USER GUIDE
SPIRIT Macro: Item Response Theory in SPSS

SPIRIT VERSION 1.0

Nicholas Rockwood*[1], Jack DiTrapani†[1] and Minjeong Jeon‡[2]

[1] *Department of Psychology, The Ohio State University*
[2] *Department of Education, University of California, Los Angeles*

*rockwood.19@osu.edu
†ditrapani.4@osu.edu
‡mjjeon@ucla.edu

Contents

1 The SPIRIT Macro

The SPIRIT macro allows researchers to conduct one-parameter item response theory (IRT) analyses through the typical SPSS syntax interface or a custom point-and-click dialogue box. It is available and easily implemented in SPSS Version 21 and above. Since a one-parameter IRT model is a specific case of a generalized linear mixed model (GLMM), the macro utilizes the GENLINMIXED function of SPSS.

One of the main advantages of SPIRIT is its flexibility in terms of model specification. Person and item covariates, multidimensional models, and multigroup designs can be specified to allow users to move beyond basic measurement models into the realm of *explanatory* IRT models. Further, SPIRIT allows the item responses to be dichotomous or ordinal, which is a feature not included in some other general purpose GLMM software packages (such as the `lme4` package for R) that are often utilized when analyzing item response data.

In addition to its flexibility, SPIRIT is intuitive and easy to use. The dialogue box allows the user to specify the model through a point-and-click interface and the syntax version requires a minimal amount of code. Concurrently, the user may request a series of output options that include estimated trait levels for each individual, item and person fit statistics, and item plots. These options are specific to IRT and, thus, not usually provided by GLMM software.

1.1 Point-and-click vs. Syntax

SPIRIT is offered in both a point-and-click version and a syntax version. In general, both versions can be used to perform the same tasks, so the decision to use one over the other can be based on personal preference. Obviously, the point-and-click version is user-friendly and intuitive, and is a popular choice among those used to the SPSS point-and-click interface. The syntax version, on the other hand, can be beneficial for documenting your analyses and automating the re-analysis of your data at a later time.

1.2 Citation

DiTrapani, J., Rockwood, N. J., & Jeon, M. (in press). IRT in SPSS using the SPIRIT macro. *Applied Psychological Measurement [Computer program exchange].*

2 System Requirements

SPIRIT is compatible with Windows and Mac Operating Systems for SPSS Version 21 and higher. For the most user-friendly output, it is recommended that SPSS Version 22 or higher is used.

2.1 Note for using SPSS Version 21

The SPIRIT Macro contains syntax to remove unnecessary additional output provided by the use of the GENLINMIXED function in SPSS. However, this syntax is not recognized by SPSS versions older than Version 22. As a result, the additional output will not be removed and the output will also contain the following error code in numerous places:

```
>Error # 1.  Command name:  OUTPUT
>The first word in the line is not recognized as an SPSS
Statistics command.
>Execution of this command stops.
```

This error is simply due to attempts to modify the output *display*, and is not an indication of error in the output *content*. Consequently, users should ignore the warnings and additional output that is presented and instead focus on the output discussed in this manual and subsequent publications and tutorials.

2.2 Note for using SPSS Version 24

There is a separate SPIRIT file for SPSS Version 24 (denoted *v24*), due to a slight modification in the SPSS macro language in which SPIRIT was developed in. Users using SPIRIT for SPSS Versions 21-23 (denoted *v21-23*) in SPSS Version 24 will receive the following error:

```
>Warning # 661.  Command name:  EXECUTE
>The SUBSTR function is deprecated.  Consider using the
CHAR.SUBSTR function.
```

Note that the output will still be correct. SPSS is simply notifying the user that one of the macro language arguments within the SPIRIT file is no longer in use, and has been replaced. Users with SPSS Version 24 should use the SPIRIT file built for Version 24 to avoid this error.

2.3 Output Language

Unfortunately, the use of SPIRIT when the SPSS *Output Language* is specified as anything other than English will result in errors. The output language can be changed to English using the Output Language drop-down menu under the *Language* header found through *Edit* → *Options...*.

3 Installation

The SPIRIT macro can be obtained by emailing any of the authors. The folder that will be provided contains this user guide, as well as two forms of the macro. The first form is syntax-based which, after installation, allows the user to operate the macro using an SPSS syntax command window. The second form is a point-and-click interface which allows the user to operate the macro using an SPSS dialogue box. While both options provide the same features, they should be viewed as separate entities in that users wishing to use both forms must install them separately. However, use of the syntax version does not require the installation of the dialogue box, and use of the point-and-click version does not require the syntax installation.

3.1 Point-and-Click Version

The point-and-click version of the SPIRIT macro can be installed by double-clicking on the *.spd* file and then clicking *Install*. The macro will now be ready for use under the *Analyze* → *SPIRIT* heading in SPSS.

The dialogue box can also be installed through the SPSS menu bar by clicking on *Utilities* → *Custom Dialogs* → *Install Custom Dialog*. Then, the *.spd* file can be found in the file browser window. It is important to note that, unlike the syntax version, **the dialogue box only needs to be installed once**.

Installation of an updated version of SPIRIT follows the previously described procedure, with the addition of selecting "OK" when prompted with the statement "There is already a dialog by this name installed. Do you want to overwrite the previous version?"

3.2 Syntax Version

The syntax version is installed using the *.sps* file. Double-clicking the file opens an SPSS syntax file containing all of the code that defines the SPIRIT function. To load the function into the system, the code must be highlighted and "Run." This can be done by selecting *Run* → *All* at the top of the syntax window. The syntax window can then be closed and the macro will remain available for use within the current session (that is, until SPSS has been closed). Users wishing to use the syntax version of the macro in a new SPSS session will have to re-install the macro using the steps described above at the beginning of each session.

4 Preparing to Use the Macro

4.1 Model Viewer/Pivot Tables and Charts

In order for the macro to run appropriately and provide useful output, the SPSS Output Display must be switched from the default of *Model Viewer* to *Pivot tables and charts*. This can be done in SPSS by selecting *Edit* → *Options*. Within the *Output* tab at the top of the *Options* dialogue box is a section titled *Output Display*. Within this section, click on the radio button titled *Pivot tables and charts*. This option will remain selected across multiple SPSS sessions, so the user does not have to repeat this step before each use of the SPIRIT macro.

4.2 Data Layout

For SPSS to correctly run its GENLINMIXED function, data must be presented in long form. This means that every *item response* has its own row, as opposed to wide form, where every *respondent* has his or her own row. See Table 1 for an example of data in long (Table 1a) and wide formats (Table 1b).

The long format is helpful for GLMMs since it allows both item and person variables to be entered into the data. The wide format does not allow for item covariates because there is not a row for each observation. Despite these advantages, the long format is not particularly common in IRT research, and the wide format is more frequently required for IRT software. It is relatively straightforward to use the *VARSTOCASES* function in SPSS to restructure a dataset from wide to long form. Example syntax for converting wide data to long data is provided below:

Subject	Item	Resp
1	1	0
1	2	1
1	3	0
2	1	0
2	2	0
2	3	1

(a) An example dataset in the long format. Each item response has its own row. Since this example contains two respondents and three items, there are six rows.

Subject	Item 1	Item 2	Item 3
1	0	1	0
2	0	0	1

(b) An example dataset in the wide format. Each subject has its own row. Since this example contains two respondents and three items, there are two rows and a column for each of the three items.

Table 1: Example item response data in long and wide formats.

```
VARSTOCASES
/MAKE response FROM firstitem TO lastitem
/INDEX=itemID(response)
/KEEP=personID othervars
/NULL=KEEP.
```

where *firstitem* is replaced by the variable name for the first item listed in the dataset, *lastitem* is replaced by the variable name for the last item listed in the dataset, *PersonID* is replaced by the name of the variable that contains the person IDs, and *othervars* is replaced by a list of additional variables that should transfer to the long form dataset. Variables not listed will be removed from the dataset. Running this syntax will create a new variable called *response* which contains each individual's response to each item, and a variable called *itemID* which contains the name of each item variable before the wide-to-long data transformation.

Note that using the "*firstitem* TO *lastitem*" line in the syntax only works if all of the item variables are listed together in the dataset. That is, there are no additional variables listed in between the first and last items. If there are additional (non-item) variables between the items, the dataset can either be rearranged before running the syntax, or the snytax can list each individual item. For example, if there were only 4 items called *item1*, *item2*, *item3*, and *item4*, the syntax would be:

```
VARSTOCASES
/MAKE response FROM item1 item2 item3 item4
/INDEX=itemID(response)
/KEEP=personID
/NULL=KEEP.
```

The use of this syntax is safer in that the ordering of the item names does not matter, but it may be more time consuming if there is a large amount of items to list.

5 Macro Arguments

Figure 1 displays the dialogue window for the point-and-click version of SPIRIT, and Table 2 provides an overview of each of the corresponding macro arguments for the syntax version. Following is a detailed description of each macro argument. Note that the section title for each argument corresponds to the labels in the dialogue window, and the numbers in Figure 1 correspond to the section labels. The argument in parentheses corresponds to the code needed to specify the argument using the syntax version.

