Vertical equating

Two hundred students at each of three grade levels, grades four, six, and eight, were given grade-appropriate versions of a 20-item arithmetic examination. Items 19 and 20 appear in the grade 4 and 6 forms; items 37 and 38 appear in the grade 6 and 8 forms. Because each item is assigned a unique column in the data records, a FORM command is not required.

The data file contains, the answer key, not-presented key and raw data. Two lines of information are given per examinee as shown below. The answer key contains 56 entries, each equal to 1. If an item has not been presented, its presence in the data will be indicated with ' '.

```
KEY    11111111111111111111111111111111111111111111111111111111
NOT
001 1 11111112221211222212
002 1 21121211121111121212
003 1 11112112211222212212
004 1 11111112121111111211
005 1 21111112221212121222
```

No items are assigned to the TEST command using the INAMES or INUMBERS keywords. By default, it is assumed that all items are assigned to the test. Although the test name (TNAME=MATH) is not enclosed in single quotes, the group names are as these names contain blanks as part of the name.

The distributions of ability are assumed to be normal at each grade level (NORMAL on the CALIB command). Grade 6 serves as the reference group in the calibration of the items (REFERENCE=2). EAP estimates of ability are calculated using the information in the posterior distributions from Phase 2. The ability estimates are rescaled to a mean of 0 and standard deviation of 1 by specifying RSCTYPE=3 on the SCORE command.

```
EXAMPL05.BLM - VERTICAL EQUATING OF TEST FORMS OVER THREE GRADE LEVELS

>GLOBAL  DFNAME='EXAMPL05.DAT', NPARM=2, SAVE;
>SAVE    SCORE='EXAMPL05.SCO', PARM='EXAMPL05.PAR';
>LENGTH  NITEMS=56;
>INPUT   NTOT=56, NGROUPS=3, NIDCH=3,
         KFNAME='EXAMPL05.DAT', NFNAME='EXAMPL05.DAT';
>ITEMS   INUM=(1(1)56), INAME=(M01(1)M56);
>TEST    TNAME=MATH;
>GROUP1  GNAME='GRADE 4', LENGTH=20, INUM=(1(1)20);
>GROUP2  GNAME='GRADE 6', LENGTH=20, INUM=(19(1)38);
>GROUP3  GNAME='GRADE 8', LENGTH=20, INUM=(37(1)56);
(3A1,1X,I1,1X,56A1)
>CALIB   NQPT=51, NORMAL, CYCLE=30, TPRIOR, REFERENCE=2;
>SCORE   METHOD=2, IDIST=3, NOPRINT, RSCTYPE=3;
```

Phase 1 output

In this example, items assigned to the three groups of examinees are selected from the following set. The items are selected in such a way that two items are common to groups 1 and 2 and two other items are common to groups 2 and 3. The groups, corresponding to school grades four, six, and eight are non-equivalent and require separate classical item statistics. The fact that classical item statistics are not invariant with respect to sampling from different populations is illustrated by the different results for common items in different groups.

```
        MULTIPLE GROUP SPECIFICATIONS
        =============================

        MULTIPLE GROUPS ARE DEFINED,
        BUT NEITHER DIF MODEL NOR PARAMETER DRIFT MODEL IS EMPLOYED.

        GROUP NUMBER:  1    GROUP NAME: GRADE 4
        TEST NUMBER:   1    TEST NAME: MATH
        NUMBER OF ITEMS:   20

              ITEM    ITEM
            NUMBER   NAME
            -----------------
                1   M01
                2   M02

               20   M20
            -----------------

        GROUP NUMBER:  2    GROUP NAME: GRADE 6
        TEST NUMBER:   1    TEST NAME: MATH
        NUMBER OF ITEMS:   20

              ITEM    ITEM
            NUMBER   NAME
            -----------------
               19   M19
               20   M20

               38   M38
            -----------------

        GROUP NUMBER:  3    GROUP NAME: GRADE 8
        TEST NUMBER:   1    TEST NAME: MATH
        NUMBER OF ITEMS:   20

              ITEM    ITEM
            NUMBER   NAME
            -----------------
               37   M37

               56   M56
            -----------------

        600 OBSERVATIONS READ FROM FILE:   EXAMPL05.DAT
        600 OBSERVATIONS WRITTEN TO FILE:  MF.DAT

            SUBTEST   1    MATH
              GROUP   1    GRADE 4    200 OBSERVATIONS
              GROUP   2    GRADE 6    200 OBSERVATIONS
              GROUP   3    GRADE 8    200 OBSERVATIONS
```

