Example: Application of the GR Model to an Attitudes Towards Condoms Scale,

MMLE, MULTILOG

In Chapter 7 we apply the RS model to survey data from an attitudes toward condoms instrument. We now revisit these data. Recall that our instrument contains six statements people have made about condoms. The respondents are asked how much they agree with each of the statements on a 4-point Likert response scale (1="strongly disagree," 2="disagree more than I agree," 3="agree more than I disagree," 4="strongly agree"). Because the six statements are "negatively" worded, a respondent who strongly agrees with a statement is indicating a less favorable attitude toward condom use. As we did in Chapter 7, we assume that we have evidence supporting the tenability of the unidimensionality assumption. We use MULTILOG to perform our calibration.

The MULTILOG command file for this calibration is presented in Table 1. The keyword GRADED on the TEST line indicates a GR model calibration. In addition, as we did with the PC model, we need to specify the number of categories per item by using the NC keyword. Because we are using individual case data, the item parameter estimation phase is uncoupled from the person location estimation stage. Therefore, we now show how to obtain person location estimates with case data. We first focus on obtaining model–data fit for the items. Once satisfactory model–data fit is obtained, we then turn our attention to the second stage of estimating the people. To facilitate this second stage the SAVE command is used with item parameter estimation to save the item parameter estimates to an external file. By default, this external file has the same name as the command file, but with the MLG extension replaced by PAR. In addition, MULTILOG allows one to specify item labels for ALL or a subset of the

de Ayala, R.J. (2009). *The Theory and Application of Item Response Theory*, New York: Guilford Publishing.

items through the `LABELS` command; the labels are limited to four characters. Therefore, we

introduce this capability to provide item labels for all the items via `>LABELS ALL,NAMES=`

`('I26B','I26C','I26E','I26G','I26H','I26K');`. The default values for EM

cycles and iterations are used. With the GR model the program assumes that the response codes

begin with 1. Because our data consist of the codes 1 through 4, our response code line is 1234

and we do not have to recode our responses. Therefore, unlike the case with the PC model (cf.

MULTILOG_PCMcalibrationEx.pdf), each of our response lines (i.e., the lines `111111`,

`222222`, and so on) matches the values in the response code line.

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

```
MULTILOG for Windows 7.00.2327.2
Attitudes toward condoms
>PROBLEM RANDOM, INDIVIDUAL, DATA = 'C:CONDOMS.DAT',NITEMS = 6, NGROUPS = 1,
       NEXAMINEES = 3473, NCHARS = 4;
>TEST ALL,
      GRADED,                               ← Specification of GR model
      NC = (4(0)6);
>LABELS ALL,NAMES=('I26B','I26C','I26E', 'I26G','I26H','I26K');
>SAVE ;
>END ;
4                                     ← Number of response categories
1234                                  ← The response code line
111111                                ← The 1s are coded to be 1s for all six items
222222                                ← The 2s are coded to be 2s for all six items
333333                                ← The 3s are coded to be 3s for all six items
444444                                ← The 4s are coded to be 4s for all six items
(4A1,6A1)
```

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

The output is presented in Table 2. The listing shows that the `MAXIMUM NUMBER OF EM`

`CYCLES PERMITTED` is 25 and that MULTILOG executed 23 (i.e., `FINISHED CYCLE`

`23`). Therefore, convergence is achieved. (If convergence is not achieved, then we would

modify the command file by, for example, inserting the line `>EST NC=999 IT=250;` directly

preceding the `>END;` line to increase the number of EM cycles and iterations and re-execute the

program.) The `NUMBER OF FREE PARAMETERS IS: 24` and reflects that there are 3 δ_{x_j} s

plus one α_j for each of the six items.

