



## Differential item functioning

This example is based on an example in Thissen, Steinberg & Wainer (1993). 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, an answer key is not required.

The groups in this example are the two sexes and this is indicated by the `NGROUP` keyword on the `INPUT` command. The same four items are presented to both groups on a single test form. The format statement following the second `GROUP` command describes the position and order of data in **exampl02.dat**. The group indicator is found in column 3 of the data records and is read in integer format. A form indicator is not required in the data records because there is only one form. The data have been sorted into answer patterns, and the frequencies are found in columns 10-11 of the data (2A1). These frequencies serve as case weights in the analysis. The `TYPE=2` and `NWGHT=3` keywords describe this type of data. The value assigned to the keyword `NWGHT` requests the use of weighting in both the statistics and calibration (by default, no weights would be applied).

A 1-parameter logistic model is requested using the `NPARM` keyword on the `GLOBAL` command. The `LOGISTIC` option on the `GLOBAL` command indicates that the natural metric of the logistic response function will be assumed in all calculations. If this keyword is not present, the logit is, by default, multiplied by 1.7 to obtain the metric of the normal response function.

The `SAVE` option on the `GLOBAL` command indicates that a `SAVE` command will follow directly after the `GLOBAL` command. On the `SAVE` command, the item parameter estimates are saved to an external file **exampl02.par** and the DIF analysis results are written to an external file **exampl02.dif**.

The total number of unique items is described using the `NTOTAL` keyword on the `INPUT` command while the `NITEMS` keyword on the `LENGTH` command is set to 4 to indicate that all 4 items are to be used in the single subtest.

The `ITEMS` command lists the four items in the order that they will be read from the data records. The `INAMES` and `INUMBERS` keywords assign each item a name and a corresponding number. Because there is only one form, the `NFORM` keyword is not required in the `INPUT` command and a `FORM` command is not required. Because examinees in both groups are presented all the items listed in the `ITEMS` command, the `TEST` and `GROUP` commands need contain only the test name and the group names, respectively.

A DIF analysis is requested through the use of the `DIF` option on the `INPUT` command.

The `REFERENCE=1` keyword on the `CALIB` command designates males as the reference group. The convergence criterion is set to 0.005 instead of the default 0.01 using the `CRIT` keyword. When `NGROUP >1`, 20 quadrature

points will be used for each group. Setting the NQPT keyword to 10 implies that 10 points will be used for each group, as fewer points are needed when the number of items is small.

No SCORE command is included in the command file, as DIF models cannot be scored.

```
EXAMPL02.BLM - MALE VS FEMALE DIFFERENTIAL ITEM FUNCTIONING
  SPELLING, GIRDER ITEM 4, OTHER 3 ITEMS 1-3
>GLOBAL  NPARAM=1, LOGISTIC, SAVE, NWGHT=3, DFNAME='EXAMPL02.DAT';
>SAVE    PARM='EXAMPL02.PAR', DIF='EXAMPL02.DIF';
>LENGTH  NITEMS=4;
>INPUT   NTOTAL=4, NGROUPS=2, DIF, NIDCHAR=2, TYPE=2;
>ITEMS   INAMES=(SP1(1)SP4), INUMBERS=(1(1)4);
>TEST    TNAME=SPELL;
>GROUP1  GNAME=MALES;
>GROUP2  GNAME=FEMALES;
(2A1,I1,T10,F2.0,T5,4A1)
>CALIB   NQPT=10, CYCLES=15, CRIT=0.005, NEWTON=2, REFERENCE=1, PLOT=1;
```

