Modelling Polytomous Items with the Generalised Partial Credit and Bock Nominal Response Models

Louise Ockwell. 12 August 2015

As discussed in 'Note 6: Score Estimation and Generalised Partial Credit Models', ConQuest can estimate scoring parameters for a wide range of models with polytomous data where item responses are categorical values, including multidimensional forms of the two-parameter family of models such as the multidimensional generalised partial credit models (Muraki, 1992). In addition, ConQuest can also estimate scoring parameters for models with polytomous data where item responses are in the form of nominal categories, such as Bock's nominal response model (Bock, 1972). In this tutorial, the use of ConQuest to fit the generalised partial credit and Bock nominal response models is illustrated through two sets of sample analyses. Both analyses use the same cognitive items: in the first the generalised partial credit model is fitted to the data; and in the second, the Bock nominal response model is fitted.

FITTING THE GENERALISED PARTIAL CREDIT MODEL

The data for the first sample analysis are the responses of 515 students to a test of science concepts related to the Earth and space previously used in Tutorial 2.

The files used in this sample analysis are:

`Ex11a.cqc`	The command statements.
`ex2a.dat`	The data.
`ex2a.lab`	The variable labels for the items on the partial credit test.
`Ex11a.shw`	The results of the generalised partial credit analysis.
`Ex11a.itn`	The results of the traditional item analyses.

(The last two files are created when the command file is executed.)

The data have been entered into the file `ex2a.dat`, using one line per student. A unique identification code has been entered in columns 2 through 7, and the students' response to each of the items has been recorded in columns 10 through 17. In this data, the upper-case alphabetic characters A, B, C, D, E, F, W, and X have been used to indicate the different kinds of responses that students gave to these items. The code Z has been used to indicate data that cannot be analysed. For each item, these codes are scored (or, more correctly, mapped onto performance levels) to indicate the level of quality of the response. For example, in the case of the first item (the item in column 10), the response coded A is regarded as the best kind of response and is assigned to level 2, responses B and C are assigned to level 1, and responses W and X are assigned to level 0. An extract of the file `ex2a.dat` is shown in Figure 1.

```
          1         2
12345678901234567890123 (column numbers)
2110104ZHWBDCBBCBEABBBB
2110106ZEACDBXBCXXXXXXX
2110109ZHBWBBBWCAXAXXXX
2110113ZIWBWBXWCXXXABBB
2110115ZHWBFBBWCWXAWAXX
2110121ZHWWEBWBBCAABABA
2110123YIBWWBEWBWXABABB
2110305ZHCBABABAABACCCA
2110313YBBCFBDBCXXXXXXX
 .               .
 .               .
```

Figure 1 Extract from the Data File `ex2a.dat`

The contents of the command file for this sample analysis are shown in Figure 2.

1. Gives a title for this analysis. The text supplied after the command `title` will appear on the top of any printed ConQuest output. If a title is not provided, the default, `ConQuest: Generalised Item Response Modelling Software`, will be used.

2. Indicates the name and location of the data file. Any name that is valid for the operating system you are using can be used here.

3. The `format` statement describes the layout of the data in the file `ex2a.dat`. This format indicates that a field called `name` is located in columns 2 through 7 and that the responses to the items are in columns 10 through 17 (the response block) of the data file.

```
1.  Title Generalised Partial Credit Model: What happened last
    night;
2.  data ex2a.dat;
3.  format name 2-7 responses 10-17;
4.  labels << ex2a.lab;
5.  codes 3,2,1,0;
6.  set constraints=cases;
7.  recode (A,B,C,W,X) (2,1,1,0,0)             !item(1);
8.  recode (A,B,C,W,X) (3,2,1,0,0)             !item(2);
9.  recode (A,B,C,D,E,F,W,X) (3,2,2,1,1,0,0,0)!item(3);
10. recode (A,B,C,W,X) (2,1,0,0,0)             !item(4);
11. recode (A,B,C,D,E,W,X) (3,2,1,1,1,0,0)     !item(5);
12. recode (A,B,W,X) (2,1,0,0)                 !item(6);
13. recode (A,B,C,W,X) (3,2,1,0,0)             !item(7);
14. recode (A,B,C,D,W,X) (3,2,1,1,0,0)         !item(8);
15. model item + item*step!scoresfree;
16. estimate;
17. show !estimates=latent >> ex11a.shw;
18. itanal >> ex11a.itn;
19. plot expected;
20. plot mcc;
```

Figure 2 Sample Command File for fitting the Generalised Partial Credit Model

4. A set of labels for the items are to be read from the file `ex2a.lab`. If you take a look at these labels, you will notice that they are quite long. ConQuest labels can be of any length, but most ConQuest printouts are limited to displaying many fewer characters

than this. For example, the tables of parameter estimates produced by the `show` statement will display only the first 11 characters of the labels.

