IRTFIT: A Macro for Item Fit and Local Dependence Tests under IRT Models

Jakob B. Bjorner MD, PhD[1], Kevin J. Smith, PhD[1], Clement Stone PhD[2] & Xiaowu Sun PhD[1]

[1] QualityMetric Incorporated

[2] School of Education, University of Pittsburgh

August 2007

Document Version 1.0 8/24/07

Table of Contents

Overview

This SAS macro will produce a variety of indices for testing the fit of item response theory (IRT) models to dichotomous and polytomous item response data. The macro does not perform estimation of item parameters, but requires that the item parameters have been estimated in IRT model software programs (e.g., MULTILOG, PARSCALE, BILOG, OPLM, WINSTEPS).

INPUT to the macro is a SAS data set containing the item parameter estimates and a SAS data set with the item response patterns.

OUTPUT from the macro is a variety of fit statistics including extensions of the $S\text{-}X^2$ and the $S\text{-}G^2$ tests [1;2] for polytomous items and the X^2* and G^2* statistics [3;4] as well as tests for local dependencies between pairs of items. To visualize misfit, the program provides observed-expected plots of fit. Test statistics are output to the screen and as SAS data sets. These output data sets can be exported as EXCEL files. Plots are shown on the screen and can be exported in various graphics formats.

MODELS handled by the program includes the dichotomous models: the 3-parameter model, the 2-parameter model, and the 1-parameter model. For polytomous items, the program handles the graded response mode, the generalized partial credit model, the generalized rating scale model, the partial credit model, the rating scale model, and the nominal categories model.

For longer scales, the calculation of some of these statistics becomes computationally intensive. For these types of scales, we recommend running the macro overnight or on weekends.

Funding Acknowledgment
Development of IRTFIT macro was supported by the National Cancer Institute (NCI) under the contract 263-MQ-515639

How to use this Manual

Read through the sections 'How to use compiled macros' and 'Syntax' to get a sense of how to use the macro. The 'Syntax' section will probably also be useful as a future reference.

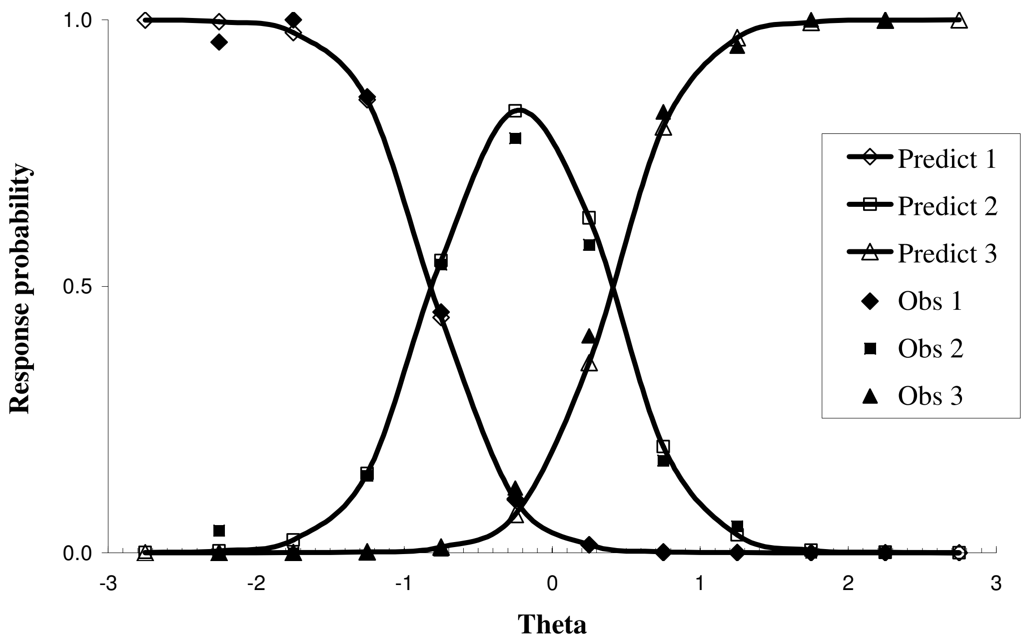
To get correct results, the item parameters need to be listed correctly in a SAS file. Read the section "Item parameters in the IRTFIT macro", that provides a detailed explanation and examples of how to identify the item parameters in output from various IRT model software programs (e.g., MULTILOG, PARSCALE, BILOG, OPLM, WINSTEPS).

Familiarity with these sections should enable you to get started, but you may want to go through one or more 'Macro Call Examples' found near then end of this manual.

The section 'Technical details' provides a background description of the fit tests provided, discusses missing data, and describes the SAS output in greater detail.

A Brief Introduction to the Fit Indices of IRTFIT

The approach to IRT fit testing taken by the fit indices in the IRTFIT macro [1-5] (and by many other IRT fit indices) is basically that of comparing expected and observed frequencies of item category responses for various levels of the IRT score (theta). The basic idea is illustrated in the figure below that shows the item response functions for an item with three response categories and the observed proportions of item responses (with the theta scale grouped into 10 categories). These differences are summarized in a test statistic.



The fundamental problem with this approach is that the theta values are latent and thus never directly observed. The traditional IRT fit statistics use estimated theta values that are treated as 'observed' data. This approach creates problems, in particular, for short scales (see e.g. [1;2;6;7]). The fit statistics used in the IRTFIT macro, address this problem in two different ways:

1. The S-X^2 and S-G^2 fit statistics [1;2] use the sum score of all items instead of theta and compare the predicted and observed response frequencies for each level of the scale sum score. Significance tests are calculated and plots can be produced comparing predicted and observed response probabilities for each level of the sum score.
2. The χ^{2*} and G^{2*} fit statistic [3-5] calculates 'pseudo-observed' score distributions for a limited number of discrete theta points to take the precision of theta estimation into account. Tests are performed for significant misfit and plots can be produced comparing predicted and 'pseudo-observed' response probabilities for each level of theta. Since the distributions of the χ^{2*} and G^{2*} have been shown to deviate from chi-square distributions, a Monte-Carlo re-sampling procedure has been recommended for hypothesis testing using the χ^{2*} and G^{2*}. This increases computing time.

Finally the IRTFIT macro can calculate tests to evaluate local item dependencies. These tests compare the observed two-way cross-tabulation table to the expected cross-tabulation table for each item pair. A chi-square test of significant differences is computed and differences between the predicted and observed polychoric correlation are calculated.

How to use the IRTFIT Macro

Installing and Using a Compiled SAS Macro

Step 1, copy the file sasmacr.sas7bdat to a folder, such as C:\IRTFITMACRO.
Step 2, copy the following 2 lines to a SAS Program window and submit them:

```
libname macro 'C:\IRTFITMACRO';  /* substitute with actual folder */
options mstored sasmstore=macro;
```

The macros can then be called like any other SAS program. The IRTFIT macro can be used as in following examples.

Syntax

The IRTFIT macro is organized as a number of statements, where some are required and some are optional (i.e. if you do not provide the particular statement, the program uses defaults).

The following statements illustrate a very simple way to call the macro:

```
%IRTFIT(DATA= datafile
,     PARFILE= parfile
,    ITEMLIST= it1-it12
,  TESTMETHOD= SUM)
```

Here is a short explanation of each statement:

DATA names the data file that contains the item responses as person level data. Each person is one row in the data file and the responses for each item is in one column (see example below).

Example of a SAS data file (items named IT1-IT12)

PERSON	IT1	IT2	IT3	IT4	IT5	IT6	IT7	IT8	IT9	IT10	IT11	IT12
1	1	2	2	2	2	3	3	3	1	1	3	3
2	2	2	2	2	2	3	3	3	3	3	3	3
3	2	2	2	2	2	3	3	3	3	3	3	3
4	2	2	2	2	2	2	3	3	3	3	2	3
5	1	1	1	2	1	1	2	2	2	3	2	2
6	2	2	1	2	2	2	3	3	2	2	2	3
7	…	…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…	…

PARFILE names a SAS data set containing the item parameters and number of response choices for each item. In the **PARFILE**, each record (line) is information for one item. The organization of the item parameter file is described further in the next section (page 9) and in detail at page 14 onwards, but an example of a PARFILE is given below.

Example of item parameter file

NAME	CHOICES	MODEL	SLOPE	THRESHOLD1	THRESHOLD2
IT1	2	DI2PL	2.76	-0.65	
IT2	2	DI2PL	3.05	-0.70	
IT3	2	DI2PL	1.52	-1.34	
IT4	2	DI2PL	2.76	-0.99	
IT5	2	DI2PL	2.22	-1.00	
IT6	3	GPCM	3.76	-0.81	0.39
IT7	3	GPCM	3.20	-1.67	-0.72
IT8	3	GPCM	2.41	-2.09	-0.99
IT9	3	GPCM	2.38	-1.24	-0.37
IT10	3	GPCM	2.12	-1.45	-0.66
IT11	3	GPCM	2.94	-1.18	-0.35
IT12	3	GPCM	1.35	-2.46	-2.48

`ITEMLIST` lists the names of the items to be evaluated. Items having no observations in `DATA` are automatically deleted from the list. The names should match exactly those in the `PARFILE;` otherwise the program will abort. However, lower or upper case may be used as the macro is not sensitive to case (all item names will be converted to upper case).

`TESTMETHOD` the method to test item fit. There are three valid options: SUM, THETA, and NONE. If SUM is chosen, the IRTFIT macro will evaluate fit based on the approach of Orlando and Thissen [1;2] and produce the $S\text{-}X^2$ and $S\text{-}G^2$ statistics. If THETA is chosen, the approach by Stone [3-5] is used to produce the X^2* and G^2* statistics. If NONE (the default) is chosen, item fit testing will not be performed.

The macro call illustrated above will run the IRTFIT macro using the program defaults for the statements that are not mentioned. The table and the call below illustrate all statements and macro options.

Overview of the statements of the IRTFIT macro

Required statements			
	DATA=		
	PARFILE=		
	ITEMLIST=		
Optional Statements			
	Statement	**Default**	**Other options**
Analysis type	TESTMETHOD=	NONE	SUM THETA
	LD_TEST=	NO	YES
Output	OUTCORE=	_IRT	[any SAS data set name]
	OUTLIB=	NONE	[any path e.g. c:\data]
	OUTPUT=	LIMITED	HUGE
	OUTFMT=	WIN	PDF RTF
	GRAPH=	NO	YES
	GRAPH_DATA=	NONE	[any SAS data set name]
Interpretation of data and item parameters	MODEL=	READ	DI1P, DI2P, DI3P GRM PCM, GPCM RSM, GRSM NOM
	D=	1	1.7
	P_MEAN=	0	[any number]
	P_STD=	1	[any number > 0]
	MINCODE=	1	0 [or any other number]
	MAXCHOICE=	7	[any number > 0]
	MISSING=	.	[any number or code]
Analysis options for TESTMETHOD= SUM	LOWRNG=	-5	[any number]
	HIGHRNG=	5	[any number]
	INTV_SUM=	0.1	[any number > 0]
	SCALE=	SUBTOTAL	TOTAL
	MIN_PRE=	5	[any number > 0]
Analysis options for TESTMETHOD= THETA	LOWRNG=	-5	[any number]
	HIGHRNG=	5	[any number]
	INTV_THETA=	0.5	[any number > 0]
	CRITCUT=	0.02	[any number > 0]
	FITBOUND=	2.02	[any number > 0]
	SAMPLETEST=	NO	YES
	N_REP=	100	[any whole number > 0]
	SEED=	-1	[any number]

The following statements illustrate a macro call using all options:

```
%IRTFIT(DATA= DATAFILE
,     PARFILE= PARFILE
,    ITEMLIST= IT1-IT12
,  TESTMETHOD= NONE
,     LD_TEST= YES
,     OUTCORE=_IRT
,      OUTLIB=NONE
,      OUTPUT=LIMITED
,      OUTFMT=WIN
,       GRAPH=NO
,  GRAPH_DATA=NONE
,       MODEL= READ
,           D=1
,      P_MEAN=0
,       P_STD=1
,     MINCODE=1
,   MAXCHOICE=7
,     MISSING=.
,      LOWRNG=-5
,     HIGHRNG=5
,    INTV_SUM= 0.1
,       SCALE= SUBTOTAL
,     MIN_PRE=5
,  INTV_THETA=0.5
,     CRITCUT=0.02
,    FITBOUND=2.02
,  SAMPLETEST=YES
,       N_REP=100
,        SEED=-1
)
```

DATA contains the item responses as person level data (no aggregate scores). Each person is one row in the data file and the responses for each item are in one column. You can use any directory reference that you would use in a SAS ‘data=’ statement. The item responses should be coded so that all items are in the same direction and the same direction as used for fitting the item response model. The item response codes should go from 1 to the max level of choices for that item. If starting from a value other than 1, say 0, you must set that value with the **MINCODE** statement (see below).

PARFILE is a SAS dataset containing the item parameters and number of response choices for each item you can use any directory reference that you would use in a ‘data=’statement. In the **PARFILE** each record (line) is information for one item. The table below shows the variables that may be included in a `PARFILE` (these exact variable names should be used. Other variables will be ignored).