5.1 Dataset Name (`dataset`)

The *dataset* argument should specify the name of the dataset being used for the analysis. Note that this is not the name of the data file. By default, *dataset* is set to *DataSet1*, which is the name given to the first dataset opened in a new SPSS session. The name of the dataset can be found in brackets after the name of the data file on the top bar of the data window.

5.2 Items (`items`)

The *items* argument should be an item identification variable which should be specified as a *string* variable.

5.3 ID (`id`)

The *id* argument should be a person identification variable, specified as a *string* variable.

5.4 Response Variable (`response`)

The *response* argument should be a variable containing each person's response on each item. If the binomial link is used (see below) this variable should be dichotomous (0,1). If the multinomial link is used, the variable can have more than two categories, but the variable "Measure" must be specified as *Ordinal* in the SPSS data viewer. This will run the rating scale model, which is the only option for polytomous data.

5.5 Covariates (`cov`)

The *cov* argument allows the specification of person, item, or person-by-item covariates, which can be either continuous or nominal. If the covariate is continuous, it must be specified in the dataset as a *numeric* variable with *Scale* measurement. If it is nominal, it must be specified in the dataset as a *string* variable with *Nominal* measurement. Currently, up to five covariates can be specified. Occasionally, an additional plot is produced in the output when multiple item covariates are included in the model. This plot can be ignored by the user.

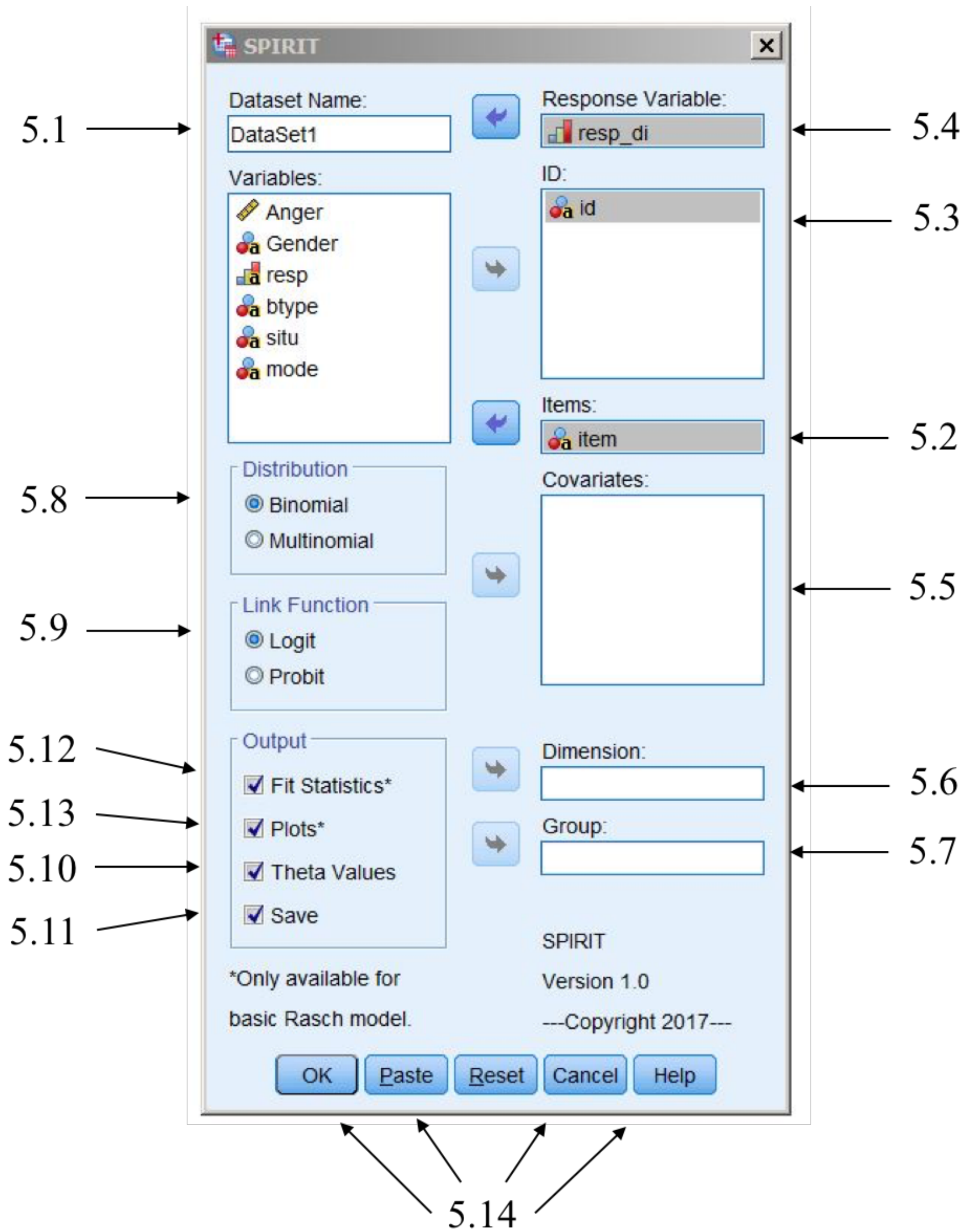

Figure 1: The SPIRIT dialogue window. The numbers correspond to the section numbers that describe each macro argument.

SPIRIT Arguments		
Argument	**Options/Variable Type**	**Description**
dataset	*DataSet1**	The name of the dataset that will be analyzed by the macro.
items	*string*†	The name of the variable that contains the item ID.
id	*string*†	The name of the variable that contains the person ID.
response	*string*†, *numeric*†	The name of the variable that contains each person's response to each item.
cov	*string*†, *numeric*†	A list of variables that are person, item, or person-by-item covariates.
dim	*string*†	For use in multidimensional models. A variable that indicates which dimension each item is a part of.
group	*string*†	For multi-group models. A variable that indicates which group each person is a part of.
dist	*binomial**, *multinomial*	The distribution of the response variable.
link	*logit**, *probit*	The link function for the model.
theta	*NO**, *YES*	Display theta values in the output.
save	*0**, *1*, *2*	Save values into the original dataset.
fit‡	*NO**, *YES*	Display item and person fit statistics in the output. If selected, the response variable must be *numeric* and *ordinal*.
plot‡	*NO**, *YES*	Display item plots in the output.

Table 2: A brief description of the available arguments for the syntax version of the SPIRIT macro. In the "Options/Variable Type" column, possible variable types are denoted with a † and an * denotes the default value. Arguments only available for the basic Rasch model (single-group, unidimensional without covariates) are denoted with a ‡.

Creating a variable to enter into SPIRIT to test for differential item functioning (DIF) is relatively easy. The variable should take on a value of zero for all items except the one being tested. The value for the item being tested should be equal to the covariate value. Example syntax that creates a variable to test for DIF of arbitrary item *item4* and covariate *Gender* with levels *M* and *F* is as follows:

```
COMPUTE DIF=0.
EXECUTE.
IF (Gender = 'M' & itemid = 'item4') DIF=1.
EXECUTE.
```

where *itemid* is replaced by the name of the item ID variable. Note that for all rows of the dataset where *itemid* does not equal *item4*, the value of the variable will be zero. For all rows where *itemid* does equal *item4*, the value will either be zero (if Gender = F) or one (if Gender = M). Once the new variable *DIF* has been constructed, it can be entered as a covariate and the grouping variable used to construct *DIF* (*Gender* here) should be entered as the group.

5.6 Dimension (`dim`)

The *dim* argument allows the specification of a multidimensional model. A *string* variable that indicates the dimension that each item pertains to should be used.

5.7 Group (`group`)

The *group* argument allows the specification of a multi-group design. A *string* variable that indicates the group that each person pertains to should be used.

5.8 Distribution (`dist`)

The *dist* argument specifies the underlying distribution of the response variable. The default value is *binomial* which should be used for dichotomous responses. If ordered responses are used, *multinomial* should be specified and the response varaible should be *ordinal*.

5.9 Link Function (`link`)

The *link* argument specifies the link function used in the model. By default, the *logit* link function is specified, but the *probit* link function can also be used.

5.10 Theta Values (`theta`)

The *theta* argument specifies whether the Theta (person ability) estimates are displayed in the output. By default, this is set to *NO*, but this function can be turned on by setting it to *YES*.

5.11 Save (`save`)

The *save* argument specifies what is saved into the original dataset. The default value is *0*, which does not save anything. The Betas, Thetas, Fit Statistics (if specified; see below), and Predicted Probabilities can be saved into the dataset using a value of *1*, while a value of *2* only saves the Beta values, and a value of *3* saves the Beta and Theta values. Checking the *Save* checkbox is equivalent to selecting *1* using syntax.

5.12 Fit Statistics (`fit`)

The *fit* argument specifies whether item and person fit statistics are displayed in the output. By default, this argument is set to *NO*, but it can be changed to *YES*. This option is only available for the basic Rasch model (unidimensional single-group model without covariates) and, if selected, the response variable must be *numeric* and *ordinal*.

5.13 Plots (`plot`)

The *plot* argument can be used to generate item characteristic and information plots for unidimensional single-group designs by specifying *YES*. By default, this argument is set to *NO*. This option is only available for the basic Rasch model (unidimensional single-group model without covariates).

