

Modelling Polytomously Scored Items With The Rating Scale and Partial Credit Models

Ray Adams and Margaret Wu. 22 August 2010

The rating scale model (Andrich, 1978; Wright and Masters, 1982) and the partial credit model (Masters, 1982; Wright and Masters, 1982) are extensions to Rasch's simple logistic model and are suitable for use when items are scored polytomously. The rating scale model was initially developed by Andrich for use with Likert-style items, while Masters' extension of the rating scale model to the partial credit model was undertaken to facilitate the analysis of cognitive items that are scored into more than two ordered categories. In this tutorial, the use of ConQuest to fit the partial credit and rating scale models is illustrated through two sets of sample analyses. In the first, the partial credit model is fit to some cognitive items; and in the second, the fit of the rating scale and partial credit models to a set of items that forms an attitudinal scale is compared.

FITTING THE 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. Previous analyses of some of these data are reported in Adams, Doig and Rosier (1991).

The files used in this sample analysis are:

<code>ex2a.cqc</code>	The command statements.
<code>ex2a.dat</code>	The data.
<code>ex2a.lab</code>	The variable labels for the items on the partial credit test.
<code>ex2a.shw</code>	The results of the partial credit analysis.
<code>ex2a.itn</code>	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
2110109ZHBWBBBWCAAXXXX
2110113ZIWBWBXWCXXXABBB
2110115ZHWBFBWBCWXAWAXX
2110121ZHWWEBWBBCAABABA
2110123YIBWWEWBWXABABB
2110305ZHCBABABAABACCCA
2110313YBBCFBDXCXXXXXXX
.
.
```

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

NOTE:

In most Rasch-type models, a one-to-one match exists between the label that is assigned to each response category to an item (the category label) and the response level (or score) that is assigned to that response category. This need not be the case with ConQuest.

In ConQuest, the distinction between a response category and a response level is an important one. When ConQuest fits item response models, it actually models the probabilities of each of the response categories for each item. The scores for each of these categories need not be unique. For example, a four-alternative multiple choice item can be modelled as a four-response category item with three categories assigned zero scores and one category assigned a score of one, or it can be modelled in the usual fashion as a two-category item where the scores identify the categories.

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 Partial Credit Model: 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. recode (A,B,C,W,X) (2,1,1,0,0) !item(1);
7. recode (A,B,C,W,X) (3,2,1,0,0) !item(2);
8. recode (A,B,C,D,E,F,W,X) (3,2,2,1,1,0,0,0) !item(3);
9. recode (A,B,C,W,X) (2,1,0,0,0) !item(4);
10. recode (A,B,C,D,E,W,X) (3,2,1,1,1,0,0) !item(5);
11. recode (A,B,W,X) (2,1,0,0) !item(6);
12. recode (A,B,C,W,X) (3,2,1,0,0) !item(7);
13. recode (A,B,C,D,W,X) (3,2,1,1,0,0) !item(8);
14. model item + item*step;
15. estimate;
16. show !estimates=latent >> ex2a.shw;
17. itanal >> ex2a.itn;
18. plot expected! gins=2;
19. plot icc! gins=2;
20. plot ccc! gins=2;

```

Figure 2 Sample Command File for a Partial Credit Test

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.-13. 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.

14. 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. This is the 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.

15. The `estimate` statement is used to initiate the estimation of the item response model.
16. 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.
17. 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, `ex2a.itn`).
- 18.-20. The `plot` statements produce a sequence of three displays for item 2 only. The first requested plot is a comparison of the observed and the modelled expected score curve. The second plot is a comparison of the observed and modelled item characteristic curves, and the third plot shows comparisons of the observed and expected cumulative item characteristic curves.

RUNNING THE PARTIAL CREDIT SAMPLE ANALYSIS

To run this sample analysis, start the GUI version. Open the file `ex2a.cqc` and choose **Run→Run All**.

ConQuest will begin executing the statements that are in the file `ex2a.cqc`; and as they are executed, they will be echoed on the screen. When ConQuest reaches the `estimate` statement, it will begin fitting the 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 51 iterations to converge.

After the estimation is complete, the two statements that produce output (`show` and `itanal`) will be processed. As in the previous sample analysis, the `show` statement will produce six separate tables. All of these tables will be in the file `ex2a.shw`. The contents of the first table were discussed in Tutorial 1. The first half of the second table, which contains information related to the parameter estimates for the first term in the `model` statement, is shown in Figure 3. The parameter estimates in this table are for the *difficulties* of each of the items. For the purposes of model identification, ConQuest constrains the difficulty estimate for the last item to ensure an average difficulty of zero. This constraint has been achieved by setting the difficulty of the last item to be the negative sum of the previous items. The fact that this item is constrained is indicated by the asterisk (*) placed next to the parameter estimate.

```

=====
Partial Credit Model: What happened last night          Tue Oct 03 11:53 2006
TABLES OF RESPONSE MODEL PARAMETER
=====
TERM 1: item
=====

```

VARIABLES			UNWEIGHTED FIT			WEIGHTED FIT		
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
1 Earth shape	-0.727	0.034	1.00 (0.88, 1.12)		-0.0	1.00 (0.90, 1.10)		-0.1
2 Earth pictu..	0.770	0.034	1.13 (0.88, 1.12)		1.9	1.11 (0.90, 1.10)		2.0
3 Falling off	0.606	0.036	0.95 (0.88, 1.12)		-0.8	0.95 (0.85, 1.15)		-0.7
4 What is Sun	0.192	0.036	0.97 (0.88, 1.12)		-0.5	0.97 (0.89, 1.11)		-0.5
5 Moonshine	-0.151	0.030	1.02 (0.88, 1.12)		0.3	1.02 (0.90, 1.10)		0.3
6 Moon and ni..	0.113	0.035	0.87 (0.88, 1.12)		-2.1	0.89 (0.90, 1.10)		-2.3
7 Night and d..	-0.219	0.030	1.03 (0.88, 1.12)		0.4	1.04 (0.90, 1.10)		0.7
8 Breathe on ..	-0.583*	0.089	1.26 (0.88, 1.12)		3.9	1.11 (0.89, 1.11)		1.8

```

=====
An asterisk next to a parameter estimate indicates that it is constrained
Separation Re
Chi-square te
^ Quick stand
=====