5. The `codes` statement is used to restrict the list of codes that ConQuest will consider valid. This meant that any character in the response block defined by the `format` statement—except a blank or a period (.) character (the default missing-response codes)—was considered valid data. In this sample analysis, the valid codes have been limited to the digits 0, 1, 2 and 3; any other codes for the items will be treated as missing-response data. It is important to note that the `codes` statement refers to the codes *after* the application of any recodes.

6. The `constraints=cases` argument of the `set` command is used to have the mean of each latent dimension set to zero, rather than the mean of the item parameters on each dimension set to zero (e.g., `constraints=items`). All item parameters are still estimated, but the mean of each of the latent dimensions is set to zero.

7.-14. The eight `recode` statements are used to collapse the alphabetic response categories into a smaller set of categories that are labelled with the digits 0, 1, 2 and 3. Each of these `recode` statements consists of three components. The first component is a list of codes contained within parentheses. These are codes that will be found in the data file `ex2.dat`, and these are called the *from* codes. The second component is also a list of codes contained within parentheses, these codes are called the *to* codes. The length of the *to* codes list must match the length of the *from* codes list. When ConQuest finds a response that matches a *from* code, it will change (or recode) it to the corresponding *to* code. The third component (the option of the `recode` command) gives the levels of the variables for which the recode is to be applied. Line 11, for example, says that, for item 6, A is to be recoded to 2, B is to be recoded to 1, and W and X are both to be recoded to 0.

Any codes in the response block of the data file that do not match a code in the *from* list will be left untouched. In these data, the Z codes are left untouched; and since Z is not listed as a valid code, all such data will be treated as missing-response data.

When ConQuest models these data, the number of response categories that will be assumed for each item will be determined from the number of distinct codes for that item. Item 1 has three distinct codes (2, 1 and 0), so three categories will be modelled; item 2 has four distinct codes (3, 2, 1 and 0), so four categories will be modelled.

15. The `model` statement for these data contains two terms (`item` and `item*step`) and will result in the estimation of two sets of parameters. The term `item` results in the estimation of a set of item difficulty parameters, and the term `item*step` results in a set of item step-parameters that are allowed to vary across the items. The option `scoresfree` results in the estimation of an additional set of item scores that are allowed to vary across the items. This is the generalised partial credit model.

In 'The Structure of ConQuest Design Matrices' in Chapter 12 of Wu, Adams, Wilson and Haldane (2007), there is a description of how the terms in the `model` statement specify different versions of the item response model. In addition, 'Note 6: Score Estimation and Generalised Partial Credit Models' describes how ConQuest estimates the score parameters in models such as the generalised partial credit model.

16. The `estimate` statement is used to initiate the estimation of the item response model.

17. The `show` statement produces a display of the item response model parameter estimates and saves them to the file `ex2a.shw`. The option `estimates=latent` requests that the displays include an illustration of the latent ability distribution.

18. The `itanal` statement produces a display of the results of a traditional item analysis. As with the `show` statement, the results have been redirected to a file (in this case, `ex11a.itn`).

19.-20. The `plot` statements produce two displays for each item in the test. The first requested plot is a comparison of the observed and the modelled expected score curve, while the second is a comparison of the observed and modelled item characteristics curves by category.

RUNNING THE GENERALISED PARTIAL CREDIT SAMPLE ANALYSIS

To run this sample analysis, start the GUI version. Open the file ex11a.cqc and choose **Run➔Run All**.

ConQuest will begin executing the statements that are in the file `ex11a.cqc`; and as they are executed, they will be echoed on the screen. When ConQuest reaches the `estimate` statement, it will begin fitting the generalised partial credit model to the data, and as it does so it will report on the progress of the estimation. This particular sample analysis will take 28 iterations to converge.

After the estimation is complete, the two statements that produce output (`show` and `itanal`) will be processed. The `show` statement will produce seven separate tables. All of these tables will be in the file `ex11a.shw`. The contents of the first and second tables were discussed in Tutorials 1 and 2, respectively. The third table (not shown here) gives the estimates of the population parameters. In this case, the mean of the latent ability distribution was constrained to 0.000, and the variance of that distribution constrained to 1.000.

The fourth table reports the reliability coefficients. Three different reliability statistics are available (Adams, 2006). In this case just the third index (the EAP/PV reliability) is reported because neither of the maximum likelihood estimates has been computed at this stage. The reported reliability is 0.746.

The fifth table was also discussed in Tutorial 2, and is a map of the parameter estimates and latent ability distribution. However, with the exception of predicted probability maps, item maps are not applicable for models with estimated scores.

