Technical Report #40 Assessment Tools for Teaching and Learning

Assessment of student number strategy development: A case study in the use of paper and pencil items

Abstract: 31 paper and pencil mental strategy assessment items were designed to elicit information about students' ability to process number problems according to six strategies of the New Zealand Number Framework. Data from end-of-year teacher assigned Numeracy Framework strategy and knowledge scores were collected from Year 5 and 6 students (*N* = 136) in one primary school. Consensus and consistency measures showed that the two instruments fundamentally provided different data, with consistently higher strategies assigned in the paper-and-pencil questionnaire. Suggested steps in improving paper-and-pencil items are provided and a model for testing the relationships of data collection mode, strategy, knowledge, and achievement is proposed.

Submitted by the Assessment Tools for Teaching and Learning team,

Auckland UniServices Ltd

University of Auckland

September 2004

Assessment of student number strategy development:

A case study in the use of paper and pencil items

This report provides the results of a study into developing paper and pencil items to assess number strategy processing stages. asTTle is funded by the Ministry of Education to Auckland Uniservices Ltd. at the University of Auckland to research and develop an assessment application for Reading, Writing, Mathematics, Pänui, Pängarau, and Tuhituhi for Years 5-10 (Levels 2-6) for New Zealand schools. We acknowledge this funding, and thank the Ministry of Education for their continued assistance in the development of this project.

A series of paper and pencil mental strategy assessment items were designed to elicit information about students' ability to process number problems according to six strategies of the New Zealand Number Framework. Data from end-of-year teacher assigned Numeracy Framework strategy and knowledge scores were collected from Year 5 and 6 students (*N* = 136) in one primary school. Consensus and consistency measures showed that the two instruments fundamentally provided different data, with consistently higher strategies assigned in the rating paper-and-pencil questionnaire. Suggested steps in improving paper-and-pencil items are provided and a model for testing the relationships of data collection mode, strategy, knowledge, and achievement is proposed

I would like to acknowledge the participation of various personnel in these studies. Mr Peter Hughes, Auckland College of Education, created the items to ensure a representative coverage of the six numeracy stages. Dr Gavin Brown managed the item development and data recording (conducted by Nikki Thompson), assisted with data analysis, and drafted this report. Ms. Sarah Martin of Bayview School provided data on children and conducted the numeracy framework interviews. The second two authors were assisted by Heidi Leeson, research analyst for asTTle, in the statistical analyses of the data.

John Hattie
Project Director, asTTle
September 2003

The bibliographic citation for this report is:

Hughes, P., Hattie, J.A., Brown, G.T.L., & Leeson, H.V. (2004, September). *Assessment of student number strategy development: A case study in the use of paper and pencil items*. asTTle Technical Report #40, University of Auckland/Ministry of Education.

Table of Contents

Number ‘strategies’ are mental or cognitive processes used to “estimate answers and solve operational problems with numbers” (Ministry of Education, 2003a, p. 1). Over 30 different mental strategies have been found in relation to learners’ mental processes for doing addition and subtraction (single digit and multi-digit often considered separately), place value, number sense, multiplication and division and fractions and decimals and that various researchers have described taxonomic-like progressions in the development or use of the strategies (Ell, 2001b). Identification of and appropriate teaching to a child’s mental arithmetic strategy has been found effective (Fennema, Carpenter, Levi, Jacobs & Empson, 1996; Jones, Thornton, Putt, Hill, Mogill, Rich & Van Zoest, 1996; Fuson, Smith & LoCicero, 1997; Thomas & Ward, 2001). Ell (2001) argued that the suggested developmental progressions within the various taxonomies are best used to guide instructional sequences and help teachers to decide ‘where to next’ for a child.