```

Figure 3 Parameter Estimates for the First Term in the model Statement

Figure 4 shows the second table, which displays the parameter estimates, standard errors and fit statistics associated with the second term in the model statement, the step parameters. You will notice that the number of step parameters that has been estimated for each item is one less than the number of modelled response categories for the item. Furthermore, the last of the parameters for each item is constrained so that the sum of the parameters for an item equals zero. This is a necessary identification constraint. In the case of item 1, for example, there are three categories, 0, 1 and 2. Two values are reported, but only the first step parameter has been estimated. The second is the negative of the first. The parameter labelled as step 1, describes the transition from category 0 to 1, where the probability of being in category 1 is greater than the probability of being in category 0, while the second step describes the transition from 1 to 2. 'The Structure of ConQuest Design Matrices' in Chapter 12 of Wu, *et al.* (2007) gives a description of why an item has two fewer step parameters than it has categories, and it discusses the interpretation of these parameters.

There is a fit statistic reported for each category. This statistic provides a comparison of the expected number of students responding in the category with the observed number responding in that category.

The Rating Scale and Partial Credit Models

TERM 2: item*step

*This part of the second table is for the term `item*step`, the second term in the `model` statement.*

VARIABLES

item	step	ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T
1 Earth shape	0			1.05	(0.88, 1.12)	0.8	1.04	(0.88, 1.12)	0.7
1 Earth shape	1	-0.234	0.095	0.97	(0.88, 1.12)	-0.5	0.98	(0.94, 1.06)	-0.6
1 Earth shape	2	0.234*		0.97	(0.88, 1.12)	-0.5	0.96	(0.93, 1.07)	-1.3
2 Earth pictu..	0			1.05	(0.88, 1.12)	0.8	1.00	(0.92, 1.08)	-0.0
2 Earth pictu..	1	-1.420	0.098	1.01	(0.88, 1.12)	0.1	1.01	(0.95, 1.05)	0.3
2 Earth pictu..	2	-0.514	0.107	1.24	(0.88, 1.12)	3.5	1.13	(0.91, 1.09)	2.5
2 Earth pictu..	3	1.934*		1.29	(0.88, 1.12)	4.2	1.01	(0.47, 1.53)	0.1
3 Falling off	0			0.84	(0.88, 1.12)	-2.6	0.89	(0.89, 1.11)	-2.1
3 Falling off	1	-2.022	0.102	0.95	(0.88, 1.12)	-0.9	0.95	(0.94, 1.06)	-1.8
3 Falling off	2	1.363	0.165	1.17	(0.88, 1.12)	2.6	1.03	(0.75, 1.25)	0.3
3 Falling off	3	0.659*		0.60	(0.88, 1.12)	-7.5	1.01	(0.58, 1.42)	0.1
4 What is Sun	0			1.03	(0.88, 1.12)	0.5	1.01	(0.92, 1.08)	0.2
4 What is Sun	1	-1.151	0.090	1.00	(0.88, 1.12)	0.1	1.00	(0.97, 1.03)	0.2
4 What is Sun	2	1.151*		0.80	(0.88, 1.12)	-3.5	0.95	(0.83, 1.17)	-0.6
5 Moonshine	0			1.07	(0.88, 1.12)	1.0	1.07	(0.91, 1.09)	1.5
5 Moonshine	1	0.308	0.094	0.97	(0.88, 1.12)	-0.4	1.00	(0.85, 1.15)	0.0
5 Moonshine	2	-0.331	0.112	1.12	(0.88, 1.12)	1.8	1.02	(0.88, 1.12)	0.4
5 Moonshine	3	0.023*		0.82	(0.88, 1.12)	-3.1	0.93	(0.90, 1.10)	-1.4
6 Moon and ni..	0			0.92	(0.88, 1.12)	-1.3	0.93	(0.92, 1.08)	-1.8
6 Moon and ni..	1	-0.748	0.090	0.97	(0.88, 1.12)	-0.4	0.98	(0.97, 1.03)	-1.6
6 Moon and ni..	2	0.748*		0.75	(0.88, 1.12)	-4.4	0.90	(0.87, 1.13)	-1.5
7 Night and d..	0			1.12	(0.88, 1.12)	0.6	1.02	(0.91, 1.09)	0.4
7 Night and d..	1			1.12	(0.88, 1.12)	-0.1	1.01	(0.85, 1.15)	0.1
7 Night and d..	2			1.12	(0.88, 1.12)	-0.3	1.00	(0.83, 1.17)	-0.0
7 Night and d..	3	-0.520*		1.01	(0.88, 1.12)	0.2	1.01	(0.90, 1.10)	0.2
8 Breathe on ..	0			1.23	(0.88, 1.12)	3.4	1.04	(0.89, 1.11)	0.8
8 Breathe on ..	1	1.240	0.148	1.28	(0.88, 1.12)	4.1	1.03	(0.75, 1.25)	0.3
8 Breathe on ..	2	1.735	0.337	1.11	(0.88, 1.12)	1.6	1.01	(0.36, 1.64)	0.1
8 Breathe on ..	3	-2.975*		1.23	(0.88, 1.12)	3.5	1.12	(0.89, 1.11)	2.1

An asterisk next to a parameter estimate indicates that it is constrained

These parameters are estimated for combinations of steps and items.

Figure 4 Parameter Estimates for the Second Term in the `model` Statement

The third table in the file (not shown here) gives the estimates of the population parameters. In this case, the mean of the latent ability distribution is -0.320, and the variance of that distribution is 0.526.

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 0.735.

The fifth table (Figure 5) is a map of the parameter estimates and latent ability distribution. For this model, the map consists of two panels, one for the latent ability distribution and one for each of the terms in the `model` statement that do not include a `step` (in this case one). In this case the leftmost panel shows the estimated latent ability distribution and the second shows the item difficulties.