Table 2. Abridged output from MULTILOG GR model calibration example.

```
    :
 ESTIMATION PARAMETERS:
  THE ITEMS WILL BE CALIBRATED--
    BY MARGINAL MAXIMUM LIKELIHOOD ESTIMATION
  MAXIMUM NUMBER OF EM CYCLES PERMITTED:  25
  NUMBER OF PARAMETER-SEGMENTS USED IS:  15
  NUMBER OF FREE PARAMETERS IS:   24
  MAXIMUM NUMBER OF M-STEP ITERATIONS IS   4 TIMES
    THE NUMBER OF PARAMETERS IN THE SEGMENT
  THE M-STEP CONVERGENCE CRITERION IS: 0.000100
  THE EM-CYCLE CONVERGENCE CRITERION IS: 0.001000
    :
 FINISHED CYCLE  23
 MAXIMUM INTERCYCLE PARAMETER CHANGE=   0.00070 P(   9)

 ITEM SUMMARY

 ITEM   1: I26B  4 GRADED CATEGORIES
         P(#) ESTIMATE (S.E.)
A         1    1.02   (0.05)
B( 1)     2   -0.55   (0.05)
B( 2)     3    0.13   (0.05)
B( 3)     4    1.00   (0.07)

@THETA:       INFORMATION:   (Theta values increase in steps of 0.2)
-3.0 - -1.6 0.073  0.087  0.103  0.120  0.139  0.159  0.181  0.202
-1.4 -  0.0 0.224  0.245  0.264  0.280  0.294  0.305  0.313  0.317
 0.2 -  1.6 0.319  0.318  0.314  0.306  0.296  0.283  0.267  0.249
 1.8 -  3.0 0.229  0.207  0.186  0.164  0.143  0.124  0.106

  OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
  CATEGORY(K):  1      2      3      4
  OBS. FREQ.   1338    492    602   1041
  OBS. PROP.  0.3853 0.1417 0.1733 0.2997
  EXP. PROP.  0.3858 0.1411 0.1725 0.3007
    :
@THETA:       INFORMATION:
-3.0 - -1.6 1.384  1.428  1.476  1.529  1.586  1.650  1.722  1.806
-1.4 -  0.0 1.907  2.031  2.189  2.393  2.653  2.974  3.347  3.739
 0.2 -  1.6 4.099  4.377  4.539  4.579  4.500  4.300  3.984  3.580
 1.8 -  3.0 3.139  2.715  2.345  2.045  1.811  1.634  1.500

@THETA:       POSTERIOR STANDARD DEVIATION:
-3.0 - -1.6 0.850  0.837  0.823  0.809  0.794  0.778  0.762  0.744
-1.4 -  0.0 0.724  0.702  0.676  0.646  0.614  0.580  0.547  0.517
 0.2 -  1.6 0.494  0.478  0.469  0.467  0.471  0.482  0.501  0.529
 1.8 -  3.0 0.564  0.607  0.653  0.669  0.743  0.782  0.816
    :
 MARGINAL RELIABILITY:    0.6706
```

The item parameter estimates for the first item are presented to show the their presentation format. The MULTILOG GR model estimates are on the logistic metric. The first line labeled A provides the item's $\hat{\alpha}$, and the following lines labeled B(1), B(2), and B(3) are the $\hat{\delta}_{11}$, $\hat{\delta}_{12}$,

and $\hat{\delta}_{13}$, respectively (i.e., $\hat{\alpha}_1 = 1.02$, $\hat{\delta}_{1_1} = -0.55$, $\hat{\delta}_{2_1} = 0.13$, and $\hat{\delta}_{3_1} = 1.00$). As would be

expected, the $\hat{\delta}_{x_j}$ s are in sequential order. Table 3 contains all the item parameter estimates.

Most items show reasonable discrimination capacity and, it appears, that the instrument should

perform well in estimating individuals in the approximate range of -1.88 to 1.23. This belief

may be verified by examining the total information function.

Table 3. GR model item parameter estimates for the attitudes towards condoms 6-item scale.

Item	$\hat{\alpha}_j$	$s_e(\hat{\alpha}_j)$	$\hat{\delta}_{1_j}$	$s_e(\hat{\delta}_{1_j})$	$\hat{\delta}_{2_j}$	$s_e(\hat{\delta}_{2_j})$	$\hat{\delta}_{3_j}$	$s_e(\hat{\delta}_{3_j})$
1	1.02	(0.05)	-0.55	(0.05)	0.13	(0.05)	1.00	(0.07)
2	2.09	(0.08)	0.39	(0.03)	0.74	(0.03)	1.13	(0.04)
3	2.15	(0.09)	0.50	(0.03)	0.87	(0.03)	1.23	(0.04)
4	0.67	(0.04)	-1.57	(0.12)	-0.34	(0.07)	1.07	(0.10)
5	0.90	(0.05)	-0.74	(0.06)	-0.15	(0.05)	0.51	(0.06)
6	0.77	(0.05)	-1.88	(0.12)	-0.93	(0.08)	0.40	(0.07)

As discussed above, the item parameter estimates are followed by two tables containing the

item's information as well as the observed and expected proportions for each category.

Comparing the expected and observed proportions shows that there is relatively close agreement

between the observed proportion within a category and what would be predicted on the basis of

the model using these parameter estimates. As mentioned in previous chapters, our fit analysis

should include a comparison of empirical and predicted ORFs. Although MULTILOG does not

provide these graphs, the program MODFIT can be used for this purpose with the GR model.

