Introduction to BILOG-MG - practically lifted from the program's help section

About BILOG-MG

BILOG-MG is an extension of the BILOG program that is designed for the efficient analysis of binary items, including multiple-choice or short-answer items scored right, wrong, omitted, or not presented. BILOG-MG is capable of large-scale production applications with unlimited numbers of items or respondents; it can perform item analysis and scoring of any number of subtests or subscales in a single program run. All the program output may be directed to text files for purposes of selecting items or preparing reports of test scores.

The BILOG-MG program implements an extension of item response theory (IRT) to multiple groups of respondents. It has many applications in test development and maintenance. The program performs the same analyses as BILOG in the one-group case.

New features in BILOG-MG 3:

The most important change is that BILOG-MG is now a Windows program. Syntax can be generated or adapted using menus and dialog boxes or, as before, in the format of syntax files in text format. The interface has menu options in the order the user would most generally use: model specification is followed by data specification and technical specifications, etc. Each of the menu options provide access to a number of dialog boxes on which specifications can be made by the user.

Phases of analysis

Phase 1: INPUT

The input routine reads formatted data records. Data for each observation consist of subject identification, optional form number, and optional group number(s), optional case weight, and item response data. Item responses of individual examinees comprise one character for each of n items. The answer key, not-presented, and omit codes are read in exactly the same format as the observations.

This section contains information on:

Omits and attempts

Omits may be scored "wrong", treated as fractionally correct, or omitted from calculations.

Classical item statistics

While preparing the item-score file, the INPUT routine also accumulates, subtest by subtest, certain item and test statistics (accumulated from the sample file when the number of cases exceeds the user-specified sampling level). These statistics consist of

1. item facilities (percent correct)
2. item-subscore correlations
3. number of respondents attempting each item.

These quantities are listed and passed to Phase 2 and Phase 3 routines to provide starting values for item parameter and respondent scale-score estimation.

Phase 2: CALIBRATE

The CALIBRATE routine fits a logistic item-response function to each item of each subscale. There are many options available to the user in this section of the program.

This section contains information on:

Item-response model

The response model may be the 1-, 2- or 3-parameter logistic response function. The scaling factor $D = 1.7$ employed to scale estimates in the normal metric may be included or omitted at the user's option. Information that assists the user in model selection is provided in the marginal log likelihood and goodness of fit indices and statistics for individual items. The user may request plots of the observed and expected item-response curves.

Marginal maximum likelihood (MML) estimation of the parameters

Estimation of item parameters by the method of marginal maximum likelihood is applicable to test of three or more items. The solution assumes the respondents are drawn randomly from a population or populations of abilities, which may be assumed to have either a normal distribution, an arbitrary distribution specified by the user, or an arbitrary distribution to be estimated jointly with item parameters. The empirical distributions of ability are represented as discrete distributions on a finite number of points (histogram).

The MML solution employs two methods of solving the marginal likelihood equations: the so-called EM method and Newton-Gauss (Fisher scoring) iterations. The default number of cycles for the EM algorithm is 10: the default for Newton steps is 2. The information matrix for all item parameters is approximated during each Newton step and then used at convergence to provide large-sample standard errors of estimation for the item parameter estimates.

Item statistics supplied by CALIBRATE

Phase 2 of the program provides the item parameters
1. lower asymptote
2. item intercept (equal to minus the product of the slope and threshold) and so-called "slope" or "discrimination" parameter
3. item threshold (location) and loading (one-factor item factor loading = $slope/\sqrt{1.0 + slope^2}$).

In Phase 2, when there is a single group, the unit and origin of the scale on which the parameters are expressed is based on the assumption that the latent ability distribution has zero mean and unit variance. This is referred to as the "0, 1" metric.

Maximum marginal a posteriori estimation of item parameters

When some items are extremely easy or extremely difficult, there may be insufficient information in the sample to estimate their parameters accurately. This will be especially true if the number of respondents is only moderate (250 or fewer). As an alternative to deleting these items, prior distributions can be placed on the item parameters. The user may specify normal priors for item thresholds, log-normal priors for slopes, and beta priors for lower asymptotes. Each item may have a different specification for its prior.

By specifying tight priors on selected item parameters, the user may hold these values essentially fixed while estimating other item parameters. This feature is useful in linking studies, where new test items are to be calibrated into an existing scale without changing parameter values for old items.

Item fit statistics

Approximate chi-square indices of fit are computed for each item following the final estimation cycle. For purposes of computing these chi-squares, the scale score continuum is divided into a number of successive intervals convenient for displaying the response proportions (maximum of 20). Each respondent is assigned to the interval that includes the EAP estimate (based on the type of prior specified by the user) of his or her score. For the item in question, the expected response probabilities corresponding to the average EAP estimate of ability of cases that fall in the interval are used as the expected proportion for the interval.

A likelihood ratio chi-square is then computed after combining extreme intervals so that the expected frequency exceeds five. Degrees of freedom are equal to the number of combined intervals. There is no reduction in degrees of freedom due to estimating the item parameters because the marginal maximum likelihood method does not place linear constraints on the residuals.

At the user's request, observed and expected item-response curves are plotted for each item.

Test of improved fit if the number of parameters is increased

When the expected frequencies of the individual response patterns are too small to justify the likelihood ratio test of goodness-of-fit, the change in likelihood ratio chi-square between the 1- and 2-parameter models, or between the 2- and 3-parameter models, is a valid large-sample test of the hypothesis that the added parameters are null. The degrees of freedom of each of these change chi-squares are equal to the number of items.

Phase 3: SCORE

The SCORE routine makes use of the master response file from Phase 1 and the item parameter estimate files from Phase 2 to compute estimated scale scores for respondents. The user may select one of the three methods described below for calculating scale scores.

Scores and standard errors for all subscales are calculated simultaneously for each respondent. Results may be printed and/or saved on an external file.

This section contains information on

Maximum likelihood (ML)

ML estimates with or without robustification are computed by the Newton-Raphson method starting from a linear transformation of the logit of the percent-correct score for the subject.

Estimates for respondents with all correct or all incorrect responses are attributed by the half-item rule. That is, respondents who score all incorrect are assigned one-half a correct response to the easiest item; respondents who score all correct are assigned one-half a correct response to the hardest item. The estimate is then computed from this modified response pattern.

Standard errors are computed as the square root of the negative reciprocal of the expected second derivative of the log likelihood at the estimate, i.e., square root of the reciprocal Fisher information.

Bayes or expected a posteriori (EAP)

EAP estimates with or without robustification are computed by quadrature using a discrete distribution on a finite number of points as the prior. The user may select the number of points and has the choice of a normal, locally uniform, or empirical prior. For the latter, the user may supply the values of the points and the corresponding empirical weights or may use the empirical weights generated in Phase 2.

The EAP estimate is the mean of the posterior distribution; the standard error is the standard deviation of the posterior distribution.

Bayes modal or maximum a posteriori (MAP)

MAP estimates with or without robustification are also computed by the Newton-Gauss method. This procedure always converges and gives estimates for all possible response patterns. A normal prior distribution with user-specified mean and variance is assumed [the default is $N(0, 1)$]. The estimate corresponds to the maximum of the posterior density function (mode); and the standard error is the square root of the negative reciprocal of the curvature of the density function at the mode.

Rescaling

The ability estimates are calculated initially in the scale of the item parameter estimates from Phase 2. In addition, however, rescaled estimates may be obtained by one of the following options:

1. the mean and standard deviation of the sample distribution of score estimates are set to arbitrary values specified by the user (default = 0, 1)
2. a linear transformation of scale is provided by the user.