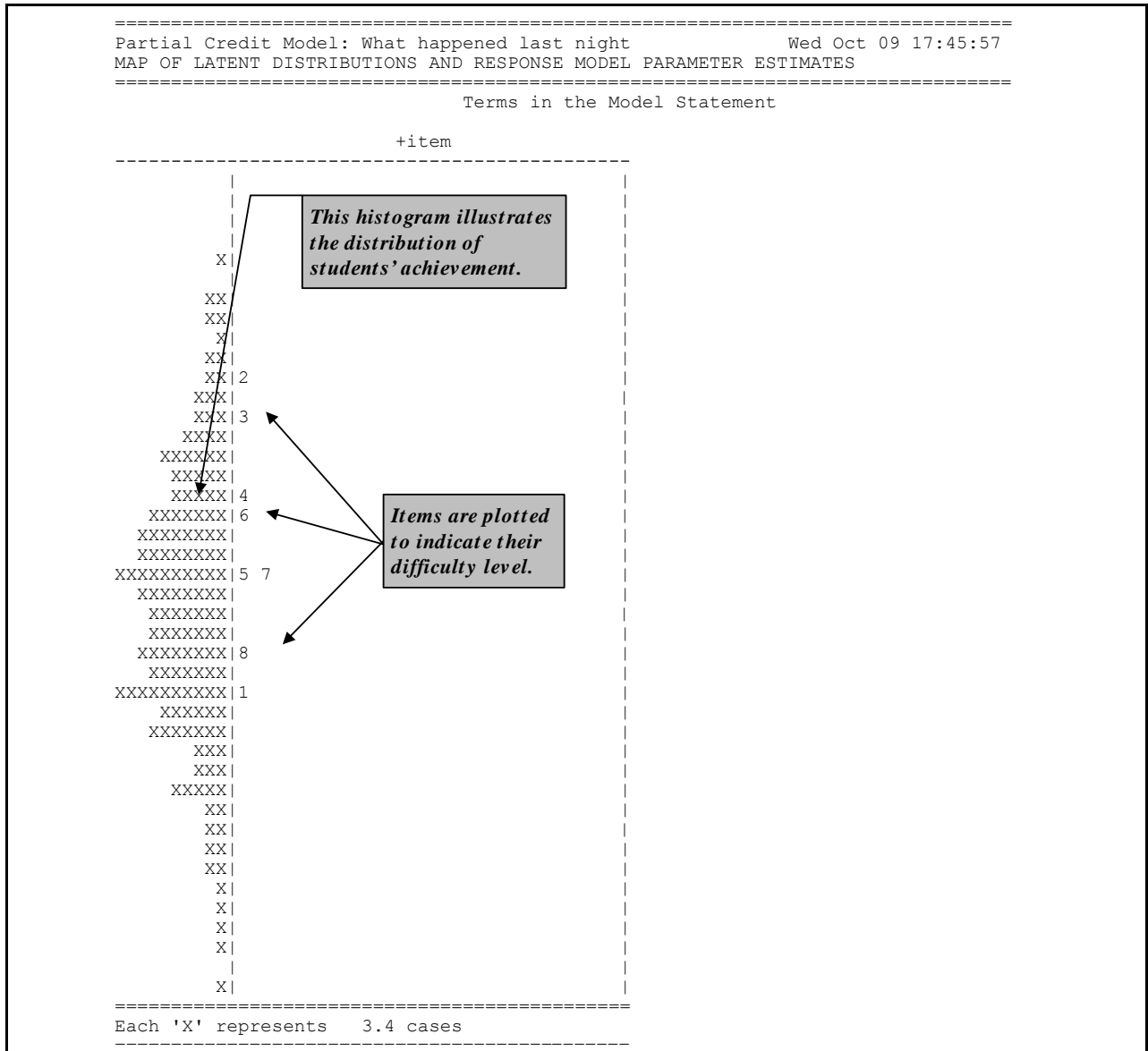


Figure 5 The Item and Latent Distribution Map for the Partial Credit Model

EXTENSION: The headings of the panels in Figure 5 are preceded by a plus sign (+). This indicates the orientation of the parameters. A plus indicates that the facet is modelled with difficulty parameters, whereas a minus sign (-) indicates that the facet is modelled with easiness parameters. This is controlled by the sign that you use in the `model` statement.

Figure 6, the sixth table from the file `ex2a.shw`, is a plot of the Thurstonian thresholds for the items. The definition of these thresholds is discussed in 'Computing Thresholds' in Wu *et al.* (2007) Briefly, they are plotted at the point where a student has a 50% chance of achieving at least the indicated level of performance on an item.