The sixth table, which contains information related to the item score estimates produced by the `scoresfree` argument in the `model` statement, is shown in Figure 3. The score parameter estimates are reported for each category of each generalised item, although for the generalised partial credit model ConQuest only estimates a single parameter for each item, shown in the final (seventh) table of the show file, discussed later.

For the first item, two score estimates have been reported, corresponding to the codes (1, 2) that this item can take in the data (code 0 will always be scored as zero). For the second item, three score estimates have been reported, corresponding to the codes (1, 2, 3) that this item can take in the data.

```
================================================================================
Generalised Partial Credit Model: What happened last night Tue Jul 21 10:18 2015
TABLE(S) OF GIN SCORES
============================================================Build: Jun 24 2015===
GIN Number   Score           Error              GIN Labels
--------------------------------------------------------------------------------
1.1          0.771           0.107             item 1  Earth shape
1.2          1.541           0.214             item 1  Earth shape
2.1          0.427           0.077             item 2  Earth picture
2.2          0.853           0.155             item 2  Earth picture
2.3          1.280           0.232             item 2  Earth picture
3.1          1.088           0.148             item 3  Falling off
3.2          2.175           0.296             item 3  Falling off
3.3          3.263           0.444             item 3  Falling off
4.1          0.922           0.127             item 4  What is Sun
4.2          1.843           0.254             item 4  What is Sun
5.1          0.726           0.097             item 5  Moonshine
5.2          1.452           0.195             item 5  Moonshine
5.3          2.178           0.292             item 5  Moonshine
6.1          1.405           0.183             item 6  Moon and night
6.2          2.810           0.367             item 6  Moon and night
7.1          0.654           0.088             item 7  Night and day
7.2          1.307           0.176             item 7  Night and day
7.3          1.961           0.264             item 7  Night and day
8.1          0.483           0.066             item 8  Breathe on moon
8.2          0.966           0.133             item 8  Breathe on moon
8.3          1.450           0.199             item 8  Breathe on moon
--------------------------------------------------------------------------------
   Average Score      1.40736
================================================================================
```

Figure 3 Item Score Parameters Estimated by the Generalised Partial Credit Model

Figure 4 shows the seventh table, which displays the Tau parameter estimates for each item and associated standard errors. This estimate is applied to each category of each generalised item to estimate the score parameter estimates that were produced in the previous table. If you compare the sixth and seventh tables, you will notice that the first score estimate for each item in the sixth table is the same as the Tau estimate for that item in the seventh table. The second score estimate (corresponding to category 2) is then double the Tau value, the third score estimate (corresponding to category 3) is triple the Tau value, and so on. Regardless of how many categories each item has, only a single Tau parameter is estimated by the model. This Tau parameter is an estimate of each item's *discrimination*. A smaller Tau value indicates lower discrimination for the item.

```
================================================================================
Generalised Partial Credit Model: What happened last night Tue Jul 21 10:18 2015
TABLE OF TAU VALUES
============================================================Build: Jun 24 2015===
 Tau 1       0.771           0.107              item Earth shape
 Tau 2       0.427           0.077              item Earth picture
 Tau 3       1.088           0.148              item Falling off
 Tau 4       0.922           0.127              item What is Sun
 Tau 5       0.726           0.097              item Moonshine
 Tau 6       1.405           0.183              item Moon and night
 Tau 7       0.654           0.088              item Night and day
 Tau 8       0.483           0.066              item Breathe on moon
--------------------------------------------------------------------------------
      Average Tau      0.80930
================================================================================
```