Item and test information tables and curves

BILOG-MG provides at the user's request a number of indices and plots concerning item and test information:

1. Plots of test information and standard error curves for each subtest
2. Tables of item information indices, including the point and value of maximum information.

Output Files

Phase 1 results appear in the *.ph1 file. They include test and item identification and classical item statistics.

Phase 2 results appear in the *.ph2 file. They include assumed prior distributions, estimated item parameters, standard errors and goodness-of-fit statistics, DRIFT parameters, estimates of differential item functioning, posterior distributions for the groups, group means and standard deviations, and estimates of their standard errors.

Phase 3 results appear in the *.ph3 file. They include assumed prior distributions of the scale scores for MAP and EAP estimation, correlations among the subtest scores, rescaling constants, rescaled item parameters, scale scores for the subjects, test information plots, and parameters of the rescaled latent distribution.

Overview of the BILOG-MG interface

This version of BILOG-MG for Windows allows the user to:
1. Create and run a new command file using menus and dialog boxes.
2. Edit generated syntax in a syntax editor with all the basic Windows options such as copy, cut, paste, etc.
3. Type in syntax and create a command file in the same way as was the case with the DOS version of BILOG-MG.
4. View, edit and print output for each phase of the analysis from within the program.

When BILOG-MG is opened the first time, a blank window is displayed with only three active options: File, View and Help. By default, however, BILOG-MG will open with the last active file displayed. In this case, or when a syntax file is opened through BILOG-MG, the main menu bar shown below is displayed.
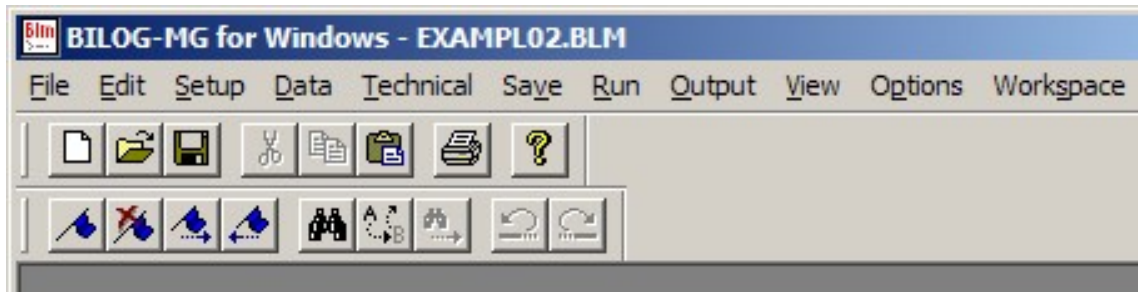
There are 13 options available on the main menu bar. The options and their main purpose are summarized in the table below.

Option	Purpose
File	Creating or opening files, printing files and exiting the program.
Edit	Standard Windows editing functions.
Setup	Model specification.
Data	Description of the data, with option to enter new data.
Technical	Specifying starting values and priors for calibration and/or scoring.
Save	Saving output to external files.
Run	Generating syntax and running one or all phases of the program; accessing the graphics procedure.
Output	Viewing output files for the current analysis.
View	Show or hide the tool bar and status bar.
Options	Changing program settings and user preferences.
Workspace	Changing the workspace required for analysis
Window	Switching between open files.
Help	Access to the online help, build number and contact information for SSI.

Getting Started with BILOG-MG

To illustrate the use of the interface in creating syntax files, the data file example1.dat is used. This problem is based on an example in Thissen, Steinberg & Wainer (1991). Other examples based on the same data can be found in the Examples section.

The data are drawn from a 100 word spelling test administered by tape recorder to psychology students at a large university. The words for the test were randomly selected from a popular word book for secretaries. Students were asked to write the words as used in a sentence on the tape recording. Responses were scored 1 if spelled correctly and 0 if spelled incorrectly. Because the items are scored 1/0 according to the defaults assumed by the program, an answer key is not required.

A single-group 3PL model is discussed in this section. Only the essential keywords are set to values through the interface; for the rest the program defaults are used.

The purpose of this section is to give the new user a quick overview of the interface and the minimum input needed to run the program. In the other examples, the syntax and keywords are discussed in detail. Additional options for model specification and the link between the interface and the syntax is discussed afterwards.

A few selected lines of the data are shown below:

```
11 0000
………
21 0001
………
31 1000
………
41 1001
………
162 1111
```

A first model: 3PL model for the Spell data

As a first example, we wish to set up a simple 3PL model for this data. To construct a syntax file, begin by selecting the New option from the File menu. The Open dialog box appears.
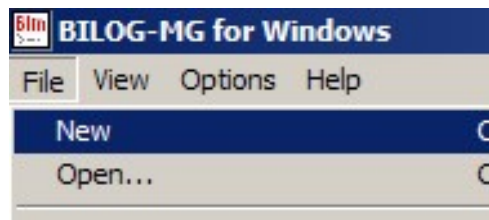
Assign a name, with the *.blm file extension, to the syntax file. In this case, the syntax file spell.blm is created in the examples folder. Click Open when done to return to the main BILOG-MG window.

Note that several options have been added on the main menu bar of the BILOG-MG window. Of interest for this example is the Setup, Data, Run and Output options. The Setup menu is used to describe the model to be fitted to the data. As a first step, select the General option from this menu to access the General dialog box as shown below.
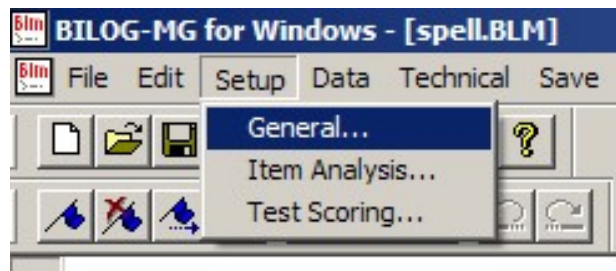
The General dialog box has four tabs, on which both required and optional keywords may be set. On the Job Description tab below, the number of items in the test is indicated as 4.