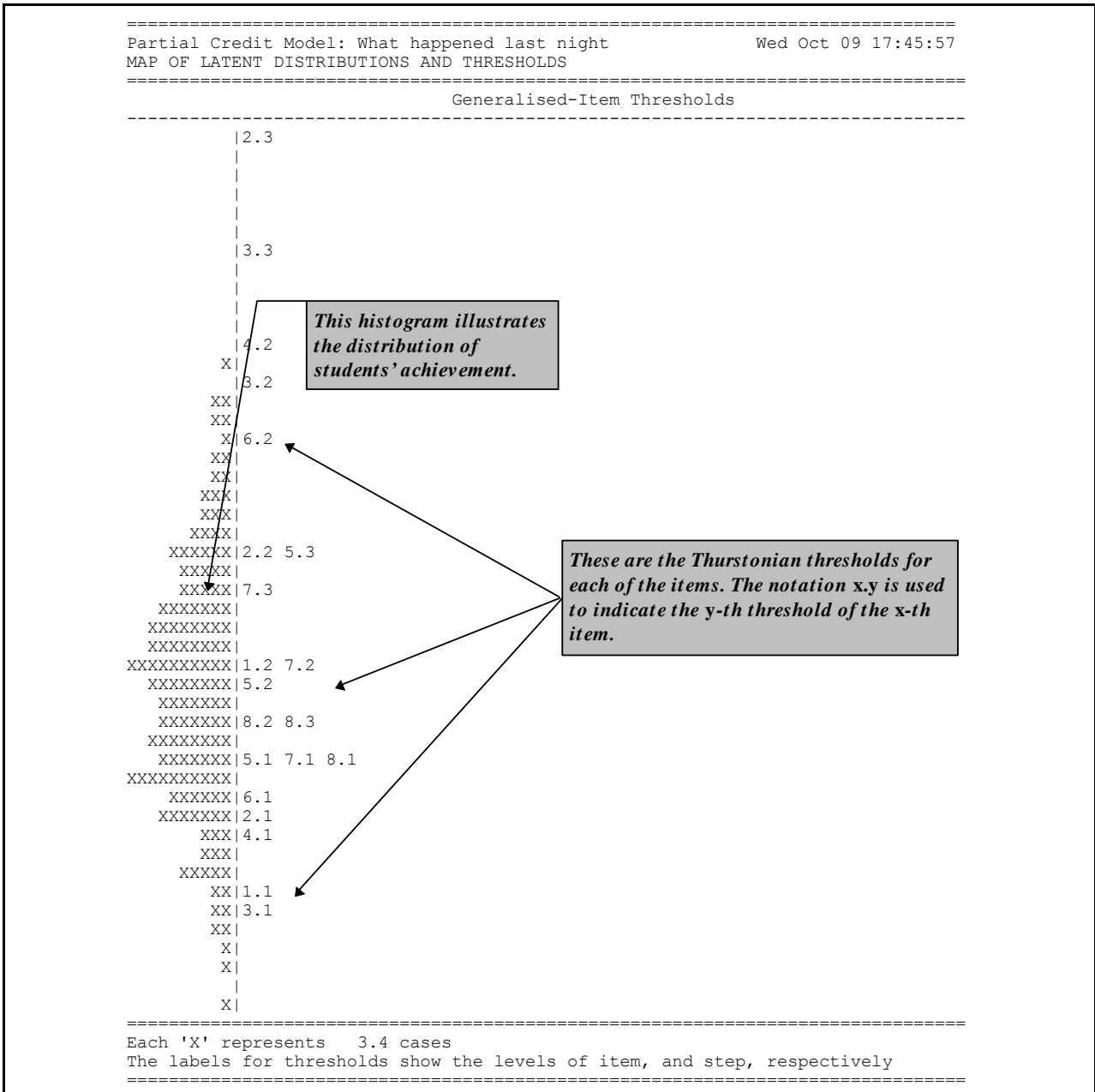


Figure 6 Item Thresholds for the Partial Credit Model

EXTENSION: The method used to construct the ability distribution is determined by the `estimates=` option used in the `show` statement. The `latent` distribution is constructed by drawing a set of plausible values for the students and constructing a histogram from the plausible values. Other options for the distribution are **EAP**, **WLE** and **MLE**, which result in histograms of expected a-posteriori, weighted maximum likelihood and maximum likelihood estimates, respectively. Details of these ability estimates are discussed elsewhere.

Label	Score	Count	% of tot	Pt Bis	t	(p)	PV1Avg:1	PV1 SD:1
0	0.00	170	33.20	-0.45	-11.30	(.000)	-0.68	0.69
1	1.00	195	38.09	0.18	4.16	(.000)	-0.17	0.64
2	2.00	135	26.37	0.22	5.03	(.000)	-0.12	0.61
3	3.00	12	2.34	0.18	4.10	(.000)	0.30	0.86

Figure 7 Extract of Item Analysis Printout for a Partial Credit Item

The `itanal` command in line 17 produces a file (`ex2a.itn`) that contains traditional item statistics (Figure 7). 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.

The three plot commands (lines, 18-20) produce the graphs shown in Figure 4.8. For illustrative purposes only plots for item 2 are requested. This item showed poor fit to the scaling model — in this case the partial credit model.

The item fit MNSQ of 1.11 indicates that this item is less discriminating than expected by the model. The first plot, the comparison of the observed and modelled expected score curves is the best illustration of this misfit. Notice how in this plot the observed curve is a little flatter than the modelled curve. This will often be the case when the MNSQ is significantly larger than 1.0.

The second plot shows the item characteristic curves, both modelled and empirical. There is one pair of curves for each possible score on the item, in this case 0, 1, 2 and 3. Note that the disparity between the observed and modelled curves for category 2 is the largest and this is consistent with the high fit statistic for this category.

The third plot is a cumulative form of the item characteristic curves. In this case three pairs of curves are plotted. The rightmost pair gives the probability of a response of 3, the next pair is for the probability of 2 or 3, and the final pairing is for the probability of 1, 2 or 3. Where these curves attain a probability of 0.5, the value on the horizontal axis corresponds to each of the three threshold parameters that are reported under the figure.