### Phase 1 output

The title and additional comments (if the optional COMMENT command has been used) are echoed to the output file. Immediately after that, Phase 1 commands and specifications of the analysis are given. Under FILE ASSIGNMENT, relevant information as read in from the GLOBAL, SAVE, LENGTH, and TEST command are listed.

```
EXAMPLE 02: MALE VS FEMALE DIFFERENTIAL ITEM FUNCTIONING
  SPELLING, GIRDER ITEM 4, OTHER 3 ITEMS 1-3
>GLOBAL  NPARAM=1, LOGISTIC, SAVE, NWGHT=3, DFNAME='EXAMPL02.DAT';

FILE ASSIGNMENT AND DISPOSITION
=====

SUBJECT DATA INPUT FILE      EXAMPL02.DAT
BILOG-MG MASTER DATA FILE    MF.DAT           WILL BE CREATED FROM DATA FILE
CALIBRATION DATA FILE        CF.DAT           WILL BE CREATED FROM DATA FILE
ITEM PARAMETERS FILE          IF.DAT           WILL BE CREATED THIS RUN
CASE SCALE-SCORE FILE         SF.DAT
CASE WEIGHTING                FOR SUBJECT STATISTICS AND
                               ITEM CALIBRATION
ITEM RESPONSE MODEL           1 PARAMETER LOGISTIC
                               LOGIT METRIC (I.E., D = 1.0)

>SAVE    PARM='EXAMPL02.PAR', DIF='EXAMPL02.DIF';

BILOG-MG SAVE FILES

[OUTPUT FILES]
ITEM PARAMETERS FILE          EXAMPL02.PAR
DIF PARAMETER FILE           EXAMPL02.DIF

>LENGTH  NITEMS=4;

TEST LENGTH SPECIFICATIONS
=====

MAIN TEST LENGTHS:           4

>INPUT   NTOTAL=4, NGROUP=2, DIF, NIDCHAR=2, TYPE=2;
```

Specification of input related keywords are echoed in the next section. The data are entered as item-score patterns (right = 1, wrong = 0) and frequencies (case weights).

DATA INPUT SPECIFICATIONS  
=====

```

NUMBER OF FORMAT LINES                1
NUMBER OF ITEMS IN INPUT STREAM        4
NUMBER OF RESPONSE ALTERNATIVES       1000
NUMBER OF SUBJECT ID CHARACTERS        2
NUMBER OF GROUPS                       2
NUMBER OF TEST FORMS                   1
TYPE OF DATA                          SINGLE-SUBJECT DATA, CASE WEIGHTS
MAXIMUM SAMPLE SIZE FOR ITEM CALIBRATION 10000000
ALL SUBJECTS INCLUDED IN RUN

```

```
>ITEMS  INAMES=(SP1(1)SP4), INUMBERS=(1(1)4);
```

TEST SPECIFICATIONS  
=====

```
>TEST  TNAME=SPELL;
```

The following lines indicate the assignment of items to the single subtest, utilizing the information on both the TEST and ITEMS commands.

```

TEST NUMBER:    1      TEST NAME: SPELL
NUMBER OF ITEMS:  4

```

ITEM NUMBER	ITEM NAME	ITEM NUMBER	ITEM NAME	ITEM NUMBER	ITEM NAME	ITEM NUMBER	ITEM NAME
1	SP1	2	SP2	3	SP3	4	SP4

Information on the forms and groups is given next. The definition of the male and female groups, and the use of the same four items for both groups are reflected below. It is also noted that a DIF model is to be employed in this analysis.

FORM SPECIFICATIONS  
=====

ITEMS READ ACCORDING TO SPECIFICATIONS ON THE ITEMS COMMAND

```
>GROUP1  GNAME=MALES;
>GROUP2  GNAME=FEMALES;
```

MULTIPLE GROUP SPECIFICATIONS  
=====

DIFFERENTIAL ITEM FUNCTIONING MODEL IS EMPLOYED.

```

GROUP NUMBER:    1      GROUP NAME: MALES
TEST NUMBER:    1      TEST NAME: SPELL
NUMBER OF ITEMS:  4

```

ITEM NUMBER	ITEM NAME
1	SP1
2	SP2
3	SP3

4 SP4

-----  
GROUP NUMBER: 2      GROUP NAME: FEMALES  
TEST NUMBER: 1      TEST NAME: SPELL  
NUMBER OF ITEMS: 4

ITEM    ITEM  
NUMBER    NAME  
-----  
1    SP1  
2    SP2  
3    SP3  
4    SP4  
-----

Following is the format statement used in reading the data and the answer, omit, and not-present keys (if any). Data for this example are item scores and they are complete; keys are not required. The case ID is read in the first 2 columns (2A1), followed by the group indicator (I1). After reading the weights (F2.0), the 4 item responses are read (4A1).

FORMAT FOR DATA INPUT IS:  
(2A1,I1,T10,F2.0,T5,4A1)

The first two cases are echoed to the output file so that the user can verify the input.

OBSERVATION #      1    WEIGHT:    22.0000    ID :    1

SUBTEST #:    1    SPELL  
GROUP #:    1    MALES

TRIED      RIGHT  
4.000      0.000

ITEM          1          2          3          4  
TRIED      1.0      1.0      1.0      1.0  
RIGHT      0.0      0.0      0.0      0.0

OBSERVATION #      2    WEIGHT:    10.0000    ID :    2

SUBTEST #:    1    SPELL  
GROUP #:    1    MALES

TRIED      RIGHT  
4.000      1.000

ITEM          1          2          3          4  
TRIED      1.0      1.0      1.0      1.0  
RIGHT      0.0      0.0      0.0      1.0

Classical item statistics for the total sample and each group sample follow. #TRIED designates the number of examinees responding to the item. For completeness, both the Pearson and biserial item-test correlations are shown. The latter has smaller bias when the percent right is extreme. The item statistics are given by group and then for the total group.

