Domain scores

This is an attempt to reconstruct the domain scores demonstration application reported in "*The Domain Score Concept and IRT: Implications for Standards Setting*" by Bock, Thissen & Zimowski (2001). We use the dataset **spell.dat**. All 100 items of the 100-word spelling test seem to be there, but there are only 660 records (instead of the 1,000 that Bock *et. al.* report). In a first run (**spell1.blm**), we calibrate all 100 items and save the parameters in an external file. The syntax is shown below.

```
SPELL1.BLM - CALIBRATION OF THE 100 WORD SPELLING TEST
            TWO-PARAMETER MODEL
>COMMENTS
We are trying first to reproduce the table with slope and location parameters for the 100
words as Bock et al. report in "The Domain Score Concept and IRT:
            Implications for Standards Setting."

The SCORE command is included to obtain the percent correct for each examinee (= the true
domain scores).

>GLOBAL DFNAME='SPELL.DAT', NPARM=2, SAVE;
>SAVE   PARM='SPELL1.PAR';
>LENGTH NITEMS=(100);
>INPUT  NTOTAL=100, NIDCHAR=10, KFNAME='SPELL.DAT';
>ITEMS  INAMES=(S001(1)S100);
>TEST1  TNAME='SPELLING', INUMBERS=(1(1)100);
(10A1,1X,25A1,1X,25A1,1X,25A1,1X,25A1)
>CALIB  NQPT=31, CYCLES=100, CRIT=0.001, NOFLOAT;
>SCORE;
```

The item parameters of the first 5 items, as reported in the item parameter file **step1blm.par,** are shown in Table 1.

Table 1: Selected item parameters from step1blm.par

Item	Slope	S.E.	Threshold	S.E.
S001	0.79494	0.07978	-0.34466	0.06899
S002	0.38723	0.07299	-3.53823	0.61667
S003	0.24041	0.04784	-3.04033	0.61821
S004	0.72020	0.07353	-0.54159	0.08115
S005	0.69253	0.07367	-1.41137	0.13523

The parameter values are in close agreement with Table 1 from Bock. *et al*. (results for the first 5 items shown in Table 2 below), showing also that we have a correct dataset, with the items in the right order (of the table), albeit not all records.

Table 2: Selected item parameters from Bock *et. al*.

Item	Slope	Threshold
S001	0.843	-0.339
S002	0.351	-3.623
S003	0.239	-3.073
S004	0.785	0.727
S005	0.269	2.273

In a second run (**spell2.blm**), we let the program compute the expected domain scores for all 660 examinees from the saved parameter file. The DOMAIN and FILE keywords on the SCORE command are used. We skip the calibration phase with the SELECT keyword on the CALIB command. The scores are saved to file by using the SCORE keyword on the SAVE command.

The contents of **spell2.blm** are shown below. All the command files and data discussed here are available to the user in the **domscore** subfolder of the BILOG-MG installation folder.

```
SPELL2.BLM - CALIBRATION OF THE 100 WORD SPELLING TEST
           TWO-PARAMETER MODEL
>COMMENTS
In a second step, we test the "DOMAIN" keyword on the score command. The item parameter file
from the SPELL1.BLM analysis has been edited and saved as SPELL2.PAR in accordance with the
FILE keyword format requirements. We save the score file.

>GLOBAL DFNAME='SPELL.DAT', NPARM=2, SAVE;
>SAVE   SCORE='SPELL2.SCO';
>LENGTH NITEMS=(100);
>INPUT  NTOTAL=100, NIDCHAR=10, KFNAME='SPELL.DAT';
>ITEMS  INAMES=(S001(1)S100);
>TEST1  TNAME='SPELLING', INUMBERS=(1(1)100);
(10A1,1X,25A1,1X,25A1,1X,25A1,1X,25A1)
>CALIB  SELECT=(0);
>SCORE  DOMAIN=100, FILE='SPELL2.PAR', METHOD=2;
```

The parameter file that we read in through the FILE keyword on the SCORE command had to be created from the saved parameter file (**spell1blm.par**) in the **spell1.blm** run. First we deleted everything before the first line with parameter estimates. Then we deleted all the columns that were not slope, threshold, or guessing parameters, leaving just those three columns and in that order. Then, we added a column with weights as the first column, in the same format. We used 1.0000, because we want all items weighed equally. We then added the variable format statement (4F10.5) as the first line in the file and renamed it to **spell1.par**.

