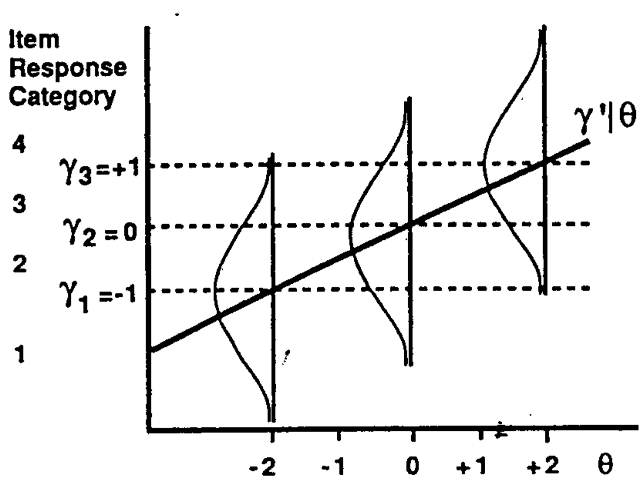
Lecture 10: Polytomous Item Response Models (I)

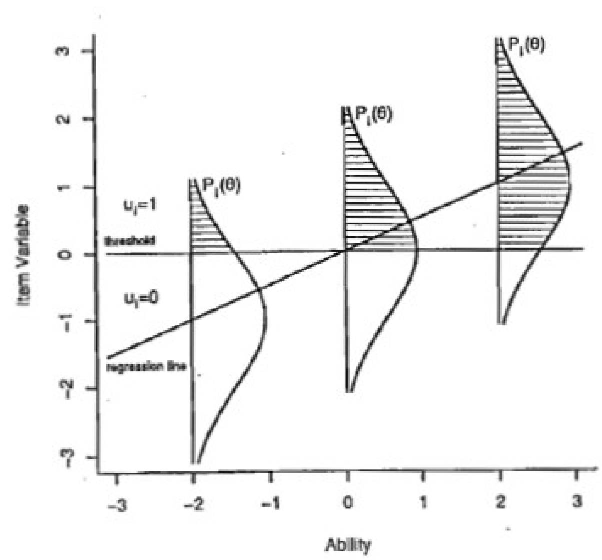
(Baker & Kim (2004), Chapters 8 & 9, Embretson & Reise (2000), Chapter 5)

<u>Graded Response Model</u> (GRM; Samejima, 1969)

The motivation for the model is very similar to the motivation for the normal ogive and 2PL models.

- Y-axis – item variable (hypothetical) γ, ranges from $-\infty$ to $+\infty$, divided into m response categories for a given item, these are ordered because of the underlying continuity.
  $k = 1,2,...m_i$ where m_i = the number of response categories for item *i*.
- *j*th examinee's response can be assigned only to one of the item response categories, and $u_{ijk} = 1$ if the response is assigned to category k, and 0 otherwise.





- The values of the item variable distinguishing the response categories are denoted by γ_k, where $\gamma_0 = -\infty$ and $\gamma_m = +\infty$.

- At each ability level, there exists a conditional normal distribution of responses of examinees. These conditional distributions are intersected by the response category boundaries. The probability of an examinee's response falling into category k is denoted by the area of conditional distributions falling between the limits γ_k and γ_{k-1} and denoted by $P_k(\theta)$.

- At each ability level, $\sum_{k=1}^{m} P_k(\theta) = 1$.

Notice that we have several thresholds instead of just one. Thus, the probability of scoring in a particular score category can be thought of as the difference between the cumulative normal evaluated at two points instead of one. The graded response model is therefore best understood by first defining *boundary characteristic curves* denoted by P_0^*, P_1^*, ..., P_k^*, assuming there are k score categories, where for a given item i

$$P_k^*(\theta) = \frac{\exp[a_i(\theta - b_{ik})]}{1 + \exp[a_i(\theta - b_{ik})]},$$

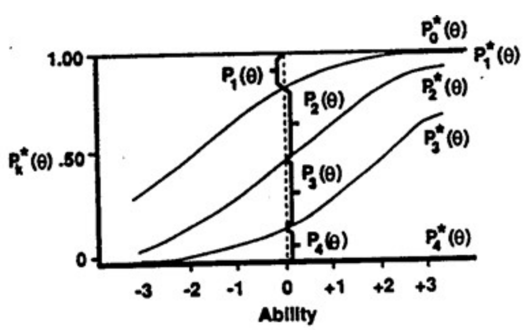
and $P_0^*(\theta) = 1.0$ and $P_k^*(\theta) = 0.0$ for all θ.