The type of model is selected on the Model tab. Choose the 3PL model. (As the default model fitted by BILOG-MG is a 2PL model). Click OK to return to the main window.
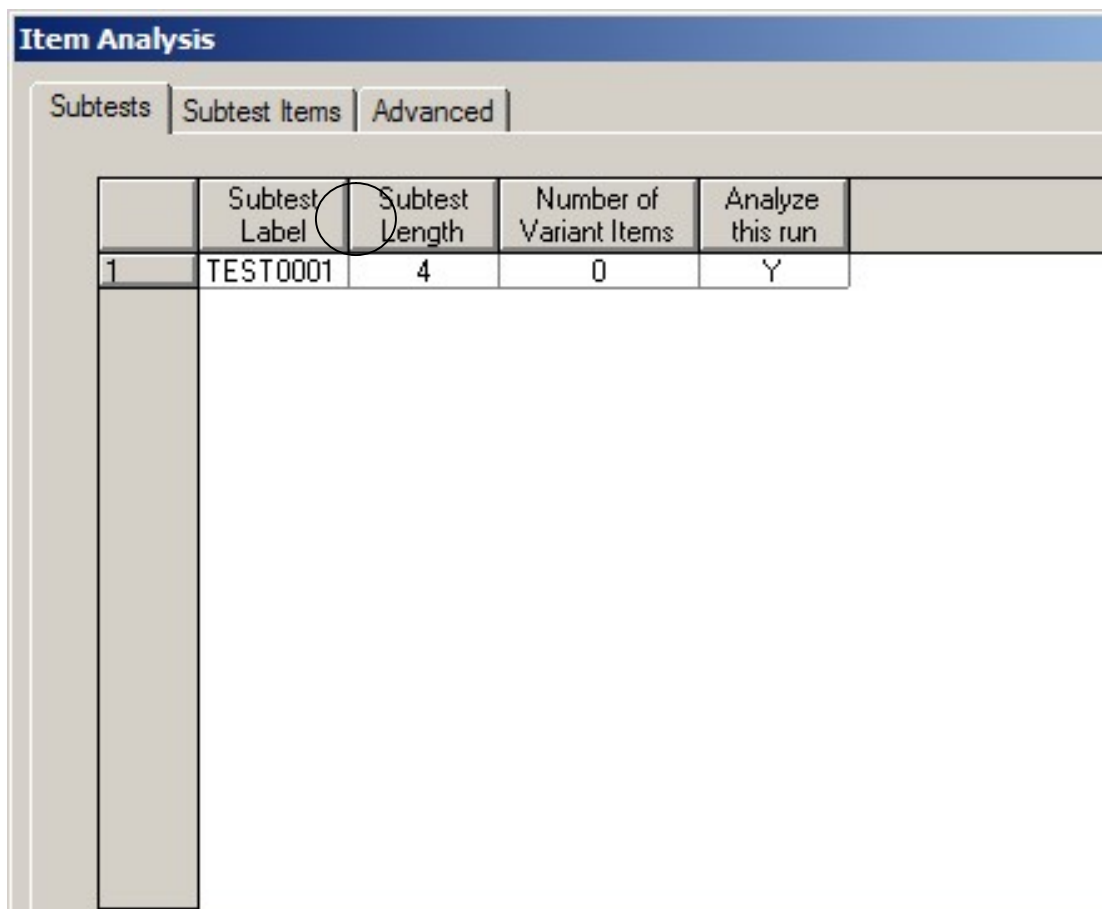
The next step in specifying the analysis is to assign the items to be calibrated to the test. This is done using the Item Analysis option on the Setup menu.

The Item Analysis dialog box appears. Change the default value of 1 under Subtest Length to 4 by clicking in this field and typing in "4". By default, all items will be analyzed, as indicated under the Analyze this run header. Click OK when done. In the Model tab, specify the 3PL model.

This completes the model specification. All that remains to be done is to provide information on the data. To do so, the Data menu is used. In this case, we have examinee data and thus the Examinee Data option is selected from the Data menu. As mentioned previously, there is no need for item keys in this example, so the Item Keys option will not be used now.

On the Examinee Data dialog box, enter the number of characters representing the examinee identification (in this case 3) in the Number of Case ID Characters field. By default, all data are used.

To provide information on the name and format of the data file, click the Data File tab.

- Use the Browse button to browse for the data file.
- Next, indicate that the data are in fixed format by clicking the Read as Fixed-Column Records radio button.
- Complete the table in the Read as Fixed-Column Records group box by clicking in the cell next to Case ID under the First header. Enter a "1" here to indicate that the examinee identification starts in column 1 of the data. Next, enter a "3" under the Last header to indicate the end of the examinee identification.
- Note that Form Number, Group Number, etc is disabled due to the information we entered from the Setup menu. The only other field to complete is the Response String fields.
- The response to the first item is in column 5 of the data, so a "5" is entered in the cell next to Response String under the First header. The response to the last item is in column 8, and a "8" is thus entered under the Last header.
- By default, BILOG-MG assumes that there is one line of data for each examinee, as indicated by the Number of Data Record per Case field. As this is the case for data in
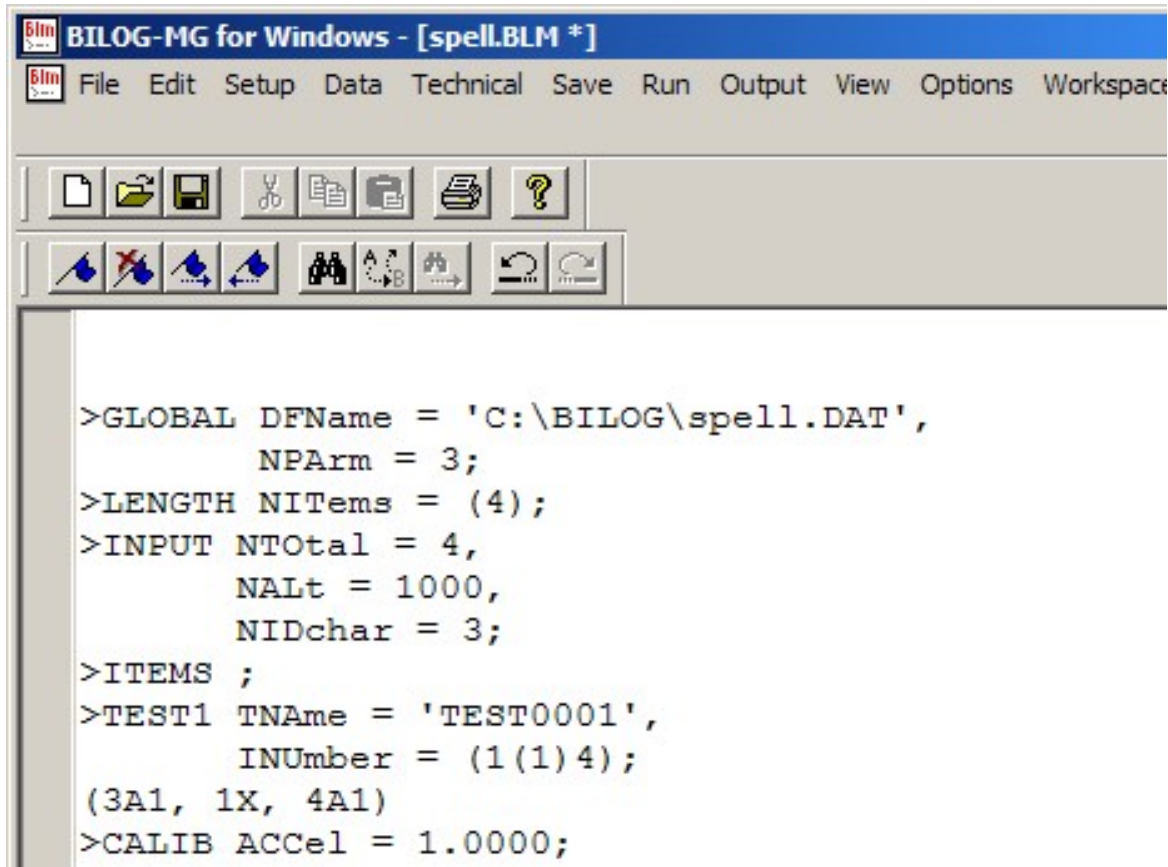
spell.dat, no further information is required. Click the Set Format button to write a format statement to the Format String field. Click OK to return to the main BILOG-MG window.