Valid PARFILE variables:

VARIABLE	**Explanation**
NAME	The name of the item. The item names provided as values for this variable should exactly match the item names in the data set with item responses. This is a required variable
CHOICES	Number of item response choices. This is a required variable
MODEL	The model for each item. Valid values are: DI1P 1 parameter (Rasch) model for dichotomous items DI2P 2 parameter model for dichotomous items DI3P 3 parameter model for dichotomous items GRM Graded Response Model PCM Partial Credit Model (including Generalized Partial Credit Model) GPCM Generalized Partial Credit Model RSM Rating Scale Model (including Generalized Rating Scale Model) GRSM Generalized Rating Scale Model NOM Nominal Categories Model If all items use the same model, this variable can be dropped and the MODEL type can be specified in the macro statements.
SLOPE	The slope / discrimination parameter. If no slope variable is provided, slope will be set to 1 for all items – except if the NOMinal categories model is specified, in which case item slopes will be read from variables named SLOPE1, SLOPE2 …
THRESHOLD1	First item threshold parameter.
THRESHOLD2	Second item threshold parameter (and so on, for an item with 6 response choices there should be 5 item threshold parameters. In case you have items with different number of response choices, just set the irrelevant item threshold parameters to missing (e.g. for an item with 3 response choices, set THRESHOLD3 and higher to missing))
GUESS	The item guessing (lower asymptote) parameter in dichotomous scored models. If no guessing variable is provided or if the model is specified as anything other than DI3P, the guessing/lower asymptote will be set to 0.
LOCATION	If no threshold parameters are provided, the IRTFIT macro will calculate the threshold parameters based on LOCATION and CATEGORY parameters.
CATEGORY1	First item category parameter. Used together with the location parameter to calculate THRESHOLD1.

VARIABLE	Explanation
CATEGORY2	Second item category parameter (and so on, for an item with 6 response choices there should be 5 item category parameters. In case you have items with different number of response choices, just set the irrelevant item category parameters to missing (e.g. for an item with 3 response choices, set CATEGORY3 and higher to missing))
SLOPE1	If a NOMinal categories model is specified, several item category slope parameters must be provided (one less than the number of item categories).
SLOPE2	Second item category slope parameter for NOMinal items (and so on, for an item with 6 response choices there should be 5 item category slope parameters. In case you have items with different number of response choices, just set the irrelevant item category slope parameters to missing (e.g. for an item with 3 response choices, set SLOPE3 and higher to missing))
INTERCEPT1	If a NOMinal categories model is specified, several item category intercept parameters must be provided (one less than the number of item categories).
INTERCEPT2	Second item category intercept parameter for NOMinal items (and so on, for an item with 6 response choices there should be 5 item category intercept parameters. In case you have items with different number of response choices, just set the irrelevant item category intercept parameters to missing (e.g. for an item with 3 response choices, set INTERCEPT3 and higher to missing))

`ITEMLIST` lists the names of the items to be evaluated. Items having no observations in `DATA` are automatically deleted from the list. The names should match exactly those in the `PARFILE;` otherwise the program will abort. However, lower or upper case may be used as the macro is not sensitive to case (all item names will be converted to upper case). As in standard SAS calls you may specify a range of items (e.g. IT1-IT12 will mean IT1, IT2, IT3, IT4, IT5, IT6, IT7, IT8, IT9, IT10, IT11, and IT12)

`TESTMETHOD` The method to test item fit. There are three valid options: `SUM, THETA`, and `NONE`. If `SUM` is chosen, the `IRTFIT` macro will evaluate fit based on the approach of Orlando and Thissen [1;2] produce the S-X^2 and S-G^2 statistics. If `THETA` is chosen, the approach by Stone [3-5] is used to produce the χ^{2*} and G^{2*} statistics. If `NONE` (the default) is chosen, fit testing will not be performed. This option can be used to plot results from earlier runs without having to repeat the analysis or if the researcher wants to test local dependence only.

`LD_TEST` If `LD_TEST = YES`, pairwise tests for test for local dependence will be performed. Default is `NO`.

The next statements control the output from the IRTFIT macro

`OUTCORE` is the core NAME for SAS output data sets. Two SAS datasets are created: `NAME_FREQ` contains predicted and observed proportion of item responses for each score level and `NAME_FIT` contains item fit statistics. Default is `_IRT` so the two file names are `_IRT_FREQ` and `_IRT_FIT`. In order to create legitimate permanent SAS data sets, the length of `OUTCORE` should not exceed 11 characters.

`OUTLIB` is the PATH name for permanent SAS output data sets. When provided, two permanent SAS datasets are created in PATH. Default is `NONE` so that no permanent SAS datasets are created. Instead, temporary data sets are created in SAS's WORK library and will be deleted when SAS is terminated.

`OUTFMT` specifies the output format for text output. Valid options are: `PDF, RTF`, and `WIN`. When `RTF` or `PDF` is specified, a corresponding `NAME.RTF` or `NAME.PDF` (based on the `OUTCORE` statement) is created in the PATH specified by the `OUTLIB` statement, and the output to SAS OUTPUT window is suppressed. The default option is `WIN`, in which case output is sent to the SAS OUTPUT window but not to any external file.

`OUTPUT` specifies the amount of output. Valid options are: `LIMITED`, and `HUGE`. When `LIMITED` (the default) is specified, only test results are written to the output file. If `HUGE` is specified, a table of expected and observed frequencies is written for each item.

`GRAPH` If `GRAPH=YES`, the IRTFIT macro outputs graphs of item response category functions for observed and expected probabilities. Default is `NO`.

`GRAPH_DATA` is used to specify an input data set containing results from a previous analysis that is to be plotted. If `TESTMETHOD= NONE` and `GRAPH=YES`, the IRTFIT macro outputs graphs of item response category functions based on results from such a previous analysis. The SAS file name of these results is specified with the `GRAPH_DATA` statement. Any legitimate SAS file name can be used (the standard name for this file is `_IRT_FREQ`).

The next statements control the way data and item parameters are interpreted by the IRTFIT macro

`MODEL` when the models of all items are same, the model's name can be provided with this option instead of having the variable of MODEL in the `PARFILE`. The default option is READ in which case the model for each item will be read from the `PARFILE`. If a `MODEL` variable is provided in the `PARFILE`, all non-missing values for this variable will be used, no matter what is specified in the `MODEL` statement..

`D` Some IRT programs scale the item parameters using a scaling parameter D=1.7 (see section below). In this case, specify `D=1.7`. Default is `D=1`.

`P_MEAN` To calculate fit statistics, the IRTFIT macro needs to assume a distribution for theta in the test population. The standard assumption is a normal distribution with mean zero and standard deviation 1. The `P_MEAN` statement can be used to specify another population mean. Default is `P_MEAN=0`. Specifying other population standard deviations are particularly relevant for Rasch type models.

`P_STD` is the population standard deviation (see above). Default is 1.

`MINCODE` is the minimum value for item responses. All items must be coded to have the same minimum value. Default is 1.

`MISSING` is the value for missing. This option can be used if you have other missing indicators than the standard SAS missing indicator Default is . (standard SAS missing indicator).

`MAXCHOICE` indicates the maximum number of response choices the macro will handle. Default is 7. If you have items with larger number of response categories, set MAXCHOICE to reflect the maximum number of response categories.

The next statements control technical aspects of the item fit tests using `TESTMETHOD= SUM` and the test for local dependence

`LOWRNG` The IRTFIT macro calculates the likelihood at a number of theta levels (often named quadrature points). The `LOWRNG` option is used to set the lower bound for the quadrature points. Default is -5.

`HIGHRNG` is the upper bound for the quadrature points. Default is 5.

`INTV_SUM` defines the distance between quadrature points for `TESTMETHOD=SUM` and test for local dependence. Default value is 0.1.

`SCALE` For `TESTMETHOD=SUM` the item response probabilities can be calculated for each level of a sum score of all items (`SCALE= TOTAL`) or for each level of a sum score of all the items except the item in question.(`SCALE=SUBTOTAL`). The default option is `SCALE=SUBTOTAL`.

`MIN_PRE` In the calculation of fit statistics for `TESTMETHOD=SUM`, cells are collapsed to avoid too small predicted cell counts. The `MIN_PRE` statement can be used to set the minimum predicted value, below which cells are collapsed. The default option is `MIN_PRE= 5`.

The next statements control technical aspects of the item fit tests using `TESTMETHOD= THETA`

`LOWRNG` The IRTFIT macro calculates the likelihood at a number of theta levels (often named quadrature points). The `LOWRNG` option is used to set the lower bound for the quadrature points. Default is -5.

HIGHRNG is the upper bound for the quadrature points. Default is 5.

INTV_THETA defines the distance between quadrature points for **TESTMETHOD=THETA**. Default value is 0.5.

CRITCUT In the calculation of fit statistics for **TESTMETHOD=THETA**, cells with expected cell frequencies below the **CRITCUT** value are excluded from the calculation of Chi-square item fit statistics. Default is 0.02.

FITBOUND defines boundaries (±**FITBOUND**) for theta range used to calculate Chi-square item fit statistics for **TESTMETHOD=THETA**. Default is 2.02.

SAMPLETEST specifies whether re-sampling procedures should be used to calculate P-values for **TESTMETHOD=THETA**. Default is **NO**, in which case only raw Chi-square values will be provided. We recommend that these raw Chi-square values are inspected before repeating the run with **SAMPLETEST=YES** (which will be more time consuming).

N_REP number of replications in re-sampling procedures for **TESTMETHOD=THETA** and **SAMPLETEST=YES**. Default is 100.

SEED specifies the value to initialize re-sampling procedures for **TESTMETHOD=THETA**. Default is -1, which uses current date and time as seed. A positive integer number is recommended to achieve reproducible results, but the seed value then needs to be changed for each new analysis.

Identifying and Compiling Item Parameters

To use the IRTFIT macro you need to specify the item parameters in a way that can be read by the program. Failure to do so will lead to wrong results – typically showing misfit for all items. We will first list the parameterization of dichotomous IRT models (which is fairly standard) and then describe parameterization of polytomous IRT models (which can be more of a problem because it can be done in numerous ways). We then discuss differences between the identification of general IRT models and Rasch type models. Finally, we provide some examples of how item parameter estimates can be found in output from standard IRT programs and input into the IRTFIT macro.

Dichotomous IRT Models

Dichotomous IRT models are relevant for items that are scored in two categories (e.g. right (1)/wrong(0) or yes(1)/no(0)). The IRTFIT macro handles the logistic IRT models. The most elaborate model for dichotomous items is the **3-parameter logistic model**, which can be written the following way (see also Figure 1, item 1):

$$P\left(X_{ij}=1\right)=c_i+(1-c_i)\frac{\exp\left\{Da_i\left(\theta_j-b_i\right)\right\}}{1+\exp\left\{Da_i\left(\theta_j-b_i\right)\right\}} \qquad (1)$$

Where:

X_{ij} is the response from person *j* to item *i*. Here, we assume that the two responses are scored as 0 and 1, which is standard in educational research. Alternative, the responses can be coded as e.g. 1 and 2 (often used in survey research). In the IRTFIT macro, the code is decided by the values of the MINCODE macro value (all items must be coded to have the same minimum code). The default MINCODE value is 1.

θ_j is the IRT score for person *j.*

a_i is the slope parameter for item *i*. This is also sometimes called the item discrimination parameter. In the item parameter file that is read by the IRTFIT macro, the slope parameter has the variable name SLOPE.

b_i is the item threshold parameter for item *i*. This is also referred to as the item difficulty parameter. In the item parameter file that is read by the IRTFIT macro, the threshold parameter has the variable name THRESHOLD1.

c_i is the lower asymptote (guessing) parameter for item *i*. In the item parameter file that is read by the IRTFIT macro, the lower asymptote parameter has the variable name ASYMPTOTE.

D is a scaling constant for a_i. Some IRT program typically use D=1.7 since that scaling constant makes the item parameters from logistic IRT model very similar to the item parameters that would be obtained in normal-ogive IRT models. If instead *d* is set to 1 (which is the default in the IRTFIT macro), a slightly simplified version of the model is achieved:

$$P\left(X_{ij}=1\right)=c+(1-c)\frac{\exp\left\{a_i\left(\theta_j-b_i\right)\right\}}{1+\exp\left\{a_i\left(\theta_j-b_i\right)\right\}} \qquad (2)$$

The **2-parameter logistic IRT model (2-PL)** is obtained by setting c_i=0. This is the default setting that the IRTFIT macro will use if no ASYMPTOTE variable is found in the item parameter file. In this case the dichotomous IRT model simplifies to the following equation:

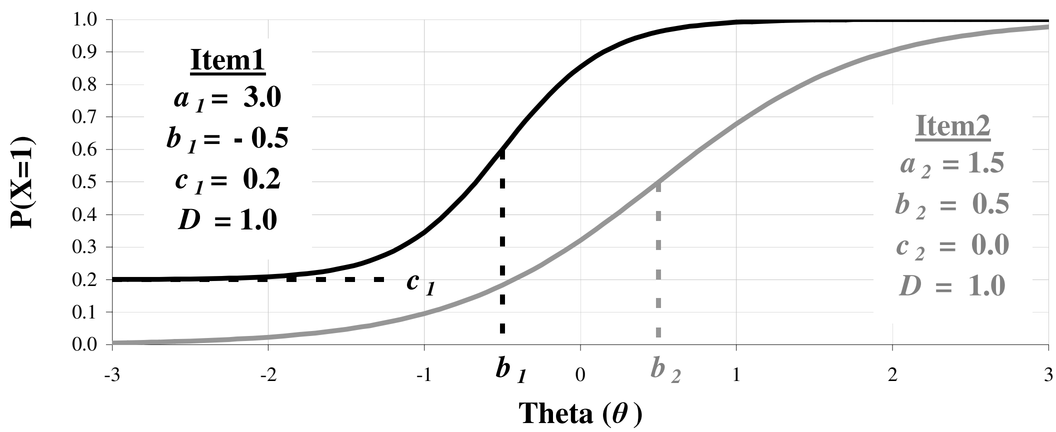
$$P\left(X_{ij}=1\right)=\frac{\exp\left\{Da_i\left(\theta_j-b_i\right)\right\}}{1+\exp\left\{Da_i\left(\theta_j-b_i\right)\right\}} \tag{3}$$

or alternatively:

$$\log\left(\frac{P\left(X_{ij}=1\right)}{P\left(X_{ij}=0\right)}\right)=Da_i\left(\theta_j-b_i\right) \tag{4}$$

This way of formulating IRT models will be useful when we discuss models for polytomous items.