Then the graded response model proposes that the probability of scoring in category k is:

$$P_k(\theta) = P_{k-1}^*(\theta) - P_k^*(\theta)$$

<u>Example</u>: Item parameters $a = .5$, $b_1 = -2.0$, $b_2 = 0.0$, $b_3 = 2.0$

θ	P_1^*	P_2^*	P_3^*	θ	P_1	P_2	P_3	P_4
-3.0								
-2.5								
-2.0								
-1.5								
-1.0								
-0.5								
0.0								
0.5								
1.0								
1.5								
2.0								
2.5								
3.0								



Constraints on parameters within the model

- The discrimination parameter, a_i used to compute the P_l^* boundary curves must stay constant across response categories for an item.
- The category thresholds are ordered ($b_1 < b_2 < ... < b_m$).

Properties of the model

- The GRM satisfies the unique maximum condition for $2 \le k \le m-1$.
- In the GRM, the modal points of the score category probabilities are ordered. In fact, the modal point for score category k $(2 < k < m-1)$ occurs at $\theta = \dfrac{b_k + b_{k-1}}{2}$.
- The GRM satisfies a collapsibility (or "additivity") property. This implies that the category response functions belong to the same model even if some categories are combined together.
- The GRM can be generalized to a continuous response model (one that has an infinite number of possible score categories).

Item Information

The information in an item can be thought of as the sum of information provided by each category multiplied by the probability of scoring in that response category. The information for an item i can therefore be computed as

$$I_i(\theta) = \sum_{k=1}^{m} I_k(\theta) P_k(\theta)$$

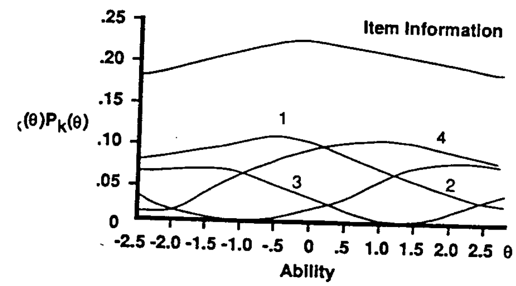
where

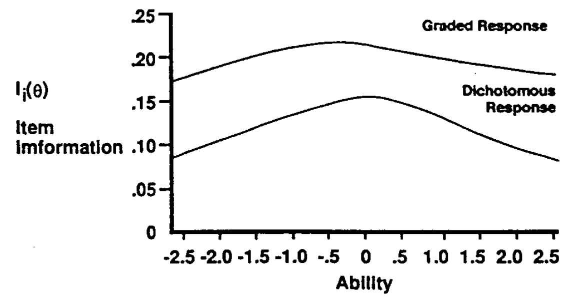
$$I_k(\theta) = -\frac{\partial^2 \log P_k(\theta)}{\partial \theta^2}$$

Applying some algebra, it becomes possible to show that $I_i(\theta)$ has a simple expression in terms of the boundary characteristic curves:

$$I_i(\theta) = \sum_{k=1}^{m} \frac{[P_{k-1}^{*'}(\theta) - P_k^{*'}(\theta)]^2}{P_{k-1}^{*}(\theta) - P_k^{*}(\theta)},$$

where $P_{k-1}^{*'}(\theta)$ and $P_k^{*'}(\theta)$ are boundary curve probabilities.





<u>Estimation of the GRM & MULTILOG (and ltm() & mirt() in R)</u>

Baker & Kim (2004) presents the joint maximum likelihood (JML) algorithm for the GRM model. However, the most familiar software package for estimating the GRM – MULTILOG (Thissen, 1991) and PARSCALE (Muraki & Bock, 1993) – both use **marginal maximum likelihood estimation (MMLE)**. As with the dichotomous models, the primary difference between these estimation methods is that JMLE estimates abilities for examinees at the same time it estimates item parameters, while MMLE integrates θ out of the likelihood equation by attaching to each response pattern a probability at each quadrature point.

