Linking Teacher Knowledge and Student Outcomes

Jenny Ward
Maths Technology Ltd
<jenny@nzmaths.co.nz>

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

This paper reports on the ongoing analysis of the validity of a teacher-knowledge assessment as a predictor of students' progress. The assessment aims to identify the teachers who have the weakest knowledge for the teaching of fractions in order to enable the provision of targeted professional development support. The relationship between teachers' scores on the assessment and student achievement gains as measured against the Number Framework was investigated. Rasch analysis was employed to obtain a measure of student gain that could be compared irrespective of starting stage. Results provide evidence of the validity of the content measures in the assessment, with students of teachers who scored highly in the content measures making significantly greater gains than students of teachers who scored poorly in the content measures. Findings for the measures based on pedagogical content knowledge were unexpected, with students of teachers scoring highly in these measures making significantly smaller mean gains than the students of teachers scoring poorly.

Background

It is well established that teachers of mathematics require a broad range of knowledge in order to be effective (Wayne & Youngs, 2003) and that professional development programmes have an important role to play in supporting the development of teacher knowledge (Putnam & Borko, 1997). Recent studies have indicated that programmes focused on the development of teacher knowledge are successful in changing teacher practice and improving student performance (Borko, 2004). The most effective professional development programmes build on teachers' existing knowledge, so an initial assessment of teachers' understandings may provide valuable information that can be used to provide targeted professional development support (Guskey, 2000).

This paper reports on the ongoing analysis of the validity of an assessment (see Appendix G, pp. 186–188) that focuses on teachers' knowledge for the teaching of fractions as a predictor of students' progress (Ward & Thomas, 2007, 2008). The assessment was developed for use in professional development programmes within the Numeracy Development Projects (NDP). The assessment is based on previous research in which measures of teachers' content knowledge for teaching were aligned with student achievement information; this research found that students with teachers in the lowest third of the knowledge distribution had significantly lower achievement (Hill, Rowan, & Ball, 2005). Based on this work, the assessment aims to identify those teachers who have the least knowledge for the teaching of fractions in order to enable the provision of targeted professional development support.

Previous analysis investigating the relationship between teacher scores on the assessment and student achievement information was limited by the non-linear scale of the stages of the Number Framework (Ward & Thomas, 2008). Students progress more quickly through the lower stages of the Framework than through the upper stages (Irwin, 2003; Irwin & Niederer, 2002; Ward & Thomas, 2002), so a simple comparison of stage gains in relation to teacher scores is not possible. The current study employs Rasch analysis to obtain a linear measure of student gain that can be compared across students, irrespective of starting stage. Similar work has been carried out previously (Irwin, 2003; Irwin & Niederer, 2002), but there was a need to update this analysis to take account of changes to the Number Framework.

The current analysis investigates the relationship between teacher scores on the assessment and Rasch measures of student achievement gain. If teachers who score poorly in the assessment have students who make significantly lower gains, this is evidence that the tool is fit for purpose; it is able to identify those teachers with the greatest need for professional development.

Method

Materials

Information on teacher knowledge was collected using the teaching of fractions assessment. The assessment is designed to collect information on teachers' content knowledge (CK) and pedagogical content knowledge (PCK) in the area of fractions. Teacher responses were evaluated using a marking criteria developed alongside the assessment. The assessment was marked out of a total of 17 points: the CK component had a total possible score of 7 points and the PCK component was worth 10 points. Further information about the development, nature, and trial of the assessment can be found in previous reports (Ward & Thomas, 2007, 2008).

Student achievement data was collected by teachers using the Numeracy Project Assessment (NumPA). The NumPA is an individual interview that provides information on students' knowledge and strategy stages aligned to the Number Framework.

Method

Participating schools received teacher assessments by post, along with a set of detailed instructions for their completion. Assessments were completed at the school and returned to the researcher for marking.

Two sets of student achievement information were collected by teachers: initial data before instruction began and final data near the end of the 2007 school year. Student data was entered into the online numeracy database.