5.14 Dialogue Buttons

The *Help* button refers the reader to this user guide. The *Reset* button will reset the dialogue window to its default settings. The *Paste* button should not be clicked, as the code it pastes will not work in the syntax window. Users wishing to use the syntax for SPIRIT should use the syntax version, rather than the point-and-click version. The *Cancel* button will exit out of the SPIRIT dialogue box, and the *OK* button will conduct the analysis.

6 Using the Macro

The previous section defined each of the available macro arguments. Use of the point-and-click interface is straightforward. The model can be specified by dragging the variables to their appropriate boxes and choosing the intended options.
The syntax version can be utilized by typing *spirit* and then listing each argument name followed by an equal sign, the argument specification, and a forward slash. A period must conclude the code. Arguments not applicable to the chosen model, as well as arguments in which the defaults will be used, can be left out of the syntax. Example syntax corresponding to the multidimensional model with two person covariates from Figure 1 is displayed below:

```
spirit dataset = DataSet1/items = item/id = id
  /response = Ans/cov = Gender Anger
  /dim = situ/theta = YES.
```

where the italicized words are variables in the dataset. The arguments can be entered in any order, but the argument specifications are case-sensitive and should match those listed in the previous section and Table 2.

7 Output

This section provides a detailed overview of each of the output tables. Some models may not result in all tables being outputted. If there are additional tables and figures that are in the output that are not described in this section, it is likely that an older version of SPSS is being used. Any table or figure that is outputted, but not described in this section, can be ignored.

7.1 Random Effects

The Random Effects table for a unidimensional single-group analysis provides the variance of the theta values (person ability levels). If the model is multidimensional, variances for each dimension, as well as dimension covariances are provided. In this case, the label for each of the dimensions is provided in the small table above the Random Effects table. For example, UN(1,1) in the Random Effects table corresponds to the variance of the first dimension in the table displayed above the Random Effects table, while UN(2,3) corresponds to the covariance between the 2nd and 3rd dimensions listed in the above table. If a multi-group model is specified, the Random Effects table is split by the grouping variable.

7.2 Fixed Effects

The Fixed Effects table contains each of the item Betas (easiness parameters), or item covariate Betas if an item covariate was used instead of individual items. If person covariates are specified, their effects are displayed in this table as well. If the covariate is categorical, it consumes multiple rows of the table, and if it is continuous it consumes only one row. For multi-group designs, the grouping variable is present in the Fixed Effect table and the value associated with it represents group mean differences.

When a covariate is included within the model, the output will include an error message looking similar to:

```
Error # 4309 in column 1024.  Text:  (End of Command) Invalid combination
of data types in an assignment.  Character strings may only be assigned
to string variables.  Numeric and logical quantities may only be assigned
to numeric variables.  Use the STRING command to declare new string
variables before assigning values to them or consider using the STRING or
NUMBER function.
Execution of this command stops.
```

This error is simply due to attempts to modify the output *display*, and is not an indication of error in the output *content*. Consequently, users should ignore this warning. We are actively trying to eliminate this from the output for future versions of SPIRIT.

7.3 Item Statistics

The Item Statistics Table is only included if *fit=YES* is included in the model syntax or the *Fit Statistics* checkbox is checked in the macro dialogue box and can only be specified for the single-group unidimensional model without covariates. This table contains the Z_3 fit statistic, as well as Infit and Outfit statistics.

7.4 Person Statistics

The Person Statistics table is included if *theta=YES* is included in the model syntax or the *Theta Values* checkbox is checked in the macro dialogue box. For unidimensional models, there is one Theta value for each person, indicating the ability level for the given trait. If the model is multidimensional, an ability level for each person on each dimension is displayed. Further, if fit statistics are requested and the basic 1PL model has been specified, the Z_3 person fit statistic for each person is included.

7.5 Plots

For unidimensional single-group models without covariates, three plots are included if *plot=YES* is included in the model syntax or the *Plots* checkbox is checked in the macro dialogue box. The first plot contains the item characteristic curve for all items, and the second plot contains the item information curve for all items. The final plot contains the "test" information curve, which consists of the sum of the item information values for the range of theta values. Note that for the item and test information plots, the values are weighted by $1/Var(\theta)$. To compute the unscaled values, multiply the values in the plot by the estimated variance of θ.

8 Questions/Support

Please note that this is a Beta Version, in its first circulation. For SPIRIT support, questions, and recommendations, please email any of the authors at the email addresses provided on the title page. Thank You.

9 Disclaimer

This software is provided "as is", without warranty of any kind, express or implied, including but not limited to the warranties of merchantability, fitness for a particular purpose and noninfringement. In no event shall the copyright holders be liable for any claim, damages or other liability, whether in an action of contract, tort, or otherwise, arising from, out of or in connection with the software or the use or other dealings in the software.

10 Examples

Different 1PL models will now be investigated and run using the SPIRIT macro. The SPIRIT specification (using the point-and-click approach) of each model, as well as screenshots of the output, are provided. These illustrations will use the *VerbAgg* dataset that is included in the SPIRIT .zip folder. This dataset contains 316 respondents self-reporting on 24 items. Each item asks whether the respondent would respond (or would want to respond) verbally aggressively in the item's given hypothetical situation. There are two different response variables in the dataset: *resp* and *resp_di*. The *resp* variable is a polytomous response variable with three possible responses: "no", "perhaps", and "yes." The *resp_di* variable is the dichotomized version of *resp*, in which a response of "1" corresponds to either a "perhaps" or a "yes" response.

A nice property of this dataset is that it contains several item covariates: the *mode* variable describes whether that particular item asks if a respondent would "do" that behavior or would "want" to do that behavior. The *situ* variable describes whether the item's hypothetical situation is the respondent's fault or the fault of somebody else. Finally, the *btype* item variable describes the type of behavior (either "shout", "curse", or "scold") that the particular item is inquiring about. These covariates allow us to answer interesting questions about the patterns of responses in the dataset; for example, we can examine if respondents are more likely to be verbally angry when the frustrating situation is someone else's fault relative to if the situation is their own fault. There are also two person covariates in the model: the *Gender* of the respondent (male or female), and an *Anger* score of each respondent, which was measured from a separate anger index.

10.1 Rasch Model for Binary Data

The first model that will be examined is the most straightforward version of the one-parameter IRT family. It contains no item or person covariates and only one dimension, making it the easiest to comprehend and the simplest to run in SPIRIT. When the response variable is dichotomous and a logit link function is used, this model is often called the one parameter logistic (1PL) model, or the Rasch model. Within the basic 1PL model, the probability that person p answers correctly or aggressively (or with a "1") on item i can be modeled as a function of the sum of the easiness/intercept of item i (β_i) and the ability, or aggressiveness relative to other people in the dataset, of person p (θ_p).

This model is very simple to specify in SPIRIT. Figure 2 shows how to specify this model using the verbal aggression dataset.

Here, the response variable is *resp_di*, since we are modeling the dichotomized version of verbal aggression responses. The column containing the item indicator of each response is entered into the "Item" box. Here, that is simply the *item* variable in the dataset. Similarly, the *id* variable is included in the "ID" box, since this column signifies the individual answering the particular response. Since we are running the basic 1PL model, the "Covariates", "Dimension", and "Group" boxes are all left empty. We also must specify "Binomial" as the distribution, since we are working with the dichotomized *resp_di* variable. All four of the optional features can be selected for this model, so we will choose all four for this example.

Output for this model contains the variance of the person latent trait (seen in Figure 3) and the estimated item intercept parameters (seen in Figure 4). In SPIRIT, IRT models are parameterized such that each item intercept represents the *easiness* of an item; when β_i increases, the probability of a "successful" response increases as well. The SPIRIT output also contains (if desired) the estimated person trait effects, or θ_p values, of each respondent (seen in Figure 5). For this dataset, θ_p represents the tendency of individual p to respond verbally aggressively relative to other respondents. If specified, the item fit statistics (seen in Figure 6), item characteristic curves (Figure 7), and information for each item (Figure 8) are outputted. Fit statistics and plots are *only* available for this basic unidimensional, single-group 1PL model with no covariates.

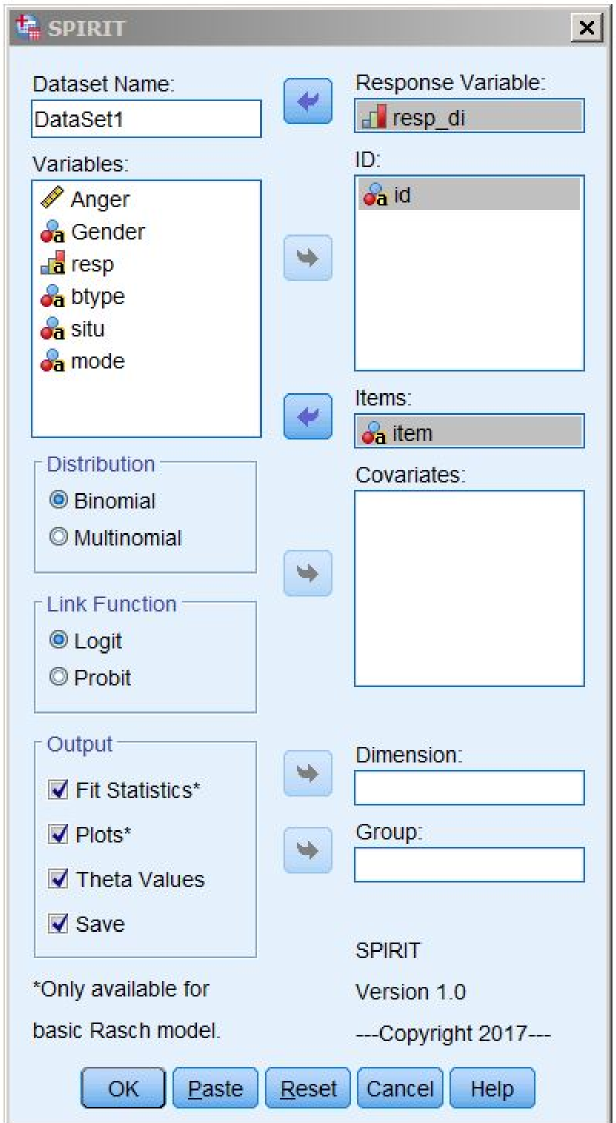

Figure 2: The appropriate SPIRIT dialogue box specification for the basic 1PL model.