MMLE of the GRM has a form of

$$L(a_i, b_{ik}) = \prod_{j=1}^{N} \int_{-\infty}^{\infty} A(\theta_m) P_{v_j}(\theta_m) d\theta_m = \prod_{j=1}^{N} P_{v_j} = \prod_{v} P_v^{r_v},$$

where v_j denotes the response pattern for examinee j, P_v the expected marginal proportion of examinees that have response patterns v, and r_v the observed number of examinees in the sample that have response pattern v.

- In the E-Step: Estimate

$$\tilde{P}_v = \sum P_v(\theta_m) A(\theta_m),$$

where $\tilde{P}_v$ is the marginal proportion of the population expected to have response pattern v based on the provisional item parameter estimates, $P_v(\theta_m)$ is the probability of observing response pattern v at θ_m, and $A(\theta_m)$ represents the quadrature weight attached to θ_m.

Also, estimate

$$\bar{r}_{ikm} = \frac{\sum_v r_v x_{kv} P_v(\theta_m) A(\theta_m)}{\tilde{P}_v}$$

where $\bar{r}_{ikm}$ is the expected number of persons scoring in category k on item i and x_{kv} is equal to 0 or 1 depending on whether response pattern v has score k on item i, and

$$\bar{N}_m = \frac{\sum_v r_v P_v(\theta_m) g(\theta_m)}{\tilde{P}_v},$$

which indicates the expected number of persons at quadrature point θ_m.

- In the M-Step: Based on estimates of $\bar{r}_{ikm}$, $\bar{N}_m$, and $\tilde{P}_v$, find estimates of the category parameters and slope parameters for each item that maximize the log-likelihood function. As before, to get these estimates, a series of Newton-Raphson iterations are performed. In MULTILOG, the maximum number of iterations used to obtain updated parameter estimates is 4, and the convergence criterion is .0001.

Then repeat the E-Step. In MULTILOG, the maximum number of EM cycles is 25 by default, and the convergence criterion is .001.

Goodness of fit

When there are a small numbers of items, MULTILOG performs a goodness-of-fit test like that in BILOG. The difference between the expected and observed numbers of each response pattern is the basis for computing a likelihood-ratio χ^2 statistic.

Regardless of the length of the test, MULTILOG also supplies for each item the expected proportion and observed proportions of responses in each score category. But there is no associated test statistic for these differences.

The function mirt() estimates GRMs and provides fit statistics for each item.

When the number of items is large, PARSCALE (Muraki & Bock) computes a chi-square goodness-of-fit statistic for each item, like in BILOG, which when summed across items provides a goodness-of-fit test for the overall test.

MULTILOG Example: 13-items measure of impulsivity

Please tell us how often you do each of the following things.

Item	When I do something	Never	Sometimes	Often	Always
1	I think about all of my choices very carefully.	1	2	3	4
2	I do the first thing that comes into my mind.	1	2	3	4
3	I compare all the good things and bad things that might happen.	1	2	3	4
4	I go with whatever feels right to me.	1	2	3	4
5	I consider what effect it will have on my health.	1	2	3	4
6	I do only what I have already been planning to do.	1	2	3	4
7	I don't even think about it; I just do it.	1	2	3	4
8	I do whatever I think will be the most fun.	1	2	3	4

9	I do the opposite of what my parents think I should do.	1	2	3	4
10	I like to do things as soon as I think about them.	1	2	3	4
11	I consider if it will be good or bad for my future.	1	2	3	4
12	I act on the spur of the moment.	1	2	3	4
13	I talk to my friends first.	1	2	3	4

```
Fitting the Graded Response Model to IDM data
>PRO RA IN NG=1 NI=13 NE=260;
>TEST ALL GR NC=(4(0)13);
>EST NC=500;
>PRIORS ALL BK=(1(1)3) PA=(0,1);
>END;
4
1234
1111111111111
2222222222222
3333333333333
4444444444444
(13A1)
```