The Ministry of Education has introduced the Number Framework (Ministry of Education, 2003a) as a teacher professional development intervention to raise student achievement in numeracy cognitive strategies and knowledge. The Number Framework has developed an eight stage strategy list and six types of knowledge requirements cross-referenced to each other capable of being used pedagogically to identify teaching and learning needs of students. The strategies, in descending order, are (8) advanced proportional, (7) advanced multiplicative, (6) advanced additive, (5) early additive, (4) advanced counting, (3) counting from one by imaging, (2) counting from one on materials, (1) one-to-one counting, and (0) emergent. The knowledge framework identifies the type of knowledge, mapped to each stage, which students must have as a precursor to progression to the next mental strategy. Knowledge stages have been identified for five categories; number identification, number sequence and order, grouping/place value, basic facts, and written recording. The Number Framework has been implemented as a professional development intervention and has been evaluated as successful in raising student level of strategy use (Higgins, 2003; Irwin, 2003; Thomas, Tagg, & Ward, 2003).

Item Development Processes

The New Zealand Number Framework identifies children's current mental strategy through interviews with children. This process of probing for children's strategies may be problematic because it involves the child's ability to describe their strategy, or the interviewer's ability to infer it and children may use different strategies in an interview setting, believing they know what the interviewer wants (Ell, 2001). Thus, it was proposed that the development of paper and pencil assessment items might indicate the type of strategy children use in completing arithmetic problems would be a useful adjunct to the current battery of interview procedures. Such items would be easier and quicker to administer and could provide more specific and reliable assessment of the strategies.

The Assessment Tools for Teaching and Learning (asTTle) project (Ministry of Education, 2003b) has developed a large number of paper and pencil assessment items calibrated to the achievement objectives of curriculum levels 2—4 for use with students in Years 5—7 (see Ell, 2001a; Thomas, Holton, Tagg, & Brown, 2002 for a description of how asTTle materials relate to the curriculum). Early work in the development of paper-and-pencil strategy related items indicated that few draft items provided meaningful information about the "strategies" used to answer the items as proposed by the New Zealand Number Framework (van Garderen, 2001). That report recommended, among others, clarifying the strategy stage each question represents, ensuring that all strategy stages are represented in the range of problems, developing questions that are difficult enough for older students that "force" them to demonstrate their working, and exploring the use of questions that either list choice a of strategies or require the student to show their working.

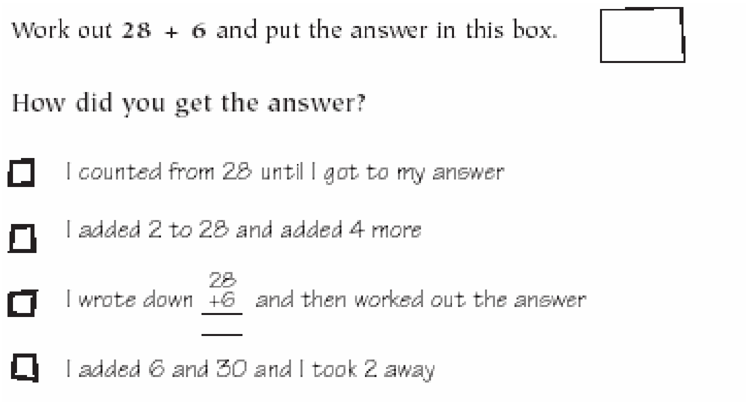
Thus, a new round of item development was undertaken to create items clearly mapped to the Number Framework strategy stages and to ensure clarity of communication with children in Years 5—6 (nominal ages 9—11). Items were drafted by the first author based on the most successful items identified in the earlier pilot study. The items were reviewed by the co-authors to ensure a diversity of item and response types. The resulting 31 items were then piloted in one high decile primary school with two focus groups to test readability. Those groups resulted in a revision of tasks to reduce reading load. The revised items were then assembled into two forms (Set A had 15 and Set B had 16 questions), so as to reduce administration time, intended to be relatively parallel in terms of content and item type.

Paper-and-Pencil Strategy Items

The paper-and-pencil strategy item were designed to be used with children in Years 5—6 who should be working in Levels 2—4 of the curriculum. Thus, the items were calibrated to just six of the eight strategy stages (i.e., (3) count all, (4) advanced counting, (5) early additive, (6) advanced additive, (7) advanced multiplicative, and (8) advanced proportional). In addition, responses could be coded where students identified that they had used an algorithmic approach. Several major types of item were used: (a) all process options correct multiple-choice (i = 16), (b) follow the provided strategy (i = 6), (c) explain a likely strategy for an exemplified answer (i = 6), and (d) identify multiple equivalent fractions or expressions (i = 3).

