Example: Application of the PC Model to a Reasoning Ability Instrument, MMLE, MULTILOG

MULTILOG (Thissen, Chen, & Bock, 2003) implements MMLE for parameter estimation for the dichotomous models that have been discussed so far, as well as for polytomous models such as the graded response, nominal response, and multiple-choice models; these polytomous models are discussed in subsequent chapters.[1]

As is the case with BILOG, estimation of the item parameters may proceed independently of estimation of the person parameters. MULTILOG 7's interface is similar to that of BILOG's in the use of menus and dialogs to facilitate the creation of an input (control) file. This file is subsequently processed by MULTILOG.

The command file for the PC calibration of the reasoning ability data is shown in Table 1. All command lines begin with the symbol "`>`," are terminated by a semicolon, and contain a series of subcommands; in general, commands and subcommands can be abbreviated to two or three characters. Briefly, in the problem line (`PROBLEM` or `PRO`) one describes some of the data's characteristics. For instance, the `INDIVIDUAL` subcommand specifies that individual case data (rather than pattern data) are being used, that there are at most 3000 examinees (`NEXAMINEES=3000`; one may use this subcommand to restrict the number of cases read), that eight items are to be calibrated (`NITEMS=8`), and the `RANDOM` subcommand indicates that one is assuming that the individuals are randomly sampled from the population of interest and that one is performing item parameter estimation. In addition, each examinee case uses 15 characters as an identification label (`NCHARS=15`).

Table 1. Command file for MULTILOG PC model calibration example.[a]

```
MULTILOG for Windows 7.00.2327.2
ALIKE, 8 ITEMS, PC                          ← title
>PROBLEM RANDOM,                            ← Problem description line
         INDIVIDUAL,
         DATA = 'Alike.DAT',
         NITEMS = 8,
         NGROUPS = 1,
         NEXAMINEES = 3000,
         NCHARS = 15;
>TEST ALL,                                  ← Instrument description line
      NOMINAL,                              ← Specification of NO model
      NC = (3(0)8),
      HIGH = (3(0)8);
>TMATRIX ALL AK POLY;                       ← constraints for PC[b]
>EQUAL ALL AK=1;                            ← constraints for PC[b]
>FIX ALL AK=2 VALUE=0.0;                    ← constraints for PC[b]
>TMATRIX ALL CK TRIANGLE;                   ← constraints for PC[b]
>EST NC=100 IT=25;
>END ;                                      ← Terminates command section
3                                           ← Number of response categories
012                                         ← The response code line
11111111                                    ← The 0s are recoded to be 1s for all eight items
22222222                                    ← The 1s are recoded to be 2s for all eight items
33333333                                    ← The 2s are recoded to be 3s for all eight items
(15A1,T1,8(A1,1X))                          ← The format for reading the data
```

[a] The text following the '←' is provided to help the reader understand the corresponding input.
[b] These lines are inserted in the main MULTILOG window.

To obtain estimates for the PC model we need to impose constraints on the nominal response model; the nominal response model is discussed in Chapter 9. Therefore, the `TEST` line instructs the program to obtain estimates for all items (`ALL`) on the instrument using the nominal response model (`NOMINAL`). Each of the eight items consists of three response categories (i.e., `NC=(3(0)8)`), and the keyword `HIGH=(3(0)8)` indicates that response category 3 is the highest category for each of the eight items. This notation of “`(3(0)8)`” is MULTILOG’s (and BILOG’s) shorthand way of indicating an integer set without having to individually indicate each value in the set. For example, `NC=(3(0)8)` indicates eight 3s and is equivalent to `NC=(3,3,3,3,3,3,3,3)`.

Subsequent to the `TEST` command is where we specify the constraints on the nominal response model to obtain the PC model. That is, the `TMATRIX`, `EQUAL`, and `FIX` lines impose the appropriate constraints on the nominal response model to obtain PC model estimates. These constraints implement the contrasts among the various nominal response model parameters to obtain PC model estimates. `TMATRIX` is MULTILOG's parlance for the The transformation matrix $\underline{\mathbf{T}}$ discussed in Chapter 9. Thissen and Steinberg (1986) as well as Muraki (1992) contain technical details on these constraints. Because the PC model states that all items have the same discrimination, we need to use the line `EQUAL ALL AK=1` to instruct MULTILOG to constrain all the $\hat{\alpha}_j$ s to be equal to one another, although not necessarily equal to 1. (If this line is omitted, each item's α_j would be estimated and the calibration model would be the generalized partial credit model [Chapter 8].)[2]

