Example: Application of the MC Model to a Science Test, MMLE, MULTILOG

As an example, we reanalyze the science test data used in the NR model example. Recall that these data involved four physical science items and responses from 1799 examinees. We observed 233 of the 256 possible patterns. Three of the items were in a multiple-choice format with four options, and the fourth item was open-ended and scored using a 4-point rubric. We use MULTILOG to perform our calibration.

Table 1 contains the command file. To instruct MULTILOG to use the MC model, the `TEST` line contains the keyword `BS` (B for Bock and S for Samejima). Given that a code of 1 represents the latent “don’t know” category, we need to recode the observed responses of 1, 2, 3, 4 to be 2, 3, 4, and 5, respectively. This implies that one needs to count the “don’t know” category when specifying the number of categories on the `TEST` line (i.e., NC=(5,5,5,5)) and when specifying the `HIGH` category for each of the items. Recoding of the responses is accomplished as done with the PC model calibration example (see MULTILOG_PCMcalibrationEx.pdf).

Table 1. Command file for MC model calibration.[a]

```
MULTILOG for Windows 7.00.2327.2
MC CALIBRATION, 4 PHYSICAL SCIENCE ITEMS
>PROBLEM RANDOM,
         PATTERNS,
         DATA = 'C:SCIENCE.PAT',
         NITEMS = 4,
         NGROUPS = 1,
         NPATTERNS = 256,
         NCHARS = 9;
>TEST ALL,
      BS,                            ⇐ Specification of MC model
      NC = (5,5,5,5),                ⇐  "Five" categories/item  counting the "don't know" category
      HIGH = (3,4,5,4);              ⇐ Ordinal position of highest frequency category with
                                     respect to "five" category scale
>EST NC=500 IT=25;
>END ;
4                                    ⇐ Specification of the number of codes in the data file,
                                     not counting the "don't category"
1234                                 ⇐ The response code line
```

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

```
2222                                ⇐ The 1s are recoded to be 2s for all four items
3333                                ⇐ The 2s are recoded to be 3s for all four items
4444                                ⇐ The 3s are recoded to be 4s for all four items
5555                                ⇐ The 4s are recoded to be 5s for all four items
(9A1,T1,4A1,F5.0)
```

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

The corresponding output is presented in Table 2. We see that the response codes are correctly recoded (see the `VECTOR OF CATEGORIES FOR CODE=` and the `CODE CATEGORY` sections). The program took 253 cycles to achieve convergence. The calibration output format is similar to that of the NR model except for the inclusion of an additional line titled `D(K)` that contains the $\hat{\phi}_{jk}$ s as well as a column labeled `D`. The item parameter estimates for item 1 are $\underline{\hat{\boldsymbol{\alpha}}}$ = (–10.81, 3.16, 3.81, 1.10, 2.75), $\underline{\hat{\boldsymbol{\gamma}}}$ = (–5.95, 1.65, 1.83, 1.31, 1.16), and $\underline{\hat{\phi}}$ = (0.24, 0.29, 0.04, 0.43) with $\sum \hat{\alpha}_{jk} = 0$, $\sum \hat{\gamma}_{jk} = 0$, and $\sum \hat{\phi}_{jk} = 1$. With respect to the $\hat{\phi}_{jk}$ s we see that 24% of the individuals who "don't know" selected the first option (labeled 2 in the output) and 29% of those who "don't know" correctly responded (labeled 3 in the output) on this item. A similar interpretation would be used for the other options. This model's unequal $\hat{\phi}_{jk}$ s indicate that examinees of very low proficiency are not randomly selecting among item 1's options. From the `OBSERVED AND EXPECTED COUNTS/PROPORTIONS` section one sees that approximately 24% of the individuals are expected to be in "category 1." All of the item parameter estimates are presented in Table 3.