The estimated domain scores from **spell2.blm** are fairly well recovered as **spell2.ph3** shows. Here are the results for the first five examinees:

```
GROUP    SUBJECT IDENTIFICATION           DOMAIN SCORE     S.E.    MARGINAL
WEIGHT    TEST      TRIED  RIGHT  PERCENT    ABILITY        S.E.     PROB
------------------------------------------------------------------------
 1  01021119001                            64.89          4.92    |
1.00 * SPELLING    100     65    65.00 |   -0.1501        0.4187 | 0.000000
 1  01041122001                            57.14          5.43    |
1.00 * SPELLING    100     56    56.00 |   -0.7839        0.4321 | 0.000000
 1  01051219001                            54.25          5.40    |
1.00 * SPELLING    100     57    57.00 |   -1.0132        0.4269 | 0.000000
 1  01061219001                            71.52          1.80    |
1.00 * SPELLING    100     69    69.00 |    0.4499        0.1768 | 0.000000
 1  01071219001                            80.77          2.68    |
1.00 * SPELLING    100     81    81.00 |    1.5475        0.4000 | 0.000000
------------------------------------------------------------------------
```

If the estimated expected domain scores are not close, something is probably wrong, so this is a good test.

In a third and final step (**step3.blm**), we take a random sample of 20 items, adapt the parameter file (**spell3.par** as described previously) and produce a new score file (**spell3.sco**).

The contents of **spell3.blm** are as follows:

```
SPELL3.BLM - CALIBRATION OF THE 100 WORD SPELLING TEST
            TWO-PARAMETER MODEL
>COMMENTS
In this third step we use a random sample of 20 items from the 100-word spelling test to
score the examinees with the item parameters from the first step. The score file is saved.

>GLOBAL DFNAME='SPELL.DAT', NPARM=2, SAVE;
>SAVE    SCORE='SPELL3.SCO';
>LENGTH NITEMS=20;
>INPUT  NTOTAL=100, NIDCHAR=10, KFNAME='SPELL.DAT';
>ITEMS  INAMES=(S001(1)S100);
>TEST1  TNAME='SPELLING', INUMBERS=(4, 9, 10, 13, 22, 26, 36, 51, 55, 65,
        69, 73, 74, 82, 83, 88, 89, 91, 94, 97);
(10A1,1X,25A1,1X,25A1,1X,25A1,1X,25A1)
>CALIB  SELECT=(0);
>SCORE DOMAIN=20, FILE='SPELL3.PAR', METHOD=2;
```

These are the results for the first five examinees:

```
GROUP    SUBJECT IDENTIFICATION           DOMAIN SCORE     S.E.    MARGINAL
WEIGHT    TEST      TRIED  RIGHT  PERCENT    ABILITY        S.E.     PROB
------------------------------------------------------------------------
 1  01021119001                            72.51          6.95    |
1.00 * SPELLING     20     14    70.00 |    0.4109        0.7214 | 0.000000
 1  01041122001                            62.76          7.94    |
1.00 * SPELLING     20     12    60.00 |   -0.4985        0.6850 | 0.000000
 1  01051219001                            63.42          7.90    |
1.00 * SPELLING     20     12    60.00 |   -0.4414        0.6870 | 0.000000
 1  01061219001                            74.25          6.61    |
1.00 * SPELLING     20     14    70.00 |    0.5971        0.7300 | 0.000000
 1  01071219001                            79.14          5.32    |
1.00 * SPELLING     20     17    85.00 |    1.2047        0.7604 | 0.000000
------------------------------------------------------------------------
```

As can be seen, a decent recovery of the "population domain scores" with the random sample of only 20 items.