All process options correct multiple-choice items: students are posed arithmetic problems and a list of correct strategies for solving the problem is presented in multiple-choice fashion. Students were asked to select the one that most closely resembled their approach to solving the problem. In Figure 1 the student first answers the question and then identifies the process that most closely matches his or her own method. Note that each process is legitimate and correct. Student responses are scored by the level of the number framework to which their selected response maps.

Figure 1
Sample All Process Options Correct Multiple-Choice Question

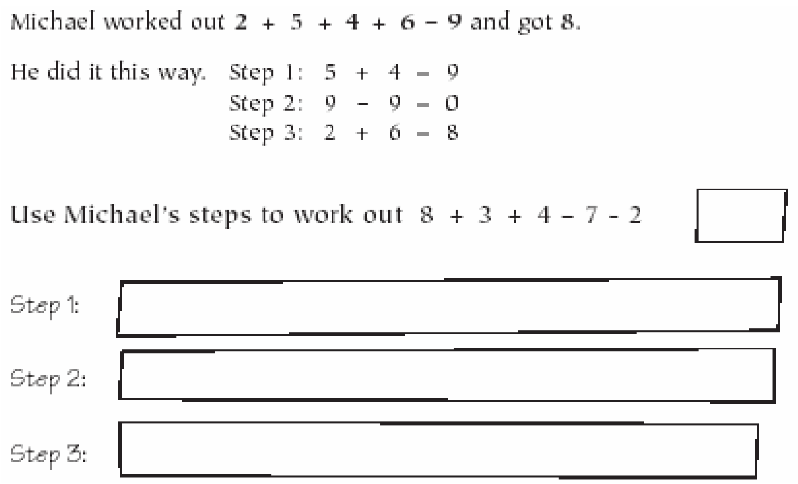

Work out 28 + 6 and put the answer in this box.

How did you get the answer?

- ☐ I counted from 28 until I got to my answer
- ☐ I added 2 to 28 and added 4 more
- ☐ I wrote down 28 +6 and then worked out the answer
- ☐ I added 6 and 30 and I took 2 away

Follow the provided strategy: students are provided with a model strategy that exemplifies a part-whole additive or multiplicative or proportional approach to solving a problem. Students are then asked to use the provided strategy to solve a parallel problem (Figure 2). Scoring is whether the student has successfully imitated the provided strategy to reach a correct answer.

Figure 2
Sample Follow the Provided Strategy Question


Michael worked out 2 + 5 + 4 + 6 − 9 and got 8.

He did it this way. Step 1: 5 + 4 = 9
Step 2: 9 − 9 = 0
Step 3: 2 + 6 = 8

Use Michael's steps to work out 8 + 3 + 4 − 7 - 2

Step 1:

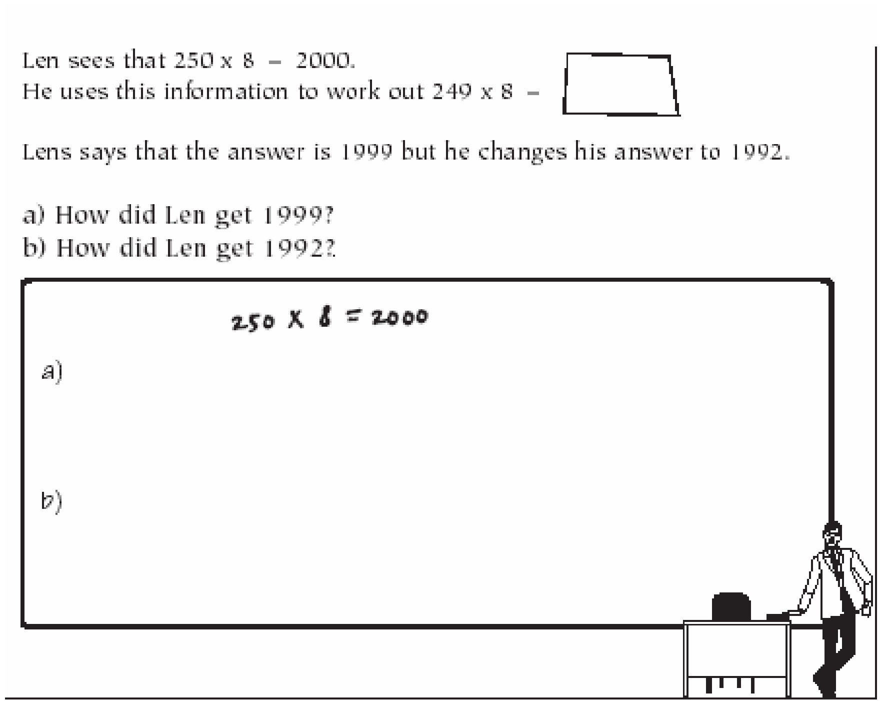
Step 2:

Step 3:

Explain a likely strategy: for an exemplified answer type questions, students are given a problem that had been solved in such a way (usually very quickly) that a correct solution could only have been arrived at by use of a part-whole or advanced arithmetic strategy. Students were asked to explain what they believed to be the most likely process would be to arrive at the solution under such conditions. Students' explanations were categorised by one of the six Number Framework strategies or by use of algorithm; answers that could not be classified onto to the list were treated as missing. In Figure 3, students were asked to explain both how a student may have got a wrong and a right answer.

Figure 3

Sample Explain a Likely Strategy Question



Identify multiple equivalent fractions or expressions: students are asked to recognise similar values without carrying out any working or calculations. These questions depend on the student's ability to recognise equivalence by the use of part-whole mental strategies. If the student answers correctly, their response is coded as the appropriate level strategy for the Number Framework. In Figure 4, a correct response indicates advanced proportional strategy use.

Figure 4
Sample Identify Multiple Equivalent Fractions or Expressions Question

Circle the two calculations that will give the same answer.
(Do not do any calculations.)

0.26 x 456 2.6% of 456 26% of 456

456 − 0.26 456 + 26% 0.26 ÷ 456

Three operational domains are identified against which the Number Framework Strategies apply (i.e., add and subtract, multiply and divide, and proportion and ratio). The number of items within the questionnaire that apply to each operational domain is shown in Table 1. By operational domain, there are 13 add and subtract, five multiply and divide, and 13 proportion and ratio questions. By strategy, there are 15 all strategies correct multiple-choice, four follow the provided strategy, nine explain a likely strategy, and three identify multiple equivalent fractions or expressions items. Thus, it is possible to analyse results from the paper and pencil questionnaire items by both operational domain and item format.

Table 1
Number of Paper and Pencil Items by Format and Operational Domain

	Operational Domain		
Item Format	Add & Subtract	Multiply & Divide	Proportion & Ratio
All Strategies Correct Multiple-Choice	7	1	7
Follow the Provided Strategy	3	1	—
Explain a Likely Strategy	3	2	4
Identify Multiple Equivalent Fractions or Expressions	—	1	2

Method

The revised test forms were trialled in one high-decile primary school with 136 Year 5 and 6 students whose Number Framework strategy level had been already ascertained by the approved interview process. The mid-year and end-of-year interviews were conducted by an experienced Number Framework facilitator, working as a teacher in the school. In addition to mental strategy, the facilitator provided a knowledge score;

although not central to this research, it is possible to compare knowledge rating with the paper-and-pencil strategy results and the interview strategy rating.

Although the Number Framework interview explores student strategy use across three operational domains (i.e., add and subtract, multiply and divide, and proportions and ratios), it could be argued that students should be classified by their highest strategy exhibited regardless of domain; lack of knowledge within any one domain may prevent exercise of a higher strategy and thus strategic ability must be rated by the display of the most sophisticated strategy independent of domain. Nevertheless, for the purposes of this study it was decided to investigate potential impact of operational domain on strategy usage.