Figure 1. Item response functions for two dichotomous items.



The **1-parameter logistic IRT model (1-PL)** can be achieved by setting a_i to the same value for all items or by setting a_i to 1. We will discuss these possibilities in more details in the end of this chapter. The specification a_i =1 is the default setting that the IRTFIT macro will use if no SLOPE variable is found in the item parameter file.

In the examples section at the end of this chapter, we show examples of how output from various programs corresponds to the values of the item parameter file for the IRTFIT macro.

Polytomous IRT Models

Similar to the differences between 1- and 2-parameter dichotomous IRT model, some polytomous IRT models do not include a slope parameter (the Partial Credit Model (PCM) and the Rating Scale Model (RSM)) while other models specify one or more slope parameters per item (e.g. the graded response model (GRM), the Generalized Partial Credit Model (GPCM), the Generalized Rating Scale Model (GRSM), and the Nominal Categories Model (NOM)). Further, the models differ in how the comparisons of categories are handled and on the restriction put on the *b* parameters. In the equations and examples below, we assume that items are coded 0, 1, 2 … m_i, where m_i is the maximum score for that item. As for the dichotomous model, the IRTFIT macro can also handle the situation where all items are coded 1, 2, 3 … (by setting set the MINCODE macro variable to 1 – the default value).

The Graded Response Model (GRM)

As illustrated in figure 2, the GRM compares the probability of scoring in category *k* or higher, with the probability of scoring lower than *k*. Thus, the simplest expression of the GRM (for *k*>0) is:

$$\log\left(\frac{P\left(X_{ij} \geq k\right)}{P\left(X_{ij} < k\right)}\right) = Da_i\left(\theta_j - b_{ik}\right) \qquad (5)$$

Where:

- θ_j is the IRT score for person *j*.
- a_i is the slope parameter for item *I* (variable name SLOPE in the item parameter file). The GRM model uses one slope parameter per item.
- b_{ik} is the item category threshold parameter for item *i*. An item with four item categories has three item category threshold parameters. In the item parameter file that is read by the IRTFIT macro, the threshold parameters have the variable names THRESHOLD1-THRESHOLD3. If the items were coded beginning with 1 (i.e. 1, 2, 3, 4) THRESHOLD1 would be the threshold between categories 1 and 2; THRESHOLD2 would be the threshold between categories 2 and 3; and THRESHOLD3 would be the threshold between categories 3 and 4.
- D is a scaling constant for a_i (either D = 1 or D = 1.7).

Figure 2. Comparison of item categories in the GRM

Category
0
1
2
3

➔

<table>
<tr><th colspan="3">GRM</th></tr>
<tr><td>0</td><td rowspan="2">0-1</td><td rowspan="3">0-2</td></tr>
<tr><td rowspan="3">1-3</td></tr>
<tr><td rowspan="2">2-3</td></tr>
<tr><td>3</td></tr>
</table>

The probability of scoring in category *k* or above is then

$$P\left(X_{ij} \geq k\right) = \frac{\exp\left\{Da_i\left(\theta_j - b_{ik}\right)\right\}}{1 + \exp\left\{Da_i\left(\theta_j - b_{ik}\right)\right\}} \qquad (6)$$

The probability of scoring in category k is:

$$P(X_{ij}=k)=P(X_{ij}\geq k)-P(X_{ij}\geq k+1)$$

$$=\frac{\exp\{Da_i(\theta_j-b_{ik})\}}{1+\exp\{Da_i(\theta_j-b_{ik})\}}-\frac{\exp\{Da_i(\theta_j-b_{ik+1})\}}{1+\exp\{Da_i(\theta_j-b_{ik+1})\}} \qquad (7)$$

Figure 3 and 4 illustrate the item category response functions in the cumulative model (equation 6) and in the standard representation of the model (equation 7). Note that the item category threshold parameter can be identified visually from the cumulative model (Figure 3) but not from the standard model (Figure 4).

Figure 3. Cumulative item category response functions in the GRM

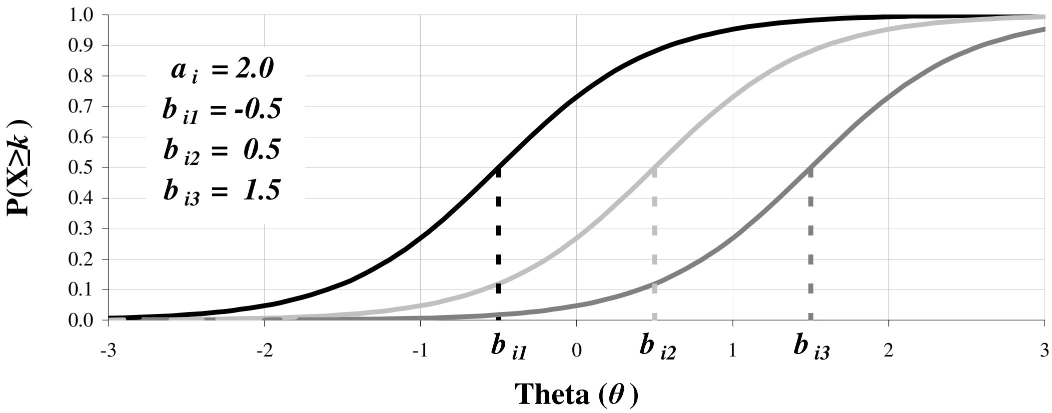
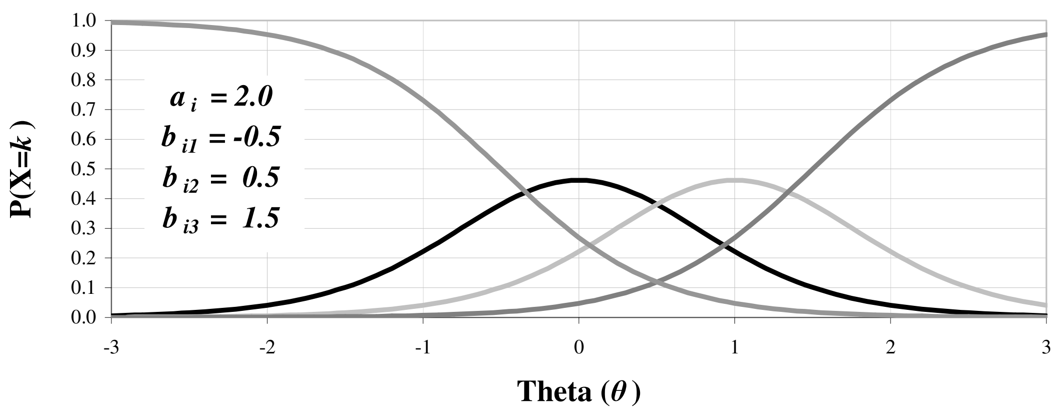


Figure 4. Item category response functions in the GRM



The GRM can be written in a slightly different way, by replacing b_{ik} with $l_i - d_{ik}$:

$$\left(\frac{P(X_{ij}\geq k)}{P(X_{ij}<k)}\right)=Da_i(\theta_j-(l_i-d_{ik})) \qquad (8)$$

Where:

l_i is named the location parameter for item *i* and is the mean of the item category threshold parameters for item *i*. In the item parameter file that is read by the IRTFIT macro, the location parameter has the variable name LOCATION.

d_{ik} is named the item category parameter for item *i* and is the difference between the location parameter and the item category threshold parameter ($d_{ik} = l_i - b_{ik}$). The sum of the item category parameters is normally set to zero. In this case, the location parameter is simply the mean of the item category threshold parameters (b_{ik}). In the item parameter file that is read by the IRTFIT macro, the item category parameters have the variable names CATEGORY1-CATEGORY3. Note that if the items were coded e.g. 1, 2, 3, 4, the item parameter file for the IRTFIT macro would still have the category parameter variable names CATEGORY1-CATEGORY3. The user can choose between providing the item category threshold parameters or providing the location and item category parameters.

The probability of scoring in category *k* is then:

$$P(X_{ij} = k) = \frac{\exp\{Da_i(\theta_j - (l_i - d_{ik}))\}}{1 + \exp\{Da_i(\theta_j - (l_i - d_{ik}))\}} - \frac{\exp\{Da_i(\theta_j - (l_i - d_{ik+1}))\}}{1 + \exp\{Da_i(\theta_j - (l_i - d_{ik+1}))\}} \qquad (9)$$

In the example in figure 3 and 4, where the item category threshold parameters were -0.5, 0.5, and 1.5, the location parameter would be 0.5, and the item category parameters would be -1, 0, and 1.

The Generalized Partial Credit Model (GPCM) and the Partial Credit Model (PCM)

As illustrated in figure 5, the GPCM compares the probability of scoring in category *k* with the probability of scoring in the category below. Thus, the simplest expression of the GPCM (for *k*>0) is:

$$\log\left(\frac{P(X_{ij} = k)}{P(X_{ij} = k - 1)}\right) = Da_i(\theta_j - b_{ik}) \qquad (10)$$

Where:

θ_j is the IRT score for person *j*.

a_i is the slope parameter for item *i* (variable name SLOPE in the item parameter file). The GRM model uses one slope parameter per item.

b_{ik} is the item category threshold parameter for item *i*. An item with four item categories has three item category threshold parameters. In the item parameter file that is read by the IRTFIT macro, the threshold parameters have the variable names THRESHOLD1-THRESHOLD3. Note that if the items were coded e.g. 1, 2, 3, 4, the item parameter file for the IRTFIT macro would still have the threshold parameter variable names THRESHOLD1-THRESHOLD3.

D is a scaling constant for a_i (either D = 1 or D = 1.7).

Figure 5. Comparison of item categories in the GPCM, PCM, GRSM, and RSM

Category
0
1
2
3

➡

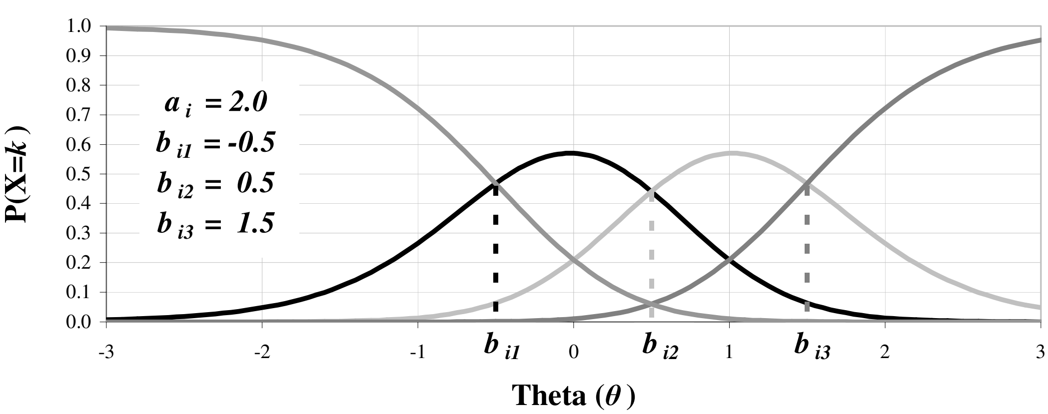
GPCM, PCM, GRSM, RSM		
0		
1	1	
	2	2
		3

The probability of scoring in category k is:

$$P\left(X_{ij}=k\right)=\frac{\exp\sum_{c=0}^{k}\left\{Da_i\left(\theta_j-b_{ic}\right)\right\}}{\sum_{h=0}^{m_i}\exp\sum_{c=0}^{h}\left\{Da_i\left(\theta_j-b_{ic}\right)\right\}} \quad , \quad \sum_{c=0}^{0}\left\{Da_i\left(\theta_j-b_{ic}\right)\right\}\equiv 0 \tag{11}$$

Figure 6 shows plots of the item category response functions. Although the item parameter have the same numeric values as for the GRM model example, the plots are slightly different. In the GPCM model, the item category threshold parameters can be read directly from the plots of item category response functions.

Figure 6. Item category response functions in the GPCM



If the item slope parameter is assumed to the equal for all items (and set to one), a simpler model is achieved, a Rasch family model named the Partial Credit Model (PCM):

$$P\left(X_{ij}=k\right)=\frac{\exp\sum_{c=0}^{k}\left\{\theta_j-b_{ic}\right\}}{\sum_{h=0}^{m_i}\exp\sum_{c=0}^{h}\left\{\theta_j-b_{ic}\right\}} \quad , \quad \sum_{c=0}^{0}\left\{\theta_j-b_{ic}\right\}\equiv 0 \tag{12}$$

This model is the default option if no slope parameter is found in the item parameter file and if D is specified as 1 (also the default option). The parameterization of Rasch family models are futher discussed in a section below.

The GPCM (or the PCM) can be written in a slightly different way, by replacing b_{ik} with $l_i - d_{ik}$:

$$\log\left(\frac{P(X_{ij}=k)}{P(X_{ij}=k-1)}\right)=Da_i\left(\theta_j-(l_i-d_{ik})\right) \tag{13}$$

Where:

l_i is named the location parameter for item i and is the mean of the item category threshold parameters for item i. In the item parameter file that is read by the IRTFIT macro, the location parameter has the variable name LOCATION.

d_{ik} is named the item category parameter for item i and is the difference between the location parameter and the item category threshold parameter ($d_{ik} = l_i - b_{ik}$). The sum of the item category parameters is normally set to zero. In this case, the location parameter is simply the mean of the item category threshold parameters (b_{ik}). In the item parameter file that is read by the IRTFIT macro, the item category parameters have the variable names CATEGORY1-CATEGORY3. Note that if the items were coded e.g. 1, 2, 3, 4, the item parameter file for the IRTFIT macro would still have the category parameter variable names CATEGORY1-CATEGORY3. The user can choose between providing the item category threshold parameters or providing the location and item category parameters.