Having completed the specification in terms of model and data, the syntax file is created by selecting the Build Syntax option from the Run menu.
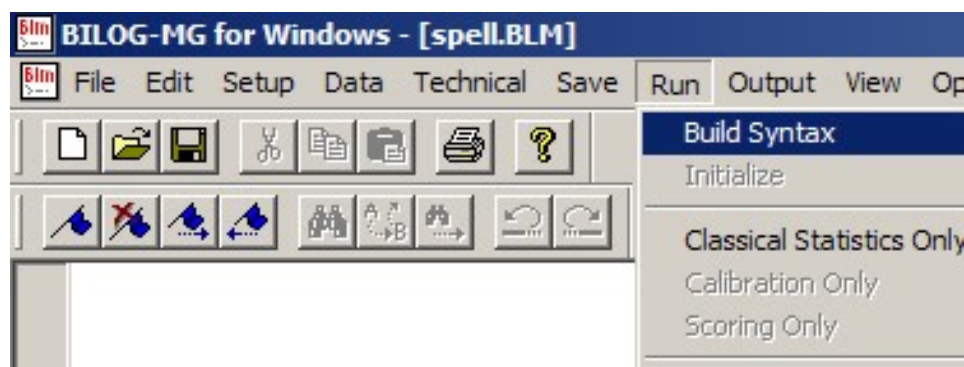
The syntax created by the program is now displayed in the main window, as shown below. Note that no options are given on the ITEMS and SCORE commands in this file, indicating that all program defaults will be used.

Save the completed syntax to file by selecting the Save option on the File menu.

The analysis is now performed by using some of the other options on the Run menu. Although the analysis can be done phase by phase (using the Classical Statistics Only, Calibration Only, and Scoring Only options) all three phases can be run in sequence by selecting the Stats, Calibration and Scoring option from this menu.
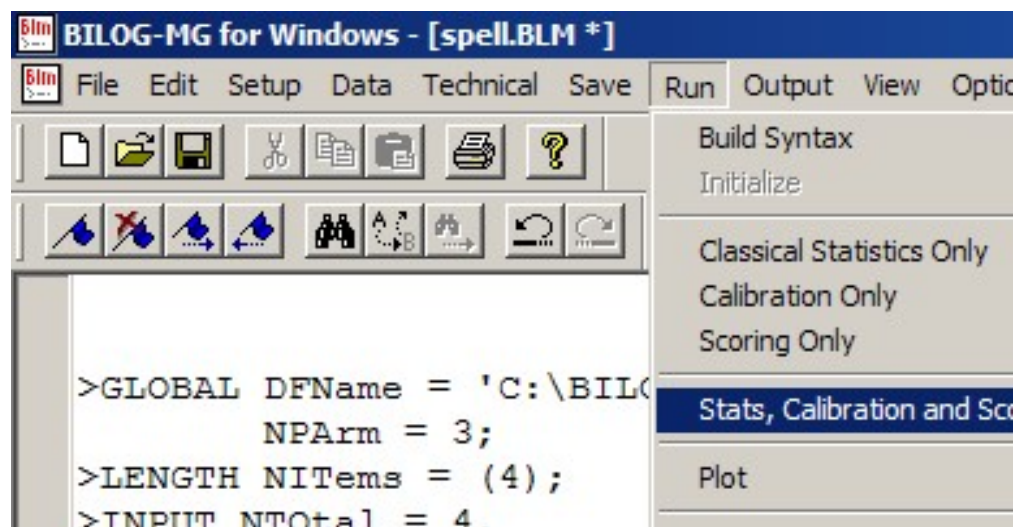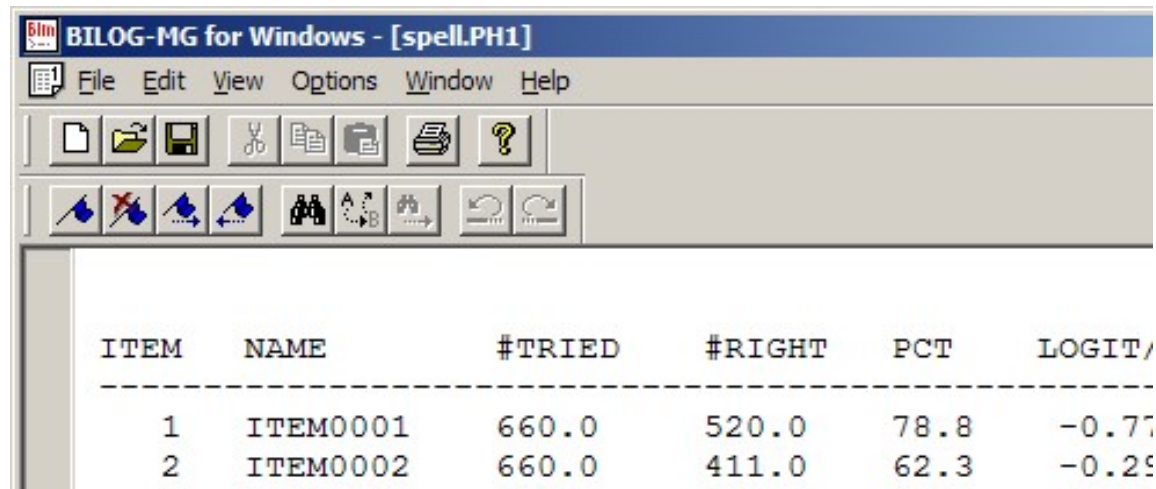
After successful completion of all three phases of the analysis, a message to this effect is displayed on the screen. If a problem was encountered during analysis, this message box will indicate that all phases were not completed successfully.
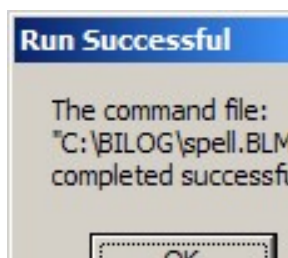
Access the output from the analysis through the Output menu. Classical statistics are given in the *.ph1 file.
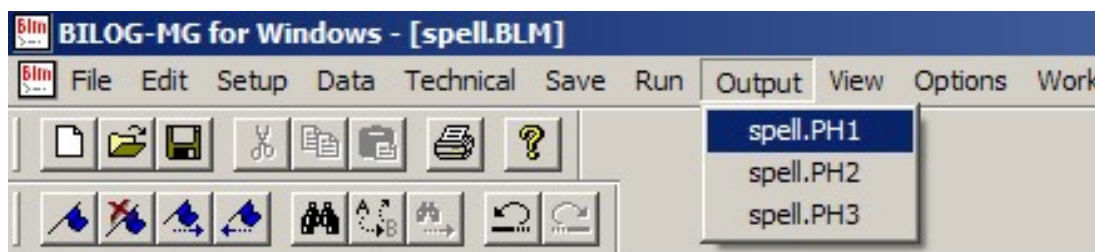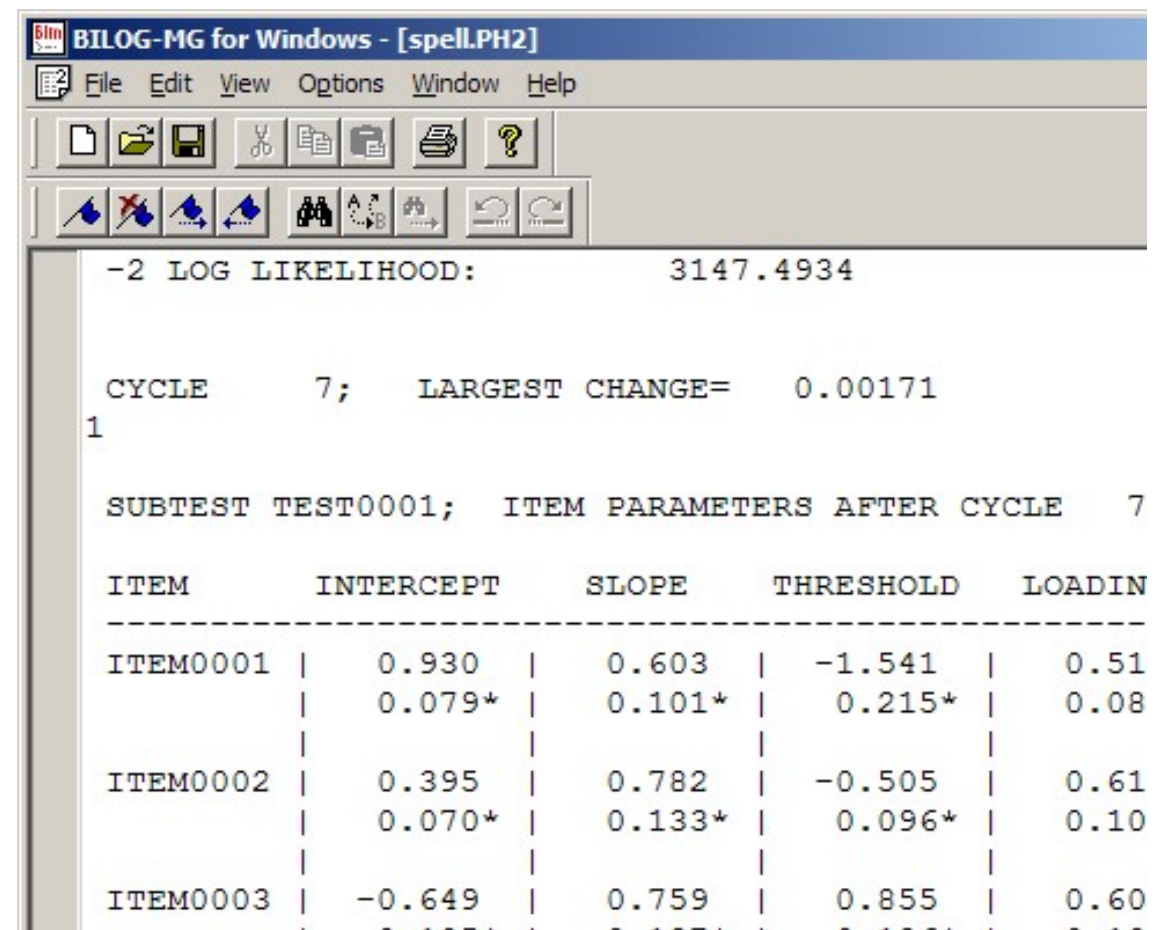
A section of this output file is shown below.

```
BILOG-MG for Windows - [spell.PH1]
File   Edit   View   Options   Window   Help


    ITEM     NAME              #TRIED      #RIGHT     PCT       LOGIT/
    ---------------------------------------------------------------
       1     ITEM0001          660.0       520.0      78.8      -0.77
       2     ITEM0002          660.0       411.0      62.3      -0.29
```