MULTILOG OUTPUT:

```
1MULTILOG---VERSION 6.30    FOR MULTIPLE CATEGORICAL ITEM RESPONSE DATA
0Fitting the Graded Response Model to IDM data
0DATA PARAMETERS:
  NUMBER OF LINES IN THE DATA FILE:  260
  NUMBER OF CATEGORICAL-RESPONSE ITEMS:  13
  NUMBER OF CONTINUOUS-RESPONSE ITEMS, AND/OR GROUPS:   1
  TOTAL NUMBER OF "ITEMS" (INCLUDING GROUPS):  14
  NUMBER OF CHARACTERS IN ID FIELDS:  0
  MAXIMUM NUMBER OF RESPONSE-CODES FOR ANY ITEM:  4
  THE MISSING VALUE CODE FOR CONTINUOUS DATA:  9.0000
  THE DATA WILL BE STORED IN MEMORY
0ESTIMATION PARAMETERS:
  THE ITEMS WILL BE CALIBRATED--
    BY MARGINAL MAXIMUM LIKELIHOOD ESTIMATION
  THERE ARE PRIOR DISTRIBUTIONS FOR (SOME) PARAMETERS
  MAXIMUM NUMBER OF EM CYCLES PERMITTED: 500
  NUMBER OF PARAMETER-SEGMENTS USED IS:  13
  NUMBER OF FREE PARAMETERS IS:   52
  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
  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
0QUADRATURE POINTS FOR MML,
  AT THETA:
   -4.500
   -3.500
   -2.500
   -1.500
   -0.500
    0.500
    1.500
    2.500
    3.500
    4.500
```

```
0NORMAL PRIORS ASSUMED FOR THE FOLLOWING PARAMETERS:
 P(####)      MEAN      S.D.
 P(   2)     0.000     1.000
 P(   3)     0.000     1.000
 P(   4)     0.000     1.000
 P(   6)     0.000     1.000
 P(   7)     0.000     1.000
 P(   8)     0.000     1.000
 P(  10)     0.000     1.000
   .          .         .
 P(  52)     0.000     1.000
1Fitting the Graded Response Model to IDM data
0READING DATA...
0KEY-
0CODE  CATEGORY

  1     1111111111111
  2     2222222222222
  3     3333333333333
  4     4444444444444
0
0FORMAT FOR DATA-

 (13A1)
0FIRST OBSERVATION AS READ-

 ITEMS 1214131442122
 NORML      0.000
0FINISHED CYCLE  29
 MAXIMUM INTERCYCLE PARAMETER CHANGE=   0.00074 P(  29)
1ITEM SUMMARY
0Fitting the Graded Response Model to IDM data

0ITEM    1:      4 GRADED CATEGORIES
        P(#) ESTIMATE  (S.E.)
 A         1    1.44   (0.23)
 B( 1)     2   -0.62   (0.16)
 B( 2)     3    0.59   (0.15)
 B( 3)     4    3.07   (0.61)
0   @THETA: -2.0    -1.5    -1.0    -0.5     0.0     0.5     1.0     1.5     2.0
  I(THETA): 0.22    0.36    0.50    0.59    0.61    0.60    0.53    0.46    0.44
0 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
  CATEGORY(K):  1       2       3      4
  OBS. FREQ.   85      90      82      3
  OBS. PROP. 0.33    0.35    0.32   0.01
  EXP. PROP. 0.34    0.31    0.32   0.03

0ITEM    2:      4 GRADED CATEGORIES
        P(#) ESTIMATE  (S.E.)
 A         5    1.69   (0.25)
 B( 1)     6   -1.36   (0.20)
 B( 2)     7    0.93   (0.16)
 B( 3)     8    2.04   (0.31)
0   @THETA: -2.0    -1.5    -1.0    -0.5     0.0     0.5     1.0     1.5     2.0
  I(THETA): 0.54    0.71    0.69    0.58    0.57    0.69    0.81    0.83    0.79
0 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
  CATEGORY(K):  1       2       3      4
  OBS. FREQ.   37     164      45     14
  OBS. PROP. 0.14    0.63    0.17   0.05
  EXP. PROP. 0.17    0.57    0.18   0.08
```

```
0ITEM    3:       4 GRADED CATEGORIES
        P(#) ESTIMATE (S.E.)
 A         9    1.06  (0.18)
 B( 1)    10   -1.23  (0.28)
 B( 2)    11    0.38  (0.17)
 B( 3)    12    2.64  (0.50)
0   @THETA: -2.0     -1.5     -1.0     -0.5     0.0     0.5     1.0     1.5     2.0
  I(THETA): 0.24     0.29     0.32     0.33     0.33    0.33    0.32    0.31    0.31
0 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