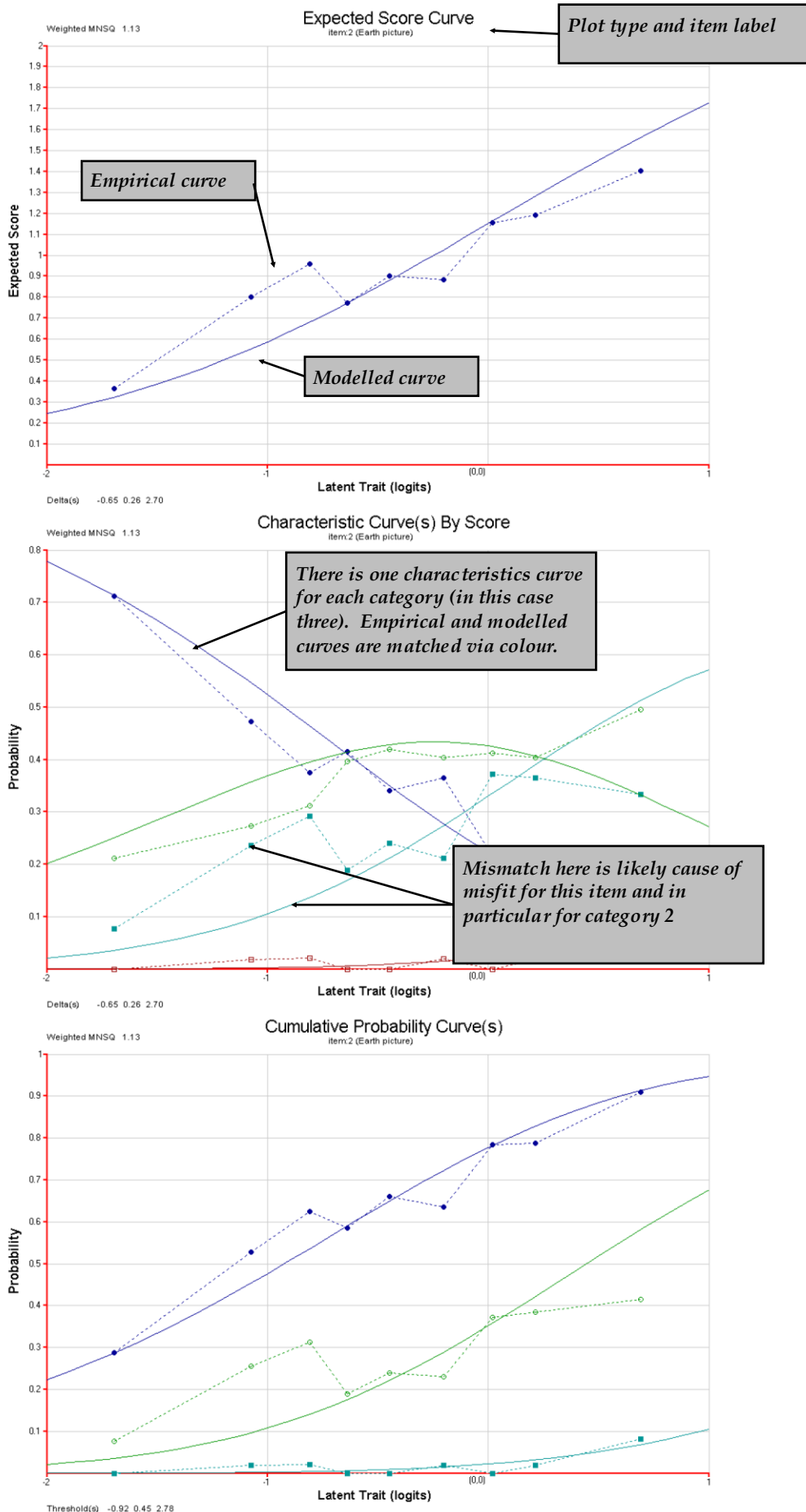


Figure 8 Plots for Item 2

THE PARTIAL CREDIT AND RATING SCALE MODELS: A COMPARISON OF FIT

A key feature of ConQuest is its ability to fit alternative Rasch-type models to the same data set. Here a rating scale model and a partial credit model are fit to a set of items that were designed to measure the importance placed by teachers on adequate resourcing and support to the success of bilingual education programs. The data come from a study undertaken by Zammit (1997).

The data consist of the responses of 582 teachers to the 10 items listed in Figure 9. Each item was presented with a Likert-style response format; and in the data file, strongly agree was coded as 1, agree as 2, uncertain as 3, disagree as 4, and strongly disagree as 5.

- | | |
|-----|--|
| 1. | A bilingual teaching program is only successful when curriculum materials are adequate. |
| 2. | A bilingual teaching program should not be implemented before there are enough bilingual teachers. |
| 3. | A bilingual teaching program requires adequate resources. |
| 4. | The staff's commitment to bilingual teaching is essential if bilingual teaching is to succeed. |
| 5. | The staff's commitment to bilingual teaching is very important in improving the students' achievement in school. |
| 6. | A bilingual program must last several years for students to achieve basic bilingual proficiency. |
| 7. | To achieve maximum success all staff should be convinced that the bilingual teaching program has achievable goals. |
| 8. | The principal's commitment to bilingual teaching is very important in improving students' achievement in school. |
| 9. | A bilingual teaching program cannot succeed without the active support of parents. |
| 10. | Bilingual teaching requires a high level of motivation from students. |

Figure 9 Items Used in the Comparison of the Rating Scale and the Partial Credit Models

The command file for fitting the rating scale and partial credit models to these data is given in Figure 10, and the files that we use are:

ex2b.cqc	The command statements.
ex2b.dat	The data.
ex2b.lab	The variable labels for the items on the rating scale.
ex2b.shw	The results of the rating scale analysis.
ex2b.itn	The results of the traditional item analyses.
ex2c.shw	The results of the partial credit analysis.

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

```
1. Title Rating Scale Analysis;
2. datafile ex2b.dat;
3. format responses 9-15,17-19;
4. codes 0,1,2;
5. recode (1,2,3,4,5) (2,1,0,0,0);
6. labels << ex2b.lab;
7. model item + step;
8. estimate;
9. Show >> ex2b.shw;
10. Itanal >> ex2b.itn;
11. reset;
12. Title Partial Credit Analysis;
13. datafile ex2b.dat;
14. format responses 9-15,17-19;
15. codes 0,1,2;
16. recode (1,2,3,4,5) (2,1,0,0,0);
17. labels << ex2b.lab;
18. model item + item*step;
19. estimate;
20. Show >> ex2c.shw;
```

Figure 10 Command File for Fitting a Rating Scale and then a Partial Credit Model to the Same Data