Random Effect

Random Effect Covariance	Estimate	Std. Error	Z	Sig.	95% Confidence Interval Lower	95% Confidence Interval Upper
Var(id)	1.696	.170	10.001	.000	1.394	2.063

Covariance Structure: Variance components
Subject Specification: (None)

Figure 3: Random Effect estimates for the basic 1PL model.

Fixed Effects

		Beta	Std. Error	t	Sig.	95% Confidence Interval Lower	95% Confidence Interval Upper
Effect	S1DoCurse	1.167	.1576	7.404	.000	.858	1.476
	S1DoScold	.373	.1485	2.512	.012	.082	.664
	S1DoShout	-.832	.1533	-5.431	.000	-1.133	-.532
	S1WantCurse	1.167	.1576	7.404	.000	.858	1.476
	S1WantScold	.540	.1496	3.613	.000	.247	.833
	S1WantShout	.077	.1476	.523	.601	-.212	.367
	S2DoCurse	.834	.1525	5.471	.000	.535	1.133
	S2DoScold	-.054	.1477	-.363	.717	-.343	.236
	S2DoShout	-1.416	.1642	-8.628	.000	-1.738	-1.094
	S2WantCurse	1.669	.1693	9.860	.000	1.338	2.001
	S2WantScold	.677	.1507	4.490	.000	.381	.972
	S2WantShout	.012	.1476	.080	.936	-.278	.301
	S3DoCurse	-.201	.1480	-1.359	.174	-.491	.089
	S3DoScold	-1.438	.1647	-8.730	.000	-1.760	-1.115
	S3DoShout	-2.840	.2241	-12.676	.000	-3.280	-2.401
	S3WantCurse	.507	.1493	3.393	.001	.214	.799
	S3WantScold	-.656	.1512	-4.337	.000	-.952	-.359
	S3WantShout	-1.459	.1652	-8.833	.000	-1.783	-1.135
	S4DoCurse	.677	.1507	4.490	.000	.381	.972
	S4DoScold	-.367	.1488	-2.465	.014	-.658	-.075
	S4DoShout	-1.911	.1788	-10.688	.000	-2.261	-1.560
	S4WantCurse	1.034	.1553	6.658	.000	.730	1.339
	S4WantScold	-.333	.1486	-2.244	.025	-.625	-.042
	S4WantShout	-.998	.1557	-6.406	.000	-1.303	-.692

Figure 4: Fixed Effect estimates for the basic 1PL model.

Person Statistics

		Person Ability	Z3 Person Fit
id	1	-1.05	-5.72
	2	-3.07	-1.16
	3	-1.22	-5.56
	4	-.65	-5.65
	5	-1.27	-4.42
	6	-.28	-2.91
	7	-.04	-3.65
	8	-1.94	-2.32
	9	-1.01	-5.65
	10	.68	-3.48
	11	-.26	-5.36
	12	1.23	.01
	13	-.90	-3.01
	14	1.06	-1.41
	15	-.85	-6.12
	16	-.94	-5.40
	17	-.53	-7.89
	18	-1.32	-4.98
	19	-3.47	.56
	20	-.90	-4.50

Figure 5: Person Statistics for the basic 1PL model.

Item Fit Statistics

		Beta	Z3 Item Fit	Infit (Chi)	Infit (t)	Outfit (Chi)	Outfit (t)
item	S1DoCurse	.000	-8.31	1.49	7.77	1.54	4.88
	S1DoScold	-.716	-11.06	1.88	9.48	2.03	5.70
	S1DoShout	-1.913	-13.89	2.98	9.59	3.52	5.79
	S1WantCurse	.137	-7.66	1.41	6.86	1.59	5.54
	S1WantScold	-.404	-9.94	1.66	8.64	1.86	5.86
	S1WantShout	-1.071	-14.65	2.33	10.95	3.02	7.80
	S2DoCurse	-.221	-9.03	1.58	8.28	1.64	5.06
	S2DoScold	-1.150	-12.60	2.27	10.15	2.35	5.52
	S2DoShout	-2.485	-12.49	3.48	8.39	3.54	4.36
	S2WantCurse	.404	-6.49	1.32	5.90	1.44	4.66
	S2WantScold	-.353	-10.31	1.67	8.91	1.91	6.29
	S2WantShout	-.987	-12.99	2.15	10.21	2.54	6.70
	S3DoCurse	-1.458	-16.64	2.88	11.68	3.40	7.11
	S3DoScold	-2.563	-13.49	3.73	8.62	3.71	4.38
	S3DoShout	-3.910	-10.09	4.60	5.53	8.82	4.41
	S3WantCurse	-.756	-16.42	2.25	12.23	2.95	9.05
	S3WantScold	-1.857	-15.57	3.17	10.54	3.55	6.01
	S3WantShout	-2.584	-14.80	3.86	8.82	5.20	5.76
	S4DoCurse	-.582	-13.56	2.01	11.22	2.20	6.92
	S4DoScold	-1.498	-14.67	2.71	10.65	3.06	6.26
	S4DoShout	-2.986	-12.53	4.00	7.49	4.70	4.37
	S4WantCurse	-.281	-13.73	1.84	11.11	2.28	8.58
	S4WantScold	-1.448	-13.92	2.61	10.47	2.77	5.75
	S4WantShout	-2.060	-14.92	3.16	9.48	4.48	6.65

Figure 6: Item Fit Statistics for the basic 1PL model.

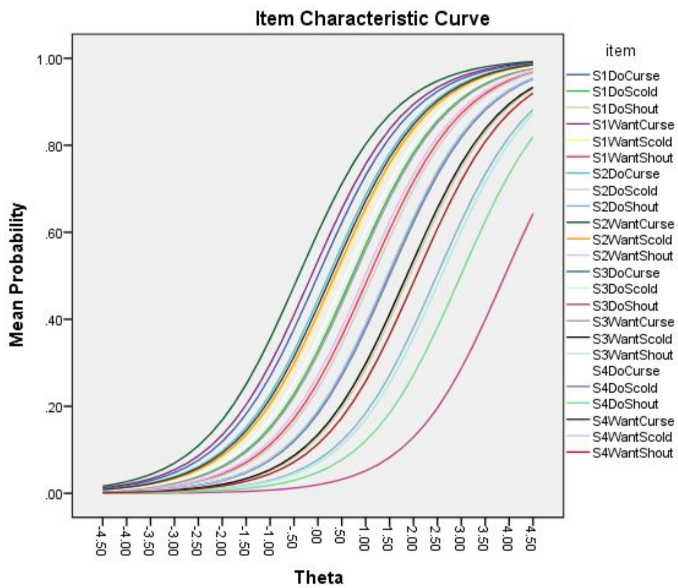

Figure 7: Item Characteristic Curves for the basic 1PL model.

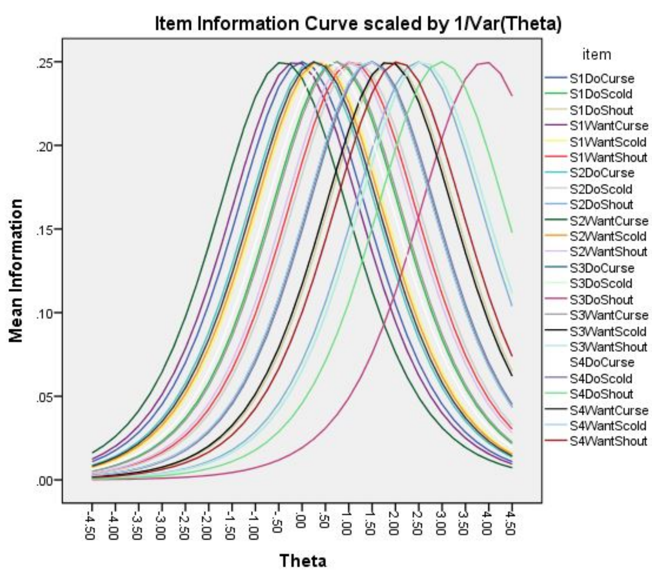

Figure 8: Item Information Curves for the basic 1PL model.

10.2 Rating Scale Model for Ordinal Data

In SPIRIT, the basic one-parameter framework can be extended for polytomous response variables by selecting the "Multinomial" option instead of "Binomial." This will run the *rating scale model*, which is currently the only option on SPIRIT for polytomous outcome variables.