Item statistics for the first 5 items of each subtest are shown below. Similar output is produced for grades 6 to 8, and for multiple groups MATH which, in this case, contains the statistics for all the grades.

```
      SUBTEST    1    MATH
       ITEM STATISTICS FOR GROUP:  1     GRADE 4
                                                     ITEM*TEST CORRELATION
       ITEM   NAME    #TRIED    #RIGHT     PCT    LOGIT/1.7  PEARSON  BISERIAL
       ------------------------------------------------------------------------
          1   M01     200.0     138.0    0.690     -0.47     0.470    0.616
      ...
         19   M19     200.0      95.0    0.475      0.06     0.520    0.652
         20   M20     200.0      67.0    0.335      0.40     0.475    0.615
       ------------------------------------------------------------------------


       ITEM STATISTICS FOR GROUP:  2     GRADE 6
                                                     ITEM*TEST CORRELATION
       ITEM   NAME    #TRIED    #RIGHT     PCT    LOGIT/1.7  PEARSON  BISERIAL
       ------------------------------------------------------------------------
         19   M19     200.0     138.0    0.690     -0.47     0.431    0.565
         20   M20     200.0     106.0    0.530     -0.07     0.512    0.643
      ...
         37   M37     200.0     104.0    0.520     -0.05     0.379    0.475
         38   M38     200.0      62.0    0.310      0.47     0.497    0.651
       ------------------------------------------------------------------------


       ITEM STATISTICS FOR GROUP:  3     GRADE 8
                                                     ITEM*TEST CORRELATION
       ITEM   NAME    #TRIED    #RIGHT     PCT    LOGIT/1.7  PEARSON  BISERIAL
       ------------------------------------------------------------------------
         37   M37     200.0     135.0    0.675     -0.43     0.420    0.546
         38   M38     200.0      96.0    0.480      0.05     0.594    0.745
      ...
         55   M55     200.0      90.0    0.450      0.12     0.471    0.592
         56   M56     200.0     111.0    0.555     -0.13     0.529    0.665
       ------------------------------------------------------------------------
```

Phase 2 output

In vertical equating over a range of age levels, the ability distributions of the groups may be widely spaced. For that reason, it is desirable to use a large number of quadrature points – in this case, 51.

The origins and unit of the ability distribution can be fixed in the calibration either by setting the mean and standard deviation of a reference group to zero and one, respectively, or, similarly, setting the mean and standard deviation of the combined groups. In this example, group 2 is selected as the reference group.

```
      >CALIB   NQPT=51, NORMAL, CYCLE=30, TPRIOR, REFERENCE=2;

       CALIBRATION PARAMETERS
       ======================
       MAXIMUM NUMBER OF EM CYCLES:              30
       MAXIMUM NUMBER OF NEWTON CYCLES:           2
       CONVERGENCE CRITERION:                   0.0100
       ACCELERATION CONSTANT:                   1.0000
       LATENT DISTRIBUTION:                     NORMAL PRIOR FOR EACH GROUP
                                                GROUP MEANS AND SDS
                                                ESTIMATED CONCURRENTLY
                                                WITH ITEM PARAMETERS
                                                REFERENCE GROUP:    2
       PLOT EMPIRICAL VS. FITTED ICC'S:         NO
       DATA HANDLING:                           DATA ON SCRATCH FILE
      …
           MEANS AND STANDARD DEVIATIONS:       PROGRAM DEFAULTS
                                                ITEM CONSTRAINTS IF PRESENT
                                                  WILL BE UPDATED EACH CYCLE
```