Table 2. Abridged output from MC model calibration example.

```
    :
<echo of command file>

 NUMBER OF CODES   4

1234
  VECTOR OF CATEGORIES FOR CODE=1
2222
  VECTOR OF CATEGORIES FOR CODE=2
3333
```

```
  VECTOR OF CATEGORIES FOR CODE=3
4444
  VECTOR OF CATEGORIES FOR CODE=4
5555
   :
  MAXIMUM NUMBER OF EM CYCLES PERMITTED: 500
  NUMBER OF PARAMETER-SEGMENTS USED IS:   4
  NUMBER OF FREE PARAMETERS IS:   44
  MAXIMUM NUMBER OF M-STEP ITERATIONS IS  25 TIMES
   :
 KEY-
 CODE  CATEGORY
  1     2222
  2     3333
  3     4444
  4     5555
   :
 FINISHED CYCLE 253
 MAXIMUM INTERCYCLE PARAMETER CHANGE=   0.00000 P(  44)
   :
 ITEM SUMMARY
   :
 ITEM   1:       5 NOMINAL CATEGORIES,  3 HIGH
  CATEGORY(K): 1      2      3      4      5
    A(K)    -10.81   3.16   3.81   1.10   2.75
    C(K)     -5.95   1.65   1.83   1.31   1.16
    D(K)             0.24   0.29   0.04   0.43

                 CONTRAST-COEFFICIENTS (STANDARD ERRORS)
  FOR:             A                   C                   D
  CONTRAST P(#)  COEFF.[ DEV.]  P(#)  COEFF.[ DEV.]  P(#)  COEFF.[ DEV.]
      1       1  13.97 (8.10)      5   7.60 (5.15)      9   0.18 (0.19)
      2       2  14.62 (8.07)      6   7.78 (5.15)     10  -1.70 (0.49)
      3       3  11.92 (8.21)      7   7.25 (5.09)     11   0.57 (0.18)
      4       4  13.56 (8.04)      8   7.11 (5.20)

 @THETA:      INFORMATION:   (Theta values increase in steps of 0.2)
 -3.0 - -1.6  0.000  0.000  0.000  0.000  0.000  0.000  0.000  0.000
 -1.4 -  0.0  0.000  0.002  0.208  5.973  4.107  0.781  1.089  0.951
  0.2 -  1.6  0.771  0.605  0.468  0.362  0.284  0.227  0.186  0.155
  1.8 -  3.0  0.132  0.114  0.099  0.086  0.076  0.066  0.058

  OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
  CATEGORY(K):  1      2      3      4      5
  OBS. FREQ.      0    456    723    245    375
  OBS. PROP.  0.0000 0.2535 0.4019 0.1362 0.2084
  EXP. PROP.  0.2412 0.2534 0.4019 0.1363 0.2084
   :
 ITEM   4:       5 NOMINAL CATEGORIES,  4 HIGH
  CATEGORY(K): 1      2      3      4      5
    A(K)     -1.83  -4.47   0.25   2.46   3.59
    C(K)      2.38  -4.28   2.36   0.23  -0.69
    D(K)             0.32   0.00   0.44   0.24

                 CONTRAST-COEFFICIENTS (STANDARD ERRORS)
  FOR:             A                   C                   D
  CONTRAST P(#)  COEFF.[ DEV.]  P(#)  COEFF.[ DEV.]  P(#)  COEFF.[ DEV.]
      1      34  -2.64 (1.50)     38  -6.66 (3.47)     42  -6.83 (****)
      2      35   2.08 (0.30)     39  -0.02 (0.15)     43   0.34 (0.17)
      3      36   4.28 (0.48)     40  -2.15 (0.51)     44  -0.27 (0.19)
      4      37   5.41 (0.50)     41  -3.07 (0.51)

 @THETA:      INFORMATION:   (Theta values increase in steps of 0.2)
 -3.0 - -1.6  0.759  0.920  0.961  0.846  0.629  0.418  0.282  0.233
 -1.4 -  0.0  0.253  0.323  0.433  0.570  0.712  0.809  0.786  0.596
  0.2 -  1.6  0.400  0.608  1.261  1.799  1.866  1.574  1.176  0.829
  1.8 -  3.0  0.575  0.403  0.290  0.215  0.164  0.128  0.101

  OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
  CATEGORY(K):  1      2      3      4      5
  OBS. FREQ.      0    288    507    551    453
  OBS. PROP.  0.0000 0.1601 0.2818 0.3063 0.2518
```

```
 EXP. PROP.  0.4490 0.1601 0.2815 0.3064 0.2521
   :
MARGINAL RELIABILITY:    0.7012
   :
NEGATIVE TWICE THE LOGLIKELIHOOD=       259.3
(CHI-SQUARE FOR SEVERAL TIMES MORE EXAMINEES THAN CELLS)
```

The overall model fit is $-2\ln L$ = 259.3 with a BIC = 589.0794 and 44 estimated parameters (i.e., for each item we have 4 $\hat{\alpha}_{jk}$ s, 4 $\hat{\gamma}_{jk}$ s, and (4 – 1) $\hat{\phi}_{jk}$ s for 11 estimated parameters or 44 parameters for the four items). This value of 44 matches the output's `NUMBER OF FREE PARAMETERS IS:` line. Compared to the NR model ($-2\ln L$= 288.4, BIC = 468.2797; see MULTILOG_NRMcalibrationEx.pdf Table 2) the difference chi-square is 288.4 – 259.3 = 29.1 with 44 – 24 = 20 degrees of freedom; alternatively, we have for the NR model *df* = 256 – 24 – 1 = 231 and for the MC model *df* = 256 – 44 –1 = 211. With a critical χ^2 of 31.4 (α = 0.05) the MC model almost provided a significantly better fit than did the NR for these data; we assume that the Full model holds for the data.