Figure 9 shows how to specify the basic one-parameter rating scale model on SPIRIT. There are only three changes: the response variable is now the polytomous *resp* instead of the dichotomous *resp_di* column, the "Multinomial" option is now selected, and the "Fit Statistics" and "Plots" options are no longer selected, since they are not possible with the rating scale model.

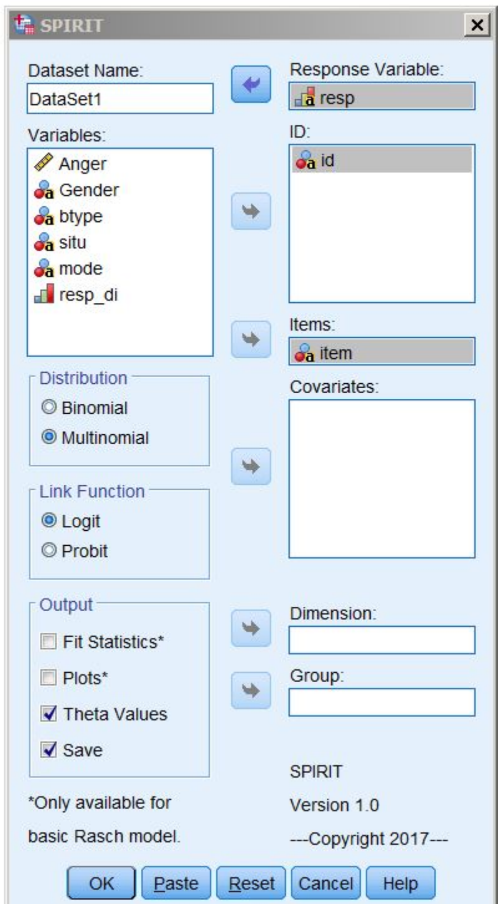

Figure 9: The appropriate SPIRIT dialogue box specification for the rating scale model.

The output of the rating scale model can be tricky to interpret at first. The output is the same as with the dichotomous 1PL, except now there are several more estimated fixed effects (as can be seen in Figure 10). The rating scale model gives $c - 1$ threshold parameters (where c is the number of possible response categories). For this model, this will result in two threshold parameters: one called "perhaps" and the other "yes." The "perhaps" threshold signifies the intercept for a response of anything "perhaps or greater" relative to exactly "no." For example, in this model's output, the "perhaps" threshold is estimated to be 1.134. When a rating scale model is run, SPIRIT specifies one particular item to be the reference item. For instance, this example model sets item *S1DoCurse* to be the reference item. This means that the intercept for answering perhaps or greater for *this specific reference item* is 1.134. All other item parameters are the *relative intercept change* for that item relative to the 1.134 reference intercept. For example, the estimated parameter for item *S1DoScold* is -0.716, suggesting that the intercept of answering perhaps or greater for this item is $1.134 - 0.716 = 0.418$.

The same logic holds for the "yes" threshold parameter, except this time the intercept is specifically for the probability that a given response to that item is exactly "yes." For example, the intercept for answering "yes" on the reference item *S1DoCurse* is -0.692, while the same intercept for item *S1DoScold* is $-0.692 - 0.716 = -1.408$.

For the rest of the appendix, all example models will be conducted on the dichotomous response variable *resp_di*. However, note that each of these models can be extended to the rating scale framework. The last example in Section 10.8 illustrates this flexibility by running a rating scale model with item covariates.

Fixed Effects

		Beta	Std. Error	t	Sig.	95% Confidence Interval Lower	95% Confidence Interval Upper
Effect	perhaps	1.134	.1323	8.565	.000	.874	1.393
	S1DoCurse	.000	.	.	.	.	.
	S1DoScold	-.716	.1569	4.564	.000	.408	1.023
	S1DoShout	-1.913	.1679	11.398	.000	1.584	2.242
	S1WantCurse	.137	.1561	-.878	.380	-.443	.169
	S1WantScold	-.404	.1560	2.589	.010	.098	.710
	S1WantShout	-1.071	.1588	6.744	.000	.760	1.382
	S2DoCurse	-.221	.1558	1.422	.155	-.084	.527
	S2DoScold	-1.150	.1593	7.215	.000	.837	1.462
	S2DoShout	-2.485	.1789	13.889	.000	2.135	2.836
	S2WantCurse	.404	.1570	-2.572	.010	-.712	-.096
	S2WantScold	-.353	.1559	2.268	.023	.048	.659
	S2WantShout	-.987	.1582	6.237	.000	.677	1.297
	S3DoCurse	-1.458	.1621	8.997	.000	1.140	1.776
	S3DoScold	-2.563	.1808	14.174	.000	2.209	2.918
	S3DoShout	-3.910	.2366	16.531	.000	3.447	4.374
	S3WantCurse	-.756	.1570	4.817	.000	.449	1.064
	S3WantScold	-1.857	.1670	11.117	.000	1.529	2.184
	S3WantShout	-2.584	.1814	14.246	.000	2.228	2.939
	S4DoCurse	-.582	.1564	3.724	.000	.276	.889
	S4DoScold	-1.498	.1625	9.220	.000	1.180	1.817
	S4DoShout	-2.986	.1933	15.449	.000	2.607	3.364
	S4WantCurse	-.281	.1558	1.805	.071	-.024	.587
	S4WantScold	-1.448	.1620	8.940	.000	1.130	1.765
	S4WantShout	-2.060	.1703	12.100	.000	1.726	2.394
	yes	-.692	.1318	-5.250	.000	-.950	-.434

Figure 10: Fixed Effect estimates for the rating scale model.

10.3 Model with Person Covariates

One-parameter IRT models can be extended to include other fixed effects covariates other than item indicator variables. IRT models containing person covariates and/or item covariates are called explanatory IRT models since the goal of the models is more focused on explaining an item response with an assortment of predictors as opposed to purely measuring person traits (De Boeck & Wilson, 2004).

One can include person-level covariates in IRT models. IRT models that contain person-level covariates are referred to as latent regression models (Mislevy, 1984, 1985). Covariates from latent regression models can be binary, categorical, or continuous, and represent properties of a respondent (such as age, gender, measured intelligence, etc.). When these covariates are included, it is possible to investigate the association between the person properties and the measured latent trait of a respondent.

Person covariates are simple to specify within SPIRIT. The specification is the same as it was for the basic 1PL model, except now the "Covariates" box must be utilized. Each person covariate that the user wants to include must simply be dragged into this box. Using the verbal aggression dataset, the effects of gender and a respondent's score on the included anger scale on the likelihood of a "successful" item response can be estimated by including *Gender* and *Anger* as person covariates. Figure 11 shows the specification of this model.

The output provided yields item easiness parameters for every item as well as the estimates for the effect of gender and a respondent's anger score. If desired, the person effects can also be outputted. However, note that these θ_p values are no longer a trait measure of the verbal aggression of person p. They should be considered residuals, since it is the aggression tendency for person p *controlling for his or her sex and anger score.*

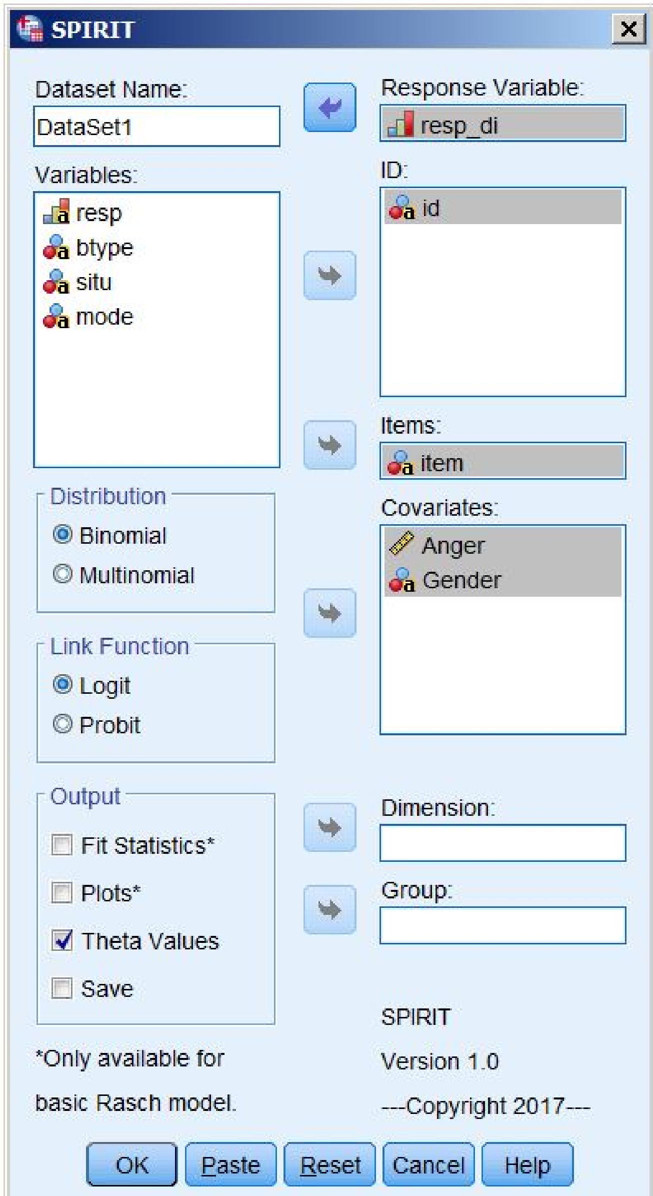

Figure 11: The appropriate SPIRIT dialogue box specification for the 1PL model with person covariates.