Further information about data collection methods can be found in previous reports (Ward & Thomas, 2007, 2008).

Participants

The final sample consisted of 17 schools: one low-decile school (1–3), 10 medium-decile schools (4–7), and 6 high-decile schools (8–10). Results from 88 teachers were included in the sample. Table 1 provides a summary of the year levels taught by these teachers. Participating teachers had between 6 months and 30 years teaching experience, with the majority of teachers' experience being at the year level of the students they were currently teaching.

Table 1
Participating Teachers

Year levels	Number of Teachers	Percentage of Teachers
1–3	27	31
4–6	18	20
7–8	32	36
9	11	13
Total	88	100

On average, each participating teacher provided complete results for 22 students, with between 2 and 49 sets of student results per teacher. Table 2 shows the distribution of participating students across year levels.

Table 2
Participating Students

Year levels	Number of Students	Percentage of Students
1	168	9
2	188	10
3	172	9
4	119	6
5	140	7
6	122	6
7	371	19
8	445	23
9	221	11
Total	1946	100

Analysis

A Rasch analysis was used to describe student gain in a three-step process. Initially, item difficulty estimates for each of the stages within the three strategy and the six knowledge domains within the Number Framework were calculated using a maximum log likelihood procedure and software developed for the purpose using R (a statistical programming language). The difficulties of all stages of all nine domains were estimated on the same logistic (Rasch) scale. The estimates were calculated using the initial results of 4000 students randomly selected from the 47 005 entries in the 2007 national numeracy results database.

Initial and final student ability estimates for the students in the sample were calculated using the item difficulty estimates and the students' initial and final stage ratings on each of the nine domains. Student gain scores were then calculated for each student by subtracting the initial ability estimate from the final ability estimate. These gain scores provided a measure of student progress in logits, which could be compared across all students irrespective of starting stage.

The relationship between student gain scores and teacher scores on the assessment was investigated. For comparison purposes, teachers were split into low-scoring and high-scoring groups. These groups were based on the distribution of scores, with one-third of the teachers placed in the low-scoring group and one-third of the teachers placed in the high-scoring group for each of the three score types: CK, PCK, and total. Table 3 shows the distribution of scores for each of these groups.

Table 3
Scoring Range of Low- and High-Scoring Comparison Groups

	Low Scoring		High Scoring	
Score	Score	Number of teachers	Score	Number of teachers
Content knowledge (CK)	≤ 4	25	7	26
Pedagogical content knowledge (PCK)	≤ 2	25	≥ 6	23
Total (CK and PCK)	≤ 7	27	≥ 12	26

T-tests were used to compare the mean gain score of students with low-scoring teachers to the mean gain score of students with high-scoring teachers for each of the three groups. For the purposes of this paper, effect sizes of 0.2 or less are described as small, effect sizes between 0.2 and 0.8 are described as medium, and effect sizes of 0.8 or higher are described as large (Cohen, cited in Coe, 2002).

Findings and Discussion

This section reports on the findings of this research under three headings. Firstly, the achievement gains of the students whose teachers attained low or high scores on the assessment are compared. This is followed by a breakdown of gains by year level. Finally, the effect of CK scores is investigated for groups in which PCK is low, and similarly, the effect of varying PCK scores is described for groups in which CK is high.

Comparison of Student Achievement Gains for Low- and High-scoring Teachers

Significant differences were found between the mean gains of students with low- and high-scoring CK teachers, with a medium effect size of 0.24. No significant differences between the mean gains of low- and high-scoring groups were found using PCK or total scores. Table 4 shows these results.