In the *.ph2 file, the results of the item calibrations are given. The item parameter estimates for the four items in the test are shown below.

```
BILOG-MG for Windows - [spell.PH2]
File   Edit   View   Options   Window   Help


    -2 LOG LIKELIHOOD:                3147.4934


    CYCLE       7;    LARGEST CHANGE=    0.00171
   1


    SUBTEST TEST0001;   ITEM PARAMETERS AFTER CYCLE    7


    ITEM           INTERCEPT      SLOPE      THRESHOLD     LOADIN
    ---------------------------------------------------------------
    ITEM0001 |      0.930    |     0.603   |    -1.541    |    0.51
             |      0.079*   |     0.101*  |     0.215*   |    0.08
             |              |            |             |
    ITEM0002 |      0.395    |     0.782   |    -0.505    |    0.61
             |      0.070*   |     0.133*  |     0.096*   |    0.10
             |              |            |             |
    ITEM0003 |     -0.649    |     0.759   |     0.855    |    0.60
```
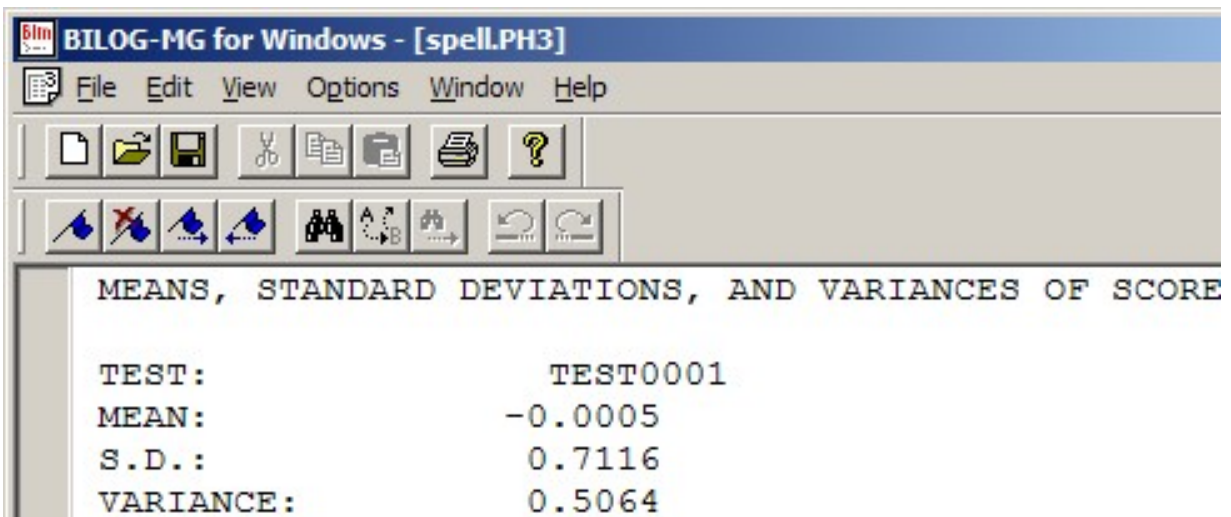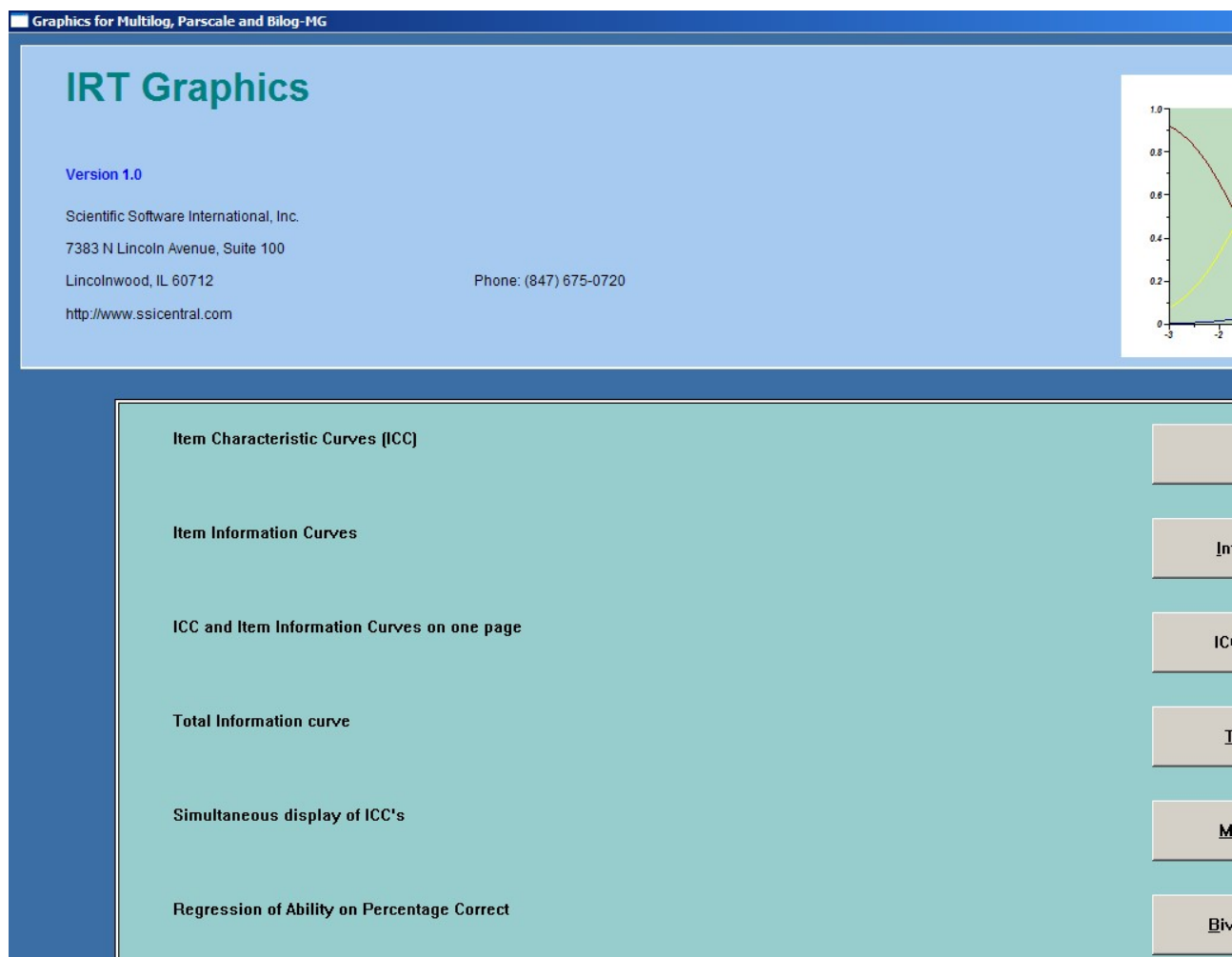
Scoring results are given in the *.ph3 file. The complete list of scores is printed to this file by default. A section of this output, showing summary statistics for the score estimates, is shown below.

```
BILOG-MG for Windows - [spell.PH3]
 File  Edit  View  Options  Window  Help

   MEANS, STANDARD DEVIATIONS, AND VARIANCES OF SCORE

   TEST:                    TEST0001
   MEAN:                    -0.0005
   S.D.:                     0.7116
   VARIANCE:                 0.5064
```