Figure 4 Tau Parameters Estimated by the Generalised Partial Credit Model

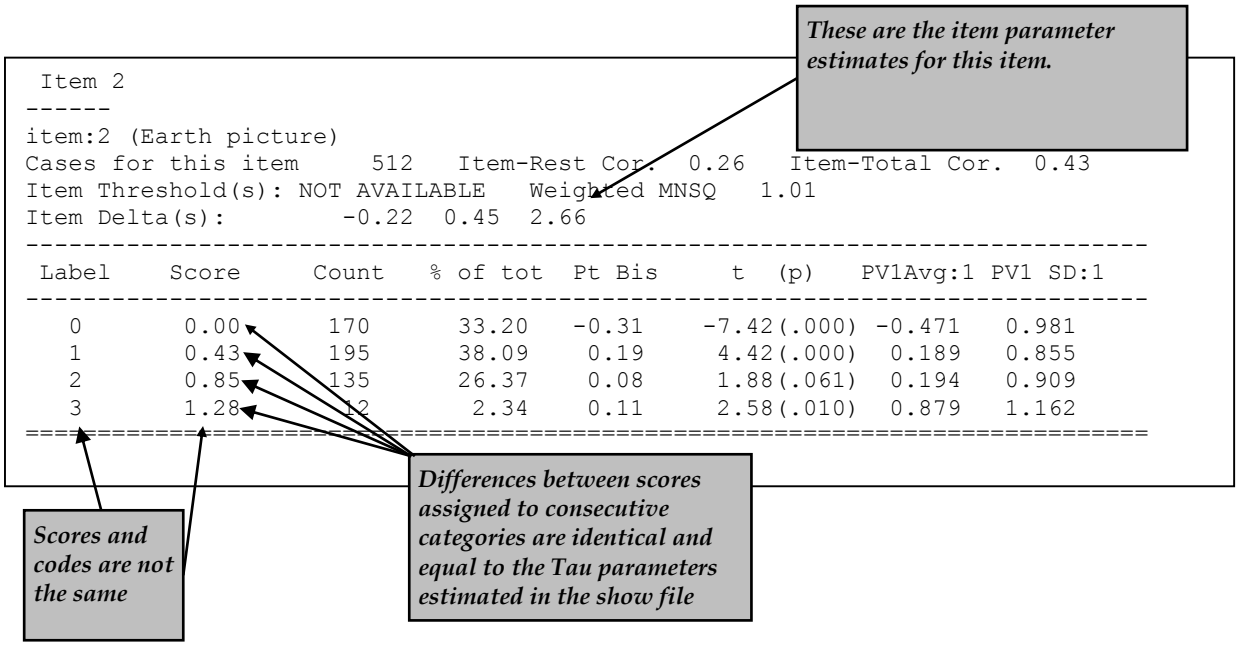

Figure 5 Extract of Item Analysis Printout for a Polytomously Scored Item Estimated with the Generalised Partial Credit Model

The `itanal` command in line 18 produces a file (`ex11a.itn`) that contains traditional item statistics (Figure 5). In this example a key statement was not used and the items use partial credit scoring. As a consequence the `itanal` results are provided at the level of scores, rather than response categories. As you can see in the output, the scores reported are those estimated by the model, not the codes that the response categories are assigned in the data. For the generalised partial credit model, the difference between the scores assigned to consecutive response categories is the same for all categories that item has, and corresponds to the Tau value estimated for that item in the show file. In this case, you can see in Figure 4 that the Tau value for item 2 is 0.427, which is equal to the difference between the scores assigned to consecutive categories shown in Figure 5.

The plot commands in line 19 and 20 produce the graphs shown in Figure 6. For illustrative purposes only plots for item 1 and 2 are shown. The second item showed poor fit to the scaling model — in this case the generalised partial credit model.

The second item's Tau value of 0.427 indicates that this item is less discriminating than the first item (Tau=0.771). The comparison of the observed and modelled expected score curves (the plots appearing on the left of the figure) is the best illustration of this lower discrimination. Notice how for the second item's plot the observed curve is a little flatter than the modelled curve. This will often be the case when the item discrimination is low.

The plots appearing on the right of the figure show the item characteristic curves, both modelled and empirical. There is one pair of curves for each possible score on the item. Note that for item 2 the disparity between the observed and modelled curves for category 2 is the largest. The second part of this tutorial will demonstrate how Conquest can estimate scores for each category of each item in the model, to determine how well each category score fits the scaling model.