Item means, initial slope estimates, and Pearson and polyserial item-test correlations are given in the next table.

## Pearson

The point biserial correlation  $r_{PB,j}$  for item  $j$  is a computationally simplified Pearson's  $r$  between the dichotomously scored item  $j$  and the total score  $x$ . It is computed as

$$r_{PB,j} = \frac{(\mu_j - \mu_x)}{\sigma_x} \sqrt{\frac{p_j}{q_j}}$$

where  $\mu_j$  is the mean total score among examinees who have responded correctly to item  $j$ ,  $\mu_x$  is the mean total score for all examinees,  $p_j$  is the item difficulty index for item  $j$ ,  $q_j$  is  $1 - p_j$ , and  $\sigma_x$  is the standard deviation of the total scores for all examinees.

## Polyserial correlation

The polyserial correlation  $r_{p,j}$  can be expressed in terms of the point polyserial correlation as

$$r_{p,j} = \frac{r_{PP,j} \sigma_j}{\sum_{k=1}^{m-1} h(z_{jk})}$$

where  $z_{jk}$  is the scoring corresponding to the cumulative proportion,  $p_{jk}$  of the  $k$ -th response category to item  $j$ ,  $\sigma_j$  is the standard deviation of items scores  $y$  for item  $j$ , and  $r_{PP,j}$  is the point-polyserial correlation.

The biserial correlation estimates the relationship between the total score and the hypothetical score on the continuous scale underlying the (dichotomous) item. The biserial correlation also assumes a normal distribution of the hypothetical scores. The reason for reporting these correlations separately for each group is that the appearance of large discrepancies between groups for a given item would suggest that the assumption of a common slope is untenable. Note that, if a biserial correlation more negative than  $-0.15$  is detected by the program during this phase of the analysis, the item in question will be assumed miskeyed and will be omitted in the Phase 2 analysis.

ITEM STATISTICS FOR GROUP: 1 MALES							
ITEM	NAME	#TRIED	#RIGHT	PCT	LOGIT	ITEM*TEST CORRELATION	
						PEARSON	BISERIAL
1	SP1	285.0	215.0	0.754	-1.12	0.243	0.332
2	SP2	285.0	181.0	0.635	-0.55	0.351	0.450
3	SP3	285.0	91.0	0.319	0.76	0.364	0.474
4	SP4	285.0	179.0	0.628	-0.52	0.360	0.461

ITEM STATISTICS FOR GROUP: 2 FEMALES							
ITEM	NAME	#TRIED	#RIGHT	PCT	LOGIT	ITEM*TEST CORRELATION	
						PEARSON	BISERIAL
1	SP1	374.0	305.0	0.816	-1.49	0.254	0.370
2	SP2	374.0	230.0	0.615	-0.47	0.295	0.376
3	SP3	374.0	109.0	0.291	0.89	0.231	0.307
4	SP4	374.0	171.0	0.457	0.17	0.306	0.385

ITEM STATISTICS FOR MULTIPLE GROUPS SPELL							
ITEM	NAME	#TRIED	#RIGHT	PCT	LOGIT	ITEM*TEST CORRELATION	
						PEARSON	BISERIAL

1	SP1	659.0	520.0	0.789	-1.32	0.239	0.337
2	SP2	659.0	411.0	0.624	-0.51	0.320	0.409
3	SP3	659.0	200.0	0.303	0.83	0.291	0.383
4	SP4	659.0	350.0	0.531	-0.12	0.324	0.406

## Phase 2 output

During calibration, a logistic item response function is fitted to each item of each subscale. In this example, a 1-parameter logistic response function is fitted (NPARM =1 on GLOBAL).