On the `EST NC=100 IT=50` line the maximum number of EM cycles that can be executed is set to 100 (`NC=100`) and the maximum number of M-step iterations is specified to be 25 (`IT=25`). Following the `END` command line is the description of the response data. There are three response categories with the codes 0, 1, and 2; therefore, these values are listed on the response code line (i.e., the line 012). Because with polytomous models MULTILOG internally uses category 0, we need to recode these response data. We recode the data by specifying that the response of 0 for each of the eight items should be recoded to 1 (i.e., the line 11111111). Similarly, we recode the response codes of 1 and 2 to be 2 and 3, respectively (i.e., the lines 22222222 and 33333333, respectively). The final line contains the FORTRAN format for reading the response data. In this example each individual's response vector is used as his or her character identification code. Therefore, the '`15A1`' in the format is associated with the `NCHARS=15` case identification subcommand on the `PROBLEM` line, whereas the '`8(A1,1X)`' is the format for reading the eight item responses; '`T1`' tells the program to return to the first column. (FORTRAN formats are briefly discussed in Appendix F “FORTRAN Formats.”)

The output presented in Table 2 begins with the specification of the number of examinees, number of items, and number of response codes, along with other logistical information. For example, the program read 2942 cases, there are eight items, and so on. The next section, `ESTIMATION PARAMETERS`, contains information concerning the calibration parameters. It shows that the maximum number of EM cycles is 100 (i.e., `NC=100`) and the maximum number of M-steps is 25 (i.e., `IT=25`). For the EM cycles and M-steps phases the convergence criteria are 0.001 and 0.0001, respectively. In addition, the line `NUMBER OF FREE PARAMETERS IS:` specifies the number of parameters estimated for this calibration. This information is useful for calculating degrees of freedom (e.g., for use with ΔG^2). In this case, there are 17 parameters being estimated (8 items*2 transition locations, δ_{j1} and δ_{j2}, per item plus the common α). Following the `ESTIMATION PARAMETERS` section is the response codes section, `CODE CATEGORY`. One should inspect these codes to ensure that the codes used are the ones that are intended to be used and that there are as many codes as there are items. Similarly, the `FIRST OBSERVATION AS READ` should be verified to be correct. Given that the value of `FINISHED CYCLE` is less than 100 (i.e., `FINISHED CYCLE` is 11) and the `MAXIMUM INTERCYCLE PARAMETER CHANGE` is less than the convergence criterion we have a converged solution.

The `ITEM SUMMARY` section contains the PC model item parameter estimates under the heading `CONTRAST-COEFFICIENTS`. From `ITEM 1` we see that the common item discrimination estimate, $\hat{\alpha}$, for all items is 0.84. Item 1’s first transition location ($\hat{\delta}_{11}$) is estimated to be –1.30 and the second one ($\hat{\delta}_{12}$) is –2.06; the standard errors for $\hat{\alpha}$, $\hat{\delta}_{11}$, and $\hat{\delta}_{12}$, are 0.01, 0.12, and 0.08, respectively. The section `OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN` contains the response frequency for each response category 0, 1, 2 (in the output these are labeled 1, 2, and 3, respectively). For instance, of the 2942 persons in the calibration sample, 201 did not receive any credit, 351 received partial credit, and 81.24% or 2390 received full credit on item 1 (i.e., the traditional item difficulty for item 1, P_1, is 2390/2942 = 0.8124). This is a relatively easy item for this sample. It is important to examine the observed frequencies (or proportions) section because it allows one to determine the distribution of (category) scores on an item, whether there are some scores that are not observed, or whether there are infrequently given scores on a particular item. In these latter cases the corresponding transition location may not be well estimated.

Returning to item 1's `CONTRAST-COEFFICIENTS` section, one sees that each estimate has a parameter number label (i.e., `P(#)`). For example, the common of 0.84 has the parameter number label of 1. Throughout the output all items that share a common $\hat{\alpha}$ (i.e., the items that are involved in the equality constraint) have the corresponding parameter estimate labeled 1. The transition location estimates for item 1, $\hat{\delta}_{11}$ and $\hat{\delta}_{12}$, have the parameter numbers 2 and 3, respectively. If item 2's estimates were presented, then $\hat{\delta}_{21}$ would have the parameter number 4 and $\hat{\delta}_{22}$ would have the parameter number 5; the common $\hat{\alpha}$ would still have the parameter number 1. This characteristic can be seen with the results for the eighth item. The common $\hat{\alpha}$ has the same parameter number as it did for the first item and the $\hat{\delta}_{81}$ =1.27 and $\hat{\delta}_{82}$ = 1.36 have parameter numbers labels of 16 and 17, respectively. That is, $\hat{\delta}_{82}$ is the 17th parameter referred to in the line `NUMBER OF FREE PARAMETERS IS:17`.