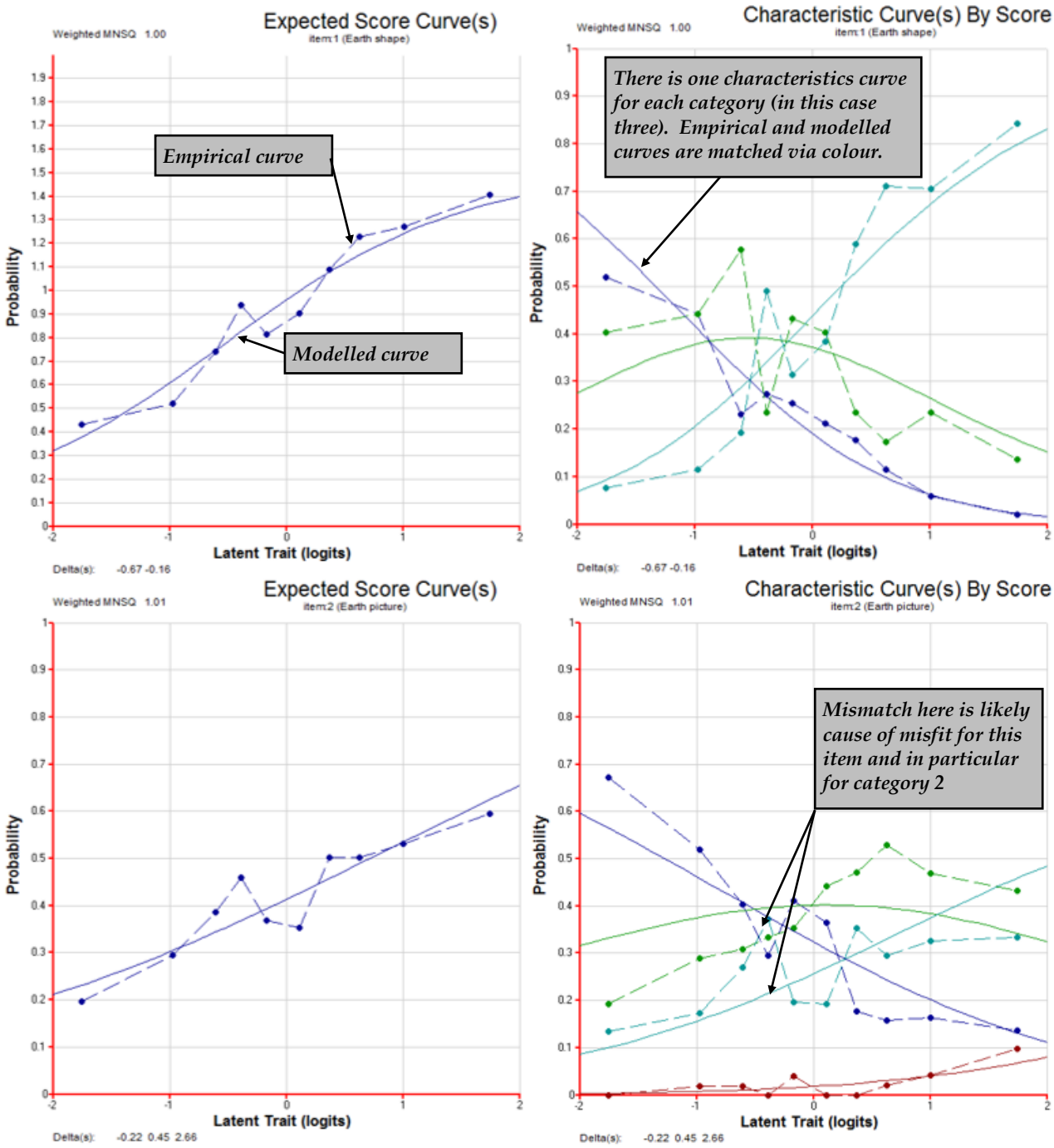

Figure 6 Plots for Items 1 and 2

BOCK'S NOMINAL RESPONSE MODEL

In the second sample analysis of this tutorial, the Bock nominal response model is fitted to the same data used in the previous analysis, to illustrate the differences between the two models. The command file for fitting the Bock nominal response model to the data is given in Figure 7, and the files that we use are:

`Ex11b.cqc`	The command statements.
`Ex2a.dat`	The data.

`Ex2a.lab`	The variable labels for the items on the test.
`Ex11b.shw`	The results of the nominal response analysis.
`Ex11b.itn`	The results of the traditional item analyses.

(The last two files are created when the command file is executed.)

```
1.  Title Bock Nominal Response Analysis: What happened last night;
2.  datafile ex2a.dat;
3.  format name 2-7 responses 10-17;
4.  labels << ex2a.lab;
5.  codes 3,2,1,0;
6.  set constraint=cases;
7.  recode (A,B,C,W,X) (2,1,1,0,0)             !items(1);
8.  recode (A,B,C,W,X) (3,2,1,0,0)             !items(2);
9.  recode (A,B,C,D,E,F,W,X) (3,2,2,1,1,0,0,0)!items(3);
10. recode (A,B,C,W,X) (2,1,0,0,0)             !items(4);
11. recode (A,B,C,D,E,W,X) (3,2,1,1,1,0,0)     !items(5);
12. recode (A,B,W,X) (2,1,0,0)                 !items(6);
13. recode (A,B,C,W,X) (3,2,1,0,0)             !items(7);
14. recode (A,B,C,D,W,X) (3,2,1,1,0,0)         !items(8);
15. model item + item*step!bock;
16. estimate;
17. show !estimates=latent >> ex11b.shw;
18. itanal >> ex11b.itn;
19. plot expected;
20. plot mcc;
```

Figure 7 Sample Command File for fitting Bock's Nominal Response Model

1. For this analysis, we are using the title `Bock Nominal Response Analysis: What happened last night`.

2.-14. The commands in these lines are exactly the same as for the generalised partial credit model analysis (see above)

15. The `model` statement for these data is exactly the same as for the generalised partial credit model analysis. The option `bock` results in the estimation of an additional set of item category scores that are allowed to vary across each of the categories of each of the items. This is the Bock nominal response model.

16.-20. The commands in these lines are exactly the same as for the generalised partial credit model analysis (see above), however the names of the show and traditional item analysis files have been changed to `ex11b.shw` and `ex11b.itn`, respectively.