The ORFs for the first item are presented in Figure 1. This figure's left panel shows that this

item ("Sex doesn't feel as good when you use a condom") is behaving primarily as a

dichotomous item. Below approximately $\theta = 0.33$ a respondent is most likely to "strongly

disagree" with the statement, and above this point would most likely "strongly agree" with the

statement. The figure's right panel shows that this item provides most of its information around

0.2, although not a large amount relative to the other items on the instrument. For example, Figure 2's right panel shows that item 2 provides substantially more information for estimating a person's location than does item 1. This difference in the amount of available information is directly related to the items' $\hat{\alpha}_j$ s; $\hat{\alpha}_1 = 1.02$ and $\hat{\alpha}_2 = 2.09$. This item ("It's embarrassing to buy condoms") also behaves primarily as a dichotomous item (the left panel). The ORFs show that individuals who hold positive attitudes toward condom use (the left side of the continuum), as well as "somewhat" neutral attitudes, are most likely not to find the purchase of condoms embarrassing. In other words, persons with θ s below approximately 0.8 are most likely to "strongly disagree" with the statement "It's embarrassing to buy condoms."

Although our examination of item 1's ORFs shows that the item is behaving primarily as a dichotomous item, this observation cannot be gleaned from the OBSERVED AND EXPECTED COUNTS/PROPORTIONS table nor from examining the $\hat{\delta}_{x_j}$ s (Table 3); this is also true for item 2. Therefore, it is recommended that standard practice with the GR model be to obtain the ORFs for all items on the instrument. For completeness, our examination of the ORFs for the remaining items on the instrument reveals the same dichotomous-like pattern of ORFs seen with items 1 and 2. If these data were from a pilot test of the instrument, then the instrument's author(s) might want to consider modifying the statements.

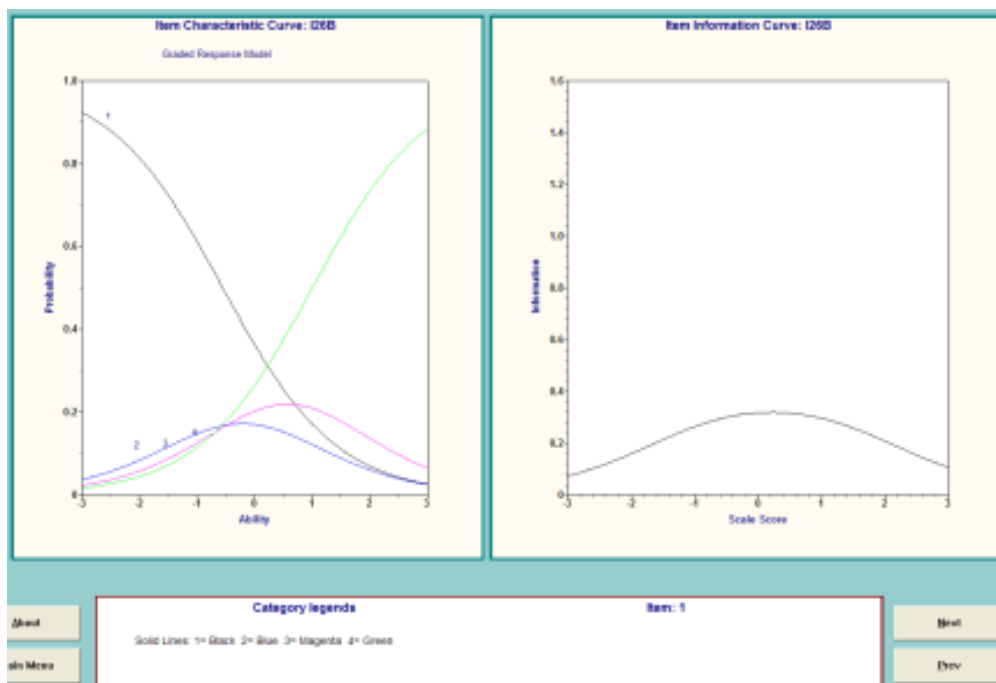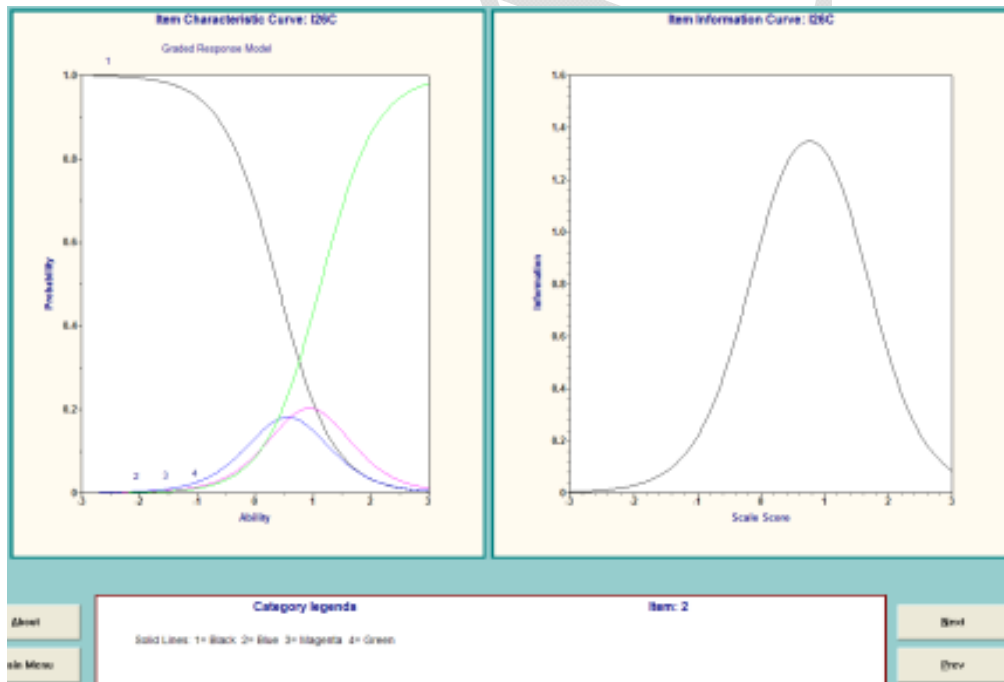Figure 1. ORFs and item information function for item 1.