The line labeled `@THETA: INFORMATION` shows item 1's information function; the `@THETA` is labeled with a θ range that specifies the corresponding location for the item information value. For example, the first line specifies the range from –3 to –1.6; therefore the information at θ = –3 is 0.352, at θ = –2.8 it is 0.404, and so on. From the θ range provided it be can be seen that item 1 provides most of its information in a neighborhood around –2 and then steadily decreases for the rest of the continuum.

Table 2. Abridged output from MULTILOG PC model calibration example.[a]

```
      :
   <echo of command file>
      :
    DATA PARAMETERS:
     NUMBER OF LINES IN THE DATA FILE: 2942
     NUMBER OF CATEGORICAL-RESPONSE ITEMS:   8
     NUMBER OF CONTINUOUS-RESPONSE ITEMS, AND/OR GROUPS:   1
     TOTAL NUMBER OF "ITEMS" (INCLUDING GROUPS):   9
     NUMBER OF CHARACTERS IN ID FIELDS: 15
     MAXIMUM NUMBER OF RESPONSE-CODES FOR ANY ITEM:  3
     THE MISSING VALUE CODE FOR CONTINUOUS DATA:  9.0000
     THE DATA WILL BE STORED IN MEMORY

    ESTIMATION PARAMETERS:
     THE ITEMS WILL BE CALIBRATED--
       BY MARGINAL MAXIMUM LIKELIHOOD ESTIMATION
     MAXIMUM NUMBER OF EM CYCLES PERMITTED: 100
     NUMBER OF PARAMETER-SEGMENTS USED IS:   9
     NUMBER OF FREE PARAMETERS IS:   17
     MAXIMUM NUMBER OF M-STEP ITERATIONS IS  25 TIMES
       THE NUMBER OF PARAMETERS IN THE SEGMENT
     THE M-STEP CONVERGENCE CRITERION IS: 0.000100
     THE EM-CYCLE CONVERGENCE CRITERION IS: 0.001000
     THE RK CONTROL PARAMETER (FOR THE M-STEPS) IS:  0.9000
     THE RM CONTROL PARAMETER (FOR THE M-STEPS) IS:  1.0000
     THE MAXIMUM ACCELERATION PERMITTED IS:  0.0000
     THETA-GROUP LOCATIONS WILL REMAIN UNCHANGED
   :
    KEY-
    CODE  CATEGORY
     0     11111111                            ← The recodings to 1 for a response of 0
     1     22222222                            ← The recodings to 2 for a response of 1
     2     33333333                            ← The recodings to 3 for a response of 2
   :
    FIRST OBSERVATION AS READ-
    ID        1 1 2 0 0 0 0 0
    ITEMS 11200000                             ← The  response pattern for the first person
    NORML      0.000

    FINISHED CYCLE  11                         ← Number of iterations
    MAXIMUM INTERCYCLE PARAMETER CHANGE=   0.00055 P(  17)




    ITEM SUMMARY

    ITEM   1:       3 NOMINAL CATEGORIES,  3 HIGH
     CATEGORY(K): 1      2      3
       A(K)     -0.84   0.00   0.84
       C(K)      0.00   1.30   3.36

              CONTRAST-COEFFICIENTS (STANDARD ERRORS)
```

```
  FOR:             A                   C
  CONTRAST P(#)  COEFF.[POLY.]  P(#)  COEFF.[ TRI.]
      1       1   0.84 (0.01)      2  -1.30 (0.12)
      2      18   0.00 (0.00)      3  -2.06 (0.08)

 @THETA:      INFORMATION:   (Theta values increase in steps of 0.2)
 -3.0 - -1.6  0.352  0.404  0.451  0.489  0.514  0.523  0.513  0.487
 -1.4 -  0.0  0.448  0.400  0.349  0.298  0.251  0.209  0.172  0.141
  0.2 -  1.6  0.115  0.093  0.076  0.062  0.051  0.042  0.034  0.028
  1.8 -  3.0  0.023  0.019  0.016  0.013  0.011  0.009  0.008

  OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
  CATEGORY(K):  1      2      3
  OBS. FREQ.    201    351   2390
  OBS. PROP.  0.0683 0.1193 0.8124
  EXP. PROP.  0.0669 0.1203 0.8128
:

 ITEM   8:       3 NOMINAL CATEGORIES,  3 HIGH
  CATEGORY(K): 1      2      3
    A(K)     -0.84   0.00   0.84
    C(K)      0.00  -1.27  -2.63

           CONTRAST-COEFFICIENTS (STANDARD ERRORS)
  FOR:             A                   C
  CONTRAST P(#)  COEFF.[POLY.]  P(#)  COEFF.[ TRI.]
      1       1   0.84 (0.01)     16   1.27 (0.06)
      2      39   0.00 (0.00)     17   1.36 (0.09)

 @THETA:      INFORMATION:   (Theta values increase in steps of 0.2)
 -3.0 - -1.6  0.017  0.020  0.023  0.028  0.033  0.040  0.047  0.057
 -1.4 -  0.0  0.068  0.081  0.096  0.115  0.137  0.163  0.192  0.226
  0.2 -  1.6  0.262  0.301  0.341  0.379  0.413  0.439  0.455  0.460
  1.8 -  3.0  0.452  0.433  0.404  0.369  0.330  0.290  0.252

  OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
  CATEGORY(K):  1      2      3
  OBS. FREQ.   2043    616    283
  OBS. PROP.  0.6944 0.2094 0.0962
  EXP. PROP.  0.6947 0.2079 0.0974

 ITEM   9: GRP1, N[MU:  0.00 SIGMA:  1.00]
    P(#);(S.E.):   43; (0.00)   44; (0.00)
:

 TOTAL TEST INFORMATION

 @THETA:      INFORMATION:
 -3.0 - -1.6  1.831  1.972  2.124  2.283  2.444  2.602  2.755  2.898
 -1.4 -  0.0  3.031  3.153  3.264  3.365  3.457  3.538  3.604  3.651
  0.2 -  1.6  3.669  3.654  3.602  3.513  3.390  3.240  3.071  2.889
  1.8 -  3.0  2.704  2.523  2.350  2.190  2.044  1.913  1.798
```

```
@THETA:      POSTERIOR STANDARD DEVIATION:
 -3.0 - -1.6  0.739  0.712  0.686  0.662  0.640  0.620  0.603  0.587
 -1.4 -  0.0  0.574  0.563  0.554  0.545  0.538  0.532  0.527  0.523
  0.2 -  1.6  0.522  0.523  0.527  0.534  0.543  0.556  0.571  0.588
  1.8 -  3.0  0.608  0.630  0.652  0.676  0.699  0.723  0.746

 MARGINAL RELIABILITY:    0.6986

 NEGATIVE TWICE THE LOGLIKELIHOOD=     -7624.1
 (CHI-SQUARE FOR SEVERAL TIMES MORE EXAMINEES THAN CELLS)
```