RUNNING THE BOCK NOMINAL RESPONSE SAMPLE ANALYSIS

To run this sample analysis, start the GUI version. Open the file ex11b.cqc and choose **Run→Run All**.

ConQuest will begin executing the statements that are in the file `ex11b.cqc`; and as they are executed, they will be echoed on the screen. When ConQuest reaches the `estimate` statement, it will begin fitting the Bock nominal response model to the data, and as it does so it will report on the progress of the estimation. This particular sample analysis will take 55 iterations to converge.

After the estimation is complete, the two statements that produce output (`show` and `itanal`) will be processed. The `show` statement will again produce seven separate tables. All of these tables will be in the file `ex11b.shw`, and are the same as those described in the generalised partial credit model (see above).

The important difference between this model and the generalised partial credit model is illustrated in the sixth and seventh tables in the show file. The sixth table, contains information related to the item score estimates produced by the `bock` option in the `model` statement, is shown in Figure 8. The score parameter estimates are reported for each category of each item, and in this case ConQuest estimates a single parameter for each category of each item (rather than a single parameter for each item, as was the case for the generalised partial credit model).

As with the generalised partial credit model, two score estimates have been reported for the first item, corresponding to the codes (1, 2) that this item can take in the data (code 0 will always be scored as zero). For the second item, three score estimates have been reported, corresponding to the codes (1, 2, 3) that this item can take in the data.

```
============================================================================
Partial Credit Model: What happened last night            Tue Jul 21 09:25 2015
TABLE(S) OF GIN SCORES
============================================================Build: Jun 24 2015===
GIN Number   Score           Error              GIN Labels
----------------------------------------------------------------------------
1.1          0.527           0.172             item 1  Earth shape
1.2          1.564           0.216             item 1  Earth shape
2.1          0.939           0.159             item 2  Earth picture
2.2          0.753           0.163             item 2  Earth picture
2.3          1.831           0.435             item 2  Earth picture
3.1          1.612           0.233             item 3  Falling off
3.2          1.740           0.330             item 3  Falling off
3.3          3.408           0.571             item 3  Falling off
4.1          0.764           0.151             item 4  What is Sun
4.2          2.119           0.300             item 4  What is Sun
5.1          0.372           0.201             item 5  Moonshine
5.2          0.994           0.224             item 5  Moonshine
5.3          2.511           0.357             item 5  Moonshine
6.1          1.220           0.205             item 6  Moon and night
6.2          2.932           0.399             item 6  Moon and night
7.1          0.660           0.203             item 7  Night and day
7.2          1.337           0.253             item 7  Night and day
7.3          1.924           0.254             item 7  Night and day
8.1          0.798           0.256             item 8  Breathe on moon
8.2          1.406           0.510             item 8  Breathe on moon
8.3          1.466           0.197             item 8  Breathe on moon
----------------------------------------------------------------------------
   Average Score      1.47022
============================================================================
```

Figure 8 Item Score Parameters Estimated by Bock's Nominal Response Model

Figure 9 shows the seventh table, which displays the Tau parameter estimates for each item and associated standard errors, as it did for the generalised partial credit model. However, you will notice that there are more values in this table than there was for the generalised partial credit model. This is because Conquest is estimating score parameters for each category of each item individually. Consequently, there is a one-to-one correspondence between the values in this table and those that were reported in the previous table. These Tau parameters provide an estimate of each item category's *discrimination*.

```
================================================================================
Partial Credit Model: What happened last night             Tue Jul 21 09:25 2015
TABLE OF TAU VALUES
============================================================Build: Jun 24 2015===
 Tau 1       0.527           0.172                item Earth shape
 Tau 2       0.939           0.159                item Earth picture
 Tau 3       1.612           0.233                item Falling off
 Tau 4       0.764           0.151                item What is Sun
 Tau 5       0.372           0.201                item Moonshine
 Tau 6       1.220           0.205                item Moon and night
 Tau 7       0.660           0.203                item Night and day
 Tau 8       0.798           0.256                item Breathe on moon
 Tau 9       1.564           0.216                item Earth shape step 1
 Tau 10      0.753           0.163                item Earth picture step 1
 Tau 11      1.831           0.435                item Earth picture step 2
 Tau 12      1.740           0.330                item Falling off step 1
 Tau 13      3.408           0.571                item Falling off step 2
 Tau 14      2.119           0.300                item What is Sun step 1
 Tau 15      0.994           0.224                item Moonshine step 1
 Tau 16      2.511           0.357                item Moonshine step 2
 Tau 17      2.932           0.399                item Moon and night step 1
 Tau 18      1.337           0.253                item Night and day step 1
 Tau 19      1.924           0.254                item Night and day step 2
 Tau 20      1.406           0.510                item Breathe on moon step 1
 Tau 21      1.466           0.197                item Breathe on moon step 2
--------------------------------------------------------------------------------
       Average Tau      1.47022
================================================================================
```