The iterative estimation procedures typically converge more slowly in nonequivalent group data than in one group or equivalent groups data. The last few iterations are shown here along with some of the resulting parameter estimates. The means of the prior distributions on item thresholds and slopes are also listed.

```
        CYCLE    19;   LARGEST CHANGE=   0.02538
        -2 LOG LIKELIHOOD =      13246.111
        UPDATED PRIOR ON LOG SLOPES; MEAN & SD =      -0.23806     0.50000
        UPDATED PRIOR ON THRESHOLDS; MEAN & SD =       0.08303     2.00000


        CYCLE    20;   LARGEST CHANGE=   0.00812

        [NEWTON CYCLES]
        UPDATED PRIOR ON LOG SLOPES; MEAN & SD =      -0.23533     0.50000
        UPDATED PRIOR ON THRESHOLDS; MEAN & SD =       0.08308     2.00000
        -2 LOG LIKELIHOOD:       13245.9542


        CYCLE    21;   LARGEST CHANGE=   0.00699


        INTERVAL COUNTS FOR COMPUTATION OF ITEM CHI-SQUARES
        --------------------------------------------------------------------
           19.    32.    56.    83.    93.   109.    82.    60.    66.
        --------------------------------------------------------------------


        INTERVAL AVERAGE THETAS
        --------------------------------------------------------------------
          -2.695 -1.942 -1.448 -0.866 -0.356  0.145  0.607  1.193  1.989
        --------------------------------------------------------------------


        SUBTEST MATH    ;   ITEM PARAMETERS AFTER CYCLE   21
        ITEM  INTERCEPT    SLOPE      THRESHOLD   LOADING   ASYMPTOTE    CHISQ  DF
                   S.E.         S.E.        S.E.        S.E.         S.E.      (PROB)
        --------------------------------------------------------------------
        M01 |   1.218  |   0.805  |  -1.512  |   0.627  |   0.000  |    3.2  5.0
            |   0.194* |   0.155* |   0.163* |   0.121* |   0.000* | (0.6741)
            |          |          |          |          |          |
        M02 |   1.149  |   0.707  |  -1.626  |   0.577  |   0.000  |    4.4  6.0
            |   0.169* |   0.129* |   0.186* |   0.105* |   0.000* | (0.6249)
            |          |          |          |          |          |
        [Similar output omitted]
        M55 |  -0.584  |   0.707  |   0.826  |   0.577  |   0.000  |    3.9  6.0
            |   0.129* |   0.126* |   0.143* |   0.103* |   0.000* | (0.6847)
            |          |          |          |          |          |
        M56 |  -0.319  |   0.849  |   0.376  |   0.647  |   0.000  |    1.1  5.0
            |   0.127* |   0.144* |   0.125* |   0.110* |   0.000* | (0.9547)
        --------------------------------------------------------------------
                                                              * STANDARD ERROR


        LARGEST CHANGE =    0.007897                     188.0 296.0
                                                          (1.0000)
        PARAMETER       MEAN  STN DEV
        ---------------------------------
        SLOPE          0.802    0.138
        LOG(SLOPE)    -0.235    0.172
        THRESHOLD      0.083    0.775
```

The within-group latent distributions are assumed normal. Their means and standard deviations are estimated relative to the reference group. In these data, the means increase over the grades (-0.722, 0.000, 0.569), but the standard deviations are relatively constant (1.069, 1.00, 1.126).

```
        GROUP:  1    GRADE 4   QUADRATURE POINTS, POSTERIOR WEIGHTS, MEAN AND S.D.:
                          1            2            3            4            5
        POINT       -0.4275E+01 -0.4105E+01 -0.3935E+01 -0.3765E+01 -0.3594E+01
        POSTERIOR    0.4299E-03  0.7062E-03  0.1119E-02  0.1717E-01  0.2558E-02
```

```
[Similar output omitted]

                     47          48          49          50          51
      POINT        0.3552E+01  0.3722E+01  0.3892E+01  0.4062E+01  0.4232E+01
      POSTERIOR    0.1899E-04  0.9879E-05  0.3535E-05  0.1816E-05  0.9055E-06


      MEAN        -0.72298
      S.E.         0.11260


      S.D.         1.06880
      S.E.         0.12631




      GROUP:  2    GRADE 6   QUADRATURE POINTS, POSTERIOR WEIGHTS, MEAN AND S.D.:
                      1           2           3           4           5
      POINT       -0.4275E+01 -0.4105E+01 -0.3935E+01 -0.3765E+01 -0.3594E+01
      POSTERIOR    0.1136E-04  0.2278E-04  0.4596E-04  0.8712E-04  0.1599E-03

      [Similar output omitted]


                     47          48          49          50          51
      POINT        0.3552E+01  0.3722E+01  0.3892E+01  0.4062E+01  0.4232E+01
      POSTERIOR    0.1172E-03  0.6346E-04  0.3291E-04  0.1689E-04  0.8409E-05


      MEAN         0.00000
      S.E.         0.00000


      S.D.         1.00000
      S.E.         0.00000


      GROUP:  3    GRADE 8   QUADRATURE POINTS, POSTERIOR WEIGHTS, MEAN AND S.D.:
                      1           2           3           4           5
      POINT       -0.4275E+01 -0.4105E+01 -0.3935E+01 -0.3765E+01 -0.3594E+01
      POSTERIOR    0.4219E-05  0.7809E-05  0.1793E-04  0.3292E-04  0.5918E-04

      [Similar output omitted]


                     47          48          49          50          51
      POINT        0.3552E+01  0.3722E+01  0.3892E+01  0.4062E+01  0.4232E+01
      POSTERIOR    0.1837E-02  0.1230E-02  0.8192E-03  0.5316E-03  0.3268E-03


      MEAN         0.56861
      S.E.         0.11855


      S.D.         1.12577
      S.E.         0.14026
```