Table 4

Comparison of Mean Student Gain for Low- and High-scoring Teacher Groups

	Mean Gain (logits)				
Score	Students of low-scoring teachers (n)	Students of high-scoring teachers (n)	Difference in means	Effect size	Significance level
Total	0.95 (583)	1.03 (546)	0.08	0.11	Not significant
CK	0.96 (540)	1.15 (582)	0.19	0.24	<0.01
PCK	1.00 (554)	1.03 (449)	0.03	0.05	Not significant

Note: n = number of students

The differences between the mean gains for low- and high-scoring CK groups seem small, but they indicate a substantial difference in gains between the groups. Students with a teacher in the high-scoring-CK group gained, on average, an additional 0.19 logits. This is 20% more than the average gain made by students of low-scoring-CK teachers. Over the course of a school year, this effectively equates to an additional eight weeks of instruction for the students of high-scoring teachers.

The lack of significant differences in the mean gains of students with teachers in the low- and high-scoring-PCK groups was unexpected, as was the lack of significant differences in the mean gains of students with teachers in the low- or high-scoring total groups. Further analysis was therefore carried out by year level.

Comparison of Student Achievement Gains by Year Level

Because the assessment was developed to assess knowledge for teaching at stages 7 and 8 of the Number Framework, it might be expected that the relationship between student gain scores and teacher scores on the assessment would be strongest at higher year levels. Comparisons were made between the student gains of low- and high-scoring teachers for all three score types (CK, PCK, and total) by year level. Significant differences were found between the means of low- and high-scoring-CK groups at years 7–8 and 9. Table 5 displays the results for CK scores. No significant differences where found between the mean gains of low- and high-scoring groups for PCK or total scores.

Table 5
Mean Gain Comparison: Low- and High-scoring Groups by Year Level

	Mean Gain (logits)				
Year	Students of low-scoring teachers (n)	Students of high-scoring teachers (n)	Difference in means	Effect size	Significance level
1–3	0.99 (206)	0.94 (117)	0.05	0.13	Not significant
4–6	0.89 (116)	0.87 (141	0.02	0.03	Not significant
7–8	1.00 (205)	1.42 (254)	0.42	0.42	<0.01
9	0.43 (13)	1.13 (70)	0.69	0.91	<0.01

Note: n = number of students

The differences in the student gains of low- and high-scoring-CK teachers in years 7–8 and 9 were substantial, with effect sizes of 0.42 in years 7–8 and 0.91 in year 9. At years 7–8, students of high-scoring-CK teachers made on average 42% more progress than students of low-scoring-CK teachers. At year 9, students of high-scoring-CK teachers made on average 160% more progress than students of low-scoring-CK teachers, although the small sample size for students of low-scoring teachers at this level indicates a need to be cautious when interpreting these results.

While the analysis by year level shows a clear relationship between measures of student gain and teachers' CK scores, teachers' PCK scores were found to be unrelated to student gain scores at any level. This may be explained, in part, by the relationship between CK and PCK scores, which makes a comparison of high and low scores problematic. A group of teachers with low CK scores will necessarily have low PCK scores, whereas a group of teachers with high CK scores may have a range of PCK scores (Ward & Thomas, 2008). In other words, those teachers who are unfamiliar with specific content are unlikely to know effective ways to teach that content, while teachers who are familiar with the content may or may not know how to teach it effectively. Conversely, a group of teachers with high PCK scores will necessarily have high CK scores, whereas a group of teachers with low PCK scores may have a range of CK scores. Figure 1 shows a plot of teachers' CK scores versus their PCK scores to illustrate this relationship. A line of best fit is included, and the numerals in brackets give the number of teachers at each of the data points.

