Mathematics Education Research Journal 2006, Vol. 18, No. 2, 56-76

Easier Analysis and Better Reporting: Modelling Ordinal Data in Mathematics Education Research

Brian Doig and Susie Groves
Deakin University

This paper presents an examination of the use of Rasch modelling in a major research project, *Improving Middle Years Mathematics and Science* (IMYMS). The project has generated both qualitative and quantitative data, with much of the qualitative data being ordinal in nature. Reporting the results of analyses for a range of audiences necessitates careful, well-designed report formats. Some useful new report formats based on Rasch modelling—the Modified Variable Map, the Ordinal Map, the Threshold Map, and the Annotated Ordinal Map—are illustrated using data from the IMYMS project. The Rasch analysis and the derived reporting formats avoid the pitfalls that exist when working with ordinal data and provide insights into the respondents' views about their experiences in schools unavailable by other approaches.

A basic requirement for any research project is the presentation of comprehensible valid and reliable results. While traditional forms of analysis can meet this requirement, other methods may be more efficacious. In this paper we present a case for the use of Rasch analysis as an approach that enables the construction of reports suitable for a range of stakeholders. We use data collected in the *Improving Middle Years Mathematics and Science: The role of subject cultures in school and teacher change* (IMYMS) project to support our case.

The IMYMS project involves four clusters of schools from urban and rural regions of Victoria to investigate the role of mathematics and science knowledge and subject cultures in mediating change processes in the middle years of schooling. In all there are five secondary and twenty-eight primary schools involved.

As is the case in many other educational research projects, the IMYMS project has provided a wealth of qualitative and quantitative data. In particular, the project researchers have used several survey instruments, and collected several sets of ordinal data. While this type of data is common in educational research, the reporting of these data often appears to ignore the mathematical properties of ordinal data, calling into question the outcomes of the research itself. It is our intention in this paper to demonstrate that the analysis of ordinal data collected using surveys and structured interviews is best achieved, through the use of Rasch modelling, and that this also provides better reporting formats.

In the following examples from the IMYMS project, we report the raw ordinal data in appropriate forms, and also in transformed form as interval data using Rasch (1960) methods and a derivative, the Masters Partial Credit Model (Masters, 1988). We also present four new Rasch-based reporting formats.

Background

The IMYMS project has its roots in the *Science in Schools* (SiS) research project, which developed a successful strategy for improving the teaching and learning of science based on two major aspects: the *SiS Components*, a framework for describing effective teaching and learning in science, and the *SiS Strategy*, a strategic process for planning and implementing change (see, Gough & Tytler, 2001). IMYMS is based on auditing the teaching of mathematics and science in each school to inform the development of school and cluster action plans. The major foci of the audit are teacher practice and beliefs, and student perceptions and learning preferences.

Based on reviews of the literature on effective teaching (Doig, 2001; 2003) and a series of interviews with fifteen effective teachers of middle years mathematics (Tytler, Waldrip, & Griffiths, 2004), the IMYMS project team extended the SiS Components to produce the *IMYMS Components of Effective Teaching and Learning* (Figure 1) to describe effective teaching and learning in mathematics and science.

1. *The learning environment promotes a culture of value and respect.*
1.1 The teacher builds positive relationships through knowing and valuing each student.
1.2 The learning environment is characterised by a sense of common purpose and collaborative inquiry.
1.3 The learning environment provides a safe place for students to take risks with their learning.
1.4 Persistence and effort are valued and lead to a sense of accomplishment.
2. *Students are encouraged to be independent and self-motivated learners.*
2.1 Students are encouraged and supported to take responsibility for their learning.
2.2 Students are encouraged to reflect on their learning.
3. *Students are challenged to extend their understandings.*
3.1 Subject matter is conceptually complex and intriguing, but accessible.
3.2 Tasks challenge students to explore, question and reflect on key ideas.
3.3 The teacher clearly signals high expectations for each student.
4. *Students are supported to develop meaningful understandings.*
4.1 Teaching strategies explore and build on students' current understandings.
4.2 Individual students' learning needs are monitored and addressed.
4.3 Students are supported to make connections between key ideas.
4.4 Teaching sequences promote sustained learning that builds over time.
4.5 Learning sequences involve an interweaving of the concrete and the abstract/conceptual.
5. *Students are encouraged to see themselves as mathematical and scientific thinkers.*
5.1 Students are explicitly supported to engage with the processes of investigation and problem solving.
5.2 Students engage in mathematical/scientific reasoning and argumentation.
6. *Mathematics and science content is linked with students' lives and interests.*
7. *Assessment is an integral part of teaching and learning.*
7.1 Learners receive feedback to support further learning.
7.2 Assessment practices reflect all aspects of the learning program.
7.3 Assessment criteria are made explicit.
8. *Learning connects strongly with communities and practice beyond the classroom.*
8.1 The learning program provides opportunities to connect with local and broader communities.
8.2 Learners engage with a rich, contemporary view of mathematics and science knowledge and practice.
9. *Learning technologies are used to enhance student learning.*

Figure 1. The IMYMS Components of Effective Teaching and Learning.

In particular, the extension of the SiS components to the *IMYMS Components of Effective Teaching and Learning* involved a number of distinct types of changes. Some of the SiS components were regarded as being equally applicable to mathematics, requiring only minor changes in wording (e.g., sub-component 2.1). Other changes, however, reflected the middle years focus of the project (e.g., sub-component 1.1); the literature review on effective teaching (e.g., sub-component 3.3); the teacher interviews (e.g., sub-component 1.4); and our previous research (e.g., sub-component 3.1, Groves & Doig, 2002).

While understanding teachers' practice is critical to understanding classroom "life", student perceptions of what they are undergoing and their attitudes to their teacher's practices are also critical. While students do not know, necessarily, why they are required to learn or practise particular aspects of mathematics and science, they nevertheless have perspectives and opinions that affect classroom "life". In other words, their perceptions do have an impact on the classroom environment and, importantly, on their learning. As van den Heuvel-Panhuizen (2005, p. 23) points out, "children give us new lenses for understanding mathematics classrooms and new ideas for improving mathematics education".

In order to provide a broader, more complete understanding of the learning environment, students in the project classrooms were surveyed with respect to their perceptions of their learning environment, as well as their views about particular pedagogical practices. The complete student survey consisted of 36 statements focussing on students' perceptions of classroom practice and their attitudes to mathematics and science, together with 24 statements related to their learning preferences. While both aspects of the student survey are important, it is the first aspect that will be discussed here, referred to as "the IMYMS student perceptions survey".

The IMYMS student perceptions survey comprised 36 statements focussed on features of the mathematics (or science) classroom environments in which the students found themselves and their attitudes to mathematics (or science). Of the 36 statements, 27 were designed to correspond to the nine IMYMS Components (three per component), with the remaining nine corresponding to three aspects of student attitudes—namely enjoyment of, aspirations for, and value to their future of, mathematics and science (again three per aspect). The survey was in a Likert-type format (Likert, 1932), with students responding by selecting one of four response categories ranging from *strongly disagree* to *strongly agree*. The complete set of student perception survey statements for mathematics can be found in Figure 5, where the numbering of statements is in the original survey order, but with statements displayed in their Component Mapping aspects of student attitudes groupings.