1. For this analysis, we are using the title `Rating Scale Analysis`.
2. The data for this sample analysis are to be read from the file `ex2b.dat`.
3. The `format` statement describes the layout of the data in the file `ex2b.dat`. This format indicates that the responses to the first seven items are located in columns 9 through 15 and that the responses to the next three items are located in columns 17 through 19.
4. The valid codes, after `recode`, are 0, 1 and 2.
5. The original codes of 1, 2, 3, 4, and 5 are recoded to 2, 1, and 0. Because 3, 4, and 5 are all being recoded to 0, this means we are collapsing these categories (uncertain, disagree, and strongly disagree) for the purposes of this analysis.
6. A set of labels for the items is to be read from the file `ex2b.lab`.
7. This is the `model` statement that corresponds to the rating scale model. The first term in the `model` statement indicates that an item difficulty parameter is modelled for each item, and the second indicates that step parameters are the same for all items.
8. The `estimate` statement is used to initiate the estimation of the item response model.
9. Item response model results are to be written to the file `ex2b.shw`.
10. Traditional statistics are to be written to the file `ex2b.itn`.
11. The `reset` statement can be used to separate jobs that are put into a single command file. The `reset` statement returns all values to their defaults. Even though many values are the same for these analyses, we advise resetting, as you may be unaware of some values that have been set by the previous statements.
- 12.-20. These lines replicate lines 1 to 9. The only difference is in the `model` statement (compare lines 18 and 7). In the first analysis, the second term of the `model` statement is `step`, whereas in the second analysis the second term is `item*step`. In the latter

case, the step structure is allowed to vary across items, whereas in the first case, the step structure is constrained to be the same across items.

RUNNING THE COMPARISON OF THE RATING SCALE AND PARTIAL CREDIT MODELS

To run this sample analysis, launch the GUI version of ConQuest and open the command file `ex2b.cqc` and choose **Run**→**Run All**.

ConQuest will begin executing the statements that are in the file `ex2b.cqc`; and as they are executed, they will be echoed on the screen. The first model, the rating scale model, will take 28 iterations to converge; and the second, the partial credit model, will take 27 iterations.

To compare the fit of the two models to these data, two tables produced by the `show` statements for each model are compared. First, the summary tables for each model are compared. These two tables are reproduced in Figure 11. From these tables we note that the rating scale model has used 12 parameters, and the partial credit model has used 21 parameters. For the rating scale model, the parameters are the mean and variance of the latent variable, nine item difficulty parameters, and a single step parameter. For the partial credit model, the parameters are the mean and variance of the latent variable, nine item difficulty parameters and 10 step parameters.

A formal statistical test of the relative fit of these models can be undertaken by comparing the deviance of the two models. Comparing the deviance in the summary tables, note that the rating scale model deviance is 67.58 greater than the deviance for the partial credit model. If this value is compared to a chi-squared distribution with 9 degrees of freedom, this value is significant and it can be concluded that the fit of the rating scale model is significantly worse than the fit of the partial credit model.

The Rating Scale and Partial Credit Models

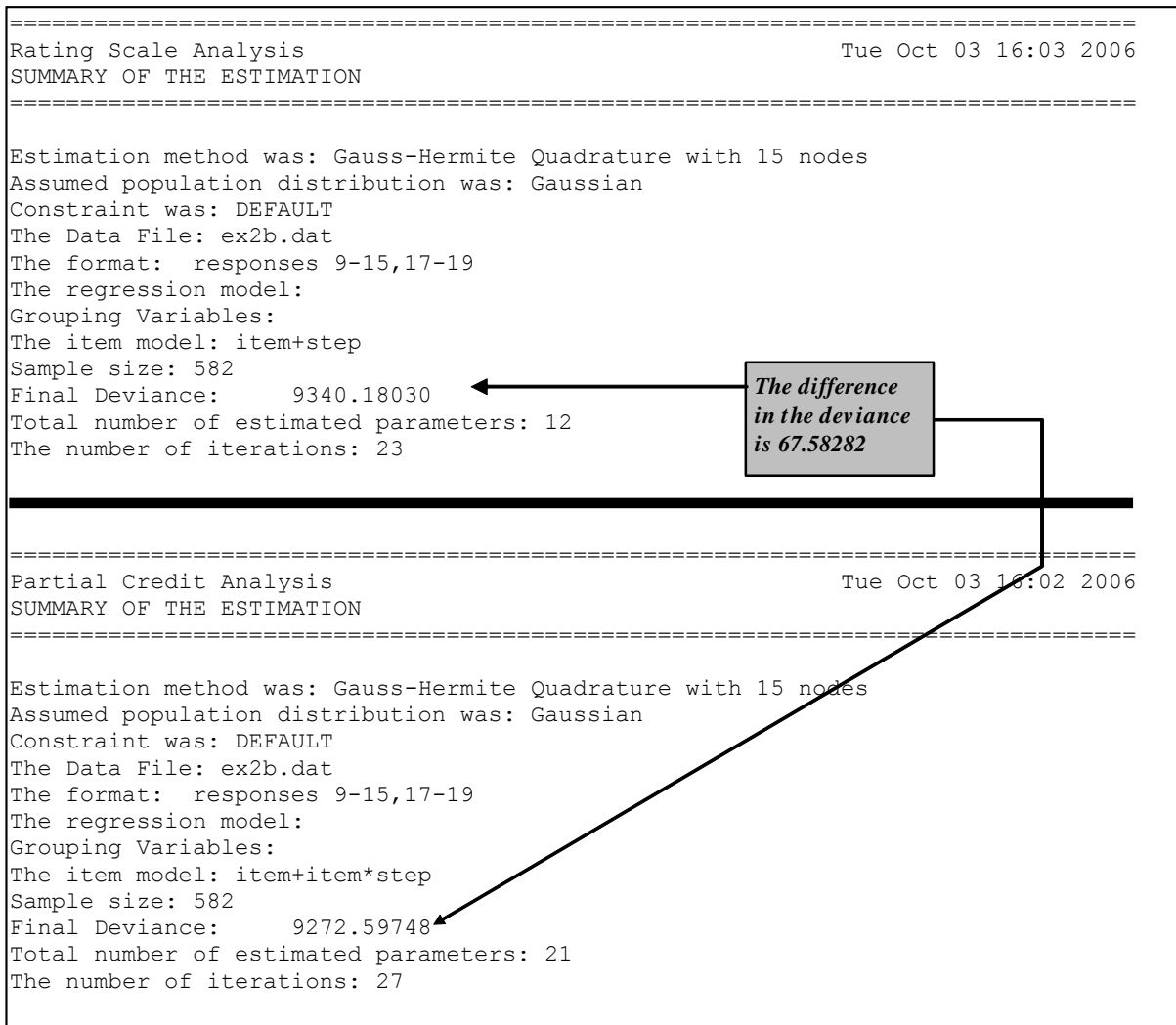


Figure 11 Summary Information for the Rating Scale and Partial Credit Analyses

The difference in the fit of these two models is highlighted by comparing the contents of Figures 12 and 13.