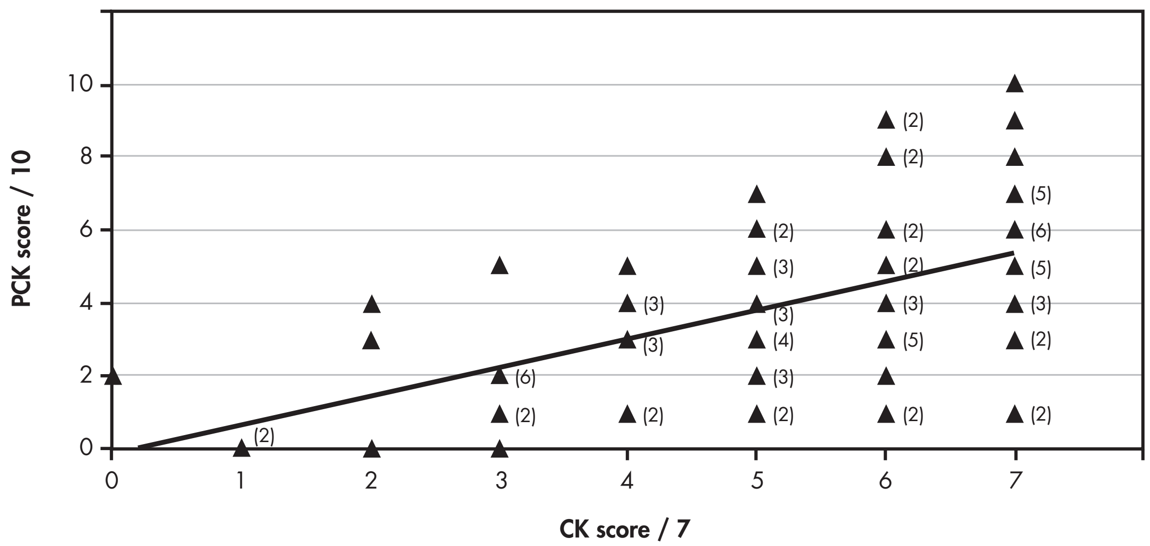

Figure 1. Teachers' CK and PCK scores

Comparisons of Individual Effects of CK and PCK

As described in the previous section, the relationship between CK and PCK scores complicates comparisons of low- and high-scoring groups. To mitigate this effect to some extent, two further analyses were carried out: firstly, an investigation of the student gains of low- and high-scoring-CK teachers whose PCK was low, and secondly, an investigation of the student gains of low- and high-scoring-PCK teachers whose CK was high. The analysis was carried out using students from years 7–8 and 9 for whom a relationship between student gain and teacher CK scores had already been established.

Three groups of year 7–9 teachers were established to enable these comparisons. Group 1 consisted of teachers who were low scoring in both CK and PCK, group 2 was made up of teachers who were low scoring in PCK and high scoring in CK, and group 3 teachers were high scoring in both CK and PCK. Figure 2 illustrates these groups.

	Low CK	High CK
High PCK		Group Three: 7 teachers 140 students
Low PCK	Group One: 6 teachers 140 students	Group Two: 2 teachers 50 students

Figure 2. CK and PCK comparison groups, years 7–9

Comparing groups 1 and 2 enables us to look at the effect of CK scores where PCK is low. Teachers in the high-scoring-CK group were found to have a significantly higher mean student gain than teachers in the low-scoring-CK group ($p < 0.01$). Figure 3 illustrates the mean gain for groups 1 and 2, using error bars with a 95% confidence interval for the mean.