Echoing of the Phase 2 commands and specification of the analysis starts the listing of Phase 2 output.

```
>CALIB NQPT=10, CYCLES=15, CRIT=0.005, NEWTON=2, REFERENCE=1;
```

Under CALIBRATION PARAMETERS, the definitions of calibration related keywords for this analysis are given:

```
CALIBRATION PARAMETERS
=====

MAXIMUM NUMBER OF EM CYCLES:          15
MAXIMUM NUMBER OF NEWTON CYCLES:      2
CONVERGENCE CRITERION:                0.0050
ACCELERATION CONSTANT:                0.5000

LATENT DISTRIBUTION:                  EMPIRICAL PRIOR FOR EACH GROUP
                                       ESTIMATED CONCURRENTLY
                                       WITH ITEM PARAMETERS
                                       REFERENCE GROUP: 1
PLOT EMPIRICAL VS. FITTED ICC'S:     YES, FOR ITEMS WITH FIT PROBABILITY
                                       LESS THAN 1.00000
DATA HANDLING:                        DATA ON SCRATCH FILE
CONSTRAINT DISTRIBUTION ON SLOPES:     NO
CONSTRAINT DISTRIBUTION ON THRESHOLDS: NO
```

MML estimation is used when tests of three or more items are specified. The solution assumes that the respondents are drawn randomly from a population or populations of abilities, which is assumed to have a normal distribution. The empirical distribution of ability is represented as a discrete distribution on a finite number of points. The quadrature points and weights used for MML estimation of the item parameters for the two groups are printed next.

```
METHOD OF SOLUTION:
EM CYCLES (MAXIMUM OF 15)
FOLLOWED BY NEWTON-RAPHSON STEPS (MAXIMUM OF 2)

QUADRATURE POINTS AND PRIOR WEIGHTS: GROUP 1 MALES
      1      2      3      4      5
POINT  -0.4000E+01 -0.3111E+01 -0.2222E+01 -0.1333E+01 -0.4444E+00
WEIGHT  0.1190E-03 0.2805E-02 0.3002E-01 0.1458E+00 0.3213E+00

      6      7      8      9     10
POINT  0.4444E+00 0.1333E+01 0.2222E+01 0.3111E+01 0.4000E+01
WEIGHT 0.3213E+00 0.1458E+00 0.3002E-01 0.2805E-02 0.1190E-03

QUADRATURE POINTS AND PRIOR WEIGHTS: GROUP 2 FEMALES
      1      2      3      4      5
POINT  -0.4000E+01 -0.3111E+01 -0.2222E+01 -0.1333E+01 -0.4444E+00
WEIGHT  0.1190E-03 0.2805E-02 0.3002E-01 0.1458E+00 0.3213E+00

      6      7      8      9     10
POINT  0.4444E+00 0.1333E+01 0.2222E+01 0.3111E+01 0.4000E+01
WEIGHT 0.3213E+00 0.1458E+00 0.3002E-01 0.2805E-02 0.1190E-03
```

The MML solution employs both the EM method and Newton-Gauss iterations to solve the marginal likelihood equations. On the CALIB command, a maximum of 15 EM cycles and 2 Newton-Gauss iterations were requested. Results for each iteration are displayed so that the extent of convergence can be judged.

In the case of nested models on the same data, the  $-2 \log$  likelihood values at convergence can be used to evaluate the fit of the models. Refitting this example, for instance as a single-group analysis will allow the comparison of non-DIF and DIF models for these data. In that way, it can be determined whether differential item functioning effects are present.

[E-M CYCLES]

-2 LOG LIKELIHOOD = 3152.375

CYCLE 1; LARGEST CHANGE= 0.17572  
 -2 LOG LIKELIHOOD = 3128.806

...

CYCLE 8; LARGEST CHANGE= 0.00486

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 on the item parameter estimates.

[FULL NEWTON CYCLES]

-2 LOG LIKELIHOOD: 3110.3990  
 CYCLE 9; LARGEST CHANGE= 0.00416

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 (the so-called “0,1” metric). In the case of multiple groups, the program provides the option of setting the mean and standard deviation of one group to 0,1 as shown here. The user may set the mean and standard deviation of the combined estimated distribution of the groups to 0 and 1 by setting the REFERENCE keyword on the CALIB command to zero. The parameter estimates can be rescaled in Phase 3 according to scale conventions selected by the user (using the RSCTYPE, SCALE and LOCATION keywords on the SCORE command). In a DIF model, no scoring is done, so use of the REFERENCE=0 specification is not pursued here.

Estimated item parameters for the two groups are given next. The INTERCEPT column contains the estimates of the item intercepts, which are defined as the product of each item’s slope and threshold. This is followed by the slope or discrimination parameters and the item threshold or location parameters. The LOADING column represents the one-factor item factor loadings given by the expression

$$\frac{\text{slope}}{\sqrt{1.0 + \text{slope}^2}}$$

For a 1PL model, no asymptotes or guessing parameters are estimated. In a 1PL model, all slopes are equal. In DIF analyses, the assumption is made that slopes are equal over the groups. This implies that items will discriminate equally well in all groups. Note that, in this example, the slopes of all items for both groups are constrained to 1.285.