[a] The text following the '⇐' is provided to help the reader understand the corresponding input.

For contrast purposes, item 8’s information is presented. This item is correctly answered by only 9.62% of the sample and its item information function peaks in the vicinity of 1.6. Given the model, its $\hat{\alpha}$ is the same as that of item 1 and its corresponding transition locations, $\hat{\delta}_{81}$ = 1.27 and $\hat{\delta}_{82}$ = 1.36, are relative to item 1’s transition locations, at the opposite end of the θ continuum. (“Item 9” is not really an item but is concerned with whether the sample comes from a [prior] normal distribution with mean 0 and variance 1.). The $\hat{\delta}_{j1}$s and $\hat{\delta}_{j2}$s correlate 0.99999 and 1.00000 with the corresponding flexMIRT estimates (see Chapter 7).

Figure 1 contains the ORFs and information functions for item 1; the PLOT menu item on the RUN menu produced this figure. Given the reversal in $\hat{\delta}$s for item 1 (i.e., $\hat{\delta}_{11}$ = –1.30 and $\hat{\delta}_{12}$ = –2.06), it is not surprising that respondents located above (approximately) –2.0 are more likely to receive full credit (ORF labeled “3”) rather than partial credit (ORF labeled “2”; left panel of Figure 1). Similarly, respondents located below –2.0 have a higher probability of receiving no credit (ORF labeled “1”) rather than partial credit. Although the OBSERVED AND EXPECTED COUNTS indicate that each of the category scores are being utilized, the ORFs show that, in effect, this item is primarily behaving like a dichotomous item. For most of the respondents this is a relatively easy item. The @THETA: INFORMATION section of the output shows that at approximately –2.0 this item provided its maximum information. This is confirmed by the item information function in the right panel of Figure 1.