Figure 9 Tau Parameters Estimated by Bock's Nominal Response Model

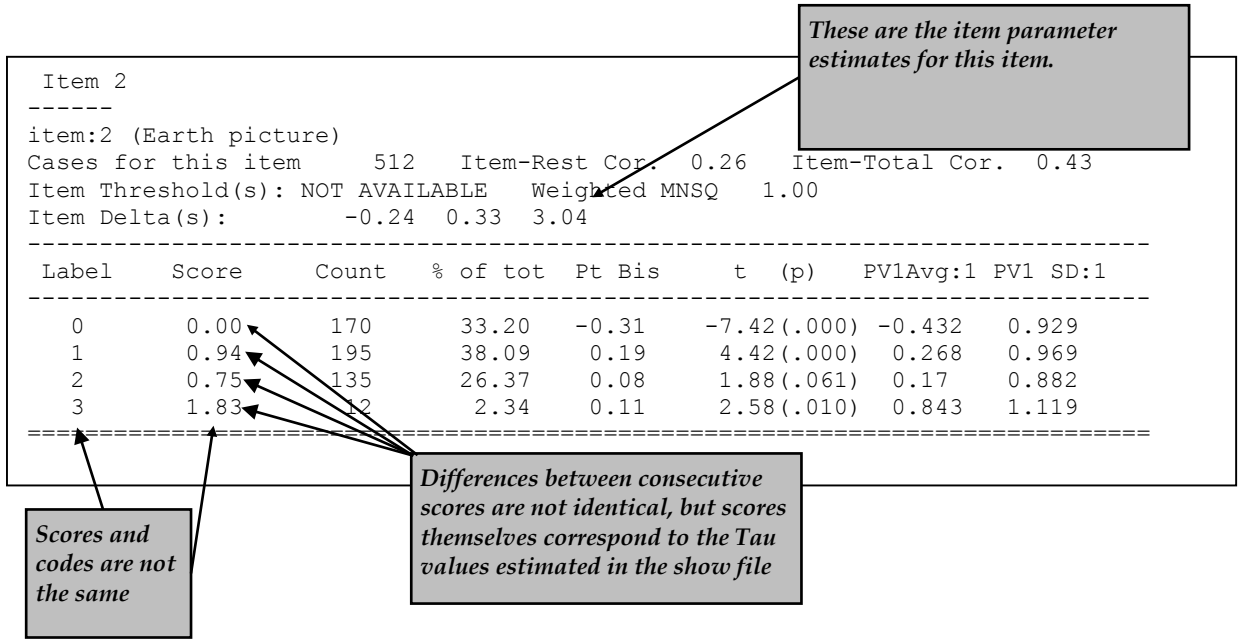

Figure 10 Extract of Item Analysis Printout for a Polytomous Item Estimated with Bock's Nominal Response Model

The `itanal` command in line 18 produces a file (`ex11b.itn`) that contains traditional item statistics (Figure 10). In this example, as with the generalised partial credit example, a key statement was not used and the items use partial credit scoring. As a consequence the `itanal` results are provided at the level of scores, rather than response categories. As you can see in the output, the scores reported are those estimated by the model, not the codes that the response

categories are assigned in the data. These scores correspond to the Tau values estimated in the show file in Figure 9, as well as the score values in Figure 8, as the Tau and score parameters are identical in the Bock nominal response model.

As you can see in both the show file and the traditional item statistics, the category scores estimated by ConQuest can differ quite substantially to the codes that were manually allocated to the data values. In an example with ordinal response data such as this, the order of the category scores estimated by Conquest should match the order of the codes that were in the data (so that a code of 2 gets a higher score than a code of 1). You can see in this example that this is not the case for item 2. The scores estimated by ConQuest for codes 1, 2 and 3 are 0.939, 0.753, and 1.831 respectively. As the score estimated for code 2 is less than that estimated for code 1, this points to a problem in the coding of the original data.

The plot commands in lines 19 and 20 produce the graphs shown in Figure 11. For illustrative purposes only plots for item 1 and 2 are shown. These graphs show a similar picture to what was shown in the generalised partial credit example. The disparity between the observed and modelled item characteristic curves for category 2 of item 2 that was noted in the generalised partial credit example is still observed here, and supported by the discrepancy between the scores estimated for this item in the show file and traditional item statistics.

