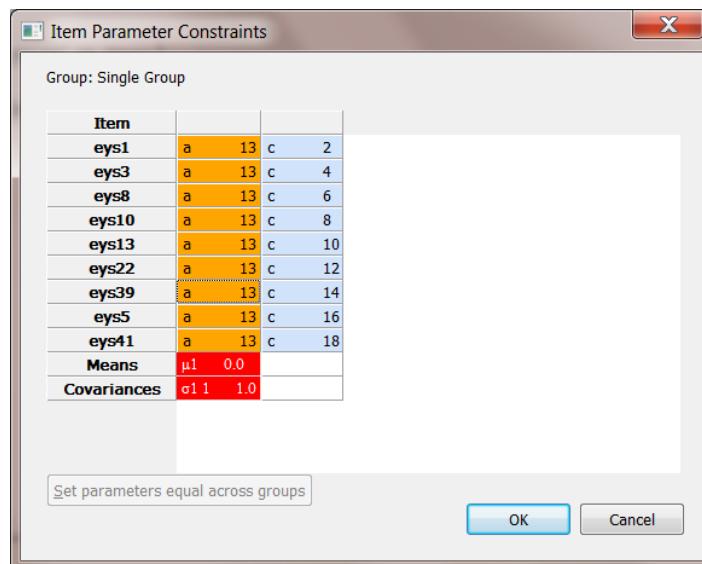
Select the same nine items that were chosen for the 2PL model, then click the **Models** tab, and recode the first seven items as discussed in the previous section:



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:  $a\theta + c$  or  $a(\theta - b)$**  ([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	eys3	9	0.83	0.10	2	1.26	0.19
3	eys8	9	0.83	0.10	3	-0.40	0.17
4	eys10	9	0.83	0.10	4	-1.68	0.20
5	eys13	9	0.83	0.10	5	0.93	0.18
6	eys22	9	0.83	0.10	6	0.21	0.16
7	eys39	9	0.83	0.10	7	0.61	0.17
8	eys5	9	0.83	0.10	8	-1.60	0.20
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	$\chi^2$	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  $-2 \log \text{likelihood}$  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 -2 log likelihood for the 1PL and 2PL models,  $1943.71 - 1918.13 = 25.58$ , which is distributed as  $\chi^2$  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.

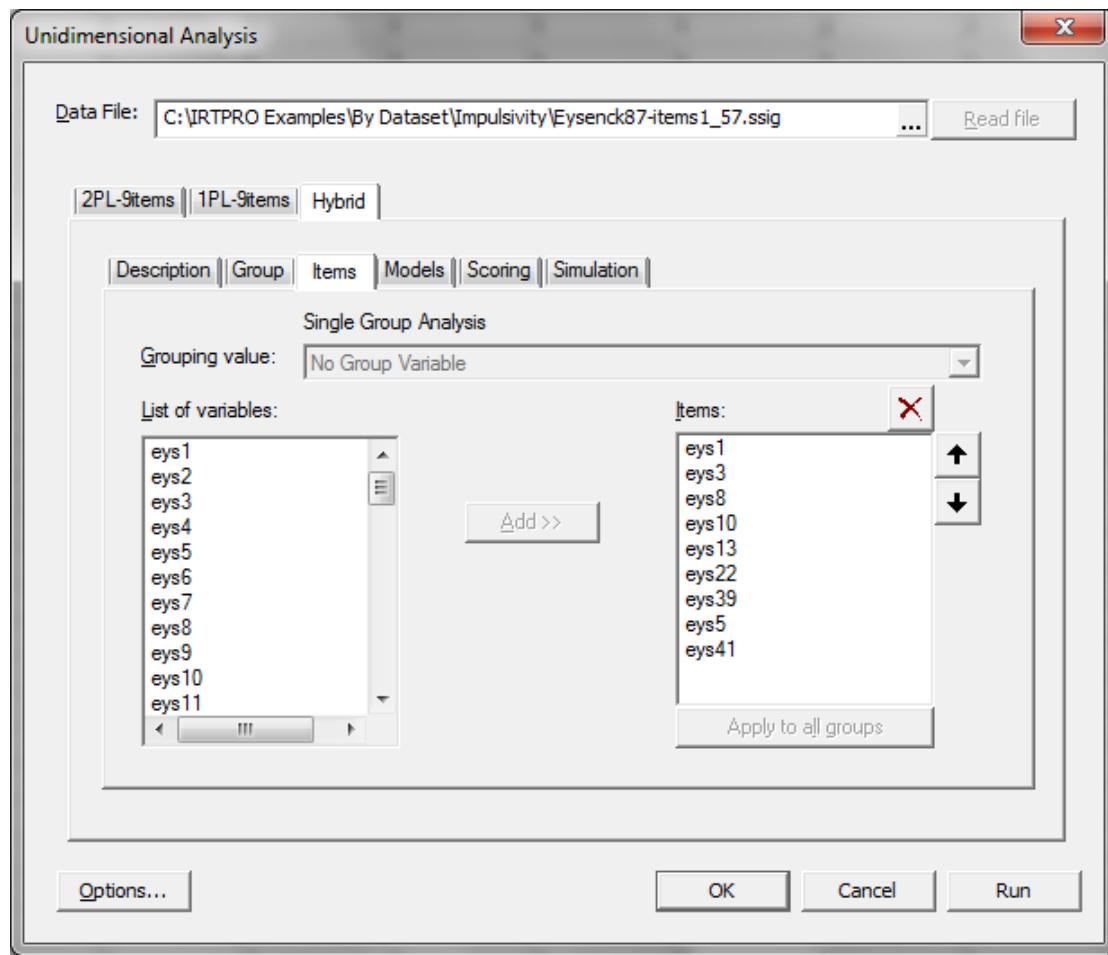
Under the hypothesis of perfect model fit, the  $S\text{-}X^2$  statistics are approximately distributed as  $\chi^2$  values with the tabulated degrees of freedom; significant values indicate lack of fit. Because a very strong model such as the 2PL rarely fits perfectly, one expects some (slightly) 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.