Figure 1. ORFs and item information function for item 1.

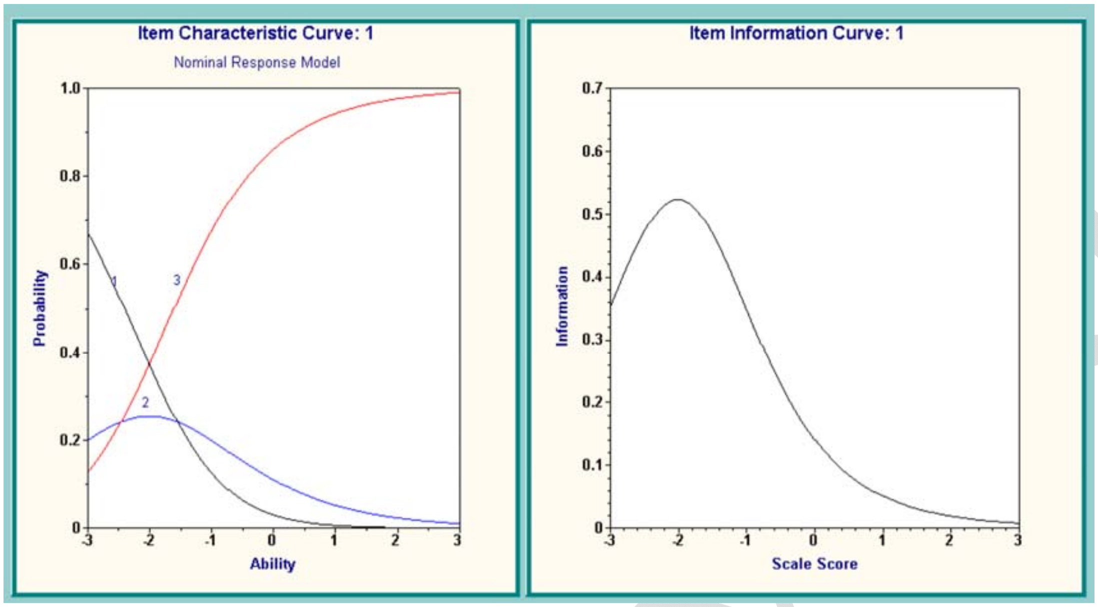
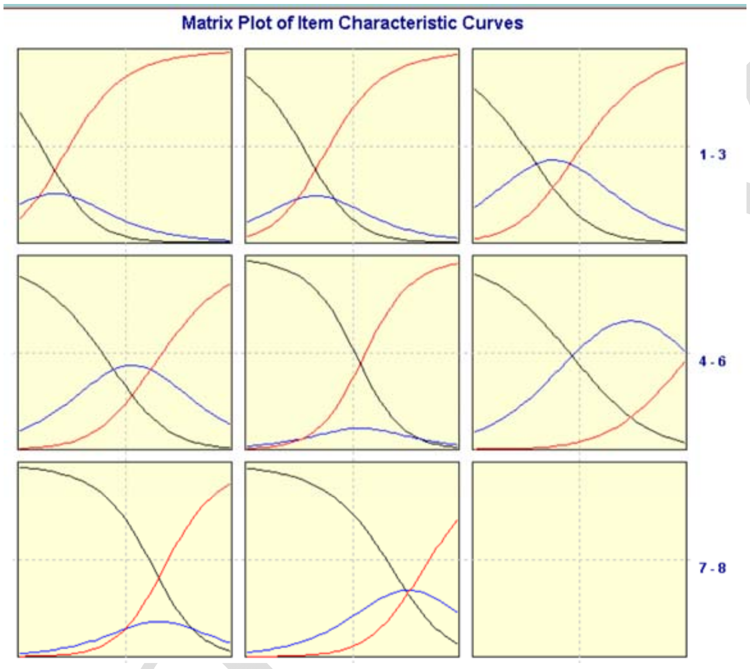


Figure 2 contains the ORFs for all the items on the instrument. Four of the items have a tendency to behave in a dichotomous way (items 1, 2, 5, and 7), whereas items 3, 4, 6, and 8 are trichotomous. It is also seen that items 1, 2, and 3 are useful for assessing people located in the lower half of the θ continuum, items 4, 5, and 7 provide information in the middle of the continuum, and items 6 and 8 are tapping the upper half of the continuum. Therefore, the instrument gives relatively broad capacity to measure reasoning ability.