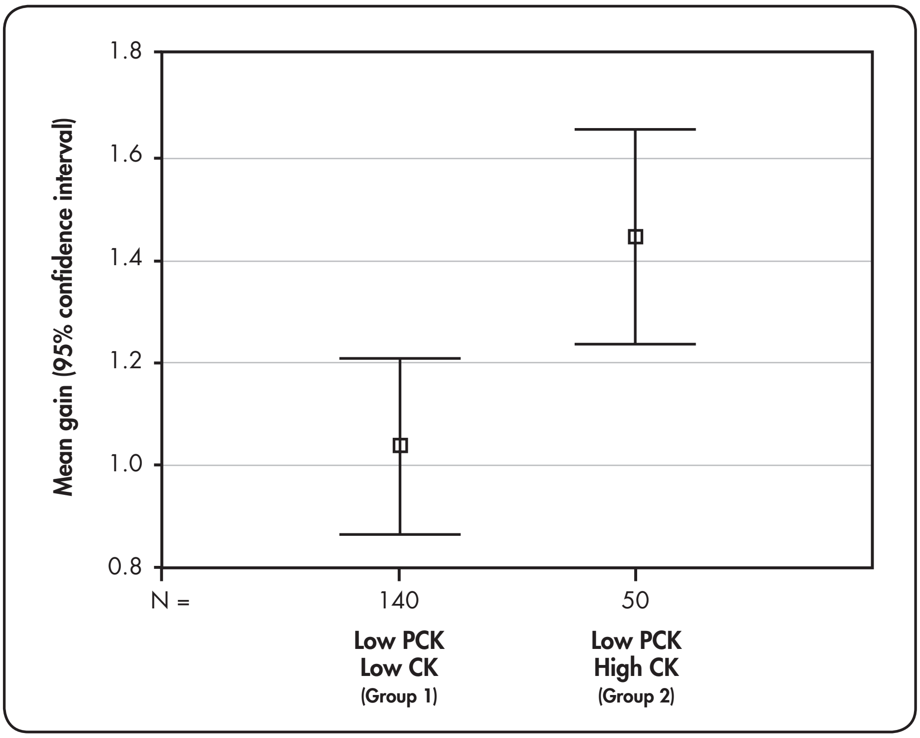

Figure 3. Effect of CK scores when PCK is low

A T-test was also carried out using student gain scores from the two groups. Students with low-scoring-CK teachers had a mean gain of 1.04 logits, while students with high-scoring-CK teachers had a mean gain of 1.44 logits. This difference was found to be significant ($p < 0.01$), with a medium effect size of 0.42. These results are in accordance with the relationship between student gain scores and CK measures already identified.

A comparison of groups 2 and 3 enables a look at the effect of PCK scores where CK is high. Unexpectedly, teachers in the low-scoring-PCK group were found to have a significantly higher mean student gain than teachers in the high-scoring-PCK group. Figure 4 illustrates the mean gain for groups 2 and 3, using error bars with a 95% confidence interval for the mean.