Phase 3 output

With nonequivalent groups, Bayes (EAP) and Bayes Modal (MAP) estimation of test scores should be carried out with respect to the Phase 2 latent distribution to which the examinee belongs. Specify IDIST=3 on the SCORE command.

```
      >SCORE    METHOD=2,IDIST=3,NOPRINT, RSCTYPE=3;


      PARAMETERS FOR SCORING, RESCALING, AND TEST AND ITEM INFORMATION
      METHOD OF SCORING SUBJECTS:              EXPECTATION A POSTERIORI
                                               (EAP; BAYES ESTIMATION)
      TYPE OF PRIOR:                           EMPIRICAL, FROM ITEM CALIBRATION
      TYPE OF RESCALING:                       IN THE SAMPLE DISTRIBUTION
      REFERENCE GROUP FOR RESCALING:           GROUP:  2
```

```
                      QUAD
TEST     NAME    GROUP POINTS
--------------------------
    1  MATH        1    51
    1  MATH        2    51
    1  MATH        3    51
--------------------------


                  RESCALING CONSTANTS
TEST     NAME         SCALE   LOCATION
------------------------------------
    1  MATH           1.000    0.000
------------------------------------
```

In this example, the scores are rescaled so that their mean and standard deviation in the total sample are zero and one, respectively. The parameter estimates are rescaled accordingly.

```
       RESCALING WITH RESPECT TO SAMPLE DISTRIBUTION
       ---------------------------------------------------

              RESCALING  CONSTANTS
       TEST        SCALE   LOCATION
       MATH        1.066      0.003

GROUP   SUBJECT IDENTIFICATION                                 MARGINAL
WEIGHT   TEST       TRIED  RIGHT  PERCENT     ABILITY    S.E.     PROB
-----------------------------------------------------------------------
1        1                              |                    |
1.00    MATH         20     11    55.00 |  -0.3055   0.3598 |  0.000000
1        1                              |                    |
1.00    MATH         20     13    65.00 |  -0.0653   0.3620 |  0.000000
-----------------------------------------------------------------------


       TEST    MATH   ;  RESCALED ITEM PARAMETERS

       ITEM        INTERCEPT   SLOPE   THRESHOLD  LOADING    ASYMPTOTE
                     S.E.       S.E.      S.E.      S.E.        S.E.
       -------------------------------------------------------------------
       M01      |   1.216 |   0.755 |  -1.610 |   0.627 |   0.000
                |   0.194* |  0.145* |   0.173* |  0.121* |   0.000*
                |        |        |        |        |
       M02      |   1.148 |   0.663 |  -1.732 |   0.577 |   0.000
                |   0.169* |  0.121* |   0.198* |  0.105* |   0.000*
                |        |        |        |        |
       [Similar output omitted]
                |        |        |        |        |
       M55      |  -0.566 |   0.670 |   0.845 |   0.581 |   0.000
                |   0.127* |  0.120* |   0.151* |  0.104* |   0.000*
                |        |        |        |        |
       M56      |  -0.298 |   0.805 |   0.370 |   0.651 |   0.000
                |   0.125* |  0.136* |   0.132* |  0.110* |   0.000*
       -------------------------------------------------------------

       PARAMETER       MEAN  STN DEV
       ----------------------------
       SLOPE           0.752    0.130
       LOG(SLOPE)     -0.299    0.172
       THRESHOLD       0.092    0.806
```

```
MEAN & SD OF SCORE ESTIMATES AFTER RESCALING


GROUP       MEAN        SD
----------------------------
   1       -0.776     1.067
   2        0.000     1.000
   3        0.608     1.118
----------------------------


MEAN & SD OF LATENT DISTRIBUTIONS AFTER RESCALING


GROUP       MEAN        SD
----------------------------
   1       -0.776     1.149
   2        0.000     1.074
   3        0.608     1.201
----------------------------
```