The ORFs for item 1 are presented in the left panel of Figure 1; the response code of 1 is represented by the line labeled 2, the response code of 2 is represented by the line labeled 3, and so on. These ORFs appear different from those seen above. For example, the correct response for this item (the "high" category with ORF labeled 3) does not have a strictly monotonically increasing ORF as seen with the NR model. Instead, its ORF contains a "dip" around –0.5. However, the "high" is always associated with the largest $\hat{\alpha}_{jk}$ s for an item. In general, whenever the $\hat{\alpha}_{jk}$ s are positive (i.e., categories 2–5) one sees that the $\hat{\phi}_{jk}$ s reflect the lower asymptotes of the ORFs; see items 1–3 in Figure 2. For instance, item 1's ORF, labeled 2, has a lower asymptote of approximately 0.24, which is the value of its $\hat{\phi}_{12}$. This is also true for this item's remaining ORFs. However, whenever one of the $\hat{\alpha}_{jk}$ s is negative for the observed responses, then the $\hat{\phi}_{jk}$ s do not reflect the lower asymptotes of the ORFs and the corresponding ORF is monotonically decreasing (see the ORFs for item 4; Figure 2).

Table 3. Item Parameter Estimates for Science Items - MC Model.

	Item 1			Item 2			Item 3			Item 4		
k	$\hat{\alpha}_{jk}$	$\hat{\gamma}_{jk}$	$\hat{\phi}_{jk}$	$\hat{\alpha}_{jk}$	$\hat{\gamma}_{jk}$	$\hat{\phi}_{jk}$	$\hat{\alpha}_{jk}$	$\hat{\gamma}_{jk}$	$\hat{\phi}_{jk}$	$\hat{\alpha}_{jk}$	$\hat{\gamma}_{jk}$	$\hat{\phi}_{jk}$
1	-10.81	-5.95		-6.69	-8.06		-12.34	-15.16		-1.83	2.38	
2	3.16	1.65	0.24	1.77	0.65	0.18	2.75	3.12	0.23	-4.47	-4.28	0.32
3	3.81	1.83	0.29	1.26	2.16	0.00	1.62	3.32	0.00	0.25	2.36	0.00
4	1.10	1.31	0.04	2.20	2.77	0.29	3.73	4.28	0.46	2.46	0.23	0.44
5	2.75	1.16	0.43	1.45	2.48	0.53	4.24	4.45	0.30	3.59	-0.69	0.24

A comparison of the item ORF plots in Figure 2 along with the corresponding item parameter estimates (Table 3) shows that the ORF shape is a complex interaction of an item's $\hat{\alpha}_{jk}$, $\hat{\gamma}_{jk}$, and $\hat{\phi}_{jk}$. If one or more of the categories are associated with $\hat{\phi}_{jk}$ s close to zero (e.g., see item 1, category 4 or item 4, category 3), then the corresponding ORF is unimodal. Moreover, it can be seen from the figure's left panel that the ORF associated with the high category (labeled 4) is bimodal, indicating that it is attractive to individuals located around –1.5, as well as to some located around 1.25. This is information that could not be gleaned from a traditional item analysis. As one would expect, the correct response category of 4 (labeled 5 in the figure) is attractive to individuals of high proficiency.