10.4 Model with Item Covariates

Explanatory IRT models including item property variables can also be run using SPIRIT. For these models, characteristics of the items themselves are treated as predictors, allowing researchers to investigate whether certain types of items are more or less likely to yield a response of "1" (a response of "yes", "correct", or however a "1" is defined). A good example of an item covariate comes from the verbal aggression dataset. Recall that half of the items from this dataset are construed as asking respondents whether they would *do* a certain behavior, while the other half ask whether the respondent would *want* to perform a certain behavior. As a result, half of the items can be thought of as "do" items and the other half as "want" items. A basic IRT model can be used to examine whether or not the type of item is relevant when modeling the probability of a "yes" response.

The simplest 1PL IRT model with item covariates is the linear logistic test model, or LLTM (Fischer, 1973). For the LLTM, there no longer is a coefficient for every individual item. Instead, there is a fixed effect parameter for each different type of item property within the dataset (with a total of K different item covariates being modeled). For example, there will now be an intercept for all "do" items and an intercept for all "want" items. Therefore, all items with that given item property will have identical "easiness" values; however, it is now possible to examine whether there are systematic differences between different types of items.

An LLTM using the want/do covariate is straightforward to specify in SPIRIT. A separate categorical variable column must be created which contains the property of the item for that specific row. The verbal dataset already contains this column, which is denoted as *mode.* To conduct this LLTM, the *mode* variable must be included into the "Covariates" box, as in Figure 12. When item covariates are included into a model, the regression coefficients for individual items (item intercept parameters) cannot be estimated. Therefore, the "Items" box of the SPIRIT interface *must* be kept empty. The LLTM model can be extended by adding more item covariates, or by including a random effect for each individual item. The model described in Section 10.8 explores both of these possibilities.

The output for this model is displayed in Figure 13. As expected, there is no longer a separate intercept for individual items. Instead, each subset of items has its own estimated intercept. For this example, there is a separate estimate

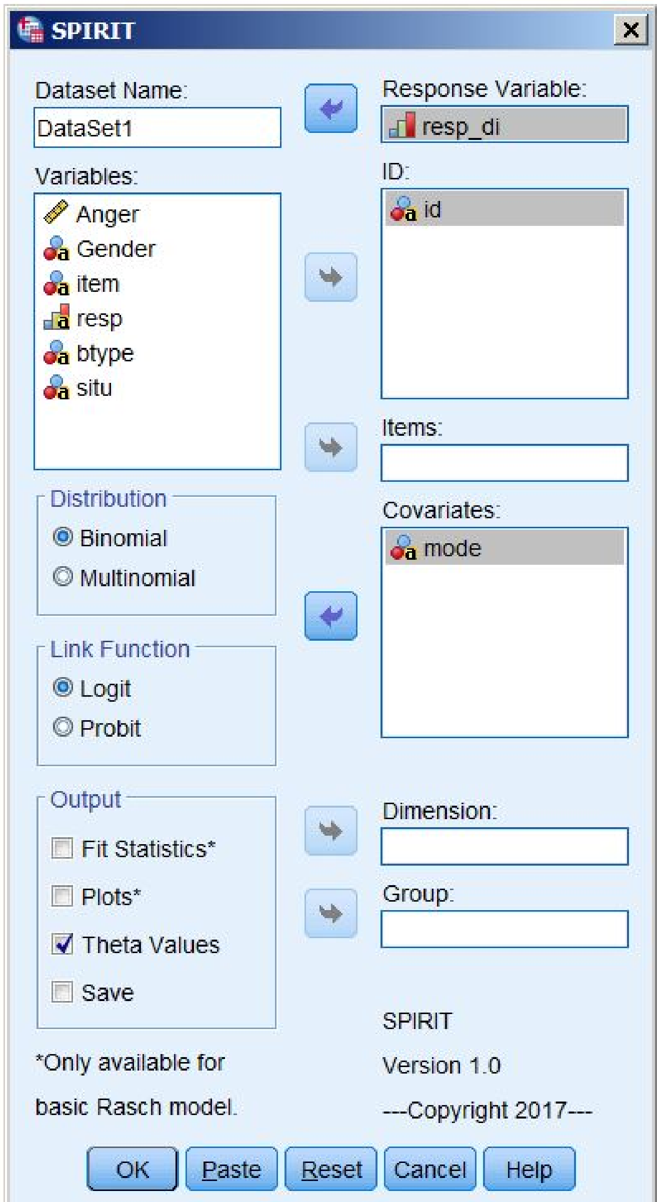

Figure 12: The appropriate SPIRIT dialogue box specification for the 1PL model with item covariates.

for "mode" and "want" items. It should be noted that it is also possible for SPSS to run a model in which one item type is treated as a reference intercept and estimates are calculated directly assessing the difference between the reference items and other item groups[1]. By running an item covariate model in this way, it is possible to use a hypothesis test to directly test whether intercept estimates from certain item groups are significantly different from the reference item group.

The person trait scores can also be outputted in this model. Contrary to the 1PL model with person covariates introduced earlier, the θ_p values obtained from the LLTM represent a latent value estimate for person p. This value can be interpreted as the tendency for person p to respond aggressively. However, for this specific model, it is assumed that a person's θ_p score is the same regardless of the type of item being answered.

Fixed Effects

		Beta	Std. Error	t	Sig.	95% Confidence Interval Lower	95% Confidence Interval Upper
Effect	do	-.382	.0695	-5.490	.000	-.518	-.245
	want	.158	.0693	2.285	.022	.022	.294

Figure 13: Fixed Effect estimates for the 1PL model with item covariates.

[1]This is not currently possible to specify within SPIRIT, although it is possible to manually create variables that accomplish this type of model.

10.5 Multidimensional Models

The 1PL models outlined up to this point assume that all items are a function of *one* latent person trait. All items in the dataset are therefore related equally to this one trait. However, one can easily think of scales or tests which have within it more than one different trait - or dimension - being measured. In this type of scale, one item may be a function of one person dimension, and another item may be related to a different person dimension.

In the IRT framework, models that contain more than one person dimension are called multidimensional IRT models. In these models, θ_p is now a vector, $\theta_{\mathbf{p}}$, with D elements, where D is the number of latent dimensions in the model. These person variables are assumed to follow a multivariate normal distribution with means of zero and an unstructured covariance matrix. Therefore, each respondent will have a separate measurement for each dimension.

Fortunately, the SPIRIT macro can conduct multidimensional models without much trouble. The one limiting condition (as least for now) is that only between-item multidimensional models can be run, where each item can only be "linked" to one dimension. *Within-item* multidimensional models in which items can load onto more than one dimension are not currently possible, but may be available on the next version of SPIRIT. However, any structure in which an item is linked to one dimension is possible. All that needs to be added to the data is a new column which denotes the dimensionality of each item.

For an example using the verbal aggression dataset, assume that researchers are interested in measuring two different latent traits for every respondent: one for a respondent's level of aggression in situations that are specifically his or her own fault ("self" items), and one for a respondent's aggression in situations in which it is someone else's fault ("other" items). Using a multidimensional IRT model, it is possible to estimate two variance parameter estimates, one for each of the latent dimensions as well as the covariance between the two dimensions. The specification of this model is very similar to other 1PL models. As Figure 14 shows, the only difference is the addition of the *situ* column into the "Dimension" box, since the *situ* column indicates whether a given response is the respondent's or someone else's fault.

The output for a multidimensional model will differ from a unidimensional model in two main ways. First, the elements of the covariance matrix of the