Each response in the asTTle questionnaire was scored according to the Number Framework strategy matching the response selected (multiple-choice type) or by the strategy required by the question. Responses not matching the Number Framework were classified as algorithm if such approach was either selected or demonstrated in workings or explanation. Other answers that could not be mapped to the Number Framework or algorithm were classified as missing.

Results

Strategy Questionnaire

There were 81 students who completed Set A of the questionnaire and 76 students completed Set B. Although assurances were given that all these students had been assessed via the strategy questionnaire, data were provided only for 62 students. A major effort was made to ensure the students had been rated by expert users of the NUMPA system; and this appears to be the case. However, the lack of variation or range within this sample restricts its utility in analysing these data.

As can be seen from the interview ratings (Table 2) by the time of the second interview, which was conducted at about the same time as the strategy questionnaire, the majority of students were at the Early or Advanced Additive stages in addition and subtraction, multiplication and division, and in proportion and ratio. The highest maximum strategy seen was in the area of proportion and ratios (advanced proportional) and the lowest maximum strategy was in the area of addition and subtraction (with nearly half the rated students using advanced additive strategies). Note that this latter result is a ceiling effect in that it is not possible for the interview to detect multiplicative or proportional thinking in the context of addition and subtraction problems.

In order to investigate measures of central tendency and variance, the number of students assigned to a strategy was multiplied by its stage value to create a weighted stage score (Table 2).

Table 2

Interview Rating of Number Strategy by Operational Domain

	Operational Domain					
	Add & subtract		Multiply & Divide		Proportion & Ratio	
Strategy	Count	Weighted score	Count	Weighted score	Count	Weighted score
3. Counting All	2	6	2	6	8	24
4. Advanced Counting	9	36	10	40	10	40
5. Early Additive	24	120	20	100	19	95
6. Advanced Additive	34	204	19	114	12	72
7. Advanced Multiplicative	—		14	98	12	84
8. Advanced Proportional	—		—		5	40
Total	69	366	65	358	66	355
Weighted Scores						
Average	5.30		5.51		5.14	
SD	.91		1.09		1.93	
Mode	6.00		5.00		5.00	
Median	5.00		5.00		5.00	
Highest	6.00		7.00		8.00	

The centre of the distribution in each domain (i.e., mean, median, and mode) is in the early additive strategy, except for the mode of the add and subtract domain being advanced additive strategy. It should also be noted that only a small number of students (between two and ten) appear to be at the lowest strategies (Counting On and Advanced Counting) across all three domains; while a similarly small number of students were rated as advanced multiplicative (*n* = 14) or advanced proportional (*n* = 5).

It is also clear from the data available for students who have two interview ratings that student numeracy strategies had generally shifted across the year. For example, in addition and subtraction only 47% of the students were similarly classified across the two interview times, 31% for multiplication and division, and 46% for proportion and ratio. Thus, the two interview times have classified about half the students similarly, which indicates a significant difference in rating between the two periods.

The average knowledge stage of this group of students at the same time as their second strategy rating is shown in Table 3. Ignoring knowledge stages with a low number of responses, the lowest knowledge stage mean was for group place values knowledge (*M* = 4.90, *SD* = 1.27), with the highest knowledge stage mean for backward number word (*M* = 5.46, *SD* = .63). Note that standard deviations for fractions, decimals, and group place values knowledge are 2 to 3 times larger than those found for forward number word knowledge, backward number word knowledge, and number identification knowledge, indicating greater potential for these categories in any correlational analysis.

Table 3
Mean Interview Ratings of Mathematics Knowledge

Knowledge Stage	*M*	*SD*	*N*
Forward Number Word Knowledge	5.43	.63	69
Backward Number Word Knowledge	5.46	.63	68
Number Identification	4.25	.50	4
Fractions	5.15	1.64	65
Decimals	5.79	1.41	24
Group Place Values	4.90	1.27	71

Paper and Pencil Questionnaire

Between 50 and 130 student responses were aggregated across the 31 items in the two questionnaire forms to provide data on student strategy use (Table 4). The mean strategy is shown by item format and operational domain. Amongst the operational domains, the highest mean scores were obtained on the proportion and ratio items (*M* = 6.62, *SD* = .84). The lowest mean scores was found on the add and subtract items (*M* = 5.08, *SD* = .57). This difference is not remarkable given the ceiling effect of the add and subtract operational domain. Nevertheless, it is worth noting that the highest strategies were found in the most demanding operational domain.