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

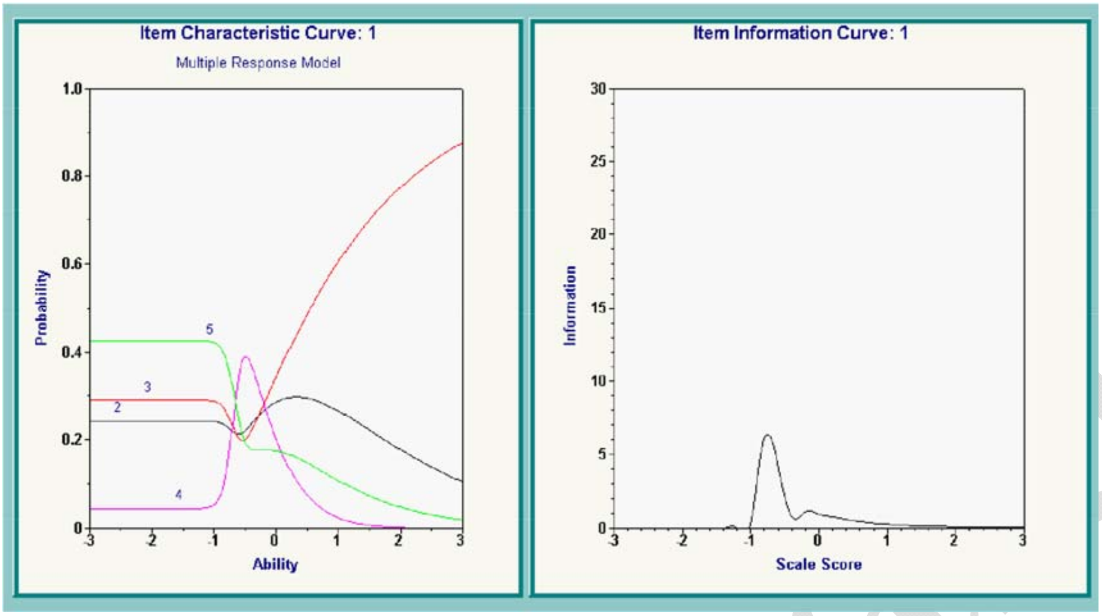
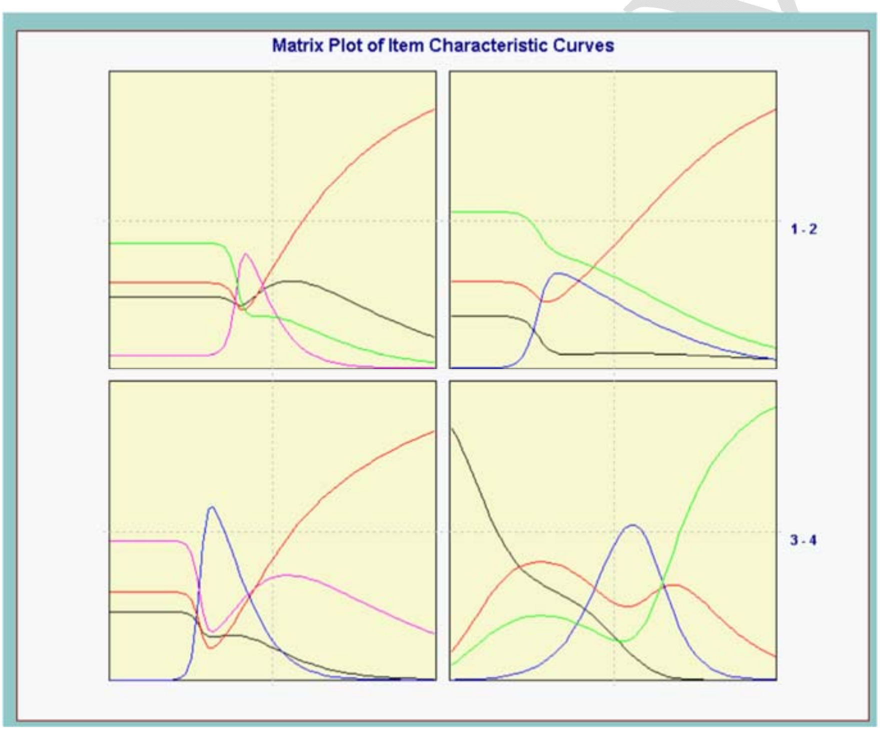
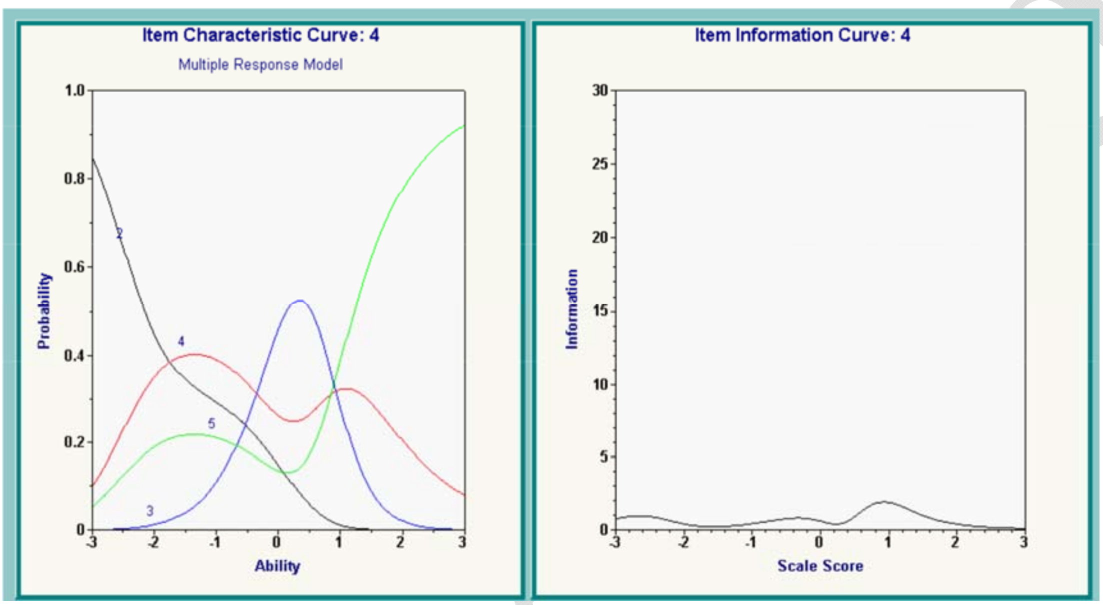