The IMYMS student survey was administered twice during the life of the project. The data analysed in this paper are from the initial administration only, where over 1600 students in Year 5 through to Year 10 (Primary $N = 731$; Secondary $N = 892$) responded to the survey for mathematics.

Analysing the IMYMS Student Perceptions Data

The IMYMS student perceptions survey is a rating scale, and therefore the data produced are ordinal. If we take the usual course of action in educational research and compute, for example, the means for the raw ordinal data for the three statements focussed on Component 1 "The learning environment promotes a culture of value and respect", we obtain the following:

- Q 1 "It is OK to say what I think in my maths class" (M = 3.1);
- Q 13 "We are encouraged to respect each other's ideas in my maths class" (M = 3.2); and
- Q 25 "My teacher values my work and ideas in maths" (M = 2.8).

What are we to make of this? At first glance we can see that, apparently, the average student response is in the *agree* category for Q 1 and Q 13, and in the *disagree* category for Q 25. How do we interpret these numbers, particularly the decimal portion? Is a mean of 3.1 for Q 1 better than a mean of 3.2 for Q 13? And what are we to make of the fact that using this approach results in an overall mean of 2.8 for all students across the entire survey? Does this indicate that students have responded mainly in category 2?

Of course, this point is moot, for as Siegal (1956) has pointed out, the flaw is in the treatment of the ordinal data at the outset. The manual for the well-known analytic tool, the Statistical Package for the Social Sciences (SPSS) (Norusis, 1990), puts it this way: "Ordering is the sole mathematical property applicable to ordinal measurements, and the use of numeric values does not imply that any other property of numbers is applicable" (p. 96).

Siegal (1956), however, is much more adamant in his presentation of the issue surrounding the analysis of rating scales and the ordinal data that they produce when he states:

> The statistic most appropriate for describing the central tendency of scores in an ordinal scale is the median, since the median is not affected by changes of any scores which are above or below it as long as the number of scores above and below remains the same. … At the risk of being excessively repetitious, the writer wishes to emphasise here that parametric statistical tests, which use means and standard deviations (i.e., which require the operations of arithmetic on the original scores), ought not to be used with data in an ordinal scale. … means and standard deviations found on the scores themselves are in error to the extent that the successive intervals (distances between classes) on the scale are not equal. … When parametric techniques of statistical inference are used with such data, any decisions about hypotheses are doubtful. Probability statements derived from the application of parametric statistical tests to ordinal data are in error to the extent that the structure of the method of collecting the data is not isomorphic to arithmetic. Inasmuch as most of the measurements made by behavioural scientists culminate in ordinal scales…this point deserves strong emphasis. (pp. 23-26)

Similar warnings can be found elsewhere in the social science research literature (e.g., Thorkildsen, 2005).

As Siegal states, the allowable operations on the ordinal data resulting from a survey such as the IMYMS student perceptions survey may be reported for every statement as:

- the median response to each category;
- the proportion of responses in each category; or
- transformed data on an interval scale.

The results of calculating and graphing the median response for each of the statements of the IMYMS student perceptions survey are shown in Figure 2.

Clearly a report offering the medians provides a minimal amount of useful information. For example, in this case it would appear that student responses to Q 8 "In maths we do things that interest me" and Q 18 "The maths we do is connected to things I am interested in outside school" are not as strongly endorsed as a statements such as Q 23 "I really want to do well in maths", which was the most strongly endorsed statement on the survey. Most other statements appear to be equally well endorsed.

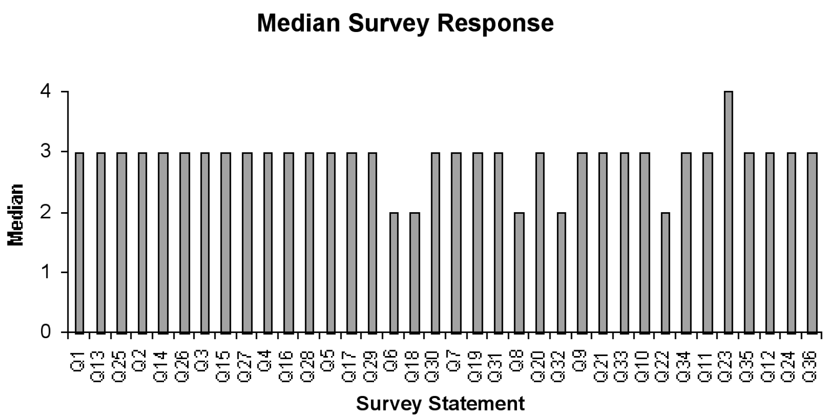

Figure 2. Median responses to the student perceptions survey statements.

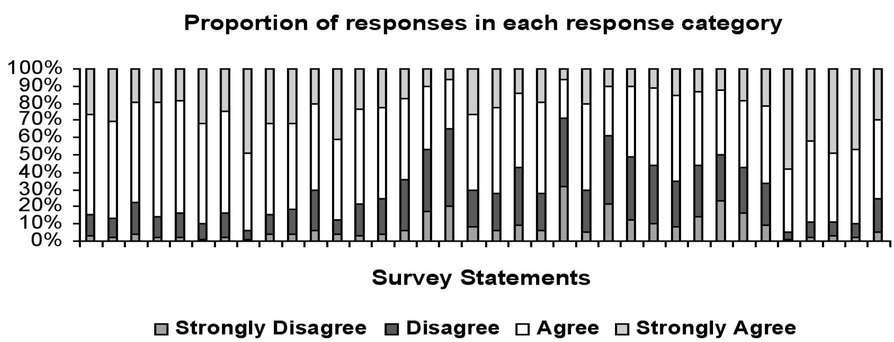

Figure 3. Proportions of responses in each of four response categories.

The major feature of the students' responses missing from the median approach is any indication of the distribution of the responses. Further, there is no way in which a particular student's response pattern can be discerned from this form of summary information. In our view, this is not a useful strategy for reporting these data.

The second alternative for reporting these data, the proportion of responses in each category for every statement, appears in Figure 3, where the IMYMS student perceptions survey statements are presented in the same order as in the median report in Figure 2.

The display of proportion of responses in each category, from *strongly disagree* at the bottom to *strongly agree* at the top of each column, provides a clear picture of the pattern of endorsement of the survey statements. For example, it appears that the statement in Q 23 "I really want to do well in maths" is most strongly endorsed, with more than half the students responding that they *strongly agree*. This statement is closely followed by Q 15 "My teacher expects everyone to do their best in maths", with about half the students responding *strongly agree*.

At the other end of the scale, the statement in Q 8 "In my maths class we work on projects outside school or have people come to talk to us" is strongly rejected as occurring in their class by about thirty percent of respondents, and rejected by a further forty percent of students. Thus, from the graph of response proportions, we can see that about three-quarters of students believe that they do not engage in projects or have guest speakers in their mathematics classes.