Table 4
Mean Number Strategy by Operational Domain and Format of Questionnaires

	Strategy Scores		
Questionnaire Dimension	*M*	*SD*	*N*
Operational Domain			
Add & Subtract	5.08	.57	120
Multiply & Divide	6.24	.96	68
Proportion & Ratio	6.62	.84	119
Item Format			
All Strategies Correct Multiple-Choice	5.80	.92	127
Follow the Provided Strategy	5.15	.33	54
Explain a Likely Strategy	6.41	.71	82
Identify Multiple Equivalent Fractions or Expressions	7.00	.00	39

Note that in the identify multiple equivalent fractions or expression items the advanced multiplicative strategy is assigned if the items were answered correctly and no alternative strategy could be assigned if the student did not answer correctly; in other words no variance in this type of item is possible and the items, in their current form, cannot be used in any correlational analysis. Note also the low number of students who could complete these items correctly. Ignoring the 'identify multiple equivalent fractions or expressions' item type, due to its small number of responses (N = 39), item format means ranged between 5.15 and 7.00 (*M* = 6.09). Follow the provided strategy item format resulted in mean scores at the early additive level (*M* = 5.15), whereas the 'explain a likely strategy' format mean scores were positioned at the advanced additive strategy level (*M* = 6.41).

Comparing Interview and Questionnaire Results

In comparison to the interview ratings, the questionnaire elicits a mean mental strategy at the advanced additive strategy level (averages across domains ranged between 5.08 and 6.62 (M = 5.98)), while the mean strategy found with the interview was early additive. Note also that the paper-and-pencil questionnaire scores also show less variance around the means than the interview ratings.

A significant method for determining the comparability of ratings is to examine the degree to which two raters or methods assign objects to the same categories (Stemler, 2004). Two approaches are commonly used: (a) the percent of cases assigned to exactly the same category and (b) the percent of cases assigned to the same category and its adjacent (plus or minus one) categories. Other research by the asTTle team comparing the consensus of two raters scoring student writing has found adjacent agreement scores in the range 70% to 90% (Brown, Glasswell, & Harland, 2004).

The consensus of strategy rating between the interview and questionnaire was examined (Table 5). The bold numbers in the central diagonal show the number of cases of exact agreement, while the adjacent agreement range is shown in the shaded boxes either side of the diagonal. The highest exact agreement consensus is seen in the early and advanced additive stages (range between 34% and 38%). On average only one in five of students (20.4%) were given the exactly the same strategy with both methods. Using the adjacent agreement approach between 80 and 90% of students in the advanced counting, early additive, and advanced additive stages received a high degree of consensus between the two approaches. However, on average only 56% of all ratings fell within the adjacent range. Given the high probability of chance agreement, these low exact and adjacent percent agreement results indicate that there is not a high consensus between these two methods.

It appears there is a systematic difference between the asTTle questionnaire items and the NUMPA interview ratings. The distribution of ratings outside the consensus range shows that at each strategy level the questionnaire places more students in a higher number strategy than the interview does. For example, at the counting on stage the questionnaire placed 24 students in a higher stage, while the interview placed four students in a higher stage. This pattern is consistent across all stages.

Table 5
Exact and Approximate Consensus Ratings between Questionnaire and Interview Ratings

	Questionnaire					
Interview Strategy	3	4	5	6	7	8
3. Counting On	**0**	2	9	10	5	0
4. Advanced Counting	0	**4**	24	11	24	2
5. Early additive	2	7	**54**	31	33	14
6. Advanced Additive	1	2	45	**43**	65	7
7. Advanced Multiplicative	1	0	21	16	**23**	7
8. Advanced Proportional	0	0	5	3	4	**3**
Total	4	15	158	114	154	33
Consensus						
% exact	.0	26.7	34.2	37.7	14.9	9.1
% adjacent	.0	86.7	77.8	78.9	59.7	30.3

A second method for determining the comparability of ratings is to examine the degree to which the scores from two raters or methods move in the same direction, despite not having similar values (Stemler, 2004). Correlations are traditionally used to determine whether ratings behave in a similar fashion, with values greater than .7 indicating that half or more of the variance in data can be attributed to a similar pattern of high and low scores between the two compared methods. Notwithstanding the low consensus ratings, correlations between two sets of ratings may be high, and so the consistency of the questionnaire operational domain and item format strategy results and the interview strategy and knowledge stage results was examined with correlations.