To obtain plots of the items, use the Plot option in the Run menu. This will open a new menu where different plots can be requested interactively.

```
BILOG-MG for Windows - [spell.BLM]
 File  Edit  Setup  Data  Technical  Save  Run  Output  View  Op

                                             Build Syntax
                                             Initialize

                                             Classical Statistics Only
                                             Calibration Only
                                             Scoring Only

   >GLOBAL  DFName = 'C:\BILC           Stats, Calibration and S
            NPArm = 3,
            LOGistic;                    Plot
```
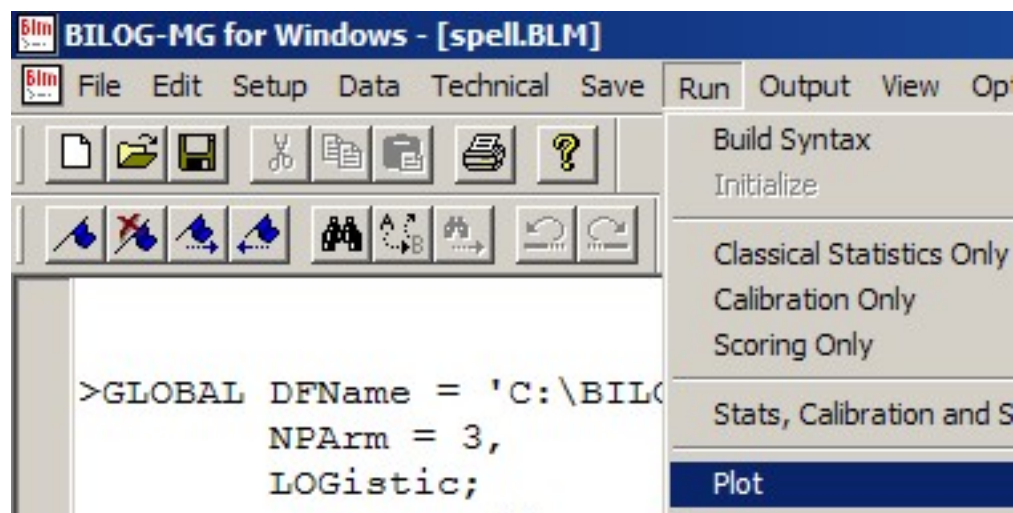
Setup Menu: Item Analysis Dialog Box: Advanced Tab

The final tab of the Item Analysis dialog box is the Advanced tab that controls the estimation of item parameters. Most of the information pertains to the CALIB command. The number of iterations and convergence criterion are set at the top of the dialog box, while the number of items and ability intervals for calculation of chi-square item fit statistics are specified in the Chi-square Item Fit Statistics group box.

At the bottom of the dialog box, prior item constraints may be requested and the estimation of the means of the prior distributions on the item parameters specified to be kept at a fixed value or to be estimated along with the parameters.

If a 3PL model is selected, all the prior check boxes in the Prior Item Constraints group box will be enabled. In the case of a 2PL model, the Prior on Guessing check box is disabled, while both the Prior on Guessing and Prior on Slope check boxes are disabled when a 1PL model is fitted to

the data. (This is what the manual says but I'm not sure if this is strictly implemented. It might be safer to manually enable/disable this option.)

Setup Menu: Test Scoring Dialog Box: General Tab

Information entered on the Test Scoring dialog box controls the type of scoring performed in Phase 3 of the analysis. The General tab of this dialog box is used to select the method of scoring and to import item parameters for scoring from previously saved files. In the latter case, the Browse button at the bottom of the tab is used to locate the file containing the item parameters to be used for scoring.