Figure 2. ORFs for all four items, MC model calibration.



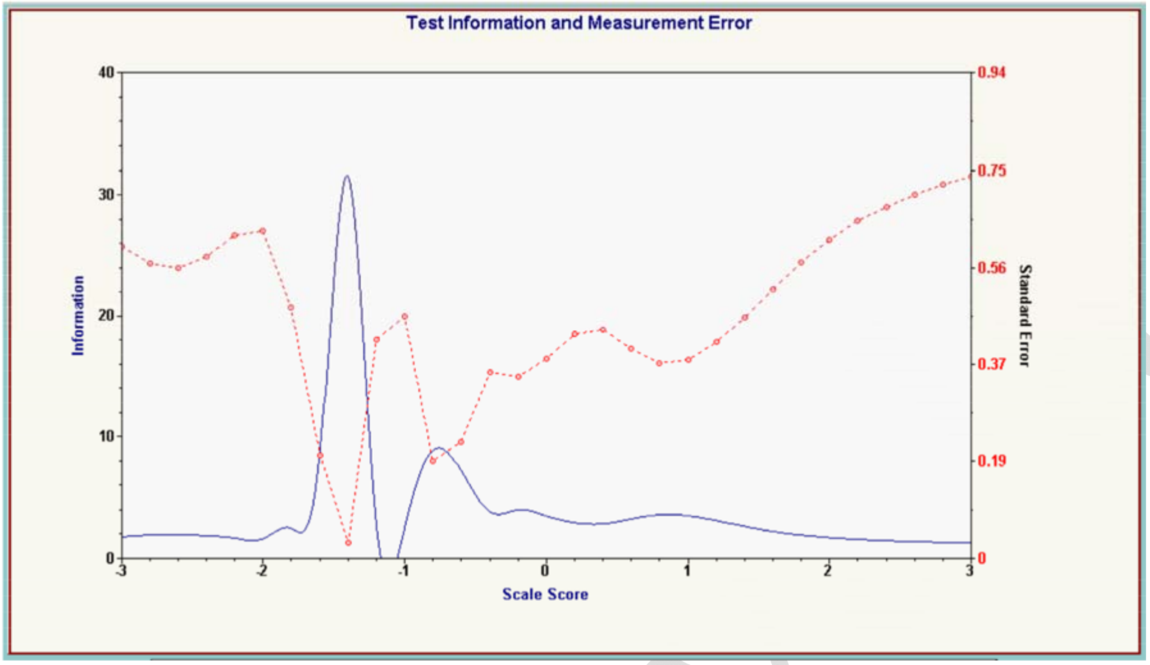
The right panel of Figure 1 shows the item information for item 1. We see that it provides most of its information in the vicinity of -0.7, but that it also provides some information around -0.2. For comparative purposes, item 4's ORFs and its information function are presented in Figure 3. This item provides information for estimating person locations at different points along the continuum, albeit not very much.

Figure 3. ORFs and item information function for item 4.



The total information for this four-item instrument is shown in Figure 4. This graph shows that this instrument provides most of its information around –1.4, although it also has a small second mode in the vicinity of –0.7.

Figure 4. Total information function for MC model calibration.[a]



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