Other inferences can be drawn from the graph of the proportions of responses since the survey statements have been grouped in such a way that those that address the same IMYMS Component are together. An example is Component 2, "Students are encouraged to be independent and self-motivated learners", that is addressed by the IMYMS student perceptions survey statements:

- Q 2 "In my maths class I am expected to make decisions about how I do my work";
- Q 14 "My teacher expects me to think about how well I understand things"; and
- Q 26 "In my maths class we are encouraged to work things out for ourselves".

The bar graph in Figure 3 shows that responses to these three statements are positive (either *agree* or *strongly agree*) for the majority of students. This indicates that these students perceive themselves as being encouraged by their teachers to become independent learners. However, while the proportions of particular responses are informative, and certainly provide much more useful information than the record of the median response shown in Figure 2 above, they do not provide any information on individual students or even about sub-groups of students. Unfortunately, as described above, we have now exhausted the possibilities (medians and proportions) available directly from the raw data, although, of course, there are non-parametric approaches that can be used to provide information on various aspects of these data.

As noted above, an alternative to using the raw ordinal data is to transform the data mathematically, using an order-preserving transformation: that is, a form of transformation that preserves the ranking of the raw data and produces an interval scale, one that allows the operations of ordinary arithmetic such as addition. Thus the transformed data can be used to provide statistics such as means and standard deviations.

In our case, we transformed the raw ordinal data into logits (log odds units) using Masters' Partial Credit Model (Wright & Masters, 1982) and the software Quest (Adams & Khoo, 1996). This approach places student ability and item difficulty on the same scale and allows one to estimate the likelihood of a student of a particular ability correctly answering an item of a particular difficulty. In the present case, where a Partial Credit analysis has been used on ordinal survey data, this means the interval scale represents both the degree of endorsement needed to respond in a particular category (the category difficulty) and the degree to which a student agrees with the survey statements (the student perception). As a direct consequence, the distances on the resulting scale have substantive meaning—a feature that we exploit to develop better, more efficacious reporting alternatives.

New Reporting Formats

In this section, four new reporting formats, derived from the transformation of ordinal data onto an interval scale using the Masters' Partial Credit Model, are described.

The Modified Variable Map

Typically output from a Rasch analysis, using software such as Quest, is represented on a *variable map*, where the distribution of students at each ability level is shown on the left of the scale, and the items are located at their difficulty level on the right hand side. Harder items and higher ability are located towards the top of the scale. Thus, the map describes graphically the relationship between person ability estimates and item difficulty estimates in a concise manner.

In order to facilitate an understanding of the IMYMS student perceptions survey data, a Modified Variable Map was constructed. In this case, as remarked above, "ability level" corresponds to students' overall endorsement, while "item difficulty" corresponds to the "difficulty" of endorsing a particular category of response. Response category thresholds were estimated using the Partial Credit Model on all student responses and these category values were then anchored. Using the anchored threshold values, both primary and secondary ability estimates were calibrated, separately, again using the Partial Credit Model. The anchoring of category estimates is an equating procedure that allows the distribution of scores of different groups to be placed on the same scale (see, Bond & Fox, 2001; Kolen, 1999, for details of these procedures). Therefore, the two variable maps produced by these two calibrations can be arranged, manually, to produce a Modified Variable Map showing distributions of both groups of students.

Figure 4 shows the Modified Variable Map for the IMYMS student perceptions survey data. In this Modified Variable Map, the distribution of primary students is on the left-hand side, and the secondary students' distribution is on the right. Students with greater agreement with, or endorsement of, the survey statements are at the higher end of the scale, and those showing lesser agreement are at the lower part of the scale. The scale is in logits, the units of the interval scale produced by the Partial Credit transformation, and has both positive and negative values.

In this Modified Variable Map, the IMYMS student perceptions survey statements are labelled as x.y where x indicates the survey statement number, and y indicates the category threshold. Thus, 10.4 is the point on the scale (the

```
     Primary students  (N = 730)                       Secondary students  (N = 891)

                            X |                                 |
 5.0                          |                                 |
                              |                                 |
                            X |                                 |
                              |                                 |
                              |                                 |
                              |                                 |
                            X |                                 |
                            X |                                 |
 4.0                          |                                 |
                           XX |                                 |
                           XX |                                 |
                           XX |                                 |
                          XXX   |  8.4  18.4                    |
                          XXX |                                 |
                         XXXX |                                 |  XX
                        XXXXX |  9.4                            | X
 3.0                    XXXXX |  6.4  21.4  32.4                |
                       XXXXXX |   22.4                          | X
                         XXXX |   10.4  19.4                    | X
                    XXXXXXXXX |   33.4                          | XXX
                      XXXXXXX |    2.4  14.4  29.4              | X
                       XXXXXX |   25.4  31.4  34.4              | XX
             XXXXXXXXXXXXXXXX |   16.4  20.4                    | XX
                    XXXXXXXXX |    3.4   5.4    7.4  11.4       | XX
 2.0     XXXXXXXXXXXXXXXXXXXX |    1.4  17.4                    | XXX
                     XXXXXXXX |   30.4                          | XXXX
     X XXXXXXXXXXXXXXXXXXXXX  |    8.3  13.4  26.4              | XXXXXX
                  XXXXX XXXXX |    4.4  27.4  36.4              | XX
           XXXXXXXXXXXXXXXXXX |                                 | XXXXXX
            X XXXXXXXXXXXXXXXX |  18.3                          | XXXXXXXX
                 XXXXXXXXXXXX |                                 | XXXXX
         XXXXXXXXXXXXXXXXXX XX | 28.4  32.3  35.4               | XXXXXXXXX
 1.0      XXXXXXXXXXXXXXXXXXX  |                                | XXXXXXXXXXXXXXX
              XXXXXXXXXXXXXXX  |  6.3  24.4                     | XXXXXXXXXXX
            XXXXXXXXXXXXXXX XX | 12.4  15.4   22.3              | XXXXXXXXXXXXXXX
            XXXXXXXXXXXXXXXXX  |  9.3                           | XXXXXXXXXXXXXXXXX
                 XXXXXXXXXXXX  | 10.3  21.3                     | XXXXXXXXXXXX
             XXXXXXXXXXXXXXXX  | 19.3  23.4  34.3               | XXXXXXXXXXXXXXXXXX
                   XXXXXXXXXX  |                                | XXXXXXXXXXXXXXXX
                      XXXXXXX  | 29.3                           | XXXXXXXXXXXXXXX
 0.0                  XXXXXXX  |  8.2  11.3  33.3               | XXXXXXXXXXXXXXX
                         XXXX  | 30.3                           | XXXXXXXXXXXXX XXXX
                    XXXXXXXXX  |  7.3  16.3  20.3   31.3        | XXXXXXXXXXXXXXXX
                       XXXXXX | 22.2  36.3                      | XXXXXXXXX
                         X  XX | 17.3  25.3                     | XXXXXXXXX
                        XXXXX |  5.3  18.2  32.2                | XXXXXXXXXXXXXXX
                            X |  4.3                            | XXXXXXXX
                            X |  6.2  34.2                      | XXXXXX
-1.0                        X |  1.3   3.3  10.2   14.3         | XXXXX
                           XX |  2.3  1 2.3  13.3  28.3         | XXXX
                            X |  9.2  27.3                      | XXX
                            X |  11.2  2 1.2  24.3  35.3        | X
                            X  | 19.2  26.3  30.2  33.2         | X
                            X |                                 | XX
                              |  23.3  31.2  36.2               | XX
                             X |  7.2  1 5.3  16.2  28.2        | X
-2.0                        X |  4.2  20. 2  27.2  29.2         | X
                            X |  1.2  12.2  25.2                |
                               | 17.2                           |
                              |                                 |
                               |  5.2  23.2  24.2               |
                            X  |  3.2  13.2  35.2               |
                               | 15.2                           |
                               |  2.2  14.2  26.2               |
-3.0                          |                                 |

      Each X is 2 students                                 Each X  is 3 students
```