GROUP 1 MALES ; ITEM PARAMETERS AFTER CYCLE 9							
ITEM	INTERCEPT	SLOPE	THRESHOLD	LOADING	ASYMPTOTE	CHISQ	DF
	S.E.	S.E.	S.E.	S.E.	S.E.	(PROB)	
SP1	1.489	1.285	-1.159	0.789	0.000	2.7	7.0
	0.168*	0.096*	0.130*	0.059*	0.000*	(0.9146)	
SP2	0.749	1.285	-0.583	0.789	0.000	20.0	7.0
	0.152*	0.096*	0.119*	0.059*	0.000*	(0.0056)	
SP3	-1.008	1.285	0.784	0.789	0.000	36.2	5.0
	0.109*	0.096*	0.085*	0.059*	0.000*	(0.0000)	
SP4	0.709	1.285	-0.552	0.789	0.000	26.6	7.0
	0.150*	0.096*	0.117*	0.059*	0.000*	(0.0004)	

\* STANDARD ERROR



LARGEST CHANGE = 0.004158 95.3 12.0  
(0.0000)

GROUP 2 FEMALES ; ITEM PARAMETERS AFTER CYCLE 9

ITEM	INTERCEPT S.E.	SLOPE S.E.	THRESHOLD S.E.	LOADING S.E.	ASYMPTOTE S.E.	CHISQ (PROB)	DF
SP1	1.887 0.168*	1.285 0.096*	-1.468 0.131*	0.789 0.059*	0.000 0.000*	11.3 (0.1261)	7.0
SP2	0.617 0.136*	1.285 0.096*	-0.480 0.106*	0.789 0.059*	0.000 0.000*	34.6 (0.0000)	7.0
SP3	-1.113 0.144*	1.285 0.096*	0.866 0.112*	0.789 0.059*	0.000 0.000*	24.5 (0.0009)	7.0
SP4	-0.203 0.133*	1.285 0.096*	0.158 0.104*	0.789 0.059*	0.000 0.000*	43.0 (0.0000)	7.0

\* STANDARD ERROR  
95.3 12.0  
(0.0000)

LARGEST CHANGE = 0.004158

NOTE: ITEM FIT CHI-SQUARES AND THEIR SUMS MAY BE UNRELIABLE  
FOR TESTS WITH LESS THAN 20 ITEMS

The item parameter estimates for each group are followed by the averages for the group thresholds. The mean threshold of the female group (Group 2) is 0.146 above that of the male or reference group. DIF is item by group interaction under the constraint that the mean thresholds of the groups are equal. The threshold adjustment sets the mean of the reference group's threshold to 1, and the mean threshold for the females is accordingly adjusted to 0.148. The unadjusted and adjusted mean thresholds for the two groups form the next section of the Phase 2 output file.

PARAMETER	MEAN	STN DEV
GROUP: 1 NUMBER OF ITEMS:		4
THRESHOLD	-0.377	0.823
GROUP: 2 NUMBER OF ITEMS:		4
THRESHOLD	-0.231	0.991

THRESHOLD MEANS

GROUP	ADJUSTMENT
1	0.000
2	0.146

MODEL FOR GROUP DIFFERENTIAL ITEM FUNCTIONING:  
ADJUSTED THRESHOLD VALUES

ITEM	GROUP	
	1	2
SP1	-1.159 0.130*	-1.614 0.131*
SP2	-0.583 0.119*	-0.626 0.106*
SP3	0.784 0.085*	0.720 0.112*
SP4	-0.552 0.117*	0.012 0.104*

\*STANDARD ERROR

The adjusted threshold values are followed by the group differences of the constrained values. The standard errors for the differences are computed as

$$s.e._{G2-G1} = \sqrt{\text{var}(G2) + \text{var}(G1)}.$$

ITEM	GROUP 2 - 1
SP1	-0.455 0.185*
SP2	-0.043 0.159*
SP3	-0.065 0.141*
SP4	0.564 0.156*

\*STANDARD ERROR

The estimated latent distributions of the groups are given next; with the origin and unit of scale set so that the mean of the reference group is 0 and the standard deviation is 1.