The probability of scoring in category k is then:

$$P(X_{ij}=k)=\frac{\exp\sum_{c=0}^{k}\left\{Da_i\left(\theta_j-(l_i-d_{ic})\right)\right\}}{\sum_{h=0}^{m_i}\exp\sum_{c=0}^{h}\left\{Da_i\left(\theta_j-(l_i-d_{ic})\right)\right\}} \tag{14}$$

In the example in Figure 6, where the item category threshold parameters were -0.5, 0.5, and 1.5, the location parameter would be 0.5, and the item category parameters would be -1, 0, and 1.

The Generalized Rating Scale Model (GRSM) and the Rating Scale Model (RSM)

The GRSM and RSM are restricted versions of the GPCM (see above) where the item category parameters are assumed to be constant over items and are then termed the category parameters d_k. The item category threshold parameters can be calculated as $b_{ik} = l_i - d_k$. The GRSM model can then be written:

$$\log\left(\frac{P(X_{ij}=k)}{P(X_{ij}=k-1)}\right)=Da_i\left(\theta_j-(l_i-d_{k})\right) \tag{15}$$

Where

$\boldsymbol{d_k}$ is named the category parameter for the set of items. The category parameters can apply to all items in a scale or to a subset of items in the scale – often termed a block. In the item parameter file that is read by the IRTFIT macro, the category parameters have the variable names CATEGORY1-CATEGORY3 (same as for the item category parameters). Note that if the items were coded e.g. 1, 2, 3, 4, the item parameter file for the IRTFIT macro would still have the category parameter variable names CATEGORY1-CATEGORY3.

The probability of scoring in category k is then:

$$P\left(X_{ij}=k\right)=\frac{\exp\sum_{c=0}^{k}\left\{Da_i\left(\theta_j-\left(l_i-d_c\right)\right)\right\}}{\sum_{h=0}^{m_i}\exp\sum_{c=0}^{h}\left\{Da_i\left(\theta_j-\left(l_i-d_c\right)\right)\right\}} \qquad (16)$$

The Nominal Categories Model (NOM)

The NOM does not assume a rank order of the items categories. In the NOM, each item category is compared to the base item category (figure 7)

The NOM can be written:

$$\log\left(\frac{P\left(X_{ij}=k\right)}{P\left(X_{ij}=0\right)}\right)=Da_{ik}\theta_j-g_{ik} \qquad (17)$$

Where:

$\boldsymbol{\theta_j}$ is the IRT score for person j.

$\boldsymbol{a_{ik}}$ is the item category slope parameter for item i. In the IRTFIT macro, an item with four item categories has three item category slope parameters (for purposes of identification the slope parameter for the category 0 is fixed at zero). In the item parameter file that is read by the IRTFIT macro, the category slope parameters have the variable names SLOPE1-SLOPE3.

$\boldsymbol{g_{ik}}$ is the item category intercept parameter for item i. In the IRTFIT macro, an item with four item categories has three item category intercept parameters (for purposes of identification the threshold parameter for the category 0 is fixed at zero). In the item parameter file that is read by the IRTFIT macro, the intercept parameters have the variable names INTERCEPT1-INTERCEPT3. Note that if the items were coded e.g. 1, 2, 3, 4, the item parameter file for the IRTFIT macro would still have the intercept parameter variable names INTERCEPT1-INTERCEPT3.

$\boldsymbol{D}$ is a scaling constant for a_i (either D = 1 or D = 1.7). Note that this constant must be the same for all IRT models in the same analysis.

Figure 7. Comparison of item categories in the NOM

Category		NOM		
0		0	0	0
1	➡	1		
2			2	
3				3

Identification of IRT and Rasch-type Models

IRT models are typically identified by letting the population distribution of IRT scores, $\phi(\theta)$, be standard normal (i.e. setting the population to zero and the population standard deviation to one). In this case, all item parameters are freely estimated. In the Rasch family models (the dichotomous Rasch model, the Partial Credit model and the Rating Scale model), identification is typically achieved another way: by setting all slope parameters to one and setting sum of all threshold parameters to zero. If the model is identified this way, the mean and standard deviation of the IRT scores can be freely estimated. Rasch family models can be represented with either of these two sets of restrictions. Both representations can be used with the IRTFIT macro.

Item Parameter Examples

The 3-PL IRT Model for Dichotomous Items

BILOG Output

Figure 7 shows an example of the typical output from the BILOG program concerning an analysis of 5 dichotomous items, analyzed using the 3-PL model. The output files from BILOG that contains item parameter estimates has the extension .PH2. The relevant columns for the purpose of running an IRTFIT analyses are labeled: SLOPE, THRESHOLD, and ASYMPTOTE. The item parameter estimates in these columns should be read into SAS using the variable names: SLOPE, THRESHOLD1, and ASYMPTOTE. An example of the final SAS file is provided in Table 1.

```
…………………… ADDITIONAL OUTPUT DELETED ………………………………………

 ITEM   INTERCEPT   SLOPE    THRESHOLD  DISPERSN   ASYMPTOTE RT MEAN SQUARE STD
          S.E.       S.E.       S.E.      S.E.       S.E.    POSTERIOR RESIDUAL
 ------------------------------------------------------------------------------
 0001 |   1.822  |   2.596  |  -0.702  |   0.385  |   0.222  |       0.361
      |   0.068* |   0.189* |   0.045* |   0.030* |   0.016* |
      |          |          |          |          |          |
 0002 |   2.350  |   3.264  |  -0.720  |   0.306  |   0.199  |       0.863
      |   0.059* |   0.233* |   0.036* |   0.024* |   0.007* |
      |          |          |          |          |          |
 0003 |   2.201  |   1.481  |  -1.486  |   0.675  |   0.230  |       0.656
      |   0.212* |   0.169* |   0.238* |   0.087* |   0.147* |
      |          |          |          |          |          |
 0004 |   2.639  |   2.504  |  -1.054  |   0.399  |   0.178  |       1.941
      |   0.085* |   0.203* |   0.061* |   0.035* |   0.030* |
      |          |          |          |          |          |
 0005 |   2.381  |   2.318  |  -1.027  |   0.431  |   0.188  |       1.894
      |   0.091* |   0.190* |   0.068* |   0.039* |   0.035* |

…………………… ADDITIONAL OUTPUT DELETED ………………………………………
```

Figure 7. Bilog output (*.PH2) 3-PL model for dichotomous items.

Table 1. Example of SAS item parameter file based on the output from Bilog (Figure 7) or Parscale (Figure 8)

NAME	CHOICES	MODEL	SLOPE	THRESHOLD1	ASYMPTOTE
IT1	2	DI3P	2.596	-0.702	0.222
IT2	2	DI3P	3.264	-0.720	0.199
IT3	2	DI3P	1.481	-1.486	0.230
IT4	2	DI3P	2.504	-1.054	0.178
IT5	2	DI3P	2.318	-1.027	0.188
…	…	…	…	…	…

The NAME variable is a text variable containing the item name. This name must exactly match the variable name in the data file that contains the item responses.
The MODEL variable is a text variable describing the IRT model for the particular item. Dichotomous models are referred to as DI3P, DI2P, and DI1P (for the 3-PL, 2-PL, and 1-PL models respectively).
The SLOPE variable contains the item slope. The THRESHOLD1 variable contains the item thresholds.
The ASYMPTOTE variable contains the lower asymptote (guessing) parameter.
The variable CHOICES reports the number of response choices for the particular item.

PARSCALE Output

Figure 8 shows the output from the Parscale program providing the same item parameters. The relevant columns in this file are: SLOPE, LOCATION, and GUESSING. Parameter estimates from these columns should be read into a SAS file such as illustrated in Table 1. Thus the variable names would be SLOPE, THRESHOLD1, and ASYMPTOTE. Item name, model and number of response choices also need to be provided.

```
……………………… ADDITIONAL OUTPUT DELETED ………………………………………

+------+-----+---------+---------+---------+---------+---------+---------+
| ITEM |BLOCK|  SLOPE  |   S.E.  |LOCATION |   S.E.  |GUESSING |   S.E.  |
+======+=====+=========+=========+=========+=========+=========+=========+
| IT1  |   1 |   2.596 |   0.189 |  -0.702 |   0.045 |   0.222 |   0.016 |
| IT2  |   1 |   3.264 |   0.233 |  -0.720 |   0.036 |   0.199 |   0.007 |
| IT3  |   1 |   1.481 |   0.169 |  -1.486 |   0.238 |   0.230 |   0.147 |
| IT4  |   1 |   2.504 |   0.203 |  -1.054 |   0.061 |   0.178 |   0.030 |
| IT5  |   1 |   2.318 |   0.190 |  -1.027 |   0.068 |   0.188 |   0.035 |
+------+-----+---------+---------+---------+---------+---------+---------+

……………………… ADDITIONAL OUTPUT DELETED ………………………………………
```

Figure 8. Parscale output (*.PH2) 3-PL model for dichotomous items.

The 1-PL (Rasch) model for Dichotomous Items

WINSTEPS Output

Figure 9 presents typical output from an analysis of 5 dichotomous items in WINSTEPS using the Rasch (1-PL) model. Regarding the item parameters (top part of figure 9), the relevant column for the purpose of running an IRTFIT analyses is labeled: MEASURE. The item parameter estimates from this column should be saved in a SAS data set with the variable name: THRESHOLD1. Table 2 shows the equivalent SAS item parameter file. The Rasch model is identified by constraining the sum of the item parameters to zero and setting all slope parameters to 1. Thus, the slope parameter should be set to 1 in the SAS file (variable name SLOPE). Alternatively, SLOPE can be set to missing or no SLOPE variable provided (as in table 2). Since the person IRT score is not centered on 0 in this version of the model, the IRTFIT macro needs to be told the mean and standard deviation of the IRT scores. These can be read from the WINSTEPS output: Person STATISTICS. The mean is 1.291 and the standard deviation is 1.505. These need to be specified directly in the macro (macro statements P_MEAN and P_STD).

```
……………………… ADDITIONAL OUTPUT DELETED ………………………………………

Item  Winsteps example  Aug 16 22:39 2006
ENTRY MEASURE STTS COUNT    SCORE  ERROR IN.MSQ IN.ZSTD OUT.MS OUT.ZSTD
    1    .556  1    28.0     13.0   .087    .78  -1.30    .72  -1.32
    2    .392  1    28.0     14.0   .088    .93   -.37    .95   -.18
    3   -.482  1    28.0     16.0   .095   1.35   1.74   1.44   1.80
    4   -.303  1    28.0     16.0   .093   1.27   1.42   1.33   1.40
    5   -.164  1    28.0     19.0   .091    .63  -1.94    .52  -1.81

……………………… ADDITIONAL OUTPUT DELETED ………………………………………

         Person STATISTICS:  MEASURE ORDER
+-----------------------------------------------------------------------------
|ENTRY    RAW                   MODEL|   INFIT  |  OUTFIT  |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD| …
|------------------------------------+----------+----------+-
……………………… ADDITIONAL OUTPUT DELETED ………………………………………
|------------------------------------+----------+----------+-
| MEAN     3.9    5.0   1.291    .550|1.01    .1| .99    .1| …
| S.D.     1.6     .0   1.505    .401| .12    .5| .16    .5| …
+-----------------------------------------------------------

……………………… ADDITIONAL OUTPUT DELETED ………………………………………
```

Figure 9. WINSTEPS output. Rasch (1-PL) model for dichotomous items.

Table 2. Example of SAS item parameter file based on the output from WINSTEPS (Figure 9) or OPLM (Figure 10)

NAME	CHOICES	MODEL	THRESHOLD1
IT1	2	DI1P	0.556
IT2	2	DI1P	0.392
IT3	2	DI1P	-0.482
IT4	2	DI1P	-0.303
IT5	2	DI1P	-0.164
…	…	…	…

OPLM Output

Figure 10 shows output from the OPLM program, again analyzing 5 dichotomous items with the Rasch model. The threshold parameters can be found in column B. The model is identified in a similar ways as the WINSTEPS model. The population standard deviation must be calculated as the square root of the variance of estimated thetas.

```
……………………… ADDITIONAL OUTPUT DELETED ………………………………………

  nr label       A      B     SE(B)     S    DF      P        M       M2       M3
---------------------------------------------------------------------------------

   1 It1         1     .556   .087   5.293  3   .152  -1.445    -.861  -1.709
   2 It2         1     .392   .088  10.956  3   .012  -3.187  -2.605  -2.939
   3 It3         1   -.482   .095  33.055  3   .000   4.754   4.875   5.134
   4 It4         1   -.303   .093   1.540  3   .673   -.224  -1.120   -.494
   5 It5         1   -.164   .091    .623  3   .891   -.018    .572   -.127
---------------------------------------------------------------------------------

……………………… ADDITIONAL OUTPUT DELETED ………………………………………

 Number of observations =              1000
 Mean (theta's) =                     1.291
 Variance (estimated theta's) =       2.266
 Variance (true theta's) =             .447
 Reliability (of est. theta's) =       .197
 Correlation (est. th.,scores) =       .994
```

Figure 10. OPLM output. Rasch (1-PL) model for dichotomous items.

Another way to present the item parameters is to rescale so the mean of the IRT scores is zero and the standard deviation is one. This is achieved by the transformation
SLOPE* = SLOPE * 1.505 (use the population standard deviation) and
THRESHOLD1* = (THRESHOLD1 – 1.291)/1.505 (use the population mean and the standard deviation). The resulting item parameters are shown in table 3. Since the slope of all items is constrained to equality across items, the model is still a 1-PL model and (except for the transformation) completely equivalent to the model in table 2.

Table 3. Example of SAS item parameter file based on the output from WINSTEPS (Figure 9) or OPLM (Figure 10). Centered on the mean of the IRT scores for the sample.

NAME	CHOICES	MODEL	SLOPE	THRESHOLD1
IT1	2	DI1P	1.505	-0.488
IT2	2	DI1P	1.505	-0.597
IT3	2	DI1P	1.505	-1.178
IT4	2	DI1P	1.505	-1.059
IT5	2	DI1P	1.505	-0.967
…	…	…	…	…

```
……………………… ADDITIONAL OUTPUT DELETED ………………………………………

ITEM   6:       3 GRADED CATEGORIES
        P(#) ESTIMATE (S.E.)
 A         11     3.94   (0.19)
 B( 1)     12    -0.81   (0.03)
 B( 2)     13     0.40   (0.03)

……………………… ADDITIONAL OUTPUT DELETED ………………………………………

ITEM   7:       3 GRADED CATEGORIES
        P(#) ESTIMATE (S.E.)
 A         14     3.40   (0.19)
 B( 1)     15    -1.73   (0.05)
 B( 2)     16    -0.69   (0.03)

……………………… ADDITIONAL OUTPUT DELETED ………………………………………

ITEM   8:       3 GRADED CATEGORIES
        P(#) ESTIMATE (S.E.)
 A         17     2.58   (0.15)
 B( 1)     18    -2.22   (0.08)
 B( 2)     19    -0.93   (0.04)

……………………… ADDITIONAL OUTPUT DELETED ………………………………………
```

Figure 11. Multilog output. GRM for items with 3 rank-scaled response categories.

The Graded Response Model (GRM)

MULTILOG Output

Figure 11 shows excerpts of the output from the MULTILOG program concerning 3 items with 3 rank-scaled categories, analyzed (together with several other items) using the GRM. The parameter estimates needed for running an IRTFIT analyses are found in the rows labeled *A* (SLOPE parameter), *B(1)* (THRESHOLD1 parameter), and *B(2)* (THRESHOLD2 parameter). An example of the final SAS file is provided below.

Table 4. Example of SAS item parameter file based on the output from Multilog (Figure 9) – GRM

NAME	CHOICES	MODEL	SLOPE	THRESHOLD1	THRESHOLD2
IT6	3	GRM	3.94	-0.81	0.40
IT7	3	GRM	3.40	-1.73	-0.69
IT8	3	GRM	2.58	-2.22	-0.93
…	…	…	…	…	…

The NAME variable is a text variable containing the item name. This name must exactly match the variable name in the data file that contains the item responses.
The MODEL variable is a text variable describing the IRT model for the particular item. The SLOPE variable contains the item slope.

The THRESHOLD1 and THRESHOLD2 variables contains the item thresholds.
The variable CHOICES contains the number of response choices for the particular item.

PARSCALE Output

Figure 12 shows excerpts of the output from the PARSCALE program concerning 3 items with 3 rank-scaled categories, analyzed (together with several other items) using the GRM. The GRM in PARSCALE is the same as the GRM in MULTILOG, but the parameters are presented in a different way, using the location and item category parameters. The item category parameters (2 for each item) are read from the output shown in the top part of Figure 12, while the slope and location parameters are read from the output shown in the lower part of figure 12. These data should be saved in a SAS item parameter file using the variable names: SLOPE, LOCATION, CATEGORY1, CATEGORY2 (see table 5). In addition the item name, the model (here GRM) and the number of response choices must be specified. Alternatively, the two item thresholds can be calculated, in which case an item parameter file similar to the one shown in table 4 can be used.

```
.......................... ADDITIONAL OUTPUT DELETED ...........................................

   ITEM BLOCK   6  BLOCK

   CATEGORY PARAMETER  :     0.603    -0.603
   S.E.                :     0.031     0.027
   ITEM BLOCK   7  BLOCK

   CATEGORY PARAMETER  :     0.523    -0.523
   S.E.                :     0.052     0.032
   ITEM BLOCK   8  BLOCK

   CATEGORY PARAMETER  :     0.644    -0.644
   S.E.                :     0.079     0.040

.......................... ADDITIONAL OUTPUT DELETED ...........................................

+------+-----+---------+---------+---------+---------+---------+---------+
| ITEM |BLOCK|  SLOPE  |   S.E.  |LOCATION |   S.E.  |GUESSING |   S.E.  |
+======+=====+=========+=========+=========+=========+=========+=========+
| IT6  |   6 |   3.943 |   0.191 |  -0.206 |   0.021 |   0.000 |   0.000 |
+------+-----+---------+---------+---------+---------+---------+---------+
| IT7  |   7 |   3.404 |   0.190 |  -1.212 |   0.030 |   0.000 |   0.000 |
+------+-----+---------+---------+---------+---------+---------+---------+
| IT8  |   8 |   2.581 |   0.149 |  -1.577 |   0.042 |   0.000 |   0.000 |
+------+-----+---------+---------+---------+---------+---------+---------+

.......................... ADDITIONAL OUTPUT DELETED ...........................................
```

Figure 12. Parscale output. GRM for items with 3 rank-scaled response categories.

Table 5. Example of SAS item parameter file based on the output from Parscale (Figure 12)

NAME	CHOICES	MODEL	SLOPE	LOCATION	CATEGORY1	CATEGORY2
IT6	3	GRM	3.943	-0.206	0.603	-0.603
IT7	3	GRM	3.404	-1.212	0.523	-0.523
IT8	3	GRM	2.581	-1.577	0.644	-0.644
…	…	…	…	…	…	…

```
……………………… ADDITIONAL OUTPUT DELETED ………………………………………

   ITEM BLOCK   6  BLOCK

   SCORING FUNCTION    :     1.000     2.000     3.000
   STEP PARAMTER       :     0.000     0.596    -0.596
   S.E.                :     0.000     0.032     0.028
   ITEM BLOCK   7  BLOCK

   SCORING FUNCTION    :     1.000     2.000     3.000
   STEP PARAMTER       :     0.000     0.474    -0.474
   S.E.                :     0.000     0.057     0.034
   ITEM BLOCK   8  BLOCK

   SCORING FUNCTION    :     1.000     2.000     3.000
   STEP PARAMTER       :     0.000     0.552    -0.552
   S.E.                :     0.000     0.089     0.043

……………………… ADDITIONAL OUTPUT DELETED ………………………………………

+------+-----+---------+---------+---------+---------+---------+---------+
| ITEM |BLOCK|  SLOPE  |   S.E.  |LOCATION |   S.E.  |GUESSING |   S.E.  |
+======+=====+=========+=========+=========+=========+=========+=========+
| IT6  |   6 |   3.756 |   0.204 |  -0.209 |   0.021 |   0.000 |   0.000 |
+------+-----+---------+---------+---------+---------+---------+---------+
| IT7  |   7 |   3.200 |   0.209 |  -1.197 |   0.031 |   0.000 |   0.000 |
+------+-----+---------+---------+---------+---------+---------+---------+
| IT8  |   8 |   2.414 |   0.167 |  -1.538 |   0.045 |   0.000 |   0.000 |
+------+-----+---------+---------+---------+---------+---------+---------+

……………………… ADDITIONAL OUTPUT DELETED ………………………………………
```

Figure 13. Parscale output. GPCM model for items with 3 rank ordered categories.

The Generalized Partial Credit Model (GPCM)

PARSCALE Output

Figure 13 shows excerpts of the output from the PARSCALE program concerning 3 items with 3 rank-scaled categories, analyzed (together with several other items) using the GPCM. The PARSCALE output for the GPCM is very similar to the output for the GRM, except that the item category parameters are named STEP PARAMTER and that the STEP PARAMTER for SCORING FUNCTION = 1 is fixed at zero to indicate the way identification is achieved in the Parscale formulation of the GPCM. The item category parameters (2 for each item) are read from the output shown in the top part of Figure 13 in the following way: the item category parameter estimates shown in the column under SCORING FUNCTION = 2 should be read into the variable CATEGORY1 and the item category parameter estimates shown in the column under SCORING FUNCTION = 3 should be read into the variable CATEGORY2. The slope and location parameters are read from the output shown in the lower part of figure 13. Thus, the data should be saved as a SAS item parameter file using the variable names: SLOPE, LOCATION, CATEGORY1, CATEGORY2 (see table 6). In addition the item name, the model, and the number of response choices must be specified. The model statement should be GPCM. In an alternative specification of the model, the two item thresholds can be calculated, in which case an item parameter file such as the one shown in table 7 can be used.

Table 6. Example of SAS item parameter file based on the output from Parscale (Figure 9)

NAME	CHOICES	MODEL	SLOPE	LOCATION	CATEGORY1	CATEGORY2
IT6	3	GPCM	3.756	-0.209	0.596	-0.596
IT7	3	GPCM	3.200	-1.197	0.474	-0.474
IT8	3	GPCM	2.414	-1.538	0.552	-0.552
…	…	…	…	…	…	…

Table 7. Example of SAS item parameter file based on the output from Parscale (Figure 9)

NAME	CHOICES	MODEL	SLOPE	THRESHOLD1	THRESHOLD2
IT6	3	GPCM	3.756	-0.805	0.386
IT7	3	GPCM	3.200	-1.671	-0.723
IT8	3	GPCM	2.414	-2.091	-0.986
…	…	…	…	…	…

The Partial Credit Model (PCM)

WINSTEPS Output

Figure 14 presents typical output from WINSTEPS concerning 3 rank-scaled items analyzed using the PCM. Regarding the item parameters (top part of figure 9), the relevant column for the purpose of running an IRTFIT analyses is labeled: MEASURE. The item parameter estimates from this column should be saved in a SAS data set with the variable names: THRESHOLD1 and THRESHOLD2. Table 8 shows the equivalent SAS item parameter file. The Rasch model is identified by constraining the sum of the item category threshold parameters to zero and setting all slope parameters to 1. Thus, the slope parameter should be set to 1 in the SAS file (variable name SLOPE), SLOPE should be set to missing or no SLOPE variable should be provided. Since the person IRT score is not centered on 0 in this version of the model, the IRTFIT macro needs to be told the mean and standard deviation of the IRT scores. These can be read from the WINSTEPS output: Person STATISTICS under the column MEASURE. The mean is 3.884 and the standard deviation is 4.165. These need to be specified directly in the macro (macro options P_MEAN and P_STD).

```
.......... ADDITIONAL OUTPUT DELETED ..................

SUMMARY OF CATEGORY STRUCTURE.  Model="R"
FOR GROUPING "1" Item NUMBER: 1  I0001

.......... ADDITIONAL OUTPUT DELETED ..................
+-----------------------------------------------------------------------------
|CATEGORY    STRUCTURE   |  SCORE-TO-MEASURE   | 50% CUM.| COHERENCE|ESTIM|
| LABEL    MEASURE  S.E. | AT CAT. ----ZONE----|PROBABLTY| M->C C->M|DISCR|
|------------------------+---------------------+---------+----------+-----+-
|   0      NONE          |( -1.76) -INF    -.66|         |  91%  78%|     |
|   1        -.66    .52 |   3.14   -.66   6.94|    -.66 |  86%  96%| 1.03|
|   2        6.94   1.02 |(  8.04)  6.94  +INF |    6.94 |   0%   0%| 1.02|
+-----------------------------------------------------------------------------

.......... ADDITIONAL OUTPUT DELETED ..................

FOR GROUPING "2" Item NUMBER: 2  I0002

.......... ADDITIONAL OUTPUT DELETED ..................
+----------------------------------------------------------------------------
|CATEGORY    STRUCTURE   |  SCORE-TO-MEASURE   | 50% CUM.| COHERENCE|ESTIM|
| LABEL    MEASURE  S.E. | AT CAT. ----ZONE----|PROBABLTY| M->C C->M|DISCR|
|------------------------+---------------------+---------+----------+-----+
|   0      NONE          |( -3.72) -INF   -2.68|         |  83%  71%|     | 0
|   1       -2.60    .59 |   -.85  -2.68    .99|   -2.63 |  61%  84%| 1.04| 1
|   2         .91    .55 |(  2.03)   .99  +INF |     .94 |  94%  76%| 1.08| 2
+----------------------------------------------------------------------------

.......... ADDITIONAL OUTPUT DELETED ..................

FOR GROUPING "3" Item NUMBER: 3  I0003

.......... ADDITIONAL OUTPUT DELETED ..................
+-------------------------------------------------------------------------------
|CATEGORY    STRUCTURE   |  SCORE-TO-MEASURE   | 50% CUM.| COHERENCE|ESTIM|
| LABEL    MEASURE  S.E. | AT CAT. ----ZONE----|PROBABLTY| M->C C->M|DISCR|
|------------------------+---------------------+---------+----------+-----+---
|   0      NONE          |( -5.83) -INF   -4.75|         |   0%   0%|     |
|   1       -4.73    .80 |  -2.29  -4.75    .17|   -4.74 |  64%  73%| 1.18|
|   2         .15    .53 |(  1.25)   .17  +INF |     .15 |  83%  83%|  .86|
+-------------------------------------------------------------------------------

.......... ADDITIONAL OUTPUT DELETED ..................

         Person STATISTICS:  MEASURE ORDER
+-------------------------------------------------------------------------------
|ENTRY    RAW                          MODEL|   INFIT  |  OUTFIT  |PTMEA|EXACT MATCH|
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ   ZSTD|MNSQ   ZSTD|CORR.| OBS%   EXP%|
|-------------------------------------+----------+----------+-----+-----------
| MEAN     4.6    3.0   3.884   2.059| .83     .1| .93     .3|     | 80.5  81.9|
| S.D.     1.7     .0   4.165    .605|1.65     .9|2.01     .8|     | 30.3  12.9|
+-------------------------------------------------------------------------------

.......... ADDITIONAL OUTPUT DELETED ..................
```

Figure 14. WINSTEPS output. PCM model for items with 3 rank ordered categories.

Table 8. Example of SAS item parameter file based on the output from WINSTEPS (Figure 14)

NAME	CHOICES	MODEL	SLOPE	THRESHOLD1	THRESHOLD2
IT6	3	PCM	1	-0.66	6.94
IT7	3	PCM	1	-2.60	0.91
IT8	3	PCM	1	-4.73	0.15
…	…	…	…	…	

OPLM Output

Figure 15 shows output from the OPLM program, again analyzing 3 rank-scaled items using the PCM. The corresponding SAS item parameter file is shown in table 9. The model is identified in a similar ways as the WINSTEPS model. The population standard deviation much be calculated as the square root of the Variance. In this example the IRT score mean is 2.470 and the standard deviation is 2.711 (square root of 7.352).

```
......................... ADDITIONAL OUTPUT DELETED .........................................

  nr label       A       B     SE(B)     S    DF      P        M       M2       M3
----------------------------------------------------------------------------


   1 It1         1     .285    .135   8.671  2    .013   -2.941  99.999   -.633
          [:2]       4.404    .285    .000  0 99.999   99.999    .189   -.054

   2 It2         1  -2.185    .197    .000  0 99.999   -1.531  99.999    .404
          [:2]        .576    .139   1.523  2    .467    -.471  -1.220   -.514

   3 It3         1  -3.090    .258    .000  0 99.999   99.999    .002   1.460
          [:2]        .009    .134   9.242  2    .010    2.359   1.161    .038
----------------------------------------------------------------------------

......................... ADDITIONAL OUTPUT DELETED .........................................

 Number of observations =              1000
 Mean (theta's) =                     2.470
 Variance (estimated theta's) =       7.352
 Variance (true theta's) =            3.667
 Reliability (of est. theta's) =       .499
 Correlation (est. th.,scores) =       .972
```

Figure 15. OPLM output. PCM model for items with 3 rank ordered categories.

Table 9. Example of SAS item parameter file based on the output from OPLM (Figure 15)

NAME	CHOICES	MODEL	SLOPE	THRESHOLD1	THRESHOLD2
IT6	3	PCM	1	0.285	4.404
IT7	3	PCM	1	-2.185	0.576
IT8	3	PCM	1	-3.090	0.009
…	…	…	…	…	

Another way to present the item parameters is to rescale so the mean of the IRT scores is zero and the standard deviation is one. This is achieved by the transformation
SLOPE* = SLOPE * 2.711 (use the population standard deviation)
THRESHOLD1* = (THRESHOLD1 – 2.470)/2.711 (use the population mean and the standard deviation). The resulting item parameters are shown in table 10. Since the slope of all items is constrained to equality across items, the model is still a PCM and (except for the transformation) completely equivalent to the model in table 9.

Table 10. Example of SAS item parameter file based on the output from OPLM (Figure 15). Centered on the mean of the IRT scores for the sample.

NAME	CHOICES	MODEL	SLOPE	THRESHOLD1	THRESHOLD2
IT6	3	PCM	2.711	-0.806	0.713
IT7	3	PCM	2.711	-1.717	-0.699
IT8	3	PCM	2.711	-2.051	-0.908
…	…	…	…	…	

The Rating & Generalized Rating Scale Models (RSM & GRSM)

PARSCALE Output

Figure 16 shows excerpts of the output from the PARSCALE program concerning 3 items with 3 rank-scaled categories, analyzed (together with several other items) using the GRSM. The PARSCALE output for the GRSM is very similar to the output for the GPCM, except that the same category parameters are used for several items (in this case all items). The BLOCK = 2 output for all three items shows that they share the same category parameters.

```
.......................... ADDITIONAL OUTPUT DELETED ............................................

   ITEM BLOCK    2  BLOCK

   SCORING FUNCTION     :      1.000      2.000      3.000
   STEP PARAMTER        :      0.000      0.499     -0.499
   S.E.                 :      0.000      0.020      0.014

.......................... ADDITIONAL OUTPUT DELETED ............................................

+------+-----+---------+---------+---------+---------+---------+---------+
| ITEM |BLOCK|  SLOPE  |   S.E.  |LOCATION |   S.E.  |GUESSING |   S.E.  |
+======+=====+=========+=========+=========+=========+=========+=========+
| IT6  |   2 |   3.663 |   0.202 |  -0.209 |   0.021 |   0.000 |   0.000 |
| IT7  |   2 |   3.156 |   0.205 |  -1.197 |   0.032 |   0.000 |   0.000 |
| IT8  |   2 |   2.356 |   0.163 |  -1.538 |   0.046 |   0.000 |   0.000 |

.......................... ADDITIONAL OUTPUT DELETED ............................................
```

Figure 16. Parscale output. GRSM model for items with 3 rank ordered categories.

The category parameters (2 for each item when the items have 3 response choices) are read from the output shown in the top part of Figure 16 in the following way: the category parameter estimates shown in the column under SCORING FUNCTION = 2 should be read into the variable CATEGORY1 and the category parameter estimates shown in the column under SCORING FUNCTION = 3 should be read into the variable CATEGORY2. The slope and location parameters are read from the output shown in the lower part of figure 13. Thus, the data should be saved as a SAS item parameter file using the variable names: SLOPE, LOCATION, CATEGORY1, CATEGORY2 (see table 11). Note that the category parameter should be listed for all items for which those parameters apply. Further, the item name, the model, and the number of response choices must be specified. The model statement should be GRSM in this case, since the slope parameters varies between items.

Table 11. Example of SAS item parameter file based on the output from Parscale (Figure 16)

NAME	CHOICES	MODEL	SLOPE	LOCATION	CATEGORY1	CATEGORY2
IT6	3	GRSM	3.663	-0.209	0.499	-0.499
IT7	3	GRSM	3.156	-1.197	0.499	-0.499
IT8	3	GRSM	2.356	-1.538	0.499	-0.499
…	…	…	…	…	…	…

In an alternative specification of the model, the item thresholds can be calculated, in which case an item parameter file such as the one shown in table 12 can be used. The model may look like a GPCM, but the uniform distance between THRESHOLD1 and THRESHOLD2 across items (0.998) shows that the model is GRSM.

Table 12. Example of SAS item parameter file based on the output from Parscale (Figure 16)

NAME	CHOICES	MODEL	SLOPE	THRESHOLD1	THRESHOLD2
IT6	3	GRSM	3.663	-0.708	0.290
IT7	3	GRSM	3.156	-1.696	-0.698
IT8	3	GRSM	2.356	-2.037	-1.039
…	…	…	…	…	…

```
…………………… ADDITIONAL OUTPUT DELETED ………………………………

ITEM   6:       3 NOMINAL CATEGORIES,  3 HIGH
  CATEGORY(K): 1      2      3
    A(K)     -3.98   0.25   3.73
    C(K)     -1.85   1.54   0.31

           CONTRAST-COEFFICIENTS (STANDARD ERRORS)
  FOR:             A                    C
  CONTRAST P(#)  COEFF.[ DEV.]  P(#)  COEFF.[ DEV.]
      1       1   4.23 (0.45)      3   3.39 (0.38)
      2       2   7.71 (0.58)      4   2.15 (0.41)

…………………… ADDITIONAL OUTPUT DELETED ………………………………

ITEM   7:       3 NOMINAL CATEGORIES,  3 HIGH
  CATEGORY(K): 1      2      3
    A(K)     -3.25  -0.11   3.36
    C(K)     -4.34   0.91   3.42

           CONTRAST-COEFFICIENTS (STANDARD ERRORS)
  FOR:             A                    C
  CONTRAST P(#)  COEFF.[ DEV.]  P(#)  COEFF.[ DEV.]
      1       5   3.15 (0.59)      7   5.25 (0.83)
      2       6   6.61 (0.65)      8   7.76 (0.86)

…………………… ADDITIONAL OUTPUT DELETED ………………………………

 ITEM   8:       3 NOMINAL CATEGORIES,  3 HIGH
  CATEGORY(K): 1      2      3
    A(K)     -2.60  -0.04   2.64
    C(K)     -4.37   0.89   3.47

           CONTRAST-COEFFICIENTS (STANDARD ERRORS)
  FOR:             A                    C
  CONTRAST P(#)  COEFF.[ DEV.]  P(#)  COEFF.[ DEV.]
      1       9   2.55 (0.59)     11   5.26 (1.01)
      2      10   5.23 (0.69)     12   7.84 (1.04)

…………………… ADDITIONAL OUTPUT DELETED ………………………………
```

Figure 17. Multilog output. NOM model for items with 3 rank ordered categories.

The Nominal Model (NOM)

MULTILOG Output

Figure 17 shows excerpts of the output from the MULTILOG program concerning 3 items with 3 categories, analyzed (together with several other items) using the NOM. The MULTILOG output for lists two ways of parameterization of the NOM. The IRTFIT macro uses the output listed under CONTRAST-COEFFICIENTS. The parameters in columns A should be read into the variables SLOPE1 and SLOPE2, the parameters in column C should be read into the variables INTERCEPT1 and INTERCEPT2 (see table 13).

Table 13. Example of SAS item parameter file based on the output from MULTILOG (Figure 17)

NAME	CHOICES	MODEL	SLOPE1	SLOPE2	INTERCEPT1	INTERCEPT2
IT6	3	NOM	4.23	7.71	3.39	2.15
IT7	3	NOM	3.15	6.61	5.25	7.76
IT8	3	NOM	2.55	5.23	5.26	7.84
…	…	…	…	…	…	…

Macro Call Examples

The following examples are meant to provide guidance on the setup and options for use of the IRTFIT macro. Data and parameters used in these examples are simulated. Graphing is an option available in both sum and theta test methods. You can create graphics at any time after running the IRTFIT macro by using the _freq dataset created by both sum and theta procedures. Please refer to the section on How To Use The IRTFIT Macro for more information and examples on item file and parameter file setup.

Examples based on sum score method (with graphs)

Generalized Partial Credit Model

Calling code

```
 Libname example 'c:\irtfit\examples';
%IRTFIT(DATA = example.sumexample
 ,    PARFILE = example.sumexaplepar
 ,   ITEMLIST = it6-it10
 ,TESTMETHOD = sum
 ,      GRAPH = yes
 ,     OUTFMT = rtf
 ,     OUTLIB = c:\irtfit\examples
 ,    OUTCORE = five_sgpcm
 );
```

Results

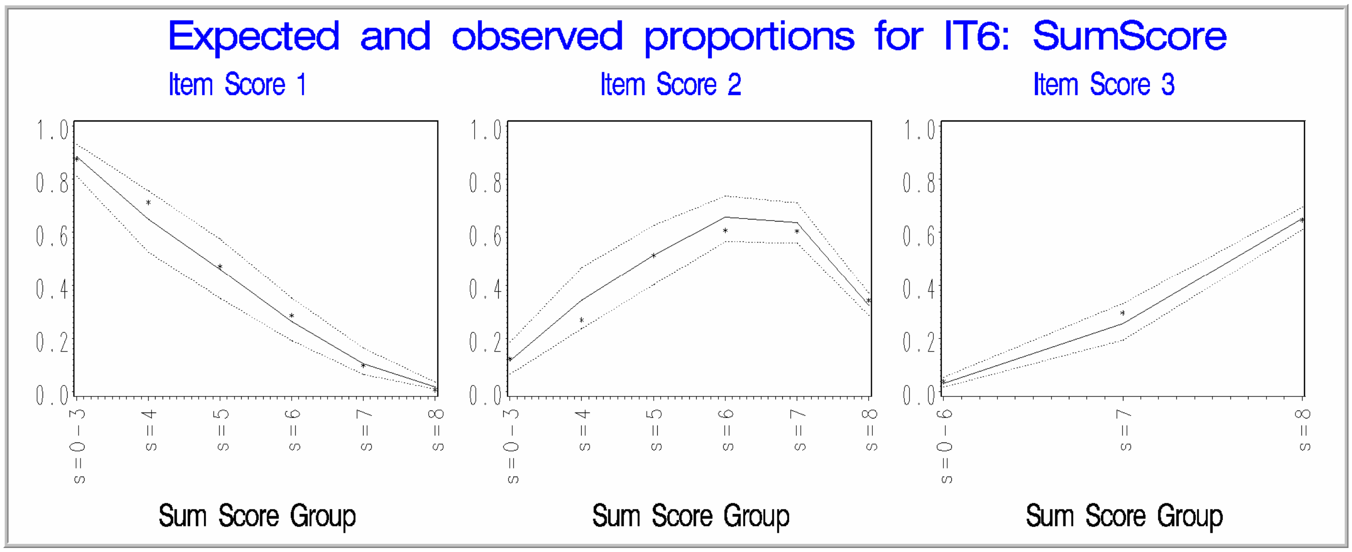
Results are written in the directory specified by OUTLIB (`c:\irtfit\examples\`). The Rich-Text-Format file five_`sgpcm.rtf` contains a table that summarizes model fit by item and graphic panels for charts of observed and expected frequencies vs. sum-score for each item response category. The summary table and a panel for Item 6 are included below.