Figure 4. The IMYMS student perceptions Modified Variable Map.

threshold) at which the most likely response for a student at that point changes from category 3 (*agree*) to category 4 (*strongly agree*). The higher on the logit scale a threshold appears, the more difficult it is for students to endorse that category. So, for example, the fact that 10.4 appears higher on the scale than 33.4 indicates that it is more difficult for a student to endorse category 4 (*strongly agree*) for statement 10 than for statement 33.

The likely student response to a survey statement can be found by aligning the student's degree of endorsement, shown by the response distributions and scale values, with the category interval at that scale point. For example, a student with an endorsement "score" of +2.0 logits aligns with Threshold 4, *strongly agree*, for statement Q 1 (shown on the Modified Variable Map as 1.4). This indicates that this student is equally likely to respond *agree* or *strongly agree* to this statement. However, for statement Q 8, this same student is most likely to respond *agree* (category 3) as the student's position on the scale is between the thresholds 8.3 and 8.4. All other combinations of response and statement can be interpreted similarly.

It can be seen from Figure 4 that the two most difficult statements for students to respond to with *strongly agree* were:

- Q 8 "In my maths class we work on projects outside school or have people come to talk to us"; and
- Q 18 "The maths we do is often connected to things I am interested in outside school".

Both of these statements have Threshold 4 situated at 3.5 on the logit scale.

Similarly, it can be seen that lowest logit score corresponding to Threshold 4 is at about 0.4 on the logit scale, so the easiest statement for students to respond with *strongly agree* is:

- Q 23 "I really want to do well in maths".

By looking at the distribution of primary and secondary respondents on the logit scale in Figure 4, it can also be seen that the primary students' scores are spread over a wider range than the secondary scores, suggesting more homongenous responses from the secondary students than the primary ones. In addition, the primary students, as a group, appear higher on the logit scale than the secondary students, indicating that, as a group, they gave a more positive response to the statements than did the secondary students. These findings will be explored further in later sections of this paper.

However, while the Modified Variable Map provides ready insight into many aspects of the survey data, it does not provide easy access to much of the information it contains. For example, the actual survey statements need to be referenced in order to understand to which statement students are responding. Further, the distances between thresholds are visible in the Modified Variable Map, but not in a way that is easy to make comparisons between these distances; nor can one easily examine the relative positions, and therefore likely responses, of individuals or groups with respect to their raw ordinal scores. This has lead to the development of the Ordinal Map discussed in the next section.

The Ordinal Map

The student perceptions survey data can also be shown visually in the Ordinal Map (Figure 5), derived from a variable map that has been rotated 90 degrees clockwise. Thus the Ordinal Map represents the ordinal data transformed onto an interval (logit) scale. It provides prima facie evidence that the original Likert scale categories formed a merely ordinal level scale. Each student perceptions survey statement lies beside a "bar" that visually represents the entire interval scale, with the boundary between categories (the threshold) indicated by a vertical line in the bar.

The bar shows the four categories of response: the first, left-most, section represents the *strongly disagree* category, while the right-most section represents *strongly agree*. The length of the four sections of each bar indicates the distance on the interval scale between each "step". Since all possible total

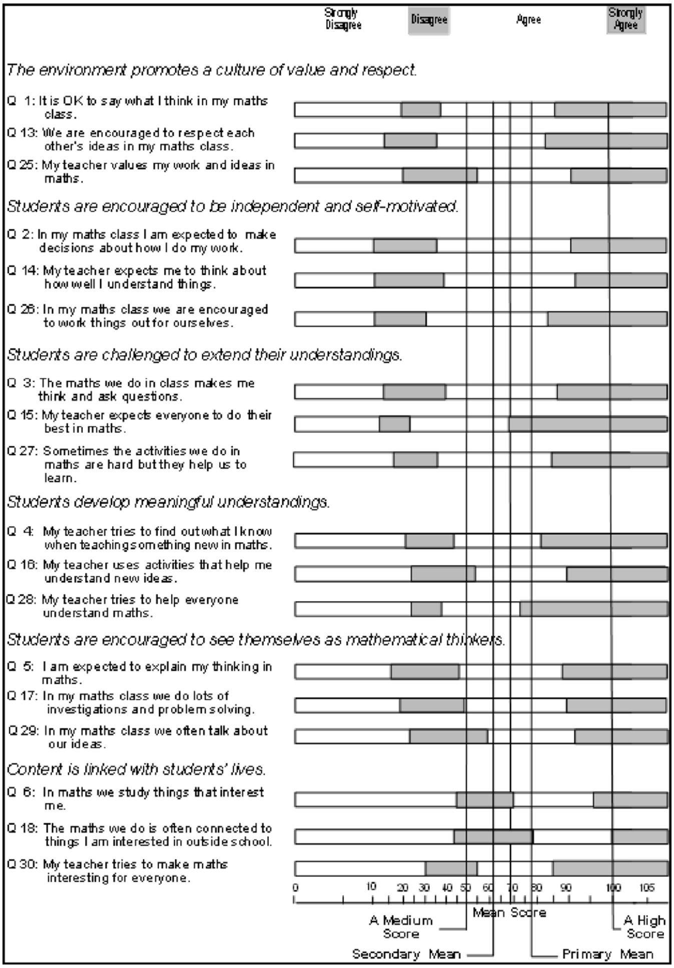

Figure 5a. The IMYMS student perceptions reported on an Ordinal Map (Part A)