GROUP: 1 MALES QUADRATURE POINTS, POSTERIOR WEIGHTS, MEAN AND S.D.:

	1	2	3	4	5
POINT	-0.3578E+01	-0.2788E+01	-0.1998E+01	-0.1208E+01	-0.4180E+00
POSTERIOR	0.1972E-03	0.4485E-02	0.4394E-01	0.1737E+00	0.2780E+00

	6	7	8	9	10
POINT	0.3720E+00	0.1162E+01	0.1952E+01	0.2742E+01	0.3532E+01
POSTERIOR	0.2647E+00	0.1724E+00	0.5526E-01	0.7020E-02	0.3483E-03

MEAN 0.00000  
S.E. 0.00000

S.D. 1.00000  
S.E. 0.00000

GROUP: 2 FEMALES QUADRATURE POINTS, POSTERIOR WEIGHTS, MEAN AND S.D.:

	1	2	3	4	5
POINT	-0.3724E+01	-0.2934E+01	-0.2144E+01	-0.1354E+01	-0.5642E+00
POSTERIOR	0.2099E-03	0.4246E-02	0.3608E-01	0.1456E+00	0.3067E+00

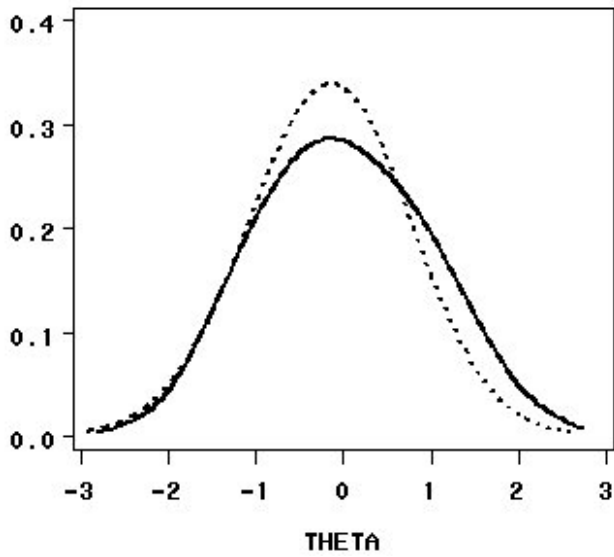
	6	7	8	9	10
POINT	0.2258E+00	0.1016E+01	0.1806E+01	0.2596E+01	0.3386E+01
POSTERIOR	0.3161E+00	0.1525E+00	0.3473E-01	0.3598E-02	0.1624E-03

MEAN -0.16191  
S.E. 0.06907

S.D. 0.89707  
S.E. 0.00845

A plot of the two estimated latent distributions are shown below. The solid line represents the estimated distribution of the male group.