Chi-square-test for Item fit
Sum score based: SUBTOTAL

item_no	*name*	*df*	*G2*	*Prob_G2*	*X2*	*Prob_X2*
1	IT6	9	8.13	0.5207	7.41	0.5946
2	IT7	10	6.33	0.7868	6.30	0.7894
3	IT8	11	7.34	0.7706	6.84	0.8116
4	IT9	12	12.29	0.4224	12.00	0.4459
5	IT10	12	7.73	0.8060	7.75	0.8045

Note: Significance denotes item misfit.

Item 6:



Note: Stars indicate observed values. Solid lines represent expected values and 95% confidence limits. Stars that fall outside of the confidence lines indicate deviations from the model.

Output datasets

Two SAS datasets are also created in the library ‘example’ that contain frequency and fit information.

five_sgpcm_freq.sas7bdat

five_sgpcm_fit.sas7bdat

Dichotomous Model

Calling code

```
Libname example ‘c:\irtfit\examples’;
%IRTFIT(DATA = example.sumdiexample
     ,    PARFILE = example.sumdiexaplepar
     ,   ITEMLIST = it1-it5
     ,TESTMETHOD = sum
     ,      SCALE = total
     ,    MINCODE = 1
     ,      GRAPH = yes
     ,     OUTFMT = rtf
     ,     OUTLIB = c:\irtfit\examples
     ,    OUTCORE = five_sdi
     );
```

Results

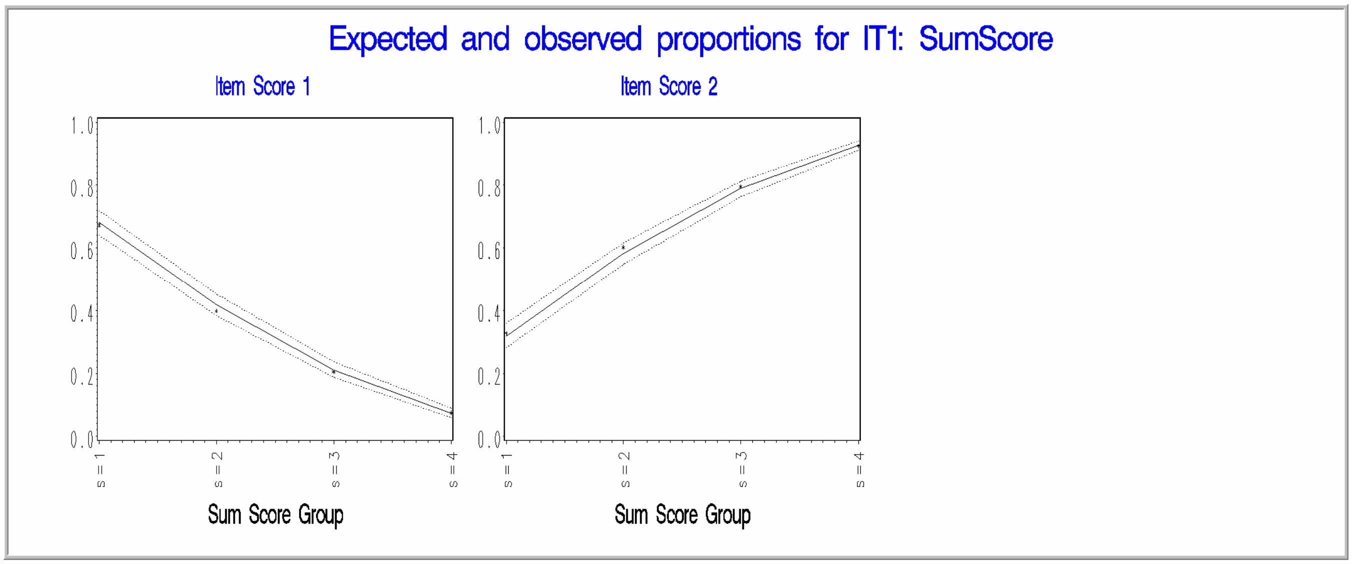
Results are written in the directory specified by OUTLIB (`c:\irtfit\examples\`). The file `five_sdi.rtf` contains a table that summarizes model fit by item and graphic panels for charts of observed and expected frequencies vs. sum-score for each item response category. The summary table and a panel for Item 1 are included below.

Chi-square-test for Item fit
Sum score based: SUBTOTAL

item_no	name	df	G2	Prob_G2	X2	Prob_X2
1	IT1	2	2.28	0.3206	2.28	0.3195
2	IT2	2	1.53	0.4664	1.57	0.4565
3	IT3	2	2.59	0.2733	2.58	0.2747
4	IT4	2	7.83	0.0200	8.96	0.0113
5	IT5	2	4.44	0.1088	4.39	0.1112

Note: Significance denotes item misfit.

Item 1:



Note: Stars indicate observed values. Solid lines represent expected values and 95% confidence limits. Stars that fall outside of the confidence lines indicate deviations from the model.

Output datasets

Two SAS datasets are also created in the library 'example' that contain frequency and fit information.

five_sdi_freq.sas7bdat
five_sdi_fit.sas7bdat

Examples based on THETA method (without graphs)

Generalized Partial Credit Model

Calling code

```
Libname example 'c:\irtfit\examples';
%IRTFIT(DATA = example.thetaexample
,    PARFILE = example.thetaexamplepar
,   ITEMLIST = it6-it10
, TESTMETHOD = theta
,     OUTFMT = rtf
,    OUTLIB = c:\irtfit\examples
,   OUTCORE = five_t
);
```

Results

A table summarizing item fit tests is written to five_t.rtf.

Chi-square-test for Item fit
Resampling based (# of replication is 100)

item_no	*name*	*G2*	*df_G2*	*Prob_G2*	*X2*	*df_X2*	*Prob_X2*
1	IT6	3.90	2.91719	0.2603	3.88	2.56886	0.2151
2	IT7	2.03	1.59491	0.2749	2.17	1.50302	0.2361
3	IT8	2.48	3.37972	0.5467	2.16	3.44120	0.6190
4	IT9	2.18	2.95761	0.5281	2.00	2.50427	0.4746
5	IT10	1.62	1.54314	0.3347	1.66	1.66167	0.3550

Note: Significance denotes item misfit.

Output datasets

Two SAS datasets are also created in the library 'example' containing frequency and fit information.

Five_T_freq.sas7bdat
Five_T_fit.sas7bdat

Dichotomous and Nominal Models (in one file)

Calling code

```
Libname example 'c:\irtfit\examples\';
%IRTFIT(DATA = example.dinomexample
,     PARFILE = example.dinomexamplepar
,    ITEMLIST = it1-it12
,  TESTMETHOD = theta
,       SCALE = total
,      OUTFMT = rtf
,     OUTLIB = c:\irtfit\examples
,    OUTCORE = twelve_di_nom
);
```

Results

A table summarizing item fit tests is written to `twelve_di_nom`.rtf.

Chi-square-test for Item fit
Resampling based (# of replication is 100)

item_no	name	G2	df_G2	Prob_G2	X2	df_X2	Prob_X2
1	IT1	2.44	3.69094	0.6064	2.37	3.55580	0.5985
2	IT2	1.18	2.98254	0.7544	1.06	2.76402	0.7473
3	IT3	0.48	2.21597	0.8264	0.48	2.26726	0.8362
4	IT4	1.24	2.66297	0.6838	1.00	2.07880	0.6257
5	IT5	0.79	1.67889	0.5912	0.75	1.55804	0.5706
6	IT6	3.41	3.42422	0.4004	4.50	2.44053	0.1486
7	IT7	3.77	3.43395	0.3526	3.73	2.16181	0.1748
8	IT8	3.45	7.62770	0.8826	3.08	6.63448	0.8525
9	IT9	8.42	5.60558	0.1779	8.68	5.59924	0.1623
10	IT10	5.26	4.55764	0.3292	4.10	3.11881	0.2666
11	IT11	4.00	4.54983	0.4866	3.75	3.76530	0.4046
12	IT12	17.18	5.87703	0.0079	14.62	4.26102	0.0069

Note: Significance denotes item misfit.

Output datasets

Two SAS datasets are also created in the library 'example' containing frequency and fit information.

`twelve_di_nom` _freq.sas7bdat
`twelve_di_nom` _fit.sas7bdat

Graphical outputs only

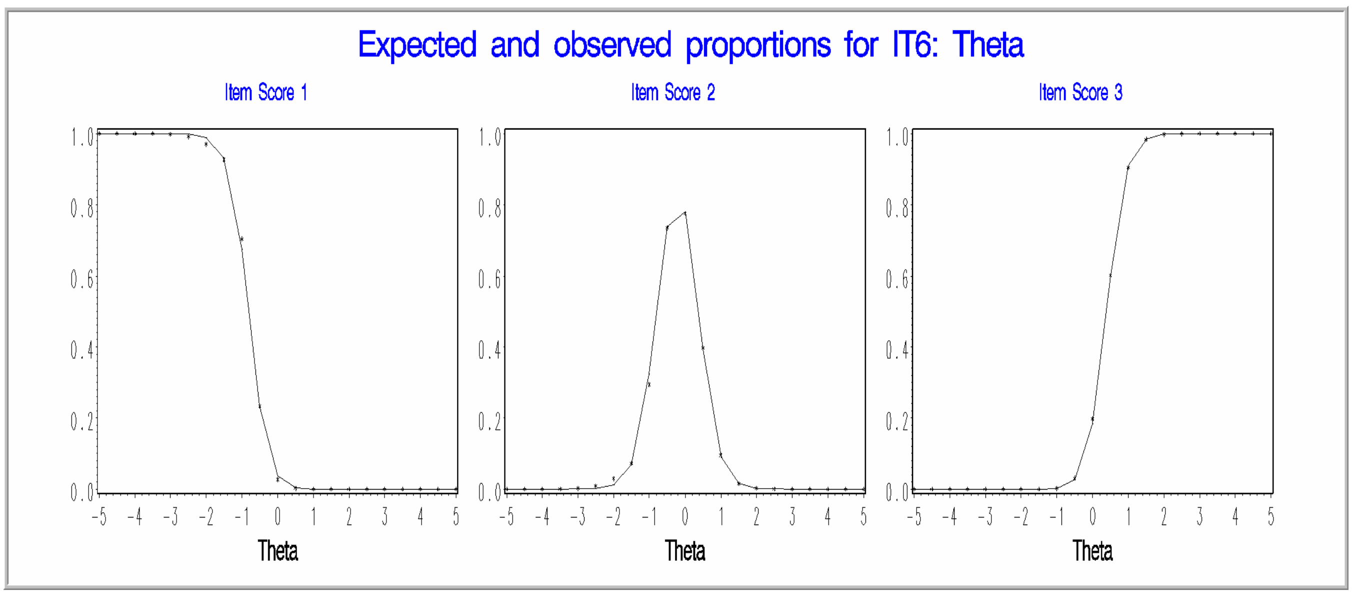
The graphs of observed and expected frequencies vs. quadrature points of theta for all item response categories are not outputted in the example above. We can utilize GH_p_freq.sas7bdat to produce the graphs at later time.

Calling code

```
Libname example ‘c:\irtfit\examples’;
%IRTFIT(GRAPH_DATA = example. five_t_freq
,   ITEMLIST = it6-it10
, TESTMETHOD = NONE
,    LD_TEST = NO
,      GRAPH = yes
,     OUTFMT = rtf
,    OUTLIB = c:\irtfit\examples
,   OUTCORE = five_t_graph
);
```

Results

The resulting RTF file is written to the directory specified by OUTLIB. The results file contains a panels of graphs, representing the In the folder of observed and expected frequencies vs. quadrature point of theta for each item category. The panel for item IT6 is displayed below.



Note: Stars indicate pseudo-observed values. Stars that fall away from the expected curves indicate deviations from the model.

Test for local dependence

Following code will suppress the tests of item fit, and only do test for local dependence.

```
Libname example 'c:\irtfit\examples';
%IRTFIT(DATA = example.ldexample
,    PARFILE = example.ldexamplepar
,   ITEMLIST = it6-it10
, TESTMETHOD = NONE
,    LD_TEST = YES
,     OUTFMT = rtf
,    OUTLIB = c:\irtfit\examples
,   OUTCORE = five_ld
);
```

Results

An output file (five_ld.rtf) containing tables summarizing tests for local dependence, chi-square and residual correlation polychoric coefficients, is written to the OUTLIB directory.

Local Dependence tests

Chi-square test		*IT6*	*IT7*	*IT8*	*IT9*
IT7	*X2*	3.41	.	.	.
	df	2	.	.	.
	p	0.1814	.	.	.
IT8	*X2*	2.21	0.01	.	.
	df	2	2	.	.
	p	0.3315	0.9943	.	.
IT9	*X2*	5.34	0.94	3.92	.
	df	2	2	2	.
	p	0.0693	0.6249	0.1412	.
IT10	*X2*	1.47	2.63	0.44	2.94
	df	2	4	2	4
	p	0.4787	0.6215	0.8012	0.5676

Residual Correlation	IT6	IT7	IT8	IT9
IT7	0.02	.	.	.
IT8	0.01	0.00	.	.
IT9	0.00	0.01	0.05	.
IT10	0.01	-0.01	0.02	0.01

Output datasets

A SAS dataset (five_ld_ld.sas7bdat) in the library 'example' containing the table data is also created.