Table 6 shows the correlations between the six knowledge stages rated in the interviews, the questionnaire operational domains, and the questionnaire item formats. The NUMPA knowledge stages had low correlations with all the operational domains ($-.30 < r < .49$). There was a moderate relationship between proportion and ratio domain and the fraction ($r = .48$) and group place value ($r = .49$) knowledge stages. Although the number identification stage had a perfect correlation with the ‘add and subtract’ domain, there were only four responses in this correlation and thus this finding is of little import.

Overall, the NUMPA knowledge stages and the questionnaire item formats were also poorly correlated ($-.25 < r < .46$). However, a moderate relationship existed between the multiple choice item format and the fraction and decimal knowledge stages ($r = .46$ and .45 respectively). Similarly, a moderate correlation was found between fraction knowledge stage and the explain a likely strategy item format (r = .40). Because of the lack of variance in the 'identify multiple equivalent fractions' item format, this correlation is not reported.

Table 6
Correlations between Interview Knowledge Stages and Questionnaire Operational Domain and Item Format

	Interview Knowledge Stages					
Questionnaire	Forward Number-Word	Back-ward Number-Word	Number Identification	Fractions	Decimals	Group Place Values
Operational Domain						
Add & Subtract	.16	.22	—	.30	.29	.24
Multiply & Divide	.26	.35	.00	-.18	-.30	-.04
Proportional & Ratio	.23	.26	.28	.48	.34	.49
Item Format						
All Strategies Correct Multiple-Choice	.13	.24	.19	.46	.45	.29
Follow the Provided Strategy	.20	.19	.00	.29	-.25	.15
Explain a Likely Strategy	.20	-.08	.00	.40	-.10	.23

Table 7 shows the correlations between the three interview strategy stages, the three questionnaire operational domains, and three questionnaire item formats. The correlations of the matching interview and questionnaire strategy stages are shown in bold. The correlations between the two modes of assessment are relatively low ($-.28 < r < .46$), with only the proportional and ratio domain showing any moderate level of degree of correlation (r = .46). Indeed, the proportional and ratio operational domain of the questionnaire has a moderate positive relationship to all the interview strategies (average r = .39). It is worth noting that a negative correlation was found between the interview and questionnaire multiply and divide strategies (r = -.28).

Table 7
Correlations for Operational Domain, Questionnaire Item Format and Interview Strategy

	Interview Strategy		
Questionnaire	Add & Subtract	Multiply & Divide	Proportional & Ratio
Operational Domain			
Add & Subtract	**.18**	.23	.18
Multiply & Divide	.37	**-.28**	-.22
Proportional & Ratio	.40	.31	**.46**
Item Format			
All Strategies Correct Multiple-Choice	.27	.31	.31
Follow the Provided Strategy	.19	.34	.24
Explain a Likely Strategy	.07	.39	.41

The correlations between the item formats and the interview strategies were positive, albeit weak to moderate (.07 < *r* < .41). The interview for 'multiply and divide' and 'proportional and ratio' strategies correlated moderately with all three item formats, with average correlations of over .30.

In contrast to these low correlations, it is worth noting that the number knowledge stages and the strategy ratings in the interview generally have strong positive correlations; .00 < *r* < .82 with an average *r* = .51. Overall, the interview strategy and knowledge stages are consistently inter-correlated; suggesting a rater or method effect. In further contrast, the internal consistency of the questionnaire operational domains with the questionnaire item formats is quite low; .00 < *r* < .49 with an average *r* = .18.