  CATEGORY(K):  1       2       3       4
  OBS. FREQ.   62      93      89      16
  OBS. PROP. 0.24    0.36    0.34    0.06
  EXP. PROP. 0.26    0.33    0.33    0.09

0ITEM    4:       4 GRADED CATEGORIES
        P(#) ESTIMATE (S.E.)
 A        13    0.82  (0.17)
 B( 1)    14   -3.45  (0.70)
 B( 2)    15   -0.36  (0.21)
 B( 3)    16    1.21  (0.33)
0   @THETA: -2.0     -1.5     -1.0     -0.5     0.0     0.5     1.0     1.5     2.0
  I(THETA): 0.18     0.18     0.19     0.20     0.20    0.20    0.19    0.18    0.16
0 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
  CATEGORY(K):  1       2       3       4
  OBS. FREQ.   12      96      77      75
  OBS. PROP. 0.05    0.37    0.30    0.29
  EXP. PROP. 0.07    0.36    0.27    0.30

0ITEM    5:       4 GRADED CATEGORIES
        P(#) ESTIMATE (S.E.)
 A        17    1.38  (0.21)
 B( 1)    18   -0.54  (0.15)
 B( 2)    19    0.53  (0.15)
 B( 3)    20    2.09  (0.35)
0   @THETA: -2.0     -1.5     -1.0     -0.5     0.0     0.5     1.0     1.5     2.0
  I(THETA): 0.20     0.32     0.45     0.54     0.57    0.57    0.55    0.54    0.52
0 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
  CATEGORY(K):  1       2       3       4
  OBS. FREQ.   93      76      72      19
  OBS. PROP. 0.36    0.29    0.28    0.07
  EXP. PROP. 0.37    0.26    0.27    0.10

0ITEM    6:       4 GRADED CATEGORIES
        P(#) ESTIMATE (S.E.)
 A        21    0.76  (0.15)
 B( 1)    22   -2.46  (0.55)
 B( 2)    23   -0.74  (0.26)
 B( 3)    24    2.65  (0.61)
0   @THETA: -2.0     -1.5     -1.0     -0.5     0.0     0.5     1.0     1.5     2.0
  I(THETA): 0.17     0.18     0.17     0.17     0.16    0.16    0.16    0.16    0.16
0 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
  CATEGORY(K):  1       2       3       4
  OBS. FREQ.   34      60     136      30
  OBS. PROP. 0.13    0.23    0.52    0.12
  EXP. PROP. 0.16    0.22    0.48    0.14

0ITEM    7:       4 GRADED CATEGORIES
        P(#) ESTIMATE (S.E.)
 A        25    2.70  (0.37)
 B( 1)    26    0.14  (0.08)
```

```
 B( 2)     27     1.28   (0.14)
 B( 3)     28     2.02   (0.28)
0   @THETA: -2.0     -1.5     -1.0     -0.5     0.0      0.5      1.0      1.5      2.0
  I(THETA): 0.02     0.09     0.31     0.94     1.80     1.81     1.89     2.10     2.01
0 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
  CATEGORY(K):  1       2       3       4
  OBS. FREQ.  143      92      20       5
  OBS. PROP. 0.55    0.35    0.08    0.02
  EXP. PROP. 0.54    0.31    0.10    0.05

0ITEM   8:       4 GRADED CATEGORIES
        P(#) ESTIMATE  (S.E.)
 A        29     2.30   (0.30)
 B( 1)    30    -1.26   (0.15)
 B( 2)    31     0.42   (0.09)
 B( 3)    32     1.23   (0.15)
0   @THETA: -2.0     -1.5     -1.0     -0.5     0.0      0.5      1.0      1.5      2.0
  I(THETA): 0.69     1.24     1.29     1.03     1.21     1.51     1.53     1.25     0.67
0 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
  CATEGORY(K):  1       2       3       4
  OBS. FREQ.   32     136      60      32
  OBS. PROP. 0.12    0.52    0.23    0.12
  EXP. PROP. 0.16    0.47    0.21    0.16

0ITEM   9:       4 GRADED CATEGORIES
        P(#) ESTIMATE  (S.E.)
 A        33     1.33   (0.23)
 B( 1)    34    -0.95   (0.19)
 B( 2)    35     1.33   (0.24)
 B( 3)    36     2.36   (0.43)
0   @THETA: -2.0     -1.5     -1.0     -0.5     0.0      0.5      1.0      1.5      2.0