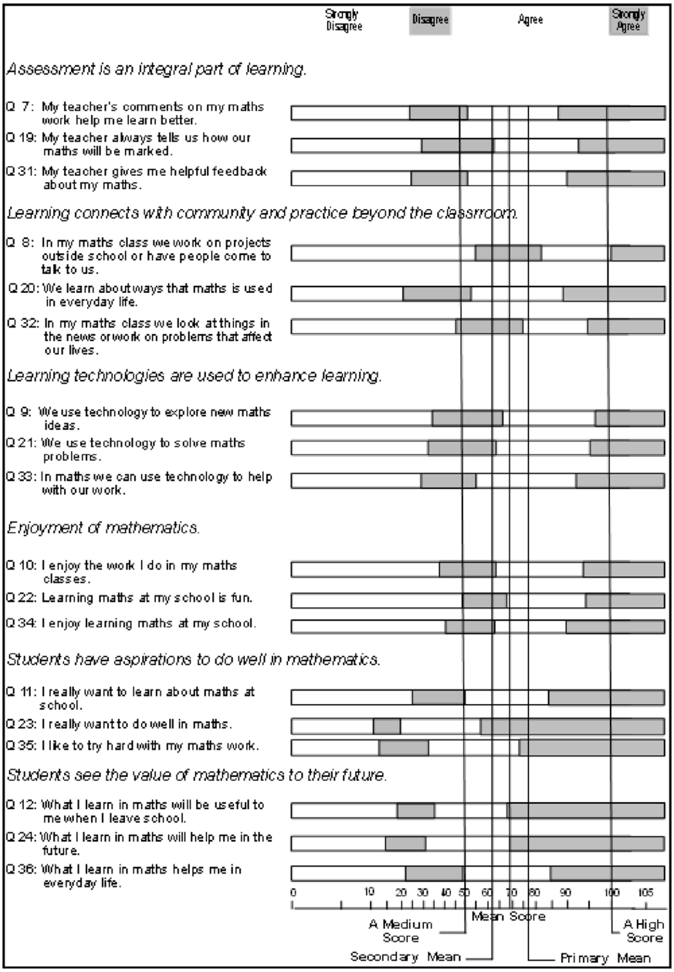

Figure 5b. The IMYMS student perceptions reported on an Ordinal Map (Part B).

raw scores have a scale value on the logit interval scale, it is possible to map raw scores onto the logit scale and vice versa. For convenience, this raw score scale has been used in Figure 5a and 5b. It should be noted that, as this is an interval scale, equal differences in the ordinal raw scores do not appear as equal length intervals on this scale. This provides succinct evidence of the effect of the Rasch measurement transformation from ordinal data to an interval scale.

When using this Ordinal Map, knowing a student's total raw score is sufficient to know the likelihood of a response to any survey statement. This can be accomplished easily by drawing a vertical line at the student's raw score position on the scale at the foot of the Ordinal Map through all the statement bars, as shown in Figure 5a and 5b.

The Ordinal Map also provides an overview of students' survey responses in that the length of each section (category) of a response bar allows one to see statements that are endorsed by students across particular score ranges.

Consider, for example, the IMYMS student perceptions survey statements:

- Q 12 "What I learn in maths will be useful to me when I leave school";
- Q 15 "My teacher expects everyone to do their best in maths";
- Q 23 "I really want to do well in maths"; and
- Q 24 "What I learn in maths will be useful in the future".

For all of these statements, students with total raw scores above 70 are most likely to *strongly agree*.

A similar, but reverse, situation is shown by the responses to the statements:

- Q 8 "In my maths class we work on projects outside school or have people come to talk to us"; and
- Q 22 "Learning maths at my school is fun".

In these two cases, students with total scores below 50 are most likely to *strongly disagree* with these statements, while students with higher total scores are more likely to respond in one of the other more favourable categories.

Alternatively, we can examine the Ordinal Map for details of the likely responses of particular groups of students. It is possible to calculate the mean total raw score for all students, or particular sub-groups. In the example, shown in Figure 5, the mean of all students is 2.6 logits (equivalent to a raw score of 68) and is shown as a vertical line. For the purposes of illustration, four other lines have been inserted: the secondary mean score of 62, the primary mean score of 77, and two arbitrary lines marked at the raw score scales of 50 (a *medium* score) and 100 (a *high* score). Each of these lines can be interpreted in a similar manner.

For example, the *medium* score passes through the *agree* category for Q 1 "It is OK for me to say what I think in my maths class" in Figure 5. This is interpreted as the most likely response to this statement for students with a raw score of 50 being *agree*. Looking further down Figure 5 to the group of statements focussing on Component 6, "Mathematics and science content is linked with students' lives and interests", we see that the most likely response from a student with a raw score of 50 is to *disagree* with all three of the student perception survey statements. All other statements can be similarly investigated for those students with a raw score of 50.

The *high* score line is marked at a raw score of 100, and it can be seen readily that students with this raw score total are most likely to endorse the *strongly agree* category for all student perception survey statements except Q 8 "In my maths class we work on projects outside school or have people come to talk to us" and Q 18 "The maths we do is often connected to things I am interested in outside school". All other total raw scores can be similarly examined for patterns of student responses.

In the IMYMS student perceptions survey, the mean score line for all students (68) indicates the most likely response for the entire group of students. It is clear from scanning Figure 5 that most students respond positively to the majority of student perceptions survey statements. Exceptions to this are statements Q 6 and Q 18 focused on Component 6, "Mathematics and science content is linked with students lives and interests" and statements Q 8 and Q 32 focused on Component 8, "Learning connects strongly with communities and practice beyond the classroom".

The difference between the mean total raw scores of primary and secondary students is easily observed in an Ordinal Map, but more interestingly, similarities and differences in the patterns of likely responses are also revealed. For example, the likely mean response of primary and secondary students is the same for many survey statements, such as all the statements under Component 1, "The learning environment promotes a culture of value and respect," which is an indication that this is a common perception of both groups of students. Other examples of this similarity of perception are found for Component 2, "Students are encouraged to be independent and self-motivated learners" and Component 3, "Students are encouraged to see themselves as mathematical and scientific thinkers". On the other hand, for all of the statements associated with the Attitude, "Enjoyment of mathematics", we find that the primary students are likely to respond *agree*, and the secondary students *disagre*e, indicating a consistent difference in their perceptions.

We argue that graphical displays like the Ordinal Map in Figure 5 give the analyst and the reader immediate access to a wealth of information about both the responses of the group and the individual, based simply on raw scores, and they afford interpretations of the data not possible with other forms of analysis. Further, this format also shows clearly the "difference" in response between any two (or more) students or groups of students. The horizontal distance between any vertical lines drawn through the raw score scale, for an individual or group, makes it very easy to see where individuals and groups agree and differ in their responses to survey statements, and whether these differences can be interpreted as being substantive. Another example of the use and advantages of the Ordinal Map can be found in Doig and Groves (2005).

The Threshold Map

A foundation principle of Rasch measurement is the invariance of item difficulties—that is, item difficulties are the same for all similar groups of respondents. In the example given here, of responses to the IMYMS student perceptions survey, constructing the item thresholds (the boundary between categories of response) on one group of respondents determines these values for other similar groups of respondents. According to Bond and Fox (2001), deviations from this theoretical principle may be deduced as evidence that:

> our measurement expectations have not been sustained in practice: that parallel versions of a test are not parallel, that some test items are biased, that some test formats discriminate against some persons or group of persons, that some items are context dependent, and so on. (p. 59)

Haertel (2004), describing the behaviour of linking items in test equating, points out that changes in common (linking) item values when tests are administered in successive years may be the result of changes in teacher practice and that "it might be that those linking items that appear anomalous from one administration to the next are precisely the ones revealing real reform effects" (p. 3). In the example used here, the IMYMS student perceptions survey, both the primary and secondary students responded to precisely the same questions. Moreover, although the number of items was relatively small (36), the number of respondents was reasonably large (approximately 1600). In this case, we would claim that deviations from item difficulty invariance are evidence of contextual factors affecting the respondents—in particular, the differences found in mathematics classroom practice in primary and secondary schools—rather than sampling variation in the response data (for examples on this issue, Baker, 2001).

In order to examine the information contained in any differences in the thresholds, for the same items and their categories, the item estimates were calculated for primary and secondary students separately. As all items were common to both sets of respondents, they all act as link items, and the equating of the two "tests" (the set of items administered to primary students

```
                 Primary                                  Secondary
  4.0 |
      |
      |
      |                                           8.4
      |                                           6.4  18.4
      |  8.4  18.4                               22.4
      |                                          32.4
      | 19.4                                      9.4  10.4
  3.0 |  9.4                                     21.4
      | 21.4
      |  6.4  32.4                               14.4  25.4  29.4
      | 33.4                                      7.4  16.4  20.4  31.4  33.4
      | 10.4  22.4                               34.4
      |  2.4                                      2.4
      |  5.4  14.4  17.4  29.4                   11.4  30.4
      | 11.4  31.4                                4.4  19.4
  2.0 |  3.4  16.4  20.4  25.4  34.4              1.4   3.4  13.4  36.4
      |  1.4                                      5.4  17.4  27.4
      |  7.4                                     26.4
      |  8.3  26.4                                8.3
      | 30.4                                     18.3  28.4
      | 13.4  36.4                               32.3
      |  4.4  18.3  27.4                         35.4
      | 19.3                                     12.4  24.4
  1.0 | 32.3  35.4                                6.3  15.4
      |                                          22.3
      |                                           9.3
      |  6.3  28.4                               10.3
      |  9.3  15.4  21.3  24.4                   34.3
      | 22.3  23.4                               21.3  23.4  29.3
      | 10.3  12.4  34.3
      | 33.3                                     30.3
  0.0 |  8.2  11.3                               11.3  16.3  33.3
      |                                           8.2  19.3  20.3
      |  7.3  29.3                               31.3
      |  5.3  17.3  20.3  31.3                    7.3  22.2  36.3
      | 16.3  22.2  32.2                         25.3
      |  3.3  18.2  30.3  36.3                    4.3  17.3  32.2
      |  2.3  19.2  27.3                          6.2  18.2
      | 14.3  25.3  34.2                          5.3  34.2
 -1.0 |  1.3   4.3   6.2  10.2                    1.3  10.2  28.3
      | 21.2  26.3                                3.3  12.3  13.3  14.3  27.3
      |  9.2  11.2  13.3  15.3  28.3              9.2
      | 33.2  35.3  12.3                         24.3  35.3
      |  7.2  23.3  24.3  27.2  31.2              2.3  11.2  21.2  30.2
      | 16.2  28.2                               26.3
      |  4.2  15.2  29.2  30.2  36.2             33.2
      | 12.2  17.2                               16.2  28.2
 -2.0 |  5.2  25.2                                7.2  19.2  20.2  23.3  29.2
      |  1.2   3.2  23.2                         31.2  36.2   4.2  25.2
      | 13.2  20.2  35.2                          1.2  12.2  15.3  27.2
      |  2.2  24.2  26.2                         17.2
      | 14.2                                     24.2
      |                                          23.2
      |                                           5.2  13.2  35.2
      |                                           3.2
 -3.0 |
```

Figure 6. The common scale Threshold Map.

and the set administered to secondary students) is achieved by the fact that the two scales constructed, under the principle of invariance, are precisely the same, within error. Thus, this equating procedure produces a common, or reference scale (Beltyukova & Fox, 2004), that enables comparisons of similarities and differences in the data to be observed.

Figure 6, which we are calling a *Threshold Map*, is a form of the more common variable map produced by Rasch analysis software such as Quest. Here, however, the student response distributions have been removed and the item thresholds, constructed from the responses of the two groups of students, have been placed in the centre of the page. The common logit scale is at the left-hand side.

The two sub-sample groups of category thresholds have been separated horizontally to make it easier to identify the thresholds relevant to the two groups of students who responded to the perceptions survey. This separation allows us to isolate similarities and differences caused by differences in the responses of the two groups of students.

As was noted earlier, it can be seen that for both primary and secondary respondents the most difficult statements to respond with *strongly agree* were:

- Q 8 "In my maths class we work on projects outside school or have people come to talk to us"; and
- Q 18 "The maths we do is often connected to things I am interested in outside school".

However, in contrast to the primary students, the secondary respondents found the following two statements amongst the most difficult statements with which to *strongly agree*:

- Q 6 "In maths we study things that interest me"; and
- Q 22 "Learning maths at my school is fun".

In both cases, the thresholds for *strongly agree* for the secondary students were about 0.8 logits higher than for the primary students.

Inspection of the Threshold Map shows that there are many other statements for which the secondary students have a considerably higher threshold (say 0.5 logits or more) for *strongly agree* than the primary students, but only one where the reverse is true, namely,

- Q 19 "My teacher always tells us how our maths will be marked".

Interestingly, all three thresholds for this statement are between 1.0 and 1.3 logits higher for primary students than secondary students.

However, as evident in an earlier version illustrating similarities and differences in primary and secondary mathematics teachers' classroom practices (see, Doig, Groves, Tytler, & Gough, 2005 for this earlier version) there are difficulties associated with such an analysis. These difficulties, which are discussed below, lead to the construction of the Annotated Ordinal Map.

The Annotated Ordinal Map

Although the Threshold Map shows differences in a diagrammatic manner, it is difficult to inspect all the thresholds for a particular statement simultaneously. Moreover, in a survey such as the one reported here, there is the additional problem of identifying statements referring to the same Component or Attitude. A more serious difficulty is that observable differences do not necessarily indicate substantive meaning, as the Threshold Map does not include any detail of measurement errors that sensibly could reduce an apparent "difference" to mere wishful thinking.

As a precaution against making such errors in inference, the difference between threshold ranges for primary and secondary respondents for each threshold was calculated for each statement using the difference between the thresholds when measurement errors were included. So, for example, Table 1 shows the thresholds and error estimates for Q 15 "My teacher expects everyone to do their best in maths".

Table 1
Thresholds and Error Estimates for Primary and Secondary Responses to Q 15

Q15	$T_2–E_2$	T_2	E_2	T_2+E_2	$T_3–E_3$	T_3	E_3	T_3+E_3	$T_4–E_4$	T_4	E_4	T_4+E_4
Prim	-2.1	-1.8	0.3	-1.4	-1.6	-1.3	0.3	-1.0	0.3	0.5	0.2	0.7
Sec	-4.3	-3.8	0.5	-3.3	-2.5	-2.3	0.3	-2.0	0.8	0.9	0.1	1.1

Note. T_n is the logit value for Threshold n. E_n is the error estimate, in logits, for Threshold n

As can be seen in Table 1, for primary students the "upper bound" produced for Threshold 4 when taking into account the error estimate is 0.7 logits, while the corresponding "lower bound" for secondary students is 0.8, suggesting that the difference between the primary and secondary threshold for students stating that they *strongly agree* with this statement is at least 0.1 logits. In other words, secondary students are finding it somewhat harder to *strongly agree* with the statement. Similarly, there is a "gap" in the opposite direction of 0.4 logits (the difference between the primary lower bound of –1.6 and the secondary upper bound of –2.0) for Threshold 3, suggesting that this time it is the primary students who are finding it harder to say that they *agree* with this statement. For Threshold 2, it is again the primary students who find it harder to *disagree* (as opposed to *strongly disagree*) to the statement, with the difference this time being even greater at 1.2 logits (= [–2.1] – [–3.3]). While there appears to be a difference between primary and secondary respondents on all of the thresholds for this statement, the difference ranges from being quite large at Threshold 2 to much smaller at Threshold 4. Similar calculations using the error estimates were carried out for each statement.

Figure 7 shows the Ordinal Map from Figure 5 annotated by an indication of those thresholds where there was a difference between primary and secondary respondents, after taking the error estimates into account. In order to provide a clear visual representation of these differences, they have been divided arbitrarily into three categories—those where the difference is

0.0 to 0.2 logits, those where it is 0.3 to 0.4 logits, and those where the difference is greater than 0.5 logits—and labeled on Figures 7a and 7b using P or S to indicate the group of respondents for whom the thresholds were higher. Using the example above, the "bar" for Q 15 has dark shading and the letter P in the section for *disagree*, to indicate that the primary respondents found it considerably more difficult to endorse the *disagree* category (Threshold 2) for this statement than the secondary respondents, while the portion corresponding to *strongly agree* has the letter S but no shading to indicate that the secondary respondents found it slightly more difficult to *strongly agree* with this statement.

The Annotated Ordinal Map provides a clear visual means of identifying where the thresholds for the two groups of respondents differ most, as well as allowing the reader to see how difficult it was overall for students to reach each threshold, and to see this information in such a way that the different statements corresponding to the same Components and Attitudes can be readily compared.

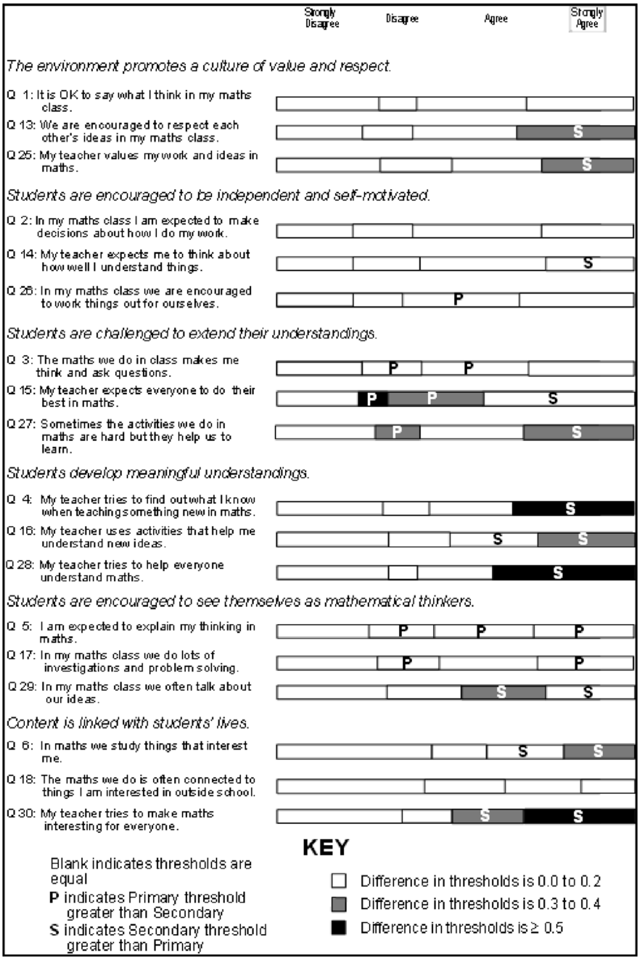

Figure 7a. The IMYMS student perceptions Annotated Ordinal Map (Part A).

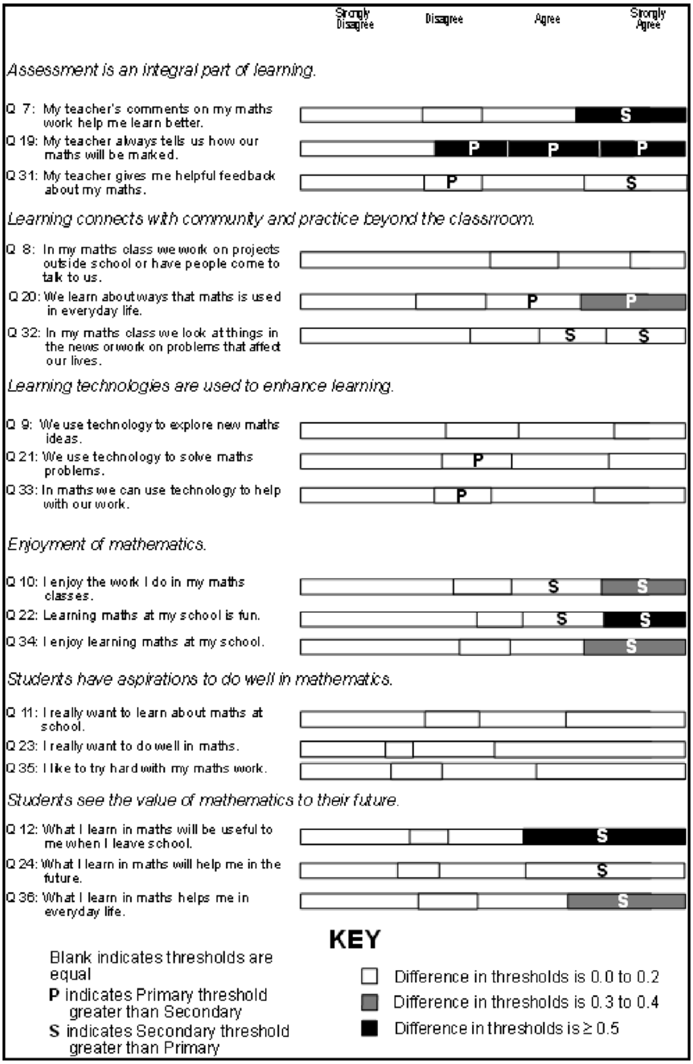

*Figure 7*b. The IMYMS student perceptions Annotated Ordinal Map (Part B).

For example, as has been observed earlier, primary students were generally more positive in their perceptions than the secondary students. This is easily visible in Figures 7a and 7b where there are only four occurrences where the secondary students' thresholds were more than 0.5 logits lower than the primary students' thresholds—all three thresholds for Q 19 "My teacher always tells us how our maths will be marked" and the lowest threshold for Q 15 "My teacher expects everyone to do their best in maths". It is interesting to note that Q 19 is one of the three statements associated with Component 7, "Assessment is an integral part of teaching and learning", and for the other two statements the primary students do not have significantly higher thresholds.