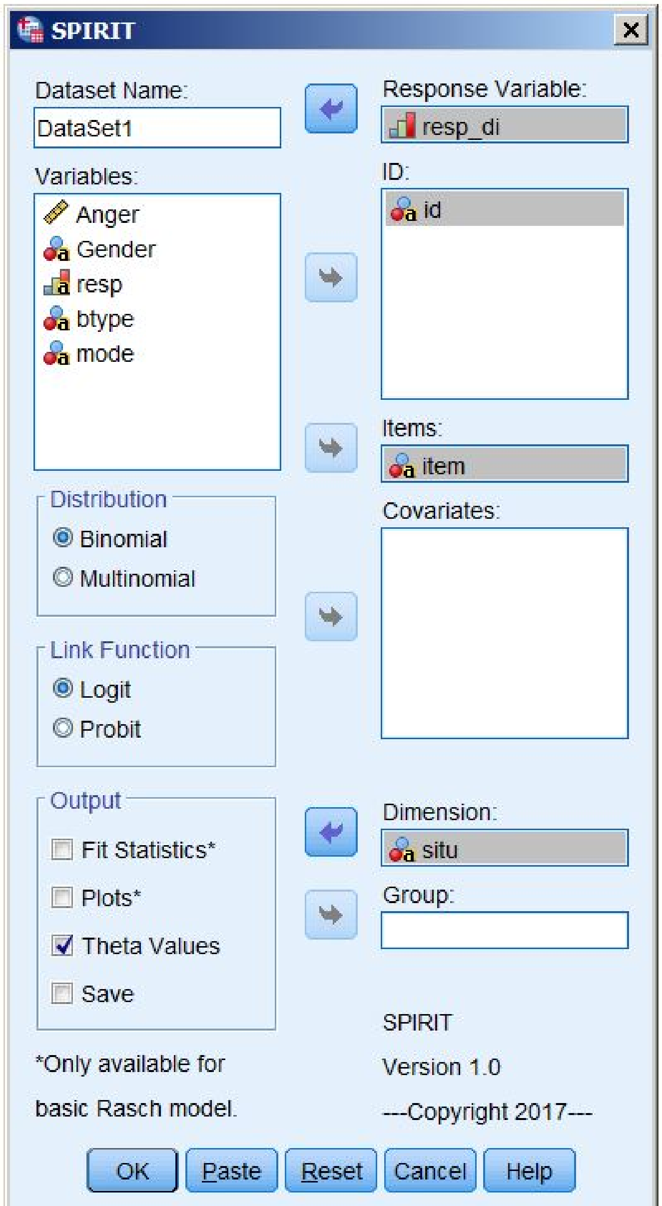

Figure 14: The appropriate SPIRIT dialogue box specification for the multi-dimensional 1PL model.

multiple dimensions will be given. Figure 15 shows what this looks like: using the box above the covariance matrix as a guide, UN(1,1) corresponds to the variance of the "self" dimension, UN(1,2) is the covariance between the two dimensions, and UN(2,2) corresponds to the variance of the "other" dimension. The other major output difference will be the given θ_p values, as there is now one for each separate dimensions. Figure 16 shows that there is now a person trait estimate for both the tendency to be angry in situations in which the respondent is at fault and situations in which someone else is at fault.

		Count
situ	self	3792
	other	3792

Random Effect

Random Effect Covariance	Estimate	Std. Error	Z	Sig.	95% Confidence Interval	
					Lower	Upper
UN (1,1)	1.555	.183	8.495	.000	1.234	1.958
UN (2,1)	1.573	.175	8.981	.000	1.230	1.916
UN (2,2)	2.265	.256	8.852	.000	1.815	2.826

Covariance Structure: Unknown
Subject Specification: id

Figure 15: Random Effect estimates for the multidimensional 1PL model.

Person Statistics

		situ	
		other	self
		Theta	Theta
id	1	-1.00	.00
	2	-2.86	-2.04
	3	.03	-.57
	4	1.36	-.10
	5	-.34	-.23
	6	-.43	.19
	7	1.36	-.10
	8	-1.78	-1.26
	9	-.71	.09
	10	2.27	1.54
	11	1.50	.16
	12	1.23	.95
	13	-.15	-.40
	14	2.09	1.24
	15	.14	.38
	16	.04	.12

Figure 16: Person Statistics for the multidimensional 1PL model.

10.6 Multigroup Models

Another application of the 1PL IRT model is the multigroup model, which allows for the mean and variance of the latent trait distribution to differ depending on the group membership of respondents. This type of model assumes that respondents are sampled from different populations - therefore, the mean response and the random effect variance will potentially be different for each group. This can be seen as an extension of the latent regression model with a categorical person covariate, since it not only calculates the mean difference between groups using a group membership variable, but it also allows for the latent trait variances to differ from one group to the next.

A straightforward example of a multigroup model using the verbal aggression data involves treating the *Gender* of the respondent as the grouping variable. We therefore are assuming that male and female respondents could be different in how aggressive they self-report to be, and will potentially differ not only in how likely they are to self-report aggressive behavior, but also in the amount of variability in the respective populations. The output for this model will therefore have two latent trait variances, one for each group.

To specify this multigroup model in the SPIRIT macro, simply insert the *Gender* variable into the "Group" box in the point-and-click interface. By doing so, a person covariate will automatically be entered into the model for the grouping variable, and the variance of the latent trait will be calculated separately for each of the groups. This model will therefore have an item easiness parameter for every item, a person covariate for *Gender*, and two latent estimate variances - one for each gender. Figure 17 displays visually how to specify this model.

The output provides a fixed effect for gender, just like when a latent regression model is run with a person covariate. As displayed in Figure 18, there is now two random effect variances in the output: one for males and one for females.

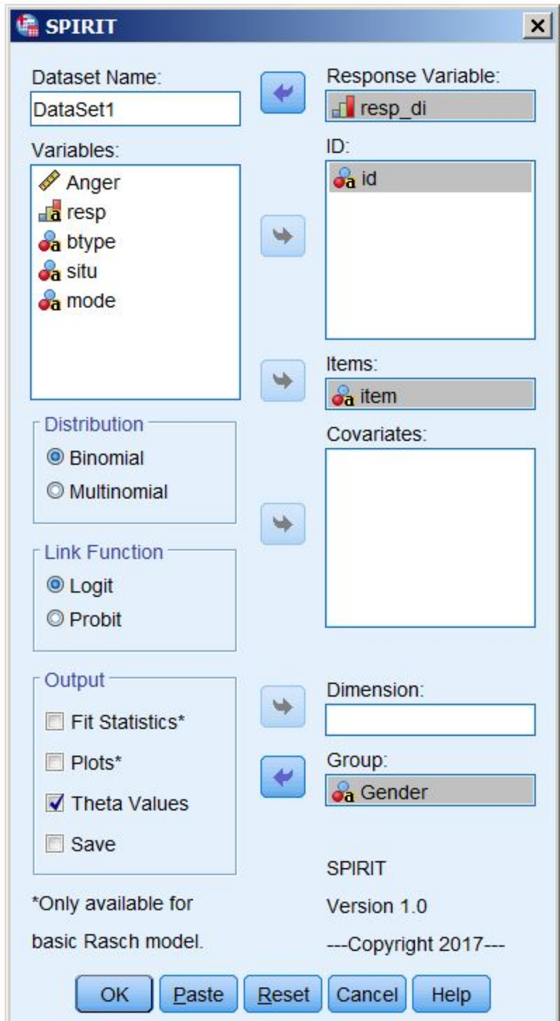

Figure 17: The appropriate SPIRIT dialogue box specification for the multigroup 1PL model.

Random Effect

Grouping	Random Effect Covariance	Estimate	Std. Error	Z	Sig.	95% Confidence Interval	
						Lower	Upper
Gender=F	Var(id)	1.772	.200	8.841	.000	1.419	2.211
Gender=M	Var(id)	1.421	.296	4.799	.000	.944	2.138

Covariance Structure: Variance components
Subject Specification: (None)

Figure 18: Random Effect estimates for the multigroup 1PL model.

10.7 DIF models

SPIRIT can also be used to examine differential item functioning - or DIF (for a discussion of the different types of differential item functioning, see Holland and Wainer (2012)). DIF occurs when the effect of an item is different for two respondents who have equal ability but differ on some person property. This item difference can be modeled by examining how the specific item in question interacts with the person covariate responsible for the difference. This interaction term indicates the quantitative estimate for DIF.

An example of DIF is relatively simple to consider. For the verbal aggression dataset, perhaps a certain item will have a differential effect on male respondents as opposed to females. Therefore, one simple β_i for the item is no longer sufficient to describe its intercept, since it functions differently depending on the gender of the respondent. A new fixed coefficient must now be added to the model that takes this difference into account. To do this, a new variable must be created in the data that is equal to "1" for responses to the item of interest answered by a member of the group of interest, and "0" otherwise. Typically, this type of model also calculates a different variance estimate for each person group, much like a multigroup model.

To illustrate a specific example, let us examine whether the "S4WantShout" item of the verbal aggression dataset contains DIF by gender. If DIF is present, it would suggest that males and females of the same overall aggression tendency (equal θ_p values) would expect to have different probabilities of answering with a "yes" to this item, due to the interaction between their gender and this specific item. Before we conduct any IRT analyses, a new variable must be constructed, which will be called d. This column is equal to 1 in the dataset for every response to item "S4WantShout" answered by a male. The variable is equal to zero for every other observation. Figure 19 shows the specification of this DIF model; note that the d column in included in the "Covariates" box, and "Gender" is being used as a grouping variable.