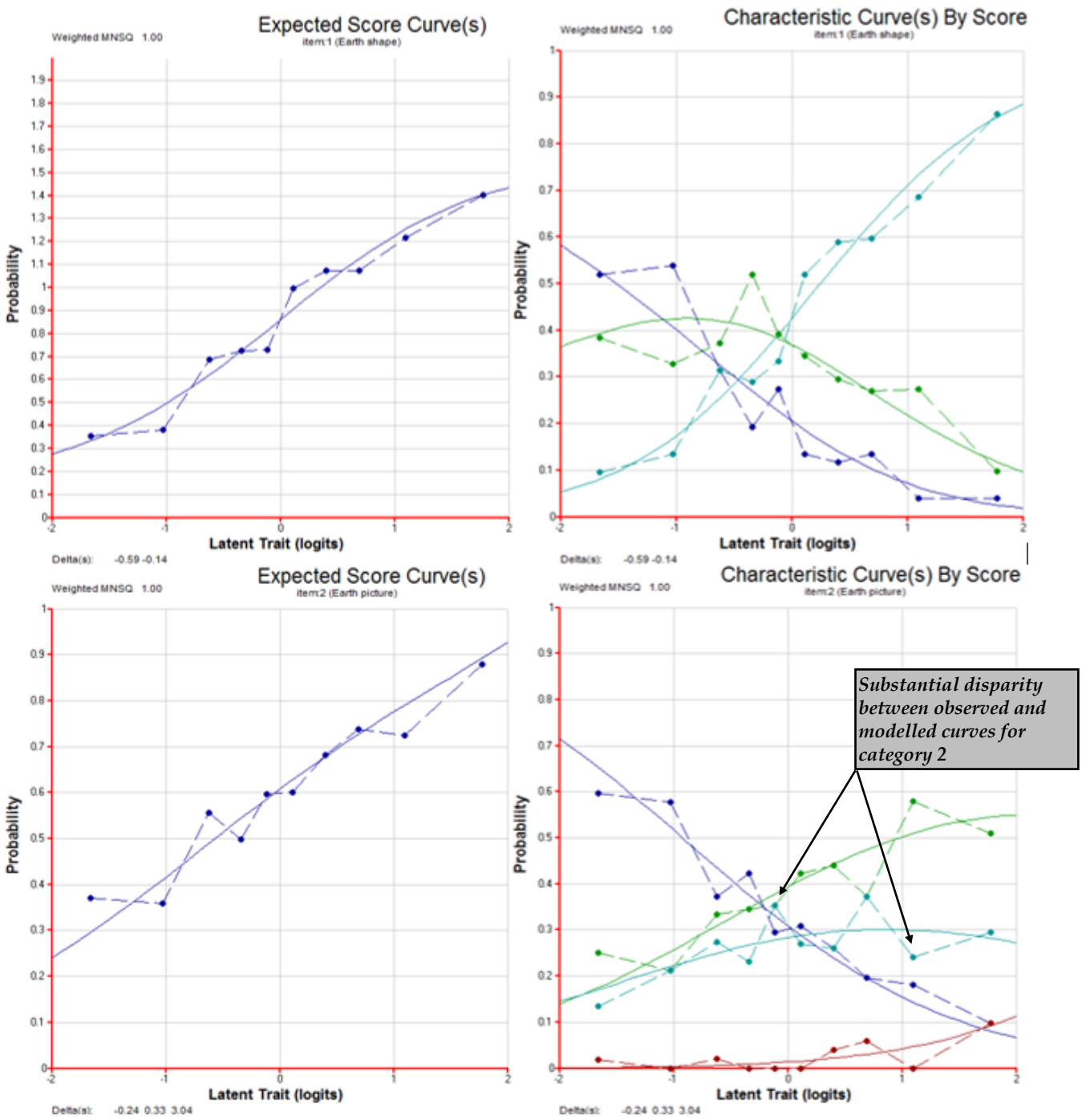

Figure 11 Plots for Item 2

SUMMARY

In this tutorial, ConQuest has been used to fit the generalised partial credit and Bock nominal response models. Some key points covered were:

- The `scoresfree` option in the `model` statement can be used to estimate a single parameter for each item in a given dataset which is used to determine scores that each item category receives (generalised partial credit model).
- The `bock` option in the `model` statement can be used to estimate a score for each category of each item in a given dataset (bock nominal response model).

- The score parameters estimated by Conquest can be used to determine item fit (generalised partial credit model) as well as item category fit (bock nominal response model)

REFERENCES

Adams, R. J. 2006. Reliability as a measurement design effect. *Studies in Educational Evaluation*, 31,162-172.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*, 29-51.

Muraki, E. (1992). A generalized partial credit model. *Applied Psychological Measurement, 16*(2), 159-176.

Wu, M. L., Adams, R. J., Wilson, M. R., Haldane, S.A. (2007). *ACER ConQuest Version 2: Generalised item response modelling software* [computer program]. Camberwell: Australian Council for Educational Research.