Figure 12 shows that, in the case of the rating scale model, the step parameter fits poorly, whereas in Figure 13 the fit statistics for the step parameters are generally small or less than their expected value (ie the t-values are negative) expected. In both cases, the difficulty parameter for item 2 does not fit well. An examination of the text of this item in Figure 9 shows that perhaps the misfit of this item can be explained by the fact that it is slightly different to the other questions in that it focuses on the conditions under which a bilingual program should be started rather than on the conditions necessary for the success of a bilingual program. Thus, although overall the partial credit model fits better than the rating scale model as discussed previously, the persistence of misfit for the difficulty parameter for this item indicates that the inclusion of this item in the scale should be reconsidered.

```

=====
Rating Scale Analysis                               Tue Oct 03 16:03 2006
TABLES OF RESPONSE MODEL PARAMETER ESTIMATES
=====
TERM 1: item
-----
VARIABLES                UNWEIGHTED FIT
-----
item      ESTIMATE  ERROR^  MNSQ    CI      T      MNSQ    CI      T
-----
1  Curriculum ..    0.716  0.054  1.12 ( 0.88, 1.12)  1.9  1.12 ( 0.89, 1.11)  2.1
2  Not Until E..   1.061  0.054  1.47 ( 0.88, 1.12)  6.8  1.51 ( 0.89, 1.11)  8.2
3  Financial R..  -0.559  0.056  0.83 ( 0.88, 1.12) -3.1  0.88 ( 0.88, 1.12) -2.1
4  Staff Commi.. -1.046  0.057  0.73 ( 0.88, 1.12) -5.0  0.79 ( 0.88, 1.12) -3.7
5  Commitment ..  -0.425  0.055  0.94 ( 0.88, 1.12) -1.0  0.94 ( 0.89, 1.11) -1.1
6  Run for som..  -0.009  0.054  1.05 ( 0.88, 1.12)  0.9  1.06 ( 0.89, 1.11)  1.1
7  Achievable ..  -0.386  0.055  0.80 ( 0.88, 1.12) -3.6  0.78 ( 0.89, 1.11) -4.0
8  Principals ..  -0.133  0.055  0.87 ( 0.88, 1.12) -2.2  0.87 ( 0.89, 1.11) -2.3
9  Parents sup..   0.232  0.054  1.06 ( 0.88, 1.12)  1.1  1.07 ( 0.89, 1.11)  1.3
10 Student mot..  0.548*  0.165  1.00 ( 0.88, 1.12)  0.1  0.98 ( 0.89, 1.11) -0.4
-----
An asterisk next to a parameter estimate indicates that it is constrained
Separation Reliability = 0.993
Chi-square test of parameter equality = 1127.17, df = 9, Sig Level = 0.000
^ Quick standard errors have been used
=====
TERM 2: step
-----
VARIABLES                UNWEIGHTED FIT
-----
step      ESTIMATE  ERROR^  MNSQ    CI      T      MNSQ    CI      T
-----
0          0          0.029  2.59 ( 0.88, 1.12) 18.8  1.75 ( 0.85, 1.11)  8.0
1      -1.184  0.029  2.39 ( 0.88, 1.12) 17.0  2.55 ( 0.88, 1.12) 18.7
2       1.184*  0.029  1.51 ( 0.88, 1.12)  7.4  1.58 ( 0.88, 1.12)  7.9
-----
An asterisk next to a parameter estimate indicates that it is constrained
^ Quick standard errors have been used
=====

```

Figure 12 Response Model Parameter Estimates for the Rating Scale Model

The Rating Scale and Partial Credit Models

Partial Credit Analysis Tue Oct 03 16:02 2006

TABLES OF RESPONSE MODEL PARAMETER ESTIMATES

TERM 1: item

VARIABLES			UNWEIGHTED FIT			WEIGHTED FIT		
item	ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T
1	Curriculum ..	0.753	0.055	1.11 (0.88, 1.12)	1.8	1.10 (0.89, 1.11)		1.8
2	Not Until E..	1.068	0.053	1.41 (0.88, 1.12)	6.0	1.37 (0.89, 1.11)		6.0
3	Financial R..	-0.524	0.058	0.82 (0.88, 1.12)	-3.2	0.87 (0.88, 1.12)		-2.3
4	Staff Commi..	-1.174	0.060	0.76 (0.88, 1.12)	-4.3	0.85 (0.88, 1.12)		-2.7
5	Commitment ..	-0.389	0.057	0.95 (0.88, 1.12)	-0.9	0.95 (0.89, 1.11)		-0.9
6	Run for som..	0.067	0.055	1.03 (0.88, 1.12)	0.6	1.02 (0.89, 1.11)		0.3
7	Achievable ..	-0.462	0.058	0.84 (0.88, 1.12)	-2.7	0.86 (0.89, 1.11)		-2.6
8	Principals ..	-0.165	0.057	0.91 (0.88, 1.12)	-1.5	0.94 (0.89, 1.11)		-1.1
9	Parents sup..	0.275	0.056	1.07 (0.88, 1.12)	1.2	1.07 (0.89, 1.11)		1.3
10	Student mot..	0.550*	0.170	1.06 (0.88, 1.12)	1.0	1.06 (0.89, 1.11)		1.1