Figure 2. ORFs and item information function for item 2.

For pedagogical reasons, we assume that we have model–data fit analysis and that there is no concern with the dichotomous-like nature of the items. Therefore, the six-item scale is used to estimate the respondents' attitudes toward condoms. In previous chapters both the MLE and EAP approaches to person location estimation have been used. This time we use *maximum a posteriori* (MAP). (MAP is discussed in Chapter 4.) Recall that like EAP, MAP provides a Bayes estimator that is typically regressed toward the mean of the prior distribution. However, whereas EAP uses the mean of the posterior distribution, MAP uses its mode. As a Bayesian approach, MAP provides person location estimates for all response patterns.

To obtain the person estimates with individual case data requires a second execution of MULTILOG. In this second execution we may choose to re-estimate the item parameters or we can provide the item parameter estimates and estimate only the people. Because in our item parameter calibration stage we saved the item parameter estimates to an external file, we estimate only the people in our second execution.

Table 4. Command file for MULTILOG GR model calibration example -MAP estimation.[a]

```
MULTILOG for Windows 7.00.2327.2
Attitudes toward condoms; used previously save item param estimates
>PROBLEM SCORE, INDIVIDUAL, DATA = 'C:CONDOMS.DAT',NITEMS = 6, NGROUPS = 1,
        NEXAMINEES = 3473, NCHARS = 4;
>TEST ALL,
       GRADED,                              ← Specification of GR model
       NC = (4(0)6);
>START ALL,PARAM='C:CONDOMS.PAR';
>END ;
4
1234
111111
222222
333333
444444
(4A1,6A1)
```

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

To estimate the person locations we modify the command file presented in Table 1. The

PROBLEM RANDOM, ... line in the command file in Table 1 is replaced with PROBLEM

SCORE, ... The subcommand SCORE on the PROBLEM line instructs MULTILOG to

calculate person location estimates. We use MAP estimation with the default normal prior with a

mean of 0.0.[1] To read the previously saved item parameter estimates file, the START command

is used. Specifically, >START ALL,PARAM='C:CONDOMS.PAR'; instructs MULTILOG to

use the item parameter values in the file CONDOMS.PAR for all the items; this file's name is

created in the item parameter estimation stage. (If the START command had been omitted, then

MULTILOG would have re-estimated the item parameters in the process of estimating the

person locations, but would not have displayed the item parameter estimates. In this context, the

use of the SAVE command would instruct MULTILOG to save the person, not item, parameter

estimates to an external file with the same file name as the command file, but with the extension

SCO.) The command file for person parameter estimation is presented in Table 4.

Table 5 contains some of the output. The file has the format previously seen, except that the

item parameter values read from the external file are echoed to the output file. These should be

checked to ensure that they are correct. The line MODAL THETAS WILL BE ESTIMATED

indicates person location estimation; with MULTILOG 7.03 this line does not indicate MAP

estimation because it is also displayed when MLE is used. In addition, the line THE MODAL

THETAS WILL BE PRINTED states that the person location estimates are displayed in the

output file as opposed to being saved to an external file (i.e., through the use of the SAVE

command).

It can be seen from the FIRST OBSERVATION AS READ line that the first person

responded "strongly disagree" to all six items (i.e., the person's response vector is 111111).

Although estimating this person's location with MLE would not have been possible, with MAP

this individual receives a location estimate of –1.883 with a standard error of 0.661.[2] This

person is almost 2 logits below the mean. In terms of the instrument's purpose this person has a

positive attitude toward condom use. In contrast, the last person responded "strongly agree" to

all items and obtained a $\hat{\theta}$ of 1.883 with the same standard error as the first person; the response

vector is determined by looking at the data file. This $\hat{\theta}$ of 1.883 indicates that the person has a

negative attitude toward condom use. The standard error is, in part, a function of the

instrument's length as well as the quality of the items comprising the scale. Given the

instrument's total information function, we would expect, by and large, that the standard errors

would be smaller for $\hat{\theta}$s around the center of the attitude toward condoms continuum than at the

extremes of this latent variable.

Table 5. Abridged MAP output from MULTILOG GR model calibration example.

```
 :
>START ALL,PARAM='C:CONDOMS.PAR';
            PARAMETER VALUES
 :
< Estimates from  external file displayed>
 :
 DATA PARAMETERS:
   NUMBER OF LINES IN THE DATA FILE: 3473
 :

 ESTIMATION PARAMETERS:
    MODAL THETAS WILL BE ESTIMATED

 OUTPUT-CONTROL PARAMETERS
  THE MODAL THETAS WILL BE PRINTED
 :
 FIRST OBSERVATION AS READ-

 ID        1
 ITEMS 111111
 NORML      0.000

 SCORING DATA...

   THETAHAT     S.E.   ITER    ID FIELD
     -1.883    0.661    6         1
     -1.883    0.661    6         2
     -1.315    0.658    4         3
     -1.883    0.661    6         4
     -1.883    0.661    6         5
      :
      1.883    0.661    6       3471
      1.883    0.661    6       3472
      1.883    0.661    6       3473
```

NOTES

1. If we want MLE person location estimates, then the PROBLEM line would also include the subcommands NOPOP and CRITERION (i.e., PROBLEM SCORE, ..., NOPOP, CRITERION;) In addition, the response vectors would need to be augmented to have a criterion value (e.g., 0.0) or at least as many blanks spaces as specified in the format statement. For example, the first case would be "ᵇᵇᵇ1111111ᵇ0.0" where the ᵇ indicates a blank space (i.e., **x** = " 1111111 0.0"). As a result, the corresponding format statement would need to include an additional field to read the criterion value, say F4.0, so that the format statement is (4A1,6A1,F4.0). Because the output file contains one or more lines for each case, we recommend that the SAVE command be used if the estimates are to be used for further analysis. When MLE is not possible for a case, then the output file contains an indication that the estimation is not possible (i.e., DID NOT CONVERGE, SUBJECT - <case number>) on one line, followed on the subsequent line by the case's response vector. If the SAVE command is used, then only the nonconverged information is presented in the output file. However, if the SAVE command is not used, then the output file contains nonconverged cases intermixed with the converged cases. For the converged cases, the $\hat{\theta}$, its standard error, the number of iterations required to achieve convergence, and the case identification are presented.

2. As is true with dichotomous models, using MLE for estimating a respondent's location in this example will not provide a finite estimate for either a response vector of all 1s or of all 4s (i.e., response vectors that consist only of the minimum or only of the maximum response categories). However, MLE respondent location estimates are available for the other (nonminimum or nonmaximum) constant response vectors of, for example, all 2s or all 3s, assuming a four-category score item. To understand why this is so, consider the ORFs for a

three-category score item with category scores of 1, 2, and 3. The ORF for $x_j = 1$ is asymptotic with 1 and 0, whereas the ORF for $x_j = 3$ is asymptotic with 0 and 1. Conceptually speaking, the likelihood function for a response vector of all 3s can be envisioned as the multiplication of the corresponding ORFs L times. As a result, the likelihood function looks like the ORF for a category score of 3 and does not have a mode. Conversely, the likelihood function for a response vector of all 1s is the multiplication of the corresponding ORFs L times and looks like the ORF for a category score of 1 and does not have a mode. In neither case can MLE provide a finite estimate of θ. However, for $x_j = 2$ we have an ORF with a mode. Therefore, when one has a response vector of 2s, the product of the corresponding ORFs results in a likelihood function that has a mode. In this case, MLE can provide a finite estimate of θ.