The sum of the item information functions is captured by the instrument's total information (in Table 2, labeled `TOTAL TEST INFORMATION`); these values have 1.0 added to them. This table, read similarly to that of the item information, indicates that the maximum information for locating individuals occurs in the neighborhood of θ = 0.2. A graphical depiction of this information function is shown in Figure 3; the `PLOT` menu item on the RUN menu produced this figure. This double-Y graph overlays the corresponding standard errors for persons as a function of θ on the total information function (the right side ordinate provides the scale for interpreting the standard errors); similar information is provided in the table entitled `@THETA: POSTERIOR STANDARD DEVIATION` in the output (Table 2). As can be seen, the accuracy of person location estimation varies as a function of the specific location. Specifically, this instrument does a comparatively better job of estimating a person's location from around –0.5 to 1.0 and a progressively poorer job as one approaches 2 or –2.

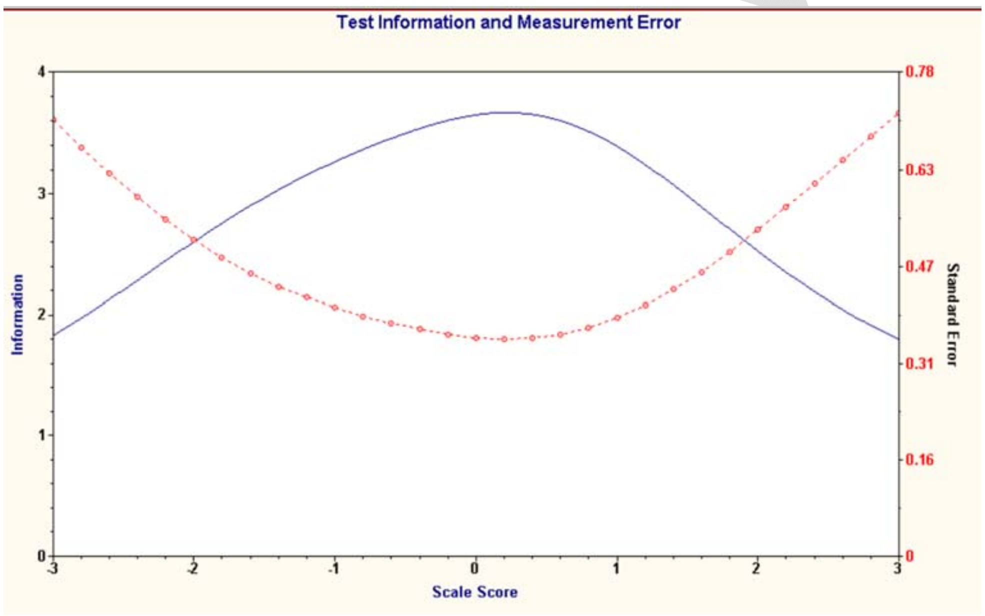
Figure 2. ORFs for all eight Alike items.[a]



[a] Items are read left to right, top to bottom. For example, for row 1 item 1 is left, item 2 is center, item 3 is right; for row 2 item 4 is left, item 5 is center, and so on.

Most practitioners and consumers of psychometrics are familiar with traditional estimators of the reliability of an instrument's scores (e.g., Cronbach (1951) coefficient alpha/Guttman λ_3 (1945)). Therefore, in some situations it may be desirable to have a single numerical value that "captures" the accuracy of the person location estimation. Green, Bock, Humphreys, Linn, and Reckase (1984) present an average or marginal measurement error index, the *marginal reliability*, that is unitless and is bounded by 0 and 1. MULTILOG presents a `MARGINAL RELIABILITY` on the output. The marginal reliability for the $\hat{\theta}$s from this instrument is 0.6986 and it reflects an average accuracy across the continuum. However, it is only when the total information function is somewhat uniformly distributed that this value accurately characterizes the precision of measurement across the continuum. In our example, the total information function is peaked and, as a result, the marginal reliability is underrepresenting the accuracy in the vicinity of the peak and overrepresenting accuracy in the tails of the information function.[3]

Figure 3. Total information for Alike reasoning exam, PC model.[a]



[a] Legend-Solid line: total information, Dotted line: Standard error.