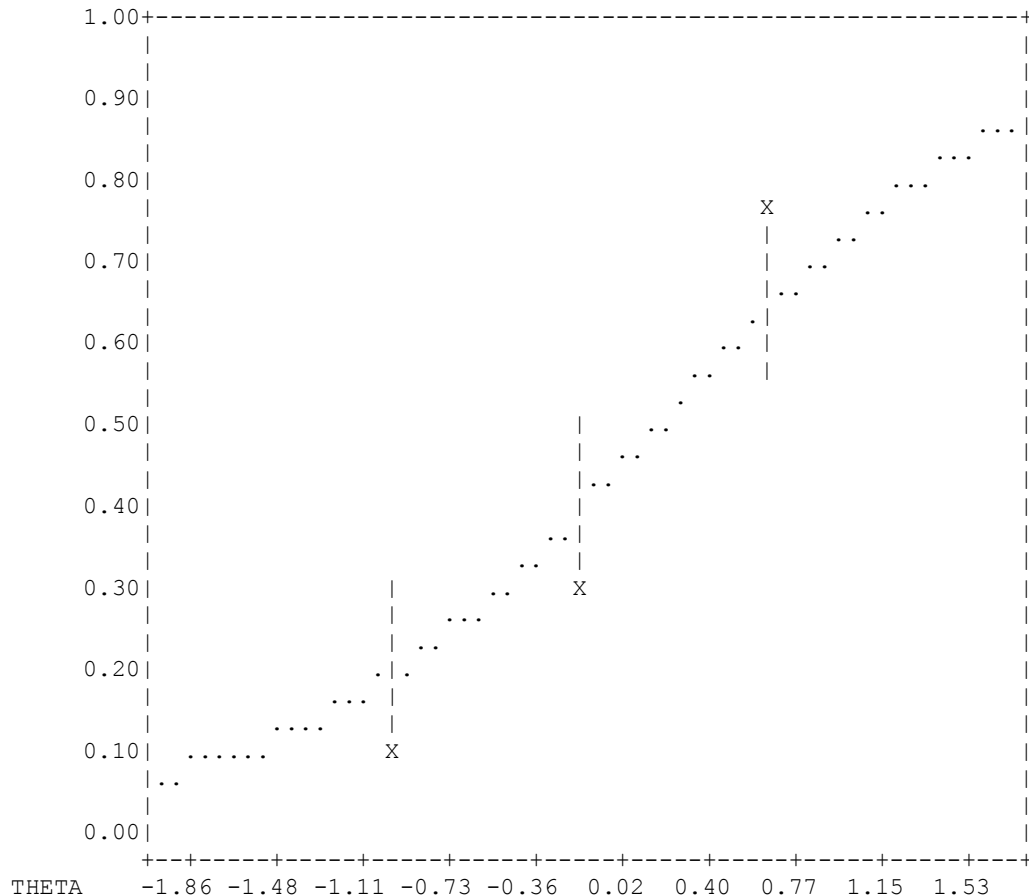
### Estimated latent distributions



BILOG-MG is also capable of producing graphic representations of a number of item and test characteristics. Using the PLOT keyword on the CALIB command, it is possible to obtain plots of the item-response functions with a significance level below the value assigned to the PLOT keyword. By default, PLOT=0 and no plots are produced. On the other hand, setting PLOT to 1.0 will lead to the display of all item response functions in the output file. One such plot, for the fourth item administered to the female group, is shown below.