Technical details

1. Item fit based on theta (X^2* and G^2* - TESTMETHOD=THETA)

Let P($C_{ij}=c \mid \theta_j, \alpha_i, \beta_i$) be the probability that person j chooses category c (numbered 0,…, m_i) on item i. This probability is assumed to depend on the person parameter, θ_j, and on item parameters, α_i, and β_i. Using Bayesian methods, the conditional probability of θ_j given a set of item responses $\overline{C_j}$:

$$P\left(\theta_j \mid \overline{C_j}\right) = \frac{\prod_i P\left(C_{ij} \mid \theta_j\right) P\left(\theta_j\right)}{P\left(\overline{C_j}\right)} \qquad (1)$$

Where $P\left(\theta_j \mid \overline{C_j}\right)$ is the posterior distribution of θ_j given the item responses, $P\left(C_{ij} \mid \theta_j\right)$ is shorthand for P($C_{ij}=c \mid \theta_j, \alpha_i, \beta_i$), P($\theta_j$) is the prior distribution of θ_j, and $P\left(\overline{C_j}\right)$ is the marginal probability of response pattern $\overline{C_j}$ for a person randomly sampled from the population. While estimation approaches such as the expected a posteriori (EAP) use the mean of the posterior as a score estimate, the approach in this fit test calculates the probability of the respondent having each of a number of θ values (evaluated across a limited number of discrete points). This procedure distributes a respondent's contribution to the item fit table over multiple θ levels, and in doing so considers the precision with which θ is estimated. These contributions are then summed across all respondents to yield pseudocounts for each response category or a pseudo-observed score distribution at each θ level. The marginal pseudocount at each level of θ is multiplied with the item response probabilities (P($C_j=c \mid \theta, \alpha_i, \beta_i$)) for the item in question to achieve the expected score distribution at each level of θ. Pearson and likelihood-ratio goodness-of-fit tests can then be computed as:

$$X^{2}* = \sum_{q=1}^{Q}\sum_{c=1}^{m}\frac{\left(O_{cq}-E_{cq}\right)^2}{E_{cq}}, \qquad where \quad 0 < E_{cq} < N_q \qquad (2)$$

and

$$G^2* = \sum_{q=1}^{Q}\sum_{c=1}^{m} O_{cq} \ln \frac{O_{cq}}{E_{cq}} \qquad where \quad 0 < E_{cq} < N \tag{3}$$

when O represents observed and E expected values. Since the expected counts for some ability levels may be 0 or very small, it may be necessary to sum over a subset of θ levels (e.g., $-2 \leq \theta \leq 2$). Further, even cells within that range may be excluded from the summation if the expected count for that cell is very low. Finally, since the probabilities underlying the pseudo-observed score distributions are not independent, a known null chi-squared distribution cannot be assumed for purposes of hypothesis testing. Stone [3] found that the fit statistics appeared to be distributed as a scaled chi-squared random variable, and described a Monte Carlo re-sampling procedure for estimating scaling corrections used to approximate a null chi-square distribution. This approach, which is implemented in the IRTFIT macro, generates a number of replications (specified by the user, default = 100) of item responses using the estimated item parameters and assuming a normal distribution for θ. For each of the replications, the fit statistics are calculated, and the mean and variance of these fit statistics across all replications are used to rescale the goodness of fit statistics. Simulation studies have found that these rescaled statistics are well behaved and exhibit sufficient power to detect minor deviation from model fit [5].

2. Item fit based on sum score (S-X^2 and S-G^2- TESTMETHOD=SUM)

Let $P(C_{ij}=c \mid \theta_j)$ be the probability that person j chooses category c (numbered $0,\ldots, m_i$) on item i. This probability is assumed to depend on the person parameter, θ_j, and on item parameters, (e.g. for the graded response model: α_i, and β_{ic}). A recursive algorithm is used to calculate the likelihood of achieving sum score t [8;9]. This algorithm builds the scale one item at a time. Let S_t be the probability of achieving sum score t for a scale of a certain number of items. With only one item, the probability $S_t = P(C=t)$. For each subsequent item that is included in the scale:

$$S_t = \sum_{t'}\sum_{c} S'_{t'} \, P(C=c) \, I(t'+c=t) \tag{4}$$

where $S'_{t'}$ is the probability of sum score t' for a scale without the item just included and $I(t'+c=t)$ is an indicator function that is 1 whenever $t' + c$ equals t and is 0 otherwise. Thus, we sum the probabilities for all the combinations of item scores and sum scores (for the previous scale) that total the sum score in question. This procedure is repeated until the probabilities for the total scale are achieved. The fit statistics originally proposed[1] was based on cross-tabulation tables of item score times total sum scale score. However, using cross-tabulation tables of item score times sum of all **other** items in the scale (subtotal) has the advantage of avoiding structural zeros in the table. Both approaches can be implemented in the macro. In the former approach, the expectations are given by:

$$E_{tci} = N_t \frac{\int S'_{t-c} \, P(C_i = c)\, \phi(\theta) d\theta}{\int S_t \, \phi(\theta) d\theta} \tag{5}$$

where $\phi(\theta)$ is the population distribution of θ and N_t is the number of people with sum score = t. In the latter approach, the expectations are given by:

$$E_{t'ci} = N_{t'} \frac{\int S'_{t'}\, P(C_i = c)\, \phi(\theta) d\theta}{\int S'_{t'}\, \phi(\theta) d\theta} \tag{6}$$

(where *T'* excludes the item in question). The integrals can be evaluated by numerical integration. Two fit test can then be defined:

$$S - X_i^{\;2} = \sum_{t=t_{\min}}^{t_{\max}} \sum_{c_i=1}^{m_i} \frac{(O_{tci} - E_{tci})^2}{E_{tci}}, \qquad where \quad 0 < E_{tci} < N_t \tag{7}$$

and

$$S - G_i^{\;2} = \sum_{t=t_{\min}}^{t_{\max}} \sum_{c_i=1}^{m_i} O_{tci} \ln \frac{O_{tci}}{E_{tci}}, \qquad where \quad 0 < E_{tci} < N_t \tag{8}$$

For long scales, the item times sum score table is often sparse – having cells with very low expected frequencies. This problem is similar to the problem encountered with the X^2* and G^2* statistics, but the approach to solving the problem is different. Before calculating the *S*-X^2 and *S*-G^2 fit statistics, cells are collapsed to achieve expected cell frequencies above a user defined threshold (default= 5). The following algorithm is used for each row (item score value): 1) The center of the row is identified by the median of the expected frequencies for the row in question, 2) Starting from the first and last cell and working towards the center, cells are collapsed with the cell closer to the center if the expected cell frequency is less than the minimum value.

3. Local dependence (LD_TEST=YES)

Based on the assumption of local independence, the joint distribution of answers to any two items in the scale is equal to the product of the marginal probabilities:

$$\mathrm{P}(C_{1j}=c_1,\ C_{2j}=c_2 \mid \theta_j) = \mathrm{P}(C_{1j}=c_1 \mid \theta_j)\ \mathrm{P}(C_{2j}=c_2 \mid \theta_j) \tag{9}$$

And the expected cross-tabulation table can be calculated as

$$E_{c_1c_2} = N \int P(C_1 = c_1 \mid \theta)\, P(C_2 = c_2 \mid \theta) \phi(\theta) d\theta \tag{10}$$

The integral can be evaluated by numerical integration. This expected cross-tabulation table can be compared with the observed cross-tabulation table. This comparison is performed or each item pair and summarized in two fit statistics: 1) a X^2 fit statistic, 2) a residual correlation. The X^2 fit statistic can be calculated as:

$$X^2 = \sum_{c_1} \sum_{c_2} \frac{(O_{c_1c_2} - E_{c_1c_2})^2}{E_{c_1c_2}} \tag{10}$$

An automated procedure is implemented to collapse cells with low expected cell frequencies. The residual correlation is based on the calculation of polychoric (for dichotomous items: tetrachoric) correlation coefficients [10] for both the observed and expected table. The residual correlation is then calculated as the difference between the expected and the observed correlation coefficient:

$$d\rho = \rho_o - \rho_e \qquad (11)$$

4. Missing data

Missing data is handled differently by each of the 3 tests.
For TESTMETHOD=SUM, only complete cases are analyzed, since both the item in question and all other items need to be answered to calculate the observed item*sum score distribution.
For TESTMETHOD=THETA, the likelihood for theta that is necessary to calculate the "pseudo-counts" can be established even if some items are unanswered. The macro will use the information from all the available items that are non-missing. If the user wants to analyze complete data only, cases with missing should be excluded from the data set, before fit testing is started.
For LD_TEST = YES, all cases with non-missing responses for the item pair in question will be used for calculating the observed cross-table, and the expected cross-table will be calculated to match that total number. Thus, the test will use all available information. If the user wants to analyze complete data only, cases with missing should be excluded from the data set, before fit testing is started.

5. Output files

(1) SAS dataset of item response category level

#	Variable	Type	Len	Label
1	Scalesco Theta	Num	8	Sum score when TestMethod is SUM Quadrature point for θ when when TestMethod is Bayesian
2	Count	Num	8	
3	Percent	Num	8	
4	GH1_O_1	Num	8	
5	GH1_O_2	Num	8	
6	GH1_O_3	Num	8	
7	GH1_O_4	Num	8	
8	GH1_O_5	Num	8	
9	GH1_E_1	Num	8	
10	GH1_E_2	Num	8	
11	GH1_E_3	Num	8	
12	GH1_E_4	Num	8	
13	GH1_E_5	Num	8	
14	GH1_n	Num	8	
...	...			

Here item 'GH1' has 5 response categories with value from 1 to 5. GH1_O_1 – GH1_O_5 are observed category frequencies related to SCALESCO or THETA, and GH1_E_1 – GH1_E_5 are corresponding expected frequencies. GH1_n is total observed count related to SCALESCO or pseudo count related to THETA. Each item has a set of such variables.

(2) SAS dataset of item fit statistics

#	Variable	Type	Len	Label
Sum score based				
1	item_no	Num	8	
2	name	Char	8	
3	n_cells	Num	8	
4	X2	Num	8	
5	G2	Num	8	
6	df	Num	8	
7	Prob_X2	Num	8	
8	Prob_G2	Num	8	
9				
Theta score based				
1	item_no	Num	8	
2	name	Char	12	
3	n_zero	Num	8	
4	X2_O	Num	8	Original X^2
5	G2_O	Num	8	Original G^2
6	n_cells	Num	8	
7	X2	Num	8	Rescaled X^2
8	df_X2	Num	8	Df of X^2 adjusted for # of parameters
9	Prob_X2	Num	8	Probability of X^2 with adjusted DF
10	df_X2_NA	Num	8	Df of X^2 not adjusted for # of parameters
11	Prob_X2_NA	Num	8	Probability of X^2 with non-adjusted DF
12	G2	Num	8	Rescaled G^2
13	df_G2	Num	8	Df of G^2 adjusted for # of parameters
14	Prob_G2	Num	8	Probability of G^2 with adjusted DF
15	df_G2_NA	Num	8	Df of G^2 not adjusted for # of parameters
16	Prob_G2_NA	Num	8	Probability of G^2 with non-adjusted DF

(3) SAS dataset for pair wise local independence tests

#	Variable	Type	Len	Label
1	Name1	Char	20	
2	Name2	Char	20	
3	item_no1	Num	8	
4	item_no2	Num	8	
5	pred_corr	Num	8	Predicted polychoric correlation
6	obs_corr	Num	8	Observed polychoric correlation
7	res_corr	Num	8	Residual polychoric correlation
8	X2	Num	8	
9	G2	Num	8	
10	df	Num	8	
11	Prob_X2	Num	8	
12	Prob_G2	Num	8	
13	warning	Char	60	

(4) Printed output

Depended on the option of **OUTFMT**, printed output is produced either within SAS or in specified folder as a RTF file or PDF file. For Bayesian procedure based item fit tests, only rescaled X2, G2, DFs adjusted for number of parameters and related probabilities are printed. Please see examples below for details.

6. Error messages

On the first run with graphics output in a SAS session, the macro may produce the following error messages:

```
ERROR: Memname GREPS is unknown.
ERROR: Memname GSEG is unknown.
ERROR: Memname GSEG is unknown.
```

These messages can be ignored.

Acknowledgements

Development of IRTFIT macro was supported by the National Cancer Institute (NCI) under the contract 263-MQ-515639. The authors wish to thank Maria Orlando and David Thissen for comments on an earlier version of the program and the manual.

References

1. Orlando M, Thissen D: Likelihood-Based Item-Fit Indices for Dichotomous Item Response Theory Models. Applied Psychological Measurement 24:50-64, 2000
2. Orlando M, Thissen D: Further Investigation of the Performance of S - X2: An Item Fit Index for Use With Dichotomous Item Response Theory Models. Applied Psychological Measurement 27:289-298, 2003

3. Stone CA: Monte Carlo based null distribution for an alternative goodness-of-fit test statistic in IRT models. J Educ Measur 37:58-75, 2000
4. Stone CA: Empirical power and type I error rates for an IRT fit statistic that considers the precision of ability estimates. Educ Psychol Measur 63:566-586, 2003
5. Stone CA, Zhang B: Assessing Goodness of Fit of Item Response Theory Models: A Comparison of Traditional and Alternative Procedures. J Educ Measur 4:331-352, 2003
6. Glas CAW, Falcón JCS: A Comparison of Item-Fit Statistics for the Three-Parameter Logistic Model. Applied Psychological Measurement 27:87-106, 2003
7. Stone CA, Hansen MA: The effect of errors in estimating ability on goodness-of-fit tests for IRT models. Educational & Psychological Measurement 60:974-991, 2000
8. Lord FM, Wingersky MS: Comparison of IRT true-score and equipercentile observed-score "euqatings". Applied Psychological Measurement 8:453-461, 1984
9. Thissen D, Pommerich M, Billeaud K, Williams VSL: Item response theory for scores on tests including polytomous items with ordered responses. Applied Psychological Measurement 19:39-49, 1995
10. Drasgow F: Polychoric and Polyserial Correlations, in Kotz S, Johnson NL, Read CB (eds): Encyclopedia of Statistical Sciences, New York, 1986, pp 68-74