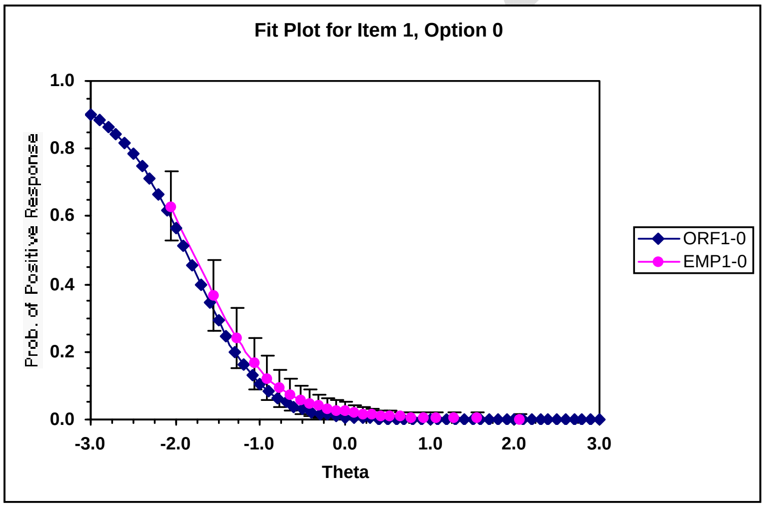
In terms of model-level fit information MULTILOG provides –2ln$L$ at the end of its output (i.e., `NEGATIVE TWICE THE LOGLIKELIHOOD`). (The value printed is negative 2 times the log of a number that is proportional to the likelihood [Thissen et al., 2003].) With individual case data, as we have in this example, the –2ln$L$ value of –7624.1 is not useful and is ignored. However, when pattern data are calibrated, this index may be used for an overall assessment of model–data fit as well as model comparisons; this assumes that most or all of the possible patterns are observed. Chapters 8 and 9 contain examples using pattern data with MULTILOG.

Because MULTILOG 7 (and earlier versions) does not provide item-level tests of fit we use another program, MODFIT (Levine, Drasgon, & Stark, 2001), to obtain this fit information. As mentioned above, the sample sizes typically seen in calibrations result in potentially powerful statistical fit tests. As such, significant statistical tests should be interpreted in conjunction with graphical displays. In this regard, MODFIT provides both the empirical and predicted ORFs, as well as the item fit statistics (e.g., χ^2). (MODFIT is limited to 40 two- to seven-option items and 3000 cases. It can produce these graphs for dichotomous as well as polytomous models; the technical details may be found in Drasgow, Levine, Tsien, Williams, & Mead [1995].)

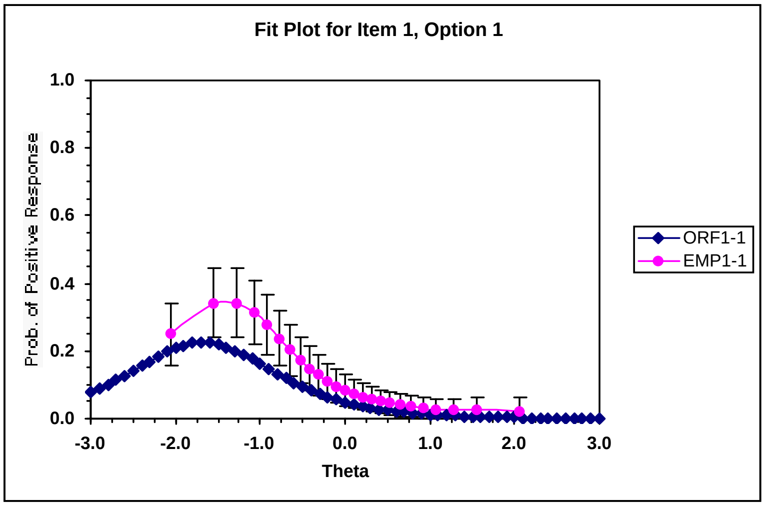
Figure 4 contains the empirical and predicted ORFs for each option for item 1 using MODFIT. As can be seen from panel A, the predicted ORF (labeled ORF 1-0; symbol = diamond) falls within the 95% error bars of the empirical ORF (labeled EMP 1- 0; symbol = circle) for the incorrect response (i.e., option 0). Similarly, the partially correct response (i.e., option 1) shows some congruence between the predicted (ORF 1-1) and empirical (EMP 1- 1) ORFs (panel B), but not as well as for option 0. The same may be said for the correct response, option 2 (panel C). Panel C shows that for part of the continuum the predicted (ORF 1-2) does not fall within the error bars of the empirical ORF (EMP 1- 2), indicating a lack of agreement. Examination of the empirical and predicted ORFs for the other items shows better agreement between the empirical and predicted ORFs than seen in Figure 4; this is true for all of the remaining items except for item 3. However, overall we consider the agreement between the empirical and predicted ORFs to be sufficiently good to believe that we have model–data fit with the PC model. (In the next chapter a model that allows items to vary in terms of their discrimination is applied to these data to see if model–data fit may be improved.)