### 1.3 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.



Proceed to the **Models** tab and recode the item scores of the first seven items from (0,1) to (1,0). See Section 1.1 where the steps to do the recoding is discussed. Click the **Constraints...** button to obtain the following display.

Unidimensional Analysis

Data File: C:\IRT PRO Examples\By Dataset\Impulsivity\Eysenck87-items1\_57.ssig

2PL-9Items | 1PL-9Items | Hybrid |

Description | Group | Items | Models | Scoring | Simulation |

Grouping value: No Group Variable

	Item List	Categories	Data Codes	Item Scores	Mo
	eys1	2	1, 2	1, 0	2PL
	eys3	2	1, 2	1, 0	2PL
	eys8	2	1, 2	1, 0	2PL
	eys10	2	1, 2	1, 0	2PL
	eys13	2	1, 2	1, 0	2PL
	eys22	2	1, 2	1, 0	2PL
	eys39	2	1, 2	1, 0	2PL
	eys5	2	1, 2	0, 1	2PL

Constraints... | DIF... | Options... | OK |

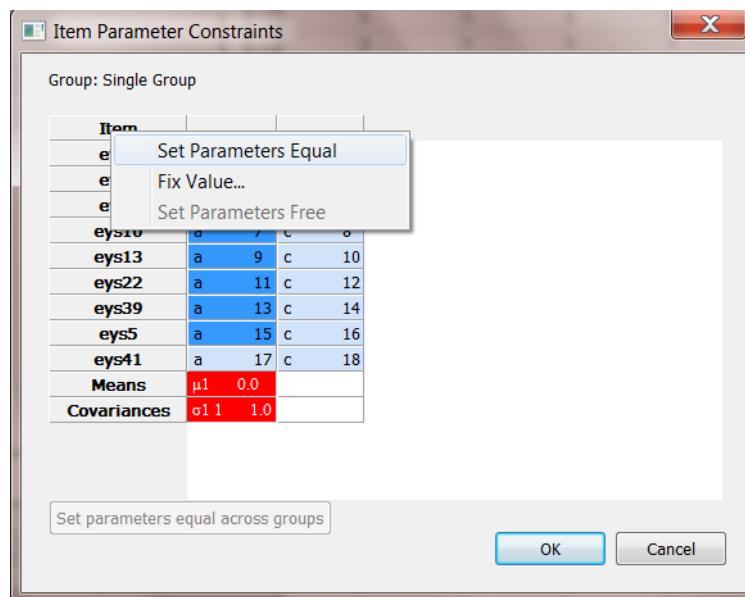
Item Parameter Constraints

Group: Single Group

Item	a	c
eys1	1	2
eys3	3	4
eys8	5	6
eys10	7	8
eys13	9	10
eys22	11	12
eys39	13	14
eys5	15	16
eys41	17	18
Means	$\mu_1$ 0.0	
Covariances	$\sigma_{11}$ 1.0	

Set parameters equal across groups |

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.