The only difference in the output (seen in Figure 20) is in the Fixed Effects table, which now contains the estimate of DIF for item "S4WantShout." In this example, the estimate for DIF is $\delta = -0.689, SE = 0.342, p = 0.044$. This estimate signifies the difference in the "S4WantShout" parameter between males and females who have the same latent trait level. Since this coefficient is significant, it suggests that DIF is present for this item, meaning this item behaves differently depending on a respondent's sex controlling for latent trait

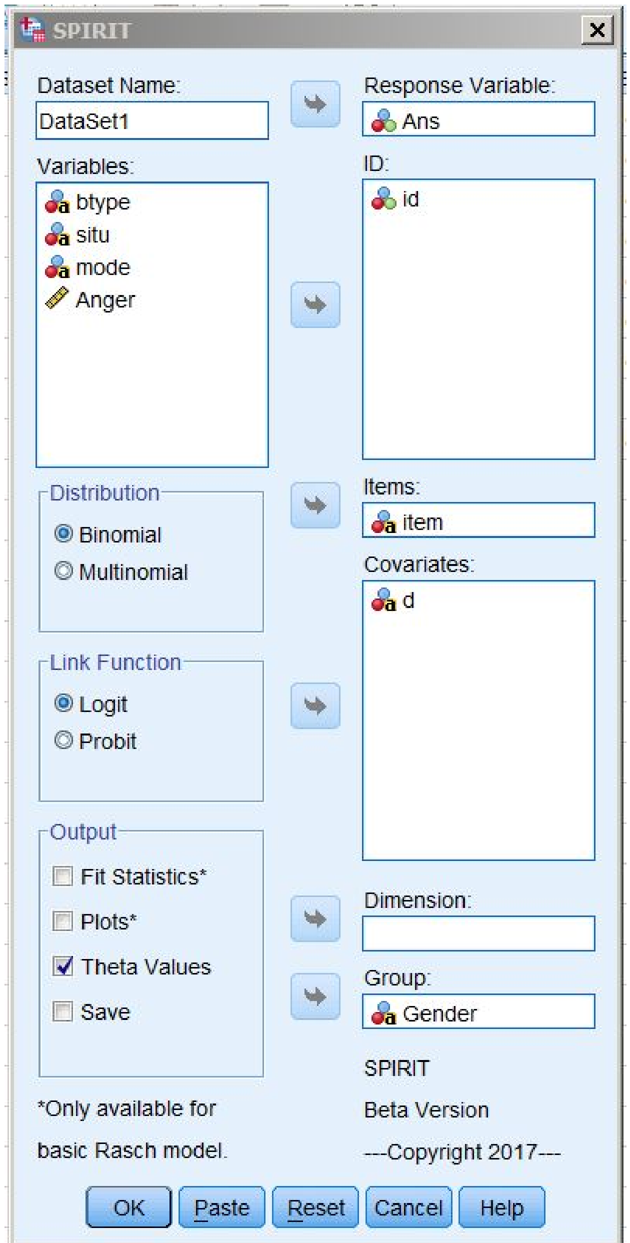

Figure 19: The appropriate SPIRIT dialogue box specification for the DIF model.

Fixed Effects

		Beta	Std. Error	t	Sig.	95% Confidence Interval Lower	95% Confidence Interval Upper
Effect	.00	.000	.	.	.	.	.
	1.00	-.689	.3415	-2.017	.044	-1.358	-.019
	F	.000	.	.	.	.	.
	M	.317	.1785	1.774	.076	-.033	.667
	S1DoCurse	1.094	.1642	6.660	.000	.772	1.416
	S1DoScold	.299	.1556	1.921	.055	-.006	.604
	S1DoShout	-.907	.1603	-5.659	.000	-1.222	-.593
	S1WantCurse	1.094	.1642	6.660	.000	.772	1.416
	S1WantScold	.467	.1566	2.979	.003	.160	.774
	S1WantShout	.003	.1549	.019	.985	-.301	.306
	S2DoCurse	.761	.1594	4.773	.000	.448	1.073
	S2DoScold	-.128	.1549	-.826	.409	-.432	.176
	S2DoShout	-1.491	.1708	-8.732	.000	-1.826	-1.157
	S2WantCurse	1.598	.1755	9.104	.000	1.254	1.942
	S2WantScold	.603	.1578	3.824	.000	.294	.913
	S2WantShout	-.063	.1548	-.404	.686	-.366	.241
	S3DoCurse	-.276	.1552	-1.776	.076	-.580	.029
	S3DoScold	-1.513	.1713	-8.832	.000	-1.849	-1.177
	S3DoShout	-2.915	.2290	-12.730	.000	-3.364	-2.467
	S3WantCurse	.433	.1564	2.767	.006	.126	.739
	S3WantScold	-.730	.1583	-4.615	.000	-1.041	-.420
	S3WantShout	-1.534	.1718	-8.931	.000	-1.871	-1.198
	S4DoCurse	.603	.1578	3.824	.000	.294	.913
	S4DoScold	-.441	.1560	-2.829	.005	-.747	-.136
	S4DoShout	-1.986	.1849	-10.740	.000	-2.348	-1.623
	S4WantCurse	.961	.1621	5.929	.000	.643	1.279
	S4WantScold	-.408	.1558	-2.619	.009	-.714	-.103
	S4WantShout	-.912	.1773	-5.142	.000	-1.259	-.564

Figure 20: Fixed Effect estimates for the DIF model.

levels. Specifically, males tend to respond significantly less aggressively to this item compared to the typical expected male response on all other items.

10.8 Complex Model: Rating Scale Model with Item Covariates

SPIRIT is very flexible in its ability to mix and match many of the models described above. To illustrate this flexibility, an example model that takes into account some of the SPIRIT model options described above will be presented here. Consider an IRT model containing two item covariates: one for the mode of behavior (do vs. want) and one for the type of situation (respondent's fault vs. other's fault). Since there are no separate fixed intercepts for individual items in this model, it would be sensible to instead include a random error term for items in the model. To do this in SPIRIT, simply add the *item* variable into the "ID" box. The *id* variable should also be included in this box to allow for random effects (θ values) for every respondent to be included. Therefore, the "ID" box can be considered the box for any type of *random* effect for a given model[2]. This type of model is similar to a random-item LLTM in which the random effects are cross-classified rather than nested (e.g., De Boeck, 2008; Janssen, Schepers, & Peres, 2004). We will model the *resp_di* polytomous response variable here. Figure 21 shows how to specify this model.

The output of this model contains the random effect variances and the Fixed Effects table, which can be seen in Figure 22. The interpretation of the values is similar to the interpretations described in section 10.2. Now though, indicator coding is used for the item covariates, where "do" and "other" items are treated as the reference items[3] (since they are the reference groups, the Beta "intercepts" for these items are constrained to be zero, as the output shows). For the *mode* variable, "want" items have a higher intercept than "do" items, suggesting that, in general, respondents are more likely to *want to* respond verbally aggressively than actually being verbally aggressive. The *situ* effect is significant: respondents are much more likely to be verbally aggressive in situations in which someone else is to blame.

As this illustration shows, there are many different models that SPIRIT can run that address a multitude of research questions. If there are any questions about how to run a particular model or how to interpret specific output, do not hesitate to contact the authors.

---

[2]Theta values can only be outputted in SPIRIT if only *one* random effect is specified.

[3]In this case, SPSS automatically chose the reference groups based on alphabetical order. Users can manually create a variable of 0's and 1's if they wish to always control which value is the reference group of a nominal predictor.

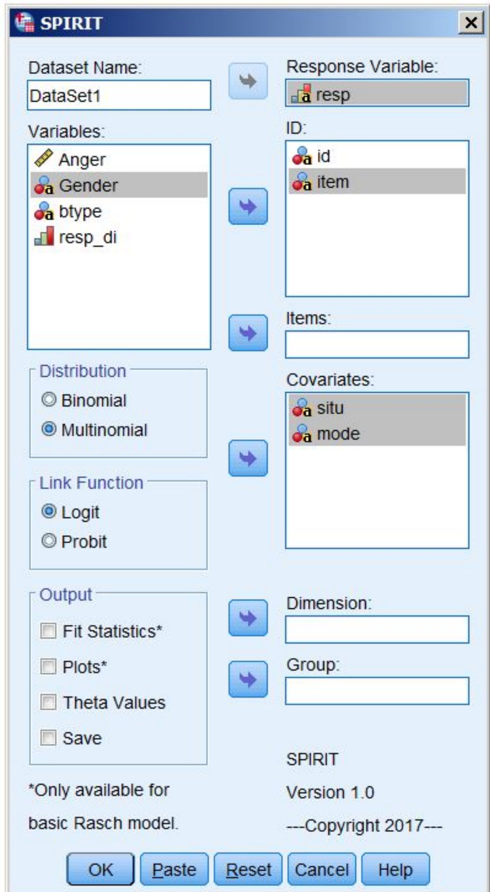

Figure 21: The appropriate SPIRIT dialogue box specification for the rating scale model with item covariates.

Fixed Effects

		Beta	Std. Error	t	Sig.	95% Confidence Interval Lower	95% Confidence Interval Upper
Effect	do	.000	.	.	.	.	.
	other	.000	.	.	.	.	.
	perhaps	.064	.3276	.196	.845	-.578	.706
	self	-1.094	.3696	2.960	.003	.369	1.818
	want	.678	.3695	-1.834	.067	-1.402	.047
	yes	-1.755	.3283	-5.347	.000	-2.399	-1.112

Figure 22: The appropriate SPIRIT dialogue box specification for the rating scale model with item covariates.

References

De Boeck, P. (2008). Random item irt models. *Psychometrika*, *73*(4), 533.

De Boeck, P., & Wilson, M. (2004). *Explanatory item response models.* New York, NY: Springer.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, *359-374*, 359-374.

Holland, P. W., & Wainer, H. (2012). *Differential item functioning.* Hillsdale, NJ: Erlbaum.

Janssen, R., Schepers, J., & Peres, D. (2004). Models with item and item group predictors. In *Explanatory item response models* (p. 189-212). Springer.

Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika*, *49*(3), 359–381.

Mislevy, R. J. (1985). Estimation of latent group effects. *Journal of the American Statistical Association*, *80*(392), 993–997.