To score test using a file supplied by the user.
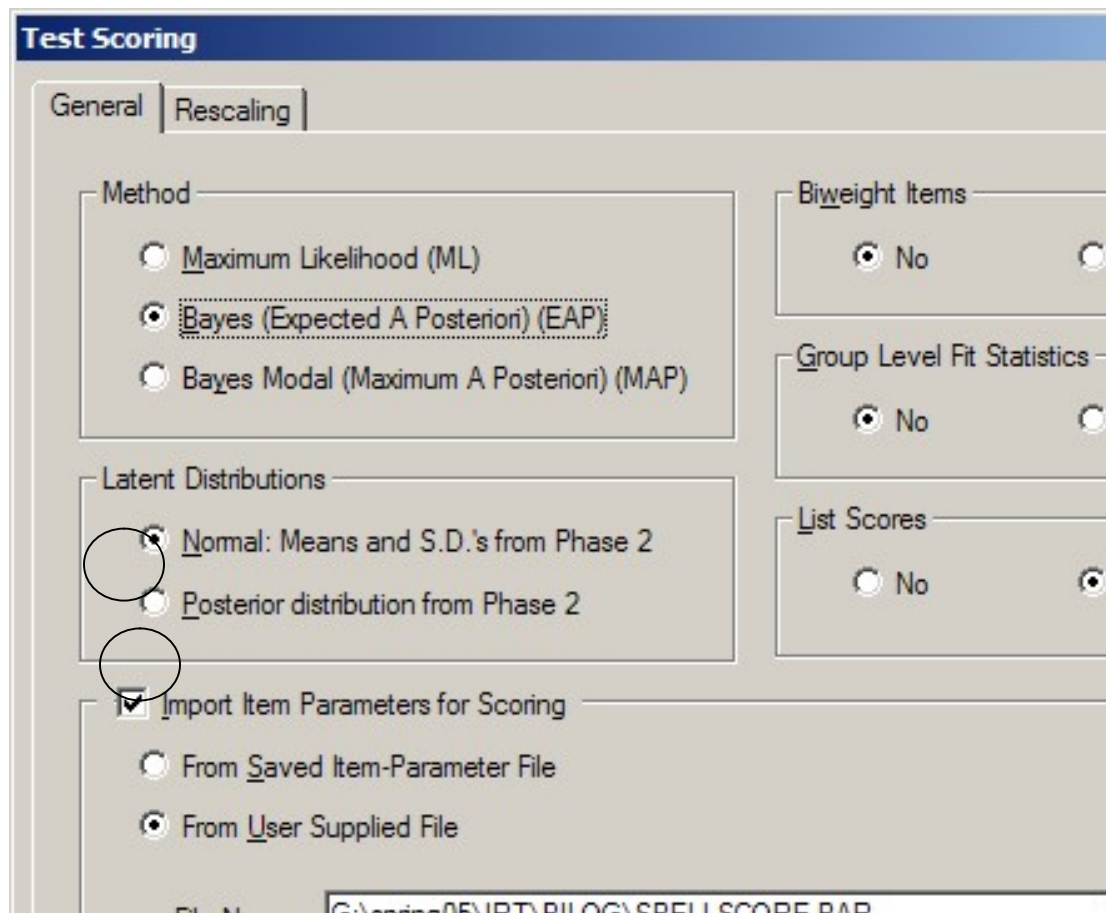
The contents of the files SPELLSCORE.PAR are as follows:

Line 1 gives the number of selected items in each subtest.

Remaining lines: The serial position of each item selected from the corresponding subtest, followed by the slope, threshold, and chance success (guessing) probability of the item. If a two-parameter model is assumed, the latter would be entered as 0.
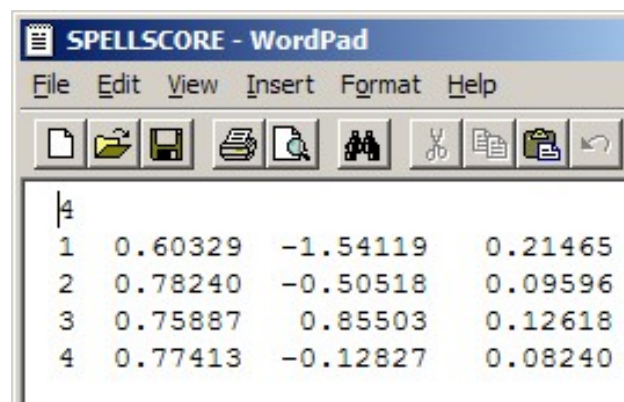
WARNING: There is a bug in the Windows interface of BILOG. This will address the problem: when you build the syntax file, edit the file by removing the first "A" from the keyword PARname before saving. Otherwise, the program won't run. Ignore the warning box when saving the file.
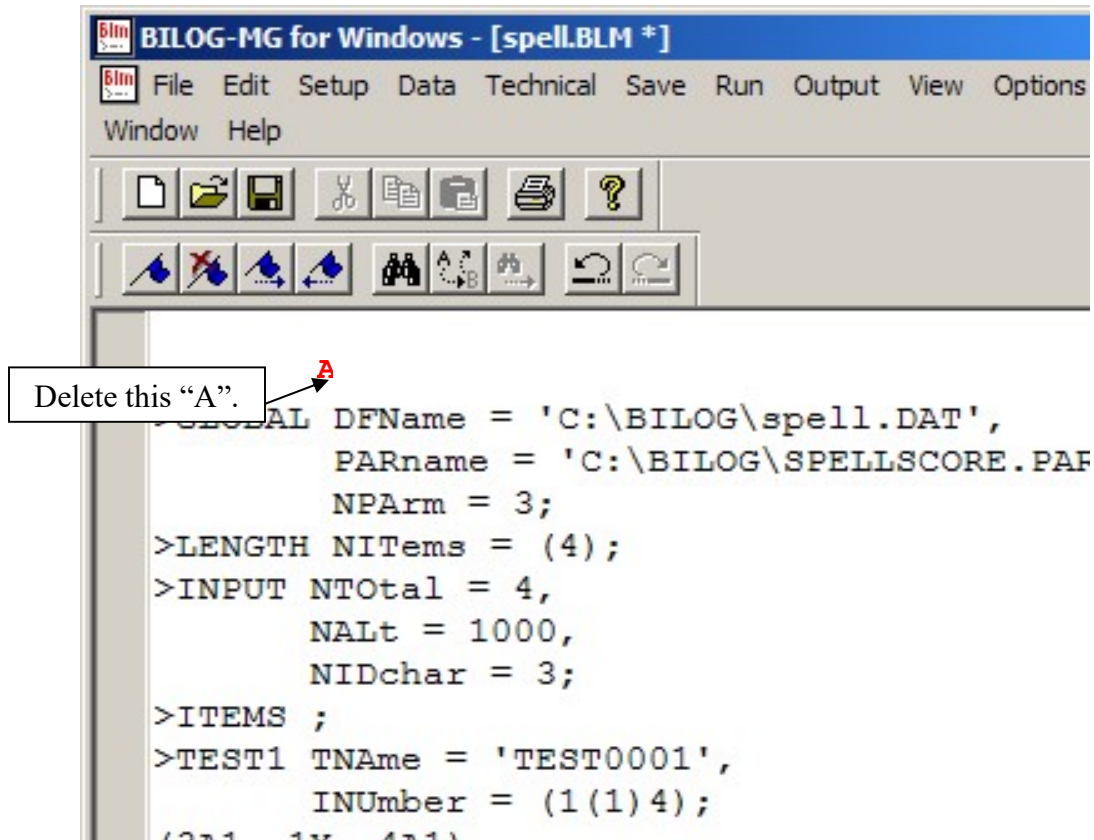
Setup Menu: Test Scoring Dialog Box: Rescaling Tab

Under Test Scoring, one can also rescale the ability estimates. For example, one can use the following linear transformation:

$$NewScore = SCALE \times OldScore + LOCATION$$

where SCALE is set at 100 and LOCATION at 500.