An asterisk next to a parameter estimate indicates that it is constrained
 Separation Reliability = 0.993
 Chi-square test of parameter equality = 1199.35, df = 9, Sig Level = 0.000
 ^ Quick standard errors have been used

TERM 2: item*step

VARIABLES		UNWEIGHTED FIT			WEIGHTED FIT				
item	step	ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T
1	Curriculum ..	0		2.03 (0.88, 1.12)		13.3	1.18 (0.89, 1.11)		3.0
1	Curriculum ..	1	-1.129	0.090	0.99 (0.88, 1.12)	-0.1	1.00 (0.95, 1.03)		0.0
1	Curriculum ..	2	1.129*		0.80 (0.88, 1.12)	-3.5	0.95 (0.89, 1.11)		-0.9
2	Not Until E..	0		2.25 (0.88, 1.12)		15.4	1.40 (0.90, 1.10)		7.1
2	Not Until E..	1	-0.626	0.093	1.04 (0.88, 1.12)	0.7	1.04 (0.94, 1.06)		1.3
2	Not Until E..	2	0.626*		1.08 (0.88, 1.12)	1.2	1.08 (0.89, 1.11)		1.4
3	Financial R..	0		1.39 (0.88, 1.12)		5.8	0.98 (0.78, 1.22)		-0.1
3	Financial R..	1	-1.184	0.096	0.80 (0.88, 1.12)	-3.5	0.87 (0.93, 1.07)		-3.9
3	Financial R..	2	1.184*		0.73 (0.88, 1.12)	-5.0	0.83 (0.91, 1.09)		-3.9
4	Staff Commi..	0		2.56 (0.88, 1.12)		18.5	0.91 (0.66, 1.34)		-0.5
4	Staff Commi..	1	-1.463	0.099	0.73 (0.88, 1.12)	-5.0	0.84 (0.92, 1.08)		-4.4
4	Staff Commi..	2	1.463*		0.73 (0.88, 1.12)	-5.0	0.83 (0.91, 1.09)		-4.0
5	Commitment ..	0		3.18 (0.88, 1.12)		23.6	1.06 (0.80, 1.20)		0.6
5	Commitment ..	1	-1.188	0.094	0.79 (0.88, 1.12)	-3.8	0.85 (0.94, 1.06)		-5.0
5	Commitment ..	2	1.188*		0.78 (0.88, 1.12)	-4.0	0.85 (0.91, 1.09)		-3.5
6	Run for som..	0		2.08 (0.88, 1.12)		13.9	1.11 (0.85, 1.15)		1.4
6	Run for som..	1	-1.018	0.092	0.89 (0.88, 1.12)	-1.8	0.93 (0.95, 1.05)		-2.7
6	Run for som..	2	1.018*		1.00 (0.88, 1.12)	0.0	0.91 (0.90, 1.10)		-2.0
7	Achievable ..	0		3.54 (0.88, 1.12)		26.2	0.97 (0.77, 1.23)		-0.2
7	Achievable ..	1	-1.457	0.094	0.80 (0.88, 1.12)	-3.6	0.85 (0.94, 1.06)		-4.7
7	Achievable ..	2	1.457*		0.75 (0.88, 1.12)	-4.6	0.82 (0.91, 1.09)		-4.2
8	Principals ..	0		1.53 (0.88, 1.12)		7.6	1.05 (0.81, 1.19)		0.5
8	Principals ..	1	-1.395	0.093	0.86 (0.88, 1.12)	-2.4	0.89 (0.94, 1.06)		-3.8
8	Principals ..	2	1.395*		0.79 (0.88, 1.12)	-3.8	0.86 (0.91, 1.09)		-3.1
9	Parents sup..	0		1.67 (0.88, 1.12)		9.3	1.14 (0.86, 1.14)		1.9
9	Parents sup..	1	-1.143	0.091	0.96 (0.88, 1.12)	-0.7	0.97 (0.95, 1.05)		-1.2
9	Parents sup..	2	1.143*		0.91 (0.88, 1.12)	-1.5	0.96 (0.90, 1.10)		-0.8
10	Student mot..	0		1.90 (0.88, 1.12)		11.9	1.08 (0.86, 1.14)		1.1
10	Student mot..	1	-1.498	0.091	0.97 (0.88, 1.12)	-0.5	0.97 (0.95, 1.05)		-1.0
10	Student mot..	2	1.498*		0.97 (0.88, 1.12)	-0.5	0.98 (0.89, 1.11)		-0.4

An asterisk next to a parameter estimate indicates that it is constrained
 ^ Quick standard errors have been used

Item 2 fits badly.

Fit looks reasonable as compared to the fit for the rating scale model.

Figure 13 Response Model Parameter Estimates for the Partial Credit Model

SUMMARY

In this tutorial, ConQuest has been used to fit partial credit and rating scale models. Some key points covered were:

- The `codes` statement can be used to provide a list of valid codes.
- The `recode` statement is used to change the codes that are given in the response block (defined in the `format` statement) for the data file.
- The number of response categories modelled by ConQuest for each item is the number of unique codes (after recoding) for that item.
- Response categories and item scores are *not* the same thing.
- The `model` statement can be used to fit different models to the same data.
- The deviance statistic can be used to choose between models.

REFERENCES

- Adams, R. J. 2006. Reliability as a measurement design effect. *Studies in Educational Evaluation*, 31,162-172.
- Adams, R. J., Doig, B. A., and Rosier, M. 1991. *Science Learning in Victorian Schools*. Hawthorn, Victoria: Australian Council for Educational Research.
- Andrich, D. A. 1978. A rating formulation for ordered response categories. *Psychometrika*, 43, 561-73.
- Masters, G. N. 1982. A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-74
- Wright, B. D., and Masters, G. N. 1982. *Rating Scale Analysis: Rasch Measurement*. Chicago: MESA Press
- 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.
- Zammit, S. A. 1997. English and home background languages in Australian primary schools. In *The Bilingual Interface Project Report*, P. McKay (ed.), pp. 111-46. Canberra: Department of Employment, Education and Training: