Secondary Numeracy Project Knowledge Test Analysis

Andrew Tagg
Maths Technology Ltd.
<andrew@nzmaths.co.nz>

Gill Thomas
Maths Technology Ltd.
<gill@nzmaths.co.nz>

Purpose

While the knowledge domains have been included in the diagnostic interview for the Early Numeracy Project (ENP), Advanced Numeracy Project (ANP), and Intermediate Numeracy Project (INP), a decision was made to use a written test to assess the four knowledge domains for the Secondary Numeracy Project (SNP). This decision was made in an effort to reduce the amount of time required to assess students, which has been identified as a barrier to the sustainability of the numeracy project (Thomas, Tagg, and Ward, 2006). While the strategy domains require students to explain how they work out answers to items, the knowledge domains assess the number knowledge that students can recall without needing to carry out calculations, so a written test taken simultaneously by the whole class is seen to be an appropriate method of assessment. The four knowledge domains assessed are number word sequence, fractions, place value, and basic facts. This chapter reports on the findings of a research project that investigated the robustness of the knowledge test with the goal of informing its future development. These research questions are addressed:

- Do the items within each domain of the SNP knowledge test get progressively more difficult?
- Are the items within each stage of each domain of the test of similar difficulty?

Sample

A random sample of six schools, stratified by decile group, was selected from the 43 schools participating in the SNP in 2005. These schools were invited to submit all of their initial student knowledge record sheets for independent analysis. These tests were completed at the beginning of the project, in Term One 2005. All of the low-decile (282) tests were marked, and similar numbers of the high-decile (350) and medium-decile (368) tests were added to take the sample to a total of 1000. The responses of each student to each item were analysed.

Rasch Analysis

Rasch item analysis allows logit scores to be calculated for both item difficulty and student ability. These scores are calculated using an iterative process that assumes that for any given item there is a probability that any given student will answer correctly; this probability can be expressed in terms of the item's difficulty and the student's ability. Both values are expressed in logits, with a negative score representing lower difficulty and a positive score representing higher difficulty. This method of calculating item difficulty takes into account which other items were answered correctly by each respondent, rather than just the proportion of correct responses. A Rasch analysis of the 1000 tests was carried out using the RASCAL program (Assessment Systems Corporation, 1996).

Figure 1 shows a scatterplot of the difficulty score for each item, with vertical lines separating the four domains. The figure clearly illustrates the increase in difficulty of items within each domain and also shows several points that appear to lie outside the trend for each domain. These will be discussed in the following section.

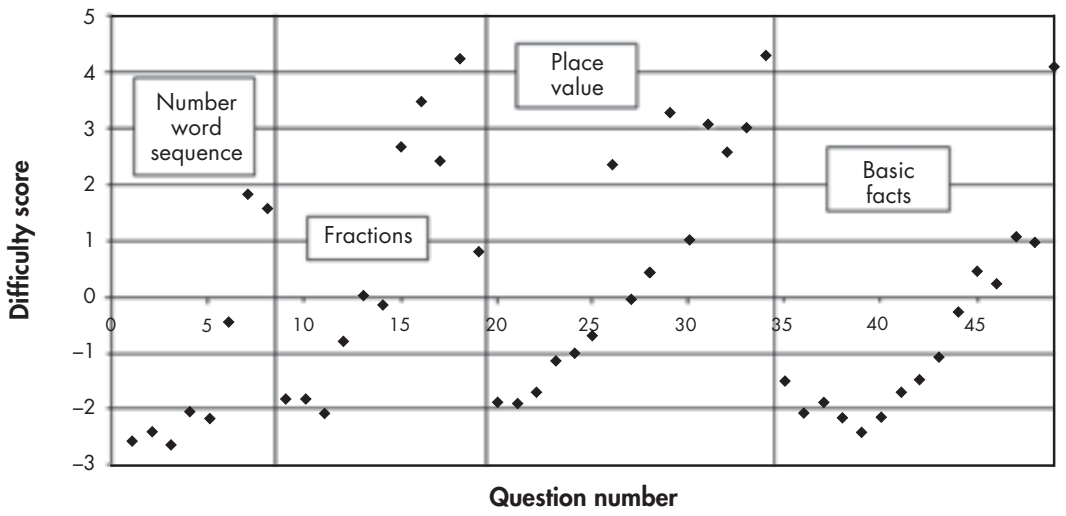

Figure 1. Difficulty scores for SNP knowledge test items

Discussion

The criteria for marking the test indicate that all items within a stage must be answered correctly for a student to be rated as having attained that stage, so it is reasonable to expect that the items within a stage should be relatively similar in difficulty. The item analysis shows some interesting results, with some items revealed as significantly easier than others within the same stage and others seemingly well out of place in terms of difficulty. It should be noted that the items for the test were chosen to be representative of the knowledge that students are thought to have at each of the described stages of the Number Framework. These decisions, while informed by professional experience, were not based on student data. This research represents the first attempt to quantify the results of individual elements of the Numeracy Project Assessment diagnostic tool (NumPA) testing. It is assumed for the following discussion that the intention of the knowledge test is solely to allocate the student with a stage for each knowledge domain and to identify the gaps in their number knowledge. Some of the items that are shown to be "misplaced" may provide additional information on students' knowledge if analysed independently.

Forwards and Backwards Number Word Sequence

It is important that teachers are able to identify the small proportion of secondary students who are not able to accurately state the number before or after a given number because an inability to do so will prevent them from making progress in their development of the most basic strategies for operating with numbers. Figure 2 illustrates the difficulty scores for the eight items in the knowledge test relating to this domain.

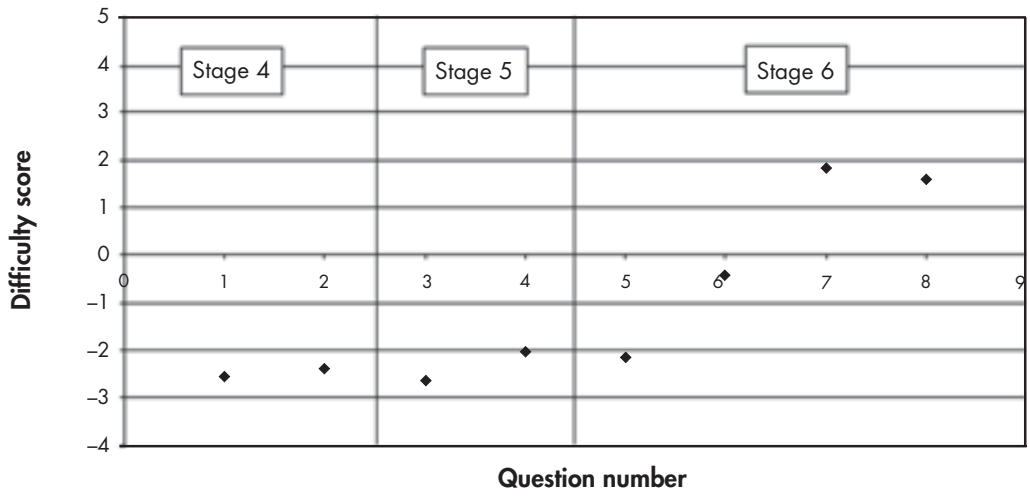

Figure 2. Difficulty scores for SNP knowledge test number word sequence items

Ninety-five percent of students tested answered the first four items in this domain correctly, giving them a rating of at least stage 5. It is the four items (5–8) that are grouped within stage 6 that are of interest; they have item difficulty scores ranging from –2.165 to 1.815. Items 5 and 6 (one more than 3049; one less than 2400) are considerably easier than items 7 and 8 (one more than 989 999; one less than 603 000) and are therefore unlikely to contribute to the outcome of the test because it is reasonably safe to assume that they will be answered correctly by all students who respond correctly to the more difficult items. In fact, a count of the data shows that 399 of the 485 students rated at stage 5 correctly answered items 5 and 6.

Fractions

The eleven items relating to the fractions domain seem to be reasonably well clustered. The difficulty scores of these items are illustrated in Figure 3.

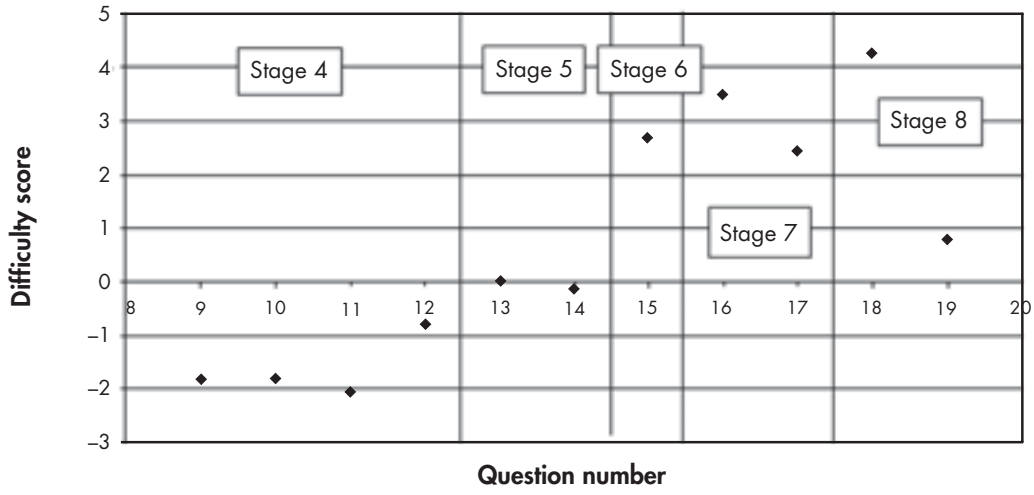

Figure 3. Difficulty scores for SNP knowledge test fractions items

There are three instances in which items in this domain do not fit well into the difficulty grouping. The relevant items discussed below are:

- item 12 Write one-quarter as a fraction.
- item 15 Which fraction is the same as eight-sixths? $\frac{6}{8}$, $\frac{12}{6}$, 1, $\frac{2}{14}$
- item 17 Which fraction is the same as three-quarters? $\frac{2}{6}$, $\frac{6}{8}$, $\frac{8}{12}$, $\frac{9}{12}$, $\frac{3}{9}$
- item 18 Which of $\frac{2}{5}$, $\frac{3}{10}$, $\frac{5}{16}$ is the smallest?
- item 19 Which of $\frac{2}{5}$, $\frac{3}{10}$, $\frac{5}{16}$ is the largest?

Firstly, item 12 appears to be more difficult than the other three items in the same stage (Write one-half, one-sixth, and one-third as a fraction). Secondly, item 15, with a difficulty score of 2.666 despite being at stage 6, is at least as difficult as stage 7's item 17, which has a difficulty score of 2.417. Finally, while item 19, with a difficulty score of 0.792, is placed at stage 8, it is rated as easier than any of the stage 6 or 7 items, which had difficulty scores between 2.417 and 3.49. In fact, 107 of the 127 students rated at stage 7 answered item 19 correctly. Only 6 of the 127 students rated at stage 7 would have been rated at stage 8 if they had answered this item correctly, with the remainder also answering item 18 incorrectly.

Place Value

The 15 place value items are those for which the difficulty scores are least well clustered. These scores are illustrated in Figure 4.

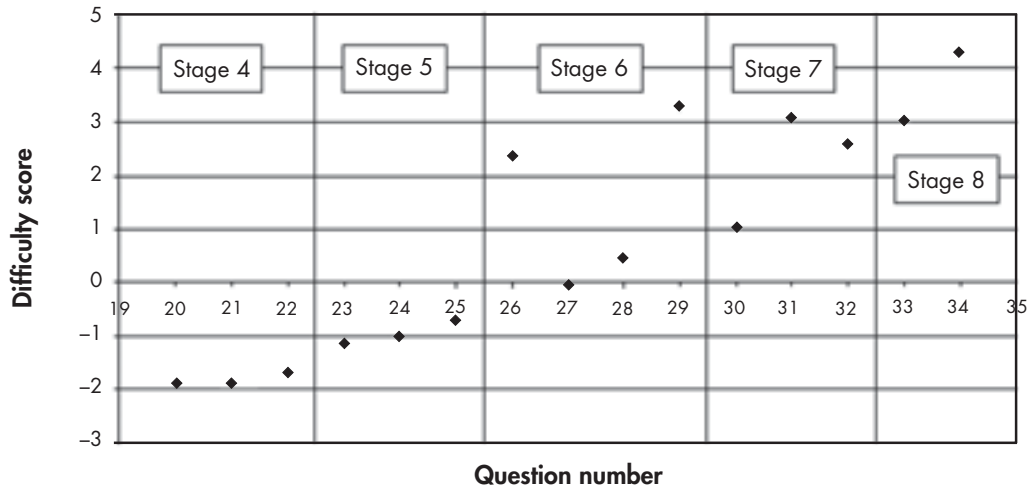

Figure 4. Difficulty scores for SNP knowledge test place value items

While the stage 4 and 5 items do seem to be grouped in terms of their difficulty, several of the other items appear to be misplaced. The relevant items discussed below are:

- item 29 Three-tenths more than 4.8.
- item 26 How many $1,000 notes in $2,408,000?
- item 30 Which of 0.478, 0.8, 0.39 is the largest?

In terms of misplaced items, item 29 in particular, which is placed at stage 6, has a difficulty score of 3.295, which is higher than any of the stage 7 items (difficulty scores between 1.019 and

3.078). Item 26, with a difficulty score of 2.366, also appears more suited to placement in stage 7. If these two items were placed in stage 7 instead of stage 6, 454 of the 578 students rated at stage 5 would have been rated at stage 6. If only item 29 were moved, 193 students would have been rated a stage higher. Conversely, item 30 appears too easy for stage 7, with a difficulty score of only 1.019, compared with 3.078 and 2.582 for the other two items in that stage. In fact, only two of the 89 students who were rated at stage 6 would have been rated at stage 7 if that item had been excluded.

Basic Facts

Figure 5 illustrates the difficulty scores of the 15 basic facts items.

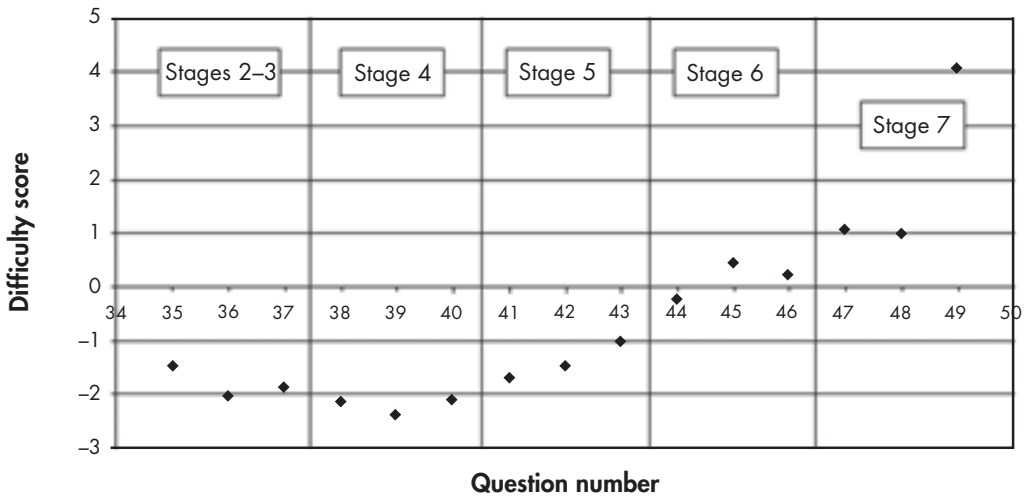

Figure 5. Difficulty scores for SNP knowledge test basic facts items

The difficulty scores of these items are particularly interesting in that they show that the three items defining stage 4 (items 38–40; difficulty scores between –2.122 and –2.404) are all easier than all of the three items defining stages 2–3 (items 35–37; difficulty scores between –1.492 and –2.081). This strongly suggests that, at least for year 9 students, addition with tens and doubles is "easier" than addition with and within 5 and within 10. In fact, of the 73 students rated at stages 0–1 because they answered at least one of items 35–37 incorrectly, 51 answered all of items 38–40 correctly. The items defining stages 5–6 have well-defined clusters of difficulty scores, with the exception of item 49 (45 divided by 6), which, with a difficulty score of 4.085, is significantly more difficult than any of the other basic facts items. The next most difficult has a difficulty score of 1.068. If item 49 were excluded, 325 of the 434 students rated at stage 6 would have been rated at stage 7.

Comparison of Domains

The criteria for rating students' knowledge stages indicate that students should be rated at a stage if they correctly answer all of the items at that stage for that domain. Table 1 shows the percentages of students rated at each stage for each knowledge domain. While there is neither a requirement nor an expectation that students will be operating at the same stage for each domain, the stages are intended to be roughly equivalent.

Table 1
Percentages of Students at Each Stage of the Four Domains

	Domain			
Stage	FNWS	Fractions	Place value	Basic facts
0–1				*7*
2–3	3	12	8	3
4	2	10	13	*7*
5	49	*37*	58	20
6	*47*	21	9	43
7		13	6	20
8		6	*7*	

It is interesting to note the relatively low proportions of students who reached stage 6 or higher on the place value domain (219) compared with the other three knowledge domains (405 or more). This disparity could be corrected by moving item 29 (What number is three-tenths more than 4.8?) from stage 6 of the place value domain to stage 7. The relatively high proportion of students at the lower stages of the fractions domain, while of concern, confirms other research focusing on students' acquisition of fraction knowledge.

Results of Findings

Changes have been made to six items in the SNP knowledge test for 2006 as a result of the release of preliminary findings from the analysis of the knowledge tests.

Items 5 and 6 (one more than 3049; one less than 2400) were made more difficult by increasing the numbers to 30 099 and 24 000 respectively. It is hoped that this will make these items similar in difficulty to items 7 and 8.

The fractions in item 19 (Which of $\frac{2}{5}$, $\frac{3}{10}$, $\frac{5}{16}$ is the largest?) have been changed to $\frac{2}{3}$, $\frac{5}{9}$, and $\frac{7}{10}$ respectively. The basis for this change is an assumption that students were answering the item correctly by applying a false assumption that a larger denominator always implies a smaller fraction. This would have resulted in students identifying $\frac{2}{5}$ as the largest fraction without taking the numerators into account. Interestingly, with the new selection of fractions, students could still answer correctly through a misconception. This time, if they believe that a larger numerator always implies a larger fraction, they will choose $\frac{7}{10}$, the correct answer, without taking the denominators into account.

The size of the numbers in item 26 have been reduced from "How many $1,000 notes are there in $2,408,000?" to "How many $100 notes are there in $26,700?". This change is intended to bring the difficulty of the item more in line with other stage 6 items.

Item 29 (What number is three-tenths more than 4.8?) has been changed to "What number is three-tenths less than 2?" This change is intended to reduce the item's difficulty, and it also, by allowing either "1.7" or "one and seven-tenths" as answers, removes the decimal symbolism from the problem.

The decimals in item 30 (Which of 0.478, 0.8, 0.39 is the largest?) have been changed to 0.76, 0.657, and 0.7 respectively. This change is based on an assumption that students were identifying 0.8 as the largest decimal by identifying it as eight-tenths, while the other two decimals were hundredths and thousandths. Knowing that tenths are bigger than hundredths or thousandths could lead to a correct answer without taking into account how many tenths, hundredths, or thousandths are involved. The change to the decimals used means that this misconception will no longer lead to a correct answer.

Concluding Comment and Key Findings

Rasch analysis of 1000 SNP knowledge tests indicated that:

- the items on the knowledge test were generally ordered from easiest to most difficult within each domain;
- there were several items that were significantly easier or more difficult than those with which they were grouped.

Six changes were made to the knowledge test for 2006 as a result of these findings.

References

Assessment Systems Corporation (1996). *RASCAL Rasch Analysis Program.* St Paul, MN: Assessment Systems Corporation.

Thomas, G., Tagg, A., and Ward, J. (2006, in press). Numeracy assessment: How reliable are teachers' judgments? *Findings from the New Zealand Numeracy Development Project 2005.* Wellington: Ministry of Education.