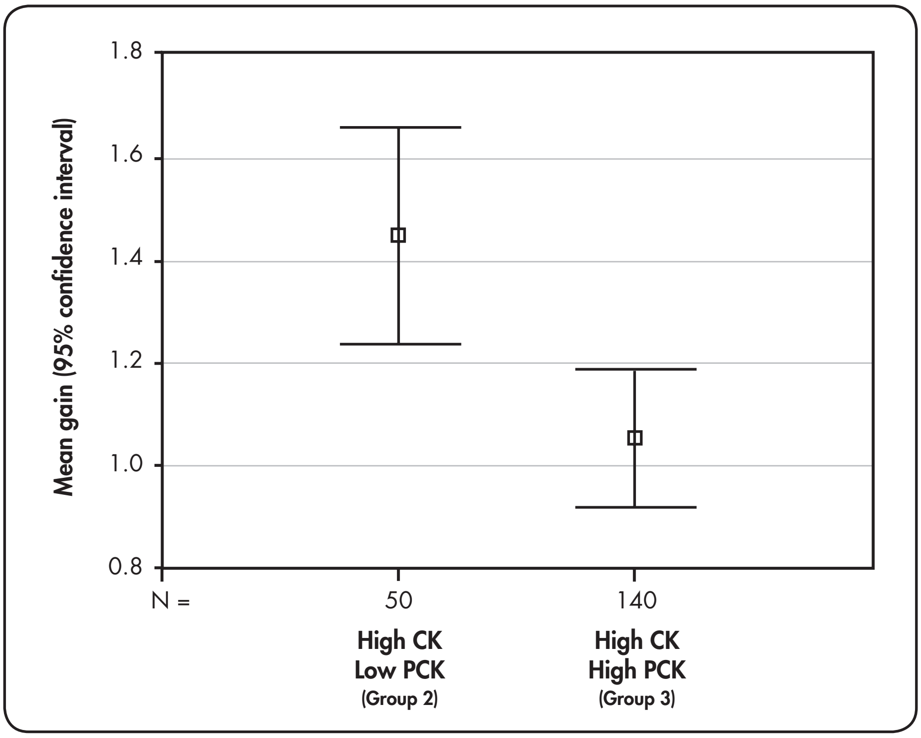

Figure 4. Effect of PCK scores when CK is high

A further T-test on the student gains of these two groups showed that students with teachers in the low-scoring-PCK group had a mean gain of 1.44 logits, while students of teachers in the high-scoring-PCK group had a mean gain of 1.05 logits. This is a difference in the means of 0.40 logits, which is significant ($p < 0.01$), with a medium effect size of –0.51. This result was unforseen and surprising; students with teachers in the high-scoring-PCK group had significantly smaller mean gains than students with teachers in the low-scoring-PCK group.

A review of the literature provides some explanation for the result in which higher PCK scores are related to diminished student gain. While there have been several attempts to measure teachers' PCK in mathematics (see, for example, Carpenter, Fennema, Peterson, & Carey, 1988 and Hill, Schilling, & Ball, 2004), attempts to align teachers' scores on these measures with student achievement data have been unsuccessful. This is in contrast to measures of teachers' content knowledge that have successfully been aligned with student achievement data (see, for example, Hill, Rowan, & Ball, 2005). The lack of alignment between measures of PCK and student achievement data has been attributed to the multidimensional nature of PCK (Alonzo, 2007; Hill, Ball, Blunk, et al., 2007; Hill, Ball, & Schilling, 2008). As a construct, PCK is problematic to measure because it consists of several elements. These include: knowledge of the development of student thinking in particular domains, knowledge of common student misunderstandings, and knowledge of effective teaching representations. The broad nature of PCK results in measurement difficulties.

Concluding Comment

Fractions are a complex and crucial area of mathematics instruction. They are recognised as both one of the most important areas in the primary school mathematics (Behr, Lesh, Post, & Silver, 1983; Lamon, 2007) and one of the most challenging to teach and learn effectively (Smith, 2002; Lamon, 2007). In addition, they represent the initial development of proportional reasoning, an area of the NDP in which both students' and teachers' understanding is of concern (Young-Loveridge, 2006, 2007; 2008; Young-Loveridge, Taylor, Hāwera, & Sharma, 2007; Ward & Thomas, 2007).

When establishing the validity of any assessment, the purpose for which items were developed and the context in which they are to be used are paramount (Kane, 2007; Lawrenz & Toal, 2007). On this basis, the question must be asked: are teachers' scores in the assessment useful for identifying those teachers whose students are making the least progress? The CK measures are valid in this regard at years 7–9, in which the content of instructional programmes is directly aligned with the content of the assessment. At these years, students of high-scoring-CK teachers make significantly greater gains than students of low-scoring-CK teachers. Contrastingly, the PCK measures cannot be regarded as valid. The unexplained result for years 7–9, in which the students of high-scoring-PCK teachers made significantly less gains than the students of low-scoring-PCK teachers, is puzzling.

Acknowledgments

The authors acknowledge the support and advice of Michael Johnston of NZQA in the use of Rasch analysis.

References

Alonzo, A. C. (2007). Challenges of simultaneously defining and measuring knowledge for teaching. *Measurement: Interdisciplinary Research & Perpectives, 5*(2–3), 131–137.

Behr, M. J., Lesh, R., Post, T. R., & Silver, E. A. (1983). Rational-number concepts. In R. Lesh & M. Landau (Eds), *Acquisition of mathematics concepts and processes* (pp. 91–126). London: Academic.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *American Educational Research Journal, 33*(8), 3–15.

Carpenter, T., Fennema, E., Peterson, P., & Carey, D. (1988). Teachers' pedagogical content knowledge of students' problem solving in elementary arithmetic. *Journal for Research in Mathematics Education, 19*(5), 385–401.