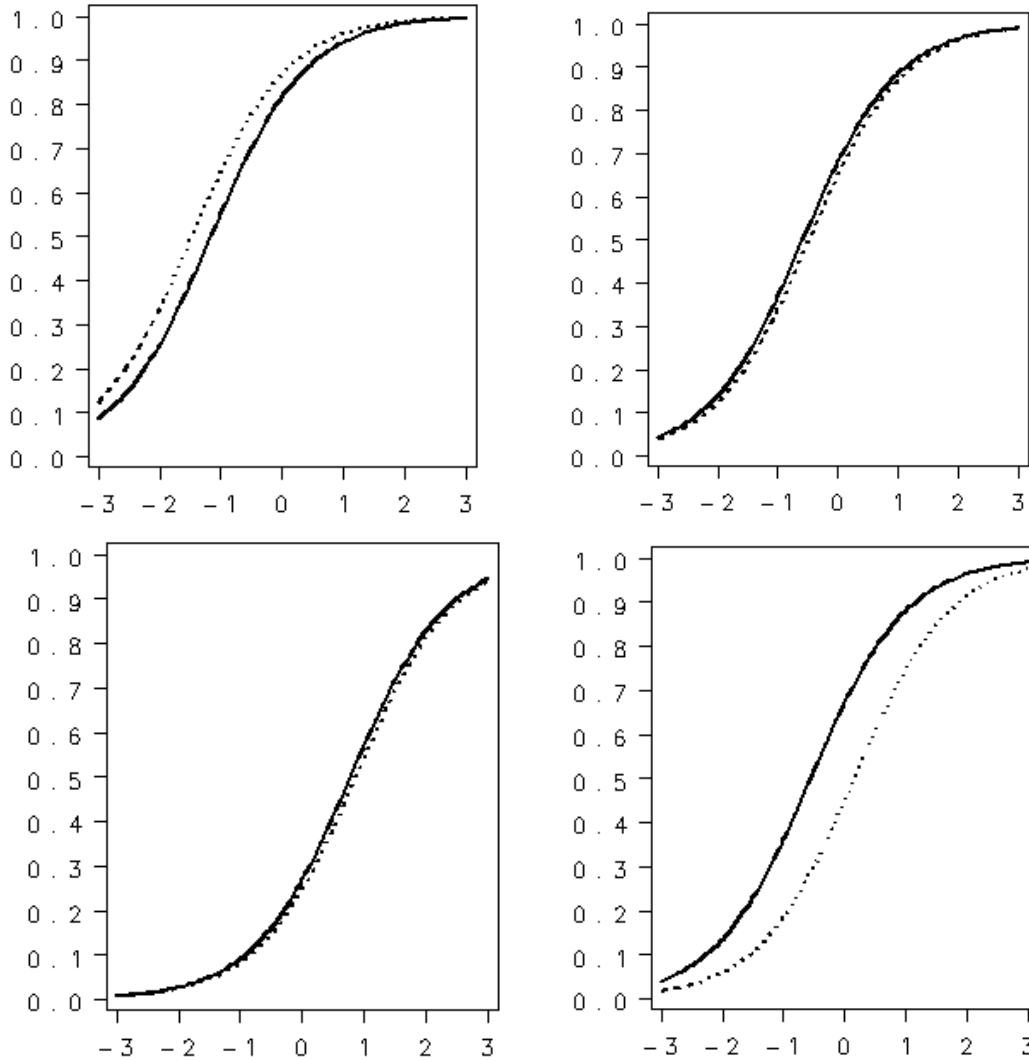
GROUP: 2 FEMALES  
SUBTEST: SPELL

ITEM: SP4 CHISQ = 43.0 DF = 7.0 PROB< 0.0000



The plot also shows 95% tolerance intervals for the observed percent correct among respondents in corresponding EAP groups, assuming the percent-correct predicted by the model is correct. Note that similar plots may be obtained using the IRTGRAPH program accessible from the **Run** menu.

By saving the estimated parameter estimates to an external file, the estimates can also be used in external packages to produce additional plots. Below, the item response functions for both groups are plotted by item.



### Single-group comparison

The data from example 2 are analyzed here as a single group. No NGROUP keyword is provided on the INPUT command and, by default, the program assumes there is only one group. No GROUP commands follow the TEST command, and the group indicator has been removed from the variable format statement.

The acceleration factor on the CALIB command has been set to its default value of 0.5 (ACCEL=0.5). The difference in the log likelihoods from the two-group and single-group solutions can be examined to determine if differential item functioning effects are present. The item parameter file obtained in the previous section is specified in the GLOBAL command to provide starting values for parameter estimation in Phase 2.

```

EXAMPL03.BLM - MALE VS. FEMALE DIFFERENTIAL ITEM FUNCTIONING
      SPELLING, GIRDER ITEM 4, OTHER 3 ITEMS 1-3
>GLOBAL  NPARAM=1, NWGHT=3, LOGISTIC, IFNAME='EXAMPL02.PAR',
      DFNAME='EXAMPL02.DAT';
>LENGTH  NITEMS=4;
>INPUT   NTOTAL=4, NIDCHAR=2, TYPE=2;
>ITEMS   INAME=(SP01,SP02,SP03,SP04), INUMBERS=(1(1)4);
>TEST    TNAME=SPELL;
      (2A1,T10,F2.0,T5,4A1)
>CALIB   EMPIRICAL, NQPT=31, CRIT=0.005, ACCEL=0.5;

```

### Phase 1 output

```

EXAMPLE 02: MALE VS. FEMALE DIFFERENTIAL ITEM FUNCTIONING
SPELLING, GIRDER ITEM 4, OTHER 3 ITEMS 1-3

```

The Phase 1 output for this example is the same as that obtained in Section 10.2, except that classical item statistics are computed only for the total sample.

### Phase 2 output

The main interest in this example is the comparison of the log likelihood of the fit of the DIF and non- DIF models. The difference,  $3138.4122 - 3110.3990 = 28.0132$ , distributed as  $\chi^2$  on four degrees of freedom, indicates significantly better fit of the DIF model.

```

-2 LOG LIKELIHOOD:      3138.4122
CYCLE      4;  LARGEST CHANGE=  0.00439

```

```

SUBTEST SPELL ; ITEM PARAMETERS AFTER CYCLE 4

```

ITEM	INTERCEPT S.E.	SLOPE S.E.	THRESHOLD S.E.	LOADING S.E.	ASYMPTOTE S.E.	CHISQ (PROB)	DF
SP01	1.688 0.124*	1.234 0.094*	-1.368 0.101*	0.777 0.059*	0.000 0.000*	15.4 (0.0015)	3.0
SP02	0.662 0.105*	1.234 0.094*	-0.536 0.085*	0.777 0.059*	0.000 0.000*	33.6 (0.0000)	3.0
SP03	-1.069 0.111*	1.234 0.094*	0.866 0.090*	0.777 0.059*	0.000 0.000*	17.9 (0.0005)	3.0
SP04	0.169 0.102*	1.234 0.094*	-0.137 0.082*	0.777 0.059*	0.000 0.000*	32.2 (0.0000)	3.0

\* STANDARD ERROR

```

LARGEST CHANGE =  0.004551      99.0 12.0
                        (0.0000)

```

NOTE: ITEM FIT CHI-SQUARES AND THEIR SUMS MAY BE UNRELIABLE FOR TESTS WITH LESS THAN 20 ITEMS