It is good practice that model–data fit analysis include examinations of the invariance and conditional independence assumptions. The corresponding evidence would aid in supporting (or not) the above conclusion of model–data fit with the PC model. Because MULTILOG does not have BILOG's random sampling capability (e.g., see Chapter 5) we would use the sampling procedure presented in Chapter 3 to investigate invariance with the PC model. Analogous to what was done in Chapter 3, this invariance investigation would use the correlations between corresponding $\hat{\delta}_{jh}$ s across the two groups.

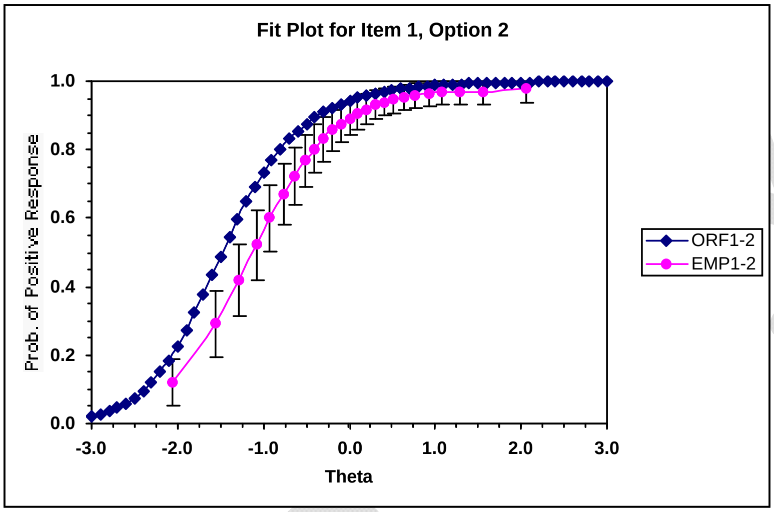
Figure 4. Empirical and predicted ORFs for item 1.
Panel A.



Panel B.



Panel C.



NOTES

1. MULTILOG does not use the scaling constant *D* in its estimation. As a result, the item parameter estimates are on the logistic metric. To place them on the normal metric we would transform the discrimination estimates using

$$\hat{\alpha}^* = \frac{\hat{\alpha}}{\zeta}$$

2. Because the constant α across items may not be equal to 1, the relationship between the MULTILOG PC model and the Masters PC model is analogous to the relationship between the 1PL and Rasch models. That is, mathematically, the MULTILOG PC model and the Masters PC model are equivalent: The values from one model can be converted into the other by appropriate rescaling. However, if one wants to directly estimate the Masters PC model, then a program such as BIGSTEPS, WINSTEPS, or ConQuest can be used.

3. Green et al. (1984) define the marginal reliability as

$$\rho_{m\arg} = \frac{\theta_\theta^2 - \theta_{em}^2}{\theta_\theta^2}$$

where θ_θ^2 is the variance of the observed person locations, θ_{em}^2 is the marginal measurement error

$$\theta_{em}^2 = \int_{-\infty}^{\infty} \sigma_e^2(\theta) g(\theta) d\theta \Bigg/ \int_{-\infty}^{\infty} g(\theta) d\theta$$

and $g(\theta)$ is the person distribution that, when $g(\theta)$ is normally distributed, can be evaluated using a Gaussian quadrature approach.

As an alternative to an index for the entire metric, one may calculate the *conditional reliability* (Green et al., 1984) at a θ point. They define conditional reliability as

$$\rho(\theta) = \frac{\theta_{\theta}^{2} - \theta_{e}^{2}(\theta)}{\theta_{\theta}^{2}} \tag{7.10}$$

where θ_{θ}^{2} is the variance of the observed person locations and $\theta_{e}^{2}(\theta)$ is the expected variance error of estimate. The term $\rho(\theta)$ specifies "the reliability if everyone were measured with the same precision as those persons" (p. 353) located at θ. In addition, the graphing of $\rho(\theta)$ as a function of θ would allow an assessment of the estimation properties of the instrument at various levels of θ on a bounded and potentially more easily interpreted (ordinate) scale from 0 to 1 than that used with the total information function. An additional reliability index is provided by Bock and Mislevy (1982). Their reliability coefficient for the EAP location estimate, $\rho = 1 - PSD(\hat{\theta})^{2}$, is based on the assumption that the latent variable is normally distributed in the population with mean 0 and variance 1.