Group: Single Group				
Item				
eys1	a	13	c	2
eys3	a	13	c	4
eys8	a	13	c	6
eys10	a	13	c	8
eys13	a	13	c	10
eys22	a	13	c	12
eys39	a	13	c	14
eys5	a	13	c	16
eys41	a	17	c	18
Means	$\mu_1$	0.0		
Covariances	$\sigma_{11}$	1.0		

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: $a\theta + c$ or $a(\theta - b)$ ([Back to TOC](#))

Item	Label	a	s.e.	c	s.e.	b	s.e.
1	eys1	<sup>9</sup> 0.98	0.12	<sup>1</sup> 1.87	0.22	-1.92	0.29
2	eys3	<sup>9</sup> 0.98	0.12	<sup>2</sup> 1.32	0.19	-1.35	0.23
3	eys8	<sup>9</sup> 0.98	0.12	<sup>3</sup> -0.42	0.17	0.43	0.18
4	eys10	<sup>9</sup> 0.98	0.12	<sup>4</sup> -1.75	0.22	1.79	0.27
5	eys13	<sup>9</sup> 0.98	0.12	<sup>5</sup> 0.97	0.18	-1.00	0.21
6	eys22	<sup>9</sup> 0.98	0.12	<sup>6</sup> 0.22	0.17	-0.22	0.17
7	eys39	<sup>9</sup> 0.98	0.12	<sup>7</sup> 0.64	0.18	-0.65	0.19
8	eys5	<sup>9</sup> 0.98	0.12	<sup>8</sup> -1.67	0.21	1.71	0.27
9	eys41	<sup>11</sup> 0.11	0.20	<sup>10</sup> 0.55	0.15	-5.26	10.03

## Summed-Score Based Item Diagnostic Tables and X<sup>2</sup>s for Group 1 [\(Back to TOC\)](#)

### S-X<sup>2</sup> Item Level Diagnostic Statistics

Item	Label	$\chi^2$	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<sup>2</sup> 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 ( $\chi^2 = 1943.71 - 1931.41 = 12.3$ , 1 d.f.,  $p = 0.0005$ ), but it does not fit significantly worse than the 2PL model ( $\chi^2 = 1931.41 - 1918.13 = 12.3$ , 7 d.f.,  $p = 0.066$ ). The values of the M<sub>2</sub> 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<sup>2</sup> 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<sup>2</sup>. The summed-score by item response table for item 7 (eys39), has a "significant" S-X<sup>2</sup> 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** [\(Back\)](#)

Score	Category 0		Category 1		
	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	2.1	2.8	26	24.9
7	0	0.6	11	10.4	
8	0	0.1	3	2.9	

The gray numbers in the S-X2 tables are observed and expected frequencies in cells that have expected frequencies less than 1.0 (or whatever the minimum value is set to in **Advance Options**). The gray numbers have been "collapsed" into adjacent red or blue numbers above or below them. For example, in the table above, the gray values 0, 0, and 1 total to the red "1" in the Score 6, Category 0 cell, and the gray numbers 0.1, 0.6 and 2.1 add up to the red 2.8 that is expected in that "collapsed" cell.

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. The bright red and bright blue values are larger in absolute value than 3.0; the less bright red and blue numbers have absolute values less than 3.0. The value of 3.0 is not particularly meaningful; it merely serves as a minimal arbitrary cut-off to make the darker values less salient, because they almost certainly do not represent any important LD. There can be black numbers in those tables, if the nominal model is fitted and the item's responses are not properly ordered.

**Item 7 S-X<sub>2</sub>(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
2		9		8.5	6
3		22		17.0	14
4		10		16.8	34
5		17		11.1	20
6		6		7.0	25
7	1	1	2.0	2.3	11
8	0	0.3			10.0
			3		2.7

Chen & Thissen (1997) proposed the LD  $\chi^2$  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  $\chi^2$  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  $\chi^2$  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.