Coe, R. (2002). *It's the effect size, stupid: What effect size is and why it is important.* Paper presented at the annual conference of the British Educational Research Association, University of Exeter, England, 12–14 September. [Available from www.leeds.ac.uk/educol]

Guskey, T. (2000). *Evaluating professional development.* Thousand Oaks: Corwin.

Hill, H. C., Ball, D. L., Blunk, M., Goffney, I. M., & Rowan, B. (2007). Validating the ecological assumption: The relationship of measure scores to classroom teaching and student learning. *Measurement: Interdisciplinary Research & Perspective, 5*(2), 107–118.

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education, 39*(4), 372–400.

Hill, H. C., Rowan, B., & Ball, L. D. (2005). Effect of teachers' mathematical knowledge for teaching on student achievement. *American Education Research Journal, 42*(2), 371–407.

Hill, H. C., Schilling, S. G., & Ball, L. D. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal, 105*(1), 11–30.

Irwin, K. C. (2003). *An evaluation of the Numeracy Project for years 7–10, 2002:* Exploring issues in mathematics education. Wellington: Ministry of Education.

Irwin, K. C., & Niederer, K. (2002). *An evaluation of the Numeracy Exploratory Study, years 7–10, 2001:* Exploring issues in mathematics education. Wellington: Ministry of Education.

Kane, M. (2007). Validating measures of mathematical knowledge for teaching. *Measurement: Interdisciplinary Research & Perspective*, *5*(2–3), 180–187.

Lamon, S. (2007). Rational numbers and proportional reasoning: Toward a theoretical framework for research. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. I, pp. 629–667). Charlotte, NC: Information Age.

Lawrenz, F., & Toal, S. (2007). Commentary: A few tweaks to the toolkit. *Measurement: Interdisciplinary Research & Perspective*, *5*(2–3), 195–198.

Putnam, R. T., & Borko, H. (1997). Teacher learning: Implications of new views of cognition. In B. J. Biddle, T. L. Good, & I. F. Goodson (Eds), *International handbook of teachers and teaching* (Vol. II, pp. 1223–1296). Dordrecht: Kluwer Academic.

Smith, J. P. III (2002). The development of students' knowledge of fractions and ratios. In National Council of Teachers of Mathematics, *Making sense of fractions, ratios, and proportions: 2002 yearbook.* Reston, VA: NCTM.

Ward, J., & Thomas, G. (2002). *An evaluation of the Count Me In Too Pilot Project:* Exploring issues in mathematics education. Wellington: Ministry of Education.

Ward, J., & Thomas, G. (2007). What do teachers know about fractions? In *Findings from the New Zealand Numeracy Development Projects 2006* (pp. 128–138). Wellington: Learning Media

Ward, J., & Thomas, G. (2008). Does teacher knowledge make a difference? In *Findings from the New Zealand Numeracy Development Projects 2007* (pp. 50–60). Wellington: Learning Media.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, *73*(1), 89–122.

Young-Loveridge, J. (2006). Patterns of performance and progress on the Numeracy Development Project: Looking back from 2005. In *Findings from the New Zealand Numeracy Development Projects 2005* (pp. 6–21). Wellington: Learning Media.

Young-Loveridge, J. (2007). Patterns of performance and progress on the Numeracy Development Projects: Findings from 2006 for years 5–9 students. In *Findings from the New Zealand Numeracy Development Projects 2006* (pp. 16–32). Wellington: Learning Media.

Young-Loveridge, J. (2008). Analysis of 2007 data from the Numeracy Development Projects: What does the picture show? In *Findings from the New Zealand Numeracy Development Projects 2007* (pp. 16–26). Wellington: Learning Media.

Young-Loveridge, J., Taylor, M., Hāwera, N., & Sharma, S. (2007). Year 7–8 students' solution strategies for a task involving addition of unlike fractions. *In Findings from the New Zealand Numeracy Development Projects 2006* (pp. 67–86). Wellington: Learning Media.