  I(THETA): 0.28     0.39     0.46     0.46     0.43     0.44     0.49     0.52     0.52
0 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
  CATEGORY(K):  1       2       3       4
  OBS. FREQ.   67     147      32      14
  OBS. PROP. 0.26    0.57    0.12    0.05
  EXP. PROP. 0.28    0.51    0.13    0.08

0ITEM  10:       4 GRADED CATEGORIES
        P(#) ESTIMATE  (S.E.)
 A        37     1.32   (0.23)
 B( 1)    38    -0.89   (0.19)
 B( 2)    39     1.36   (0.24)
 B( 3)    40     2.16   (0.38)
0   @THETA: -2.0     -1.5     -1.0     -0.5     0.0      0.5      1.0      1.5      2.0
  I(THETA): 0.27     0.38     0.45     0.46     0.43     0.44     0.48     0.52     0.51
0 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
  CATEGORY(K):  1       2       3       4
  OBS. FREQ.   71     145      26      18
  OBS. PROP. 0.27    0.56    0.10    0.07
  EXP. PROP. 0.29    0.51    0.11    0.10
0ITEM  11:       4 GRADED CATEGORIES
        P(#) ESTIMATE  (S.E.)
 A        41     1.22   (0.21)
 B( 1)    42    -0.94   (0.21)
 B( 2)    43     0.24   (0.14)
 B( 3)    44     2.20   (0.40)
0   @THETA: -2.0     -1.5     -1.0     -0.5     0.0      0.5      1.0      1.5      2.0
  I(THETA): 0.25     0.34     0.41     0.44     0.45     0.43     0.42     0.41     0.41
0 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
```

```
  CATEGORY(K):  1       2       3       4
  OBS. FREQ.   70      76      94      20
  OBS. PROP. 0.27    0.29    0.36    0.08
  EXP. PROP. 0.29    0.26    0.34    0.10

0ITEM  12:       4 GRADED CATEGORIES
        P(#) ESTIMATE (S.E.)
 A        45     1.69  (0.25)
 B( 1)    46    -0.74  (0.14)
 B( 2)    47     1.20  (0.19)
 B( 3)    48     2.36  (0.40)
0   @THETA: -2.0     -1.5     -1.0     -0.5      0.0      0.5      1.0      1.5      2.0
  I(THETA): 0.27     0.49     0.69     0.73     0.66     0.67     0.77     0.82     0.82
0 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
  CATEGORY(K):  1       2       3       4
  OBS. FREQ.   72     145      35       8
  OBS. PROP. 0.28    0.56    0.13    0.03
  EXP. PROP. 0.30    0.50    0.15    0.05

0ITEM  13:       4 GRADED CATEGORIES
        P(#) ESTIMATE (S.E.)
 A        49     0.67  (0.14)
 B( 1)    50    -3.07  (0.72)
 B( 2)    51    -0.63  (0.27)
 B( 3)    52     2.87  (0.68)
0   @THETA: -2.0     -1.5     -1.0     -0.5      0.0      0.5      1.0      1.5      2.0
  I(THETA): 0.13     0.13     0.13     0.13     0.13     0.13     0.13     0.12     0.12
0 OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
  CATEGORY(K):  1       2       3       4
  OBS. FREQ.   27      76     126      31
  OBS. PROP. 0.10    0.29    0.48    0.12
  EXP. PROP. 0.13    0.28    0.45    0.15

0ITEM  14: GRP1, N[MU:  0.00 SIGMA:  1.00]
    P(#);(S.E.):   53; (0.00)  397; (0.00)
0   @THETA: -2.0     -1.5     -1.0     -0.5      0.0      0.5      1.0      1.5      2.0
  I(THETA): 1.00     1.00     1.00     1.00     1.00     1.00     1.00     1.00     1.00
0
0TOTAL TEST INFORMATION
0   @THETA: -2.0     -1.5     -1.0     -0.5      0.0      0.5      1.0      1.5      2.0
  I(THETA):  4.5      6.1      7.1      7.6      8.6      9.0      9.3      9.2      8.4
 SE(THETA): 0.47     0.40     0.38     0.36     0.34     0.33     0.33     0.33     0.34
0MARGINAL RELIABILITY: 0.87
0NEGATIVE TWICE THE LOGLIKELIHOOD=      4619.8
 (CHI-SQUARE FOR SEVERAL TIMES MORE EXAMINEES THAN CELLS)
0MAXIMUM MEMORY USED        0 BYTES OF 16000000 AVAILABLE
```