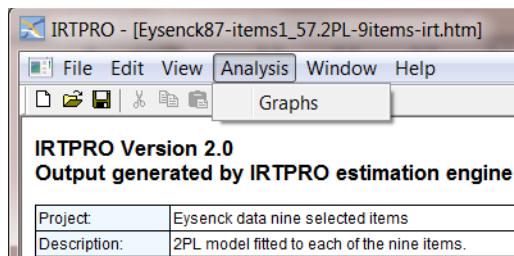
Because the standardized LD  $\chi^2$  statistic is only approximately standardized, and is known to be based on a statistics with a long-tailed ( $\chi^2$ ) distribution, we do not consider values larger than 2 or 3 to be large. Rather, we consider values larger than 10 large, indicating likely LD; 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. All of the values are relatively small, indicating no evidence of LD, and suggesting that the model fits satisfactorily.

**Marginal fit ( $\chi^2$ ) and Standardized LD  $\chi^2$  Statistics for Group 1** [\(Back to TOC\)](#)

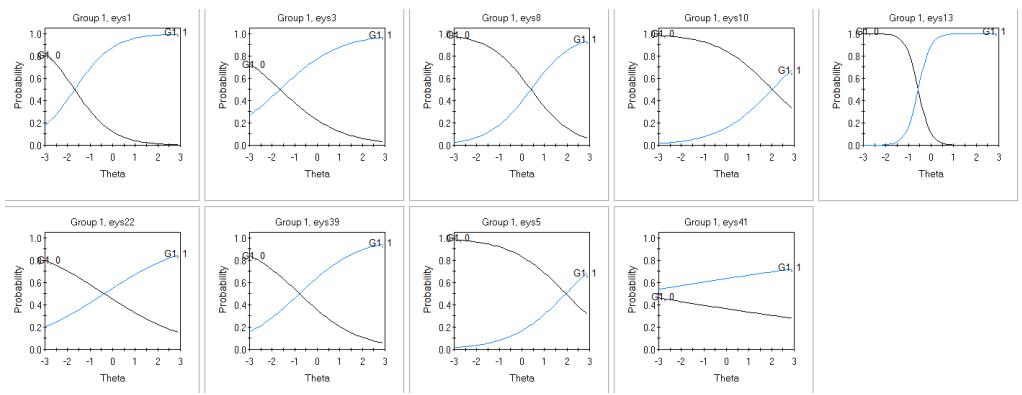
Marginal										
Item	Label	$\chi^2$	1	2	3	4	5	6	7	8
1	eys1	0.0								
2	eys3	0.0	-0.6							
3	eys8	0.0	-0.7	-0.4						
4	eys10	0.0	-0.2	1.2	-0.5					
5	eys13	0.0	4.6	-0.7	2.6	-0.6				
6	eys22	0.0	-0.5	2.7	-0.1	-0.2	-0.7			
7	eys39	0.0	-0.1	-0.7	0.6	-0.4	-0.1	-0.2		
8	eys5	0.0	6.6	-0.6	0.5	0.1	-0.0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

#### 1.4 Trace Lines and information curves for the 2PL model

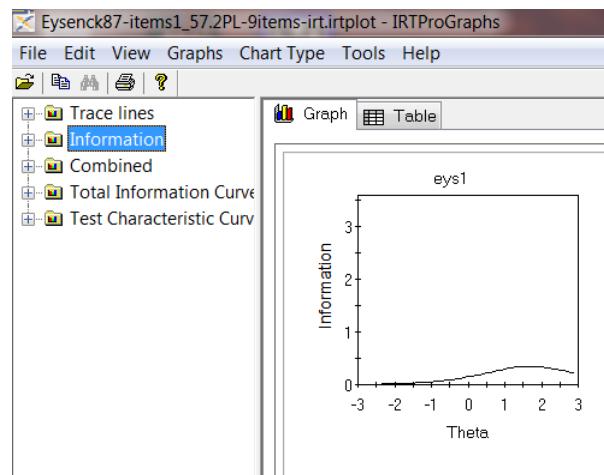
To obtain graphical representations of the trace lines and information curves for the nine impulsivity subscale items, open the output file containing the results of the 2PL analysis ([Eysenck87-items1\\_57.2PL-9items.htm](#)) and select **Analysis, Graphs** as shown below.



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.