In terms of consistency across statements associated with the same Component or Attitude, the most consistent differences occur for Component 4, "Students are supported to develop meaningful understandings", followed by the Attitude, "Enjoyment of mathematics", and the Attitude, "Students see the value of mathematics to their future", with all three of these having consistently higher thresholds for secondary students for all three statements. This is also true to a lesser extent for Component 6, "Mathematics and science content is linked with students' lives and interests", and Component 1, "The learning environment promotes a culture of value and respect", where in each case, for two of the three statements, secondary students have higher thresholds for *strongly agree* than do primary students, as well as for *agree* for Component 6. The only Components other than Component 7, mentioned above, where there is a real mix in the direction of difference between secondary and primary thresholds are Component 3, "Students are challenged to extend their understandings", Component 5, "Students are encouraged to see themselves and mathematical and scientific thinkers", and, to a lesser extent, Component 8, "Learning connects strongly with communities and practice beyond the classroom."

Conclusions

It is unarguable that it is important to take students' perceptions, or views, into account when planning learning and teaching for them. There appears, however, to be little research evidence of this happening in Middle Years learning and teaching. The *Improving Middle Years Mathematics and Science* student perceptions survey is an attempt to make visible these student viewpoints, and report them in a way that is accessible to teachers and researchers involved in the project.

In particular, we have shown that in the initial administration of the student survey, the two statements most difficult for both primary and secondary students to endorse were those related to mathematics being connected to students' interests outside school and linking with practice beyond the classroom through the use of projects outside school or people coming in to talk to students. Primary, but not secondary students, also found it very difficult to *agree* that their teachers tell them how their mathematics will be marked, while almost all other differences between primary and secondary students were in the direction of secondary students being less positive than primary students. For example, secondary students found it more difficult than primary students to *agree* with all three statements dealing with the support of students in their development of meaningful understandings, and, to a lesser extent, those related to students' enjoyment of mathematics and their seeing the value of mathematics to their future.

This type of analysis and reporting will allow a similar comparison of data from the science survey, as well as a comparison of data from the initial survey with that obtained recently in the second student survey.

As suggested earlier, the use of ordinal data is a common feature of social science, and thus, of educational research as well. However, we recognise the constraints that the use of ordinal data imposes on researchers, and have exposed the limitations of the legitimate operations on these data—the median and simple proportions. Further, we have described aspects of possible legitimate strategies, with a focus on the Rasch model. The examples of analyses of the *Improving Middle Years Mathematics and Science* student perceptions survey data have served to illustrate the usefulness of these approaches in any research endeavour that relies on questionnaires or surveys as its data collection strategy.

From this exploration of Rasch modelling of ordinal data, and the Partial Credit Model in particular, we have shown the usefulness of transforming ordinal data onto an interval scale, and the advantages of such a transformation in terms of possibilities for improved reporting. Further, we have shown how the nexus between students' interval scaled scores and the difficulty of endorsing different survey statements can be used to identify response patterns of students in greater detail than by alternative means.

We have described four newly developed formats of report—the Modified Variable Map, the Ordinal Map, the Threshold Map, and the Annotated Ordinal Map—all of which are ways of reporting ordinal data in an effective and easily comprehended way. We maintain that researchers interested in maximising the returns for their efforts in collecting ordinal data should explore the possibilities afforded by the use of Rasch modelling and these new report formats.

Acknowledgments

Improving Middle Years Mathematics and Science: The role of subject cultures in school and teacher change (IMYMS) is funded by an Australian Research Council Linkage Grant, with Industry Partner the Victorian Department of Education and Training. The Chief Investigators are Russell Tytler, Susie Groves and Annette Gough. School clusters are funded through the Victorian state government's *Schools for Innovation and Excellence* initiative.

References

Adams, R. J., & Khoo, S.-T. (1996). Quest: The interactive test analysis system (Version 2) [Computer Software]. Melbourne: ACER.

Baker, F. B. (2001). *The basics of Item Response Theory* (2nd ed.). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.

Beltyukova, S. A., & Fox, C. M. (2004). Equating student satisfaction measures. *Journal of Applied Measurement, 5*(1), 62—69.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.

Doig, B. (2001). *Summing up: Australian numeracy performances, practices, programs and possibilities*. Melbourne: ACER Press.

Doig, B. (2003). *Quality learning communities: Effective practices in the Middle Years of schooling*. (Report prepared for the Middle Years Pedagogy Research and Development Project). Melbourne: Deakin University Consultancy and Development Unit.

Doig, B., & Groves, S. (2005). Reason for hope: Primary students' mathematics learning environment. In Novotná, J. (Ed.), *Proceedings of the International Symposium Elementary Maths Teaching* (pp. 100–107). Prague, Czech Republic: Charles University.

Doig, B., Groves, S., Tytler, R., & Gough, A. (2005). Primary and secondary mathematics practice: How different is it? In P. Clarkson, A. Downton, D. Gronn, M. Horne, A. McDonough, R. Pierce & A. Roche (Eds.), *Building connections: Theory, research and practice* (Proceedings of the 28th annual conference of the Mathematics Education Research Group of Australasia, Melbourne, Vol. 1, pp. 305–312). Sydney: MERGA.

Gough, A., & Tytler, R. (2001). Researching effective teaching and learning in science: The Science in Schools research project. In *Crossing borders: New frontiers in education research* (Proceedings of the Australian Association for Research in Education, AARE 2001 International Education Research Conference, pp. 1–4). Fremantle, WA: AARE.

Groves, S., & Doig, B. (2002). Developing conceptual understanding: the role of the task in communities of mathematical inquiry. In A. D. Cockburn & E. Nardi (Eds.), *Proceedings of the 26th conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 25–32). Norwich, UK: PME.

Haertel, E. H. (2004). *The behavior of linking items in test equating*. Los Angeles: CRESST/Stanford University.

Kolen, M. J. (1999). Equating of tests. In G. N. Masters & J. P. Keeves (Eds.), *Advances in measurement in educational research and assessment* (pp. 164-185). Amsterdam: Pergamon.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology* (Number 140).

Masters, G. N. (1988). Partial Credit model. In J. P. Keeves (Ed.), *Educational research, methodology, and measurement* (pp. 292-297). Oxford, UK: Pergamon.

Norusis, M. J. (1990). *SPSS base system user's guide*. Chicago: SPSS.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests* (Expanded Edition). Chicago: University of Chicago Press.

Siegal, S. (1956). *Non-parametric statistics for the behavioral sciences* (International Student Edition). New York: McGraw-Hill.

Thorkildsen, T. A. (2005). *Fundamentals of measurement in applied research*. Boston, MA: Pearson Education.

Tytler, R., Waldrip, B., & Griffiths, M. (2004). Windows into practice: Constructing effective science teaching and learning in a school change initiative. *International Journal of Science Education*, *26*(2), 171–194.

van den Heuvel-Panhuizen, M. (2005). Children's perspectives of the mathematics classroom. In J. Novotná (Ed.), *Proceedings of the International Symposium Elementary Maths Teaching* (pp. 23 – 33). Prague, Czech Republic: Charles University.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.

Authors

Susie Groves, Associate Professor, Faculty of Education, Deakin University, 221 Burwood Highway, Burwood VIC 3125. Email: <grovesac@deakin.edu.au>

Brian Doig, Faculty of Education, Deakin University, 221 Burwood Highway, Burwood VIC 3125. Email: <badoig@deakin.edu.au>