WARNING: The syntax would not reflect the change in SCALE so it needs to be typed in. (It keeps the default value of 1 or the value previously used.) Don't forget the comma separator.

```
>GLOBAL DFName = 'C:\BILOG\spell.DAT',
        PRname = 'C:\BILOG\SPELLSCORE.PAR',
        NPArm = 3;
>LENGTH NITems = (4);
>INPUT NTOtal = 4,
        NALt = 1000,
        NIDchar = 3;
>ITEMS ;
>TEST1 TNAme = 'TEST0001',
        INUmber = (1(1)4);
(3A1, 1X, 4A1)
```
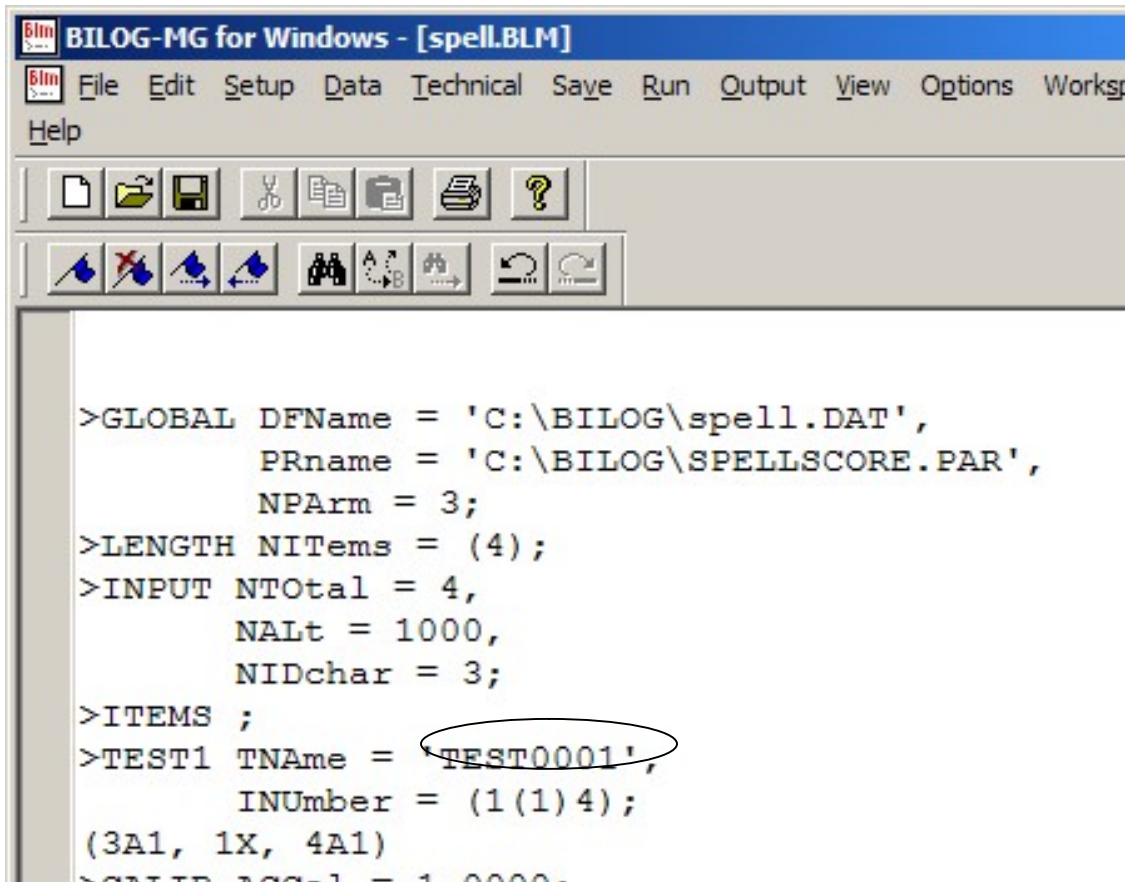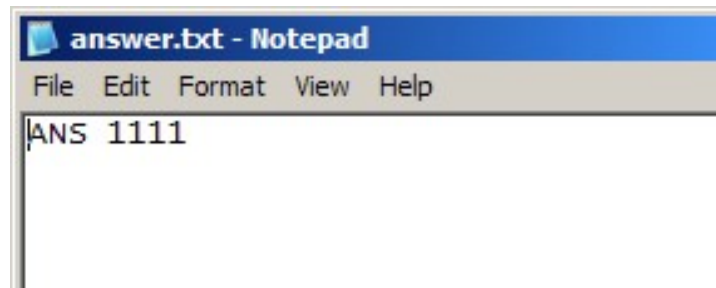
Data Menu - Item Keys Dialog Box: Answer Key Tab

The sole purpose of the Item Keys option on the Data menu is to provide the option to use answer, not-presented or omit keys. The three tabs on the Item Keys dialog box are similar. Possible key codes consist of the entire list of "possible keys". The information is taken from the Response Codes edit box on the Response tab of the General dialog box on the Setup menu.

On the first tab, Answer Key, an answer key may be read from external file. Type the file name then click the Open button.

The key can also be entered interactively in the window towards the bottom of the tab. If a key is entered interactively, the Save button may be used to save the entered information to an external file. The file used as answer key is referenced by the KFNAME keyword on the INPUT command. Data Menu:

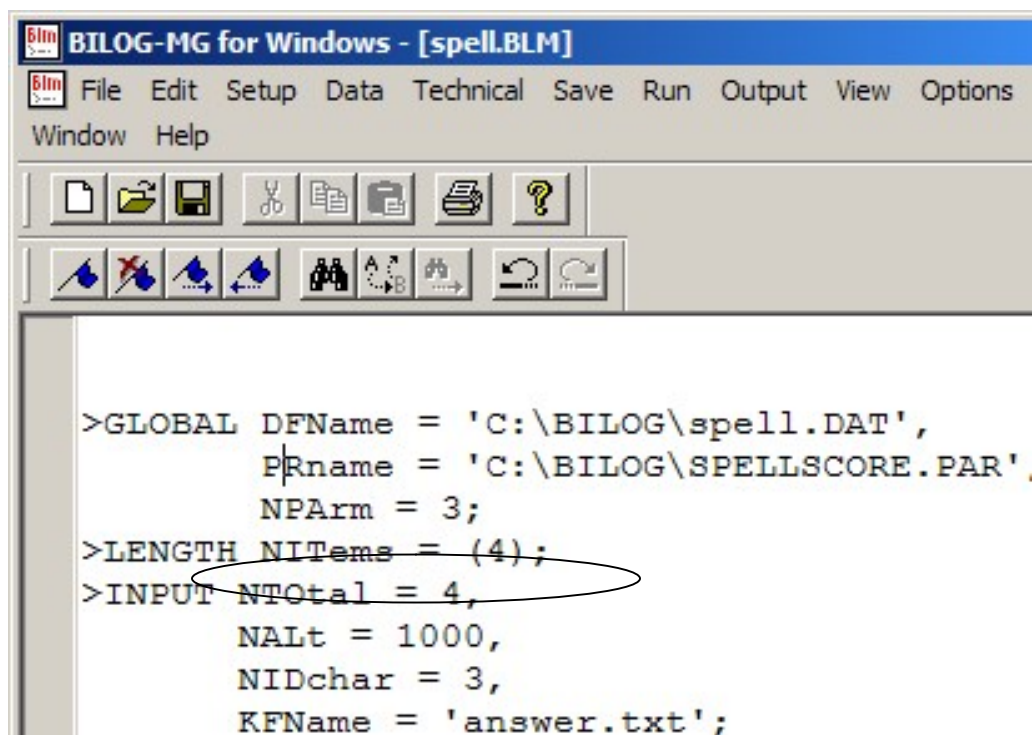In the example below, the external file answer.txt below was used as the key.