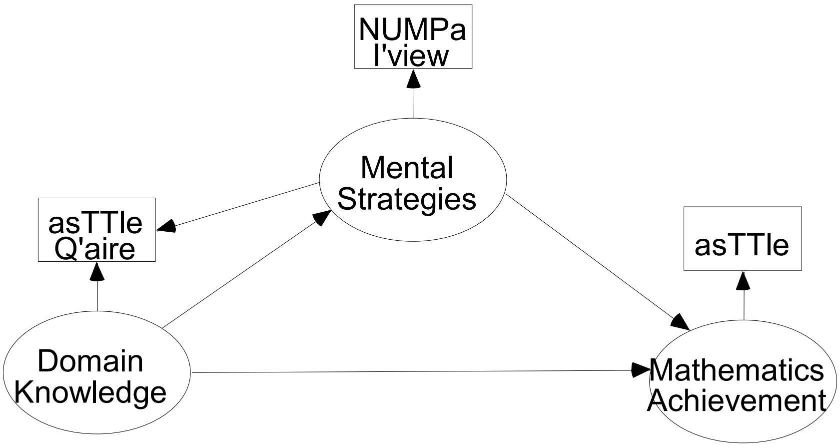
Discussion

The inter-method comparisons clearly indicate that the paper-and-pencil questionnaire and the oral interview ratings of student mental strategy in using the number framework strategies provide somewhat different information. Absolute scores are in general lower on the interview than on the questionnaire, while the consensus and consistency methods of measuring similarity clearly indicate significant difference. This may be important: The aim of the multiple-choice items and/or the interview items is to provide a robust method to best determine the highest level of processing, and it is clear from these results that the multiple-choice items are more likely to accomplish this aim. Furthermore, there are reasonably strong correlations between the interview strategy and knowledge stages, suggesting much consistency in scoring through the interview data collection method.

Of concern to this item development exercise is the large proportion of missing data and low levels of variance. This was despite efforts to ensure that the items could be easily read and answered by students in this year range and efforts to include students with a variety of skills and backgrounds. It may also be that when students of this age or year level are presented with paper-and-pencil problems, rather than using a variety of mental strategies, they resort to algorithmic processes. It is recommended that should any further work be carried out that more be done to reduce missing data and increase the variability within range of participants responding.

This research points to potential areas for improving the paper-and-pencil items. The moderate correlations between the interview knowledge stages (e.g., fractions, decimals, and group place value) and the questionnaire proportional and ratio domain items point to a potential connection of the paper-and pencil items to the number knowledge stages rather than to mental strategies. The moderate correlations between the proportion and ratio ratings in both questionnaire and interview formats suggest that more work with paper-and-pencil questions within that domain may be fruitful. The effectiveness of the multiple-choice all strategies correct items in correlating with the higher knowledge stages suggests that further research with that item format is justified. These data hint at how the various elements (i.e., knowledge—strategy, interview—paper-and-pencil) may be interrelated (Figure 5). It may well be, as the NUMPA research suggests, that knowledge stages contribute to strategy use and that developing both knowledge and strategy contributes to overall mathematics achievement. Given these data, it may be that the asTTle paper-and-pencil questionnaire items best predict knowledge stages, the NUMPA strategy interview ratings best predict strategy stages, and that both directly and indirectly predict achievement which best could be measured with asTTle's mathematics assessments.

Figure 5
Tentative Model of Relationships of Interview and Questionnaire to Number Strategy, Knowledge, and Achievement



With further refinements to the asTTle paper-and-pencil items along the lines suggested above, some steps toward testing such a theory could be taken. Any future investigations should also attempt to understand the interesting finding that higher number strategies seems to be reported in the context of paper-and-pencil items than with the interview ratings.

References

Brown, G.T.L., Glasswell, K., & Harland, D. (2004). Accuracy in the scoring of writing: The basis for confidence in using asTTle writing assessments. *Assessing Writing, 9*(2), 105-121.

Ell, F. (2001). *Maths in the New Zealand curriculum - A concept map of the curriculum document.* (Technical Report No. 11). Auckland, NZ: University of Auckland, Project asTTle.

Ell, F. (2002). *Mathematics strategies for number: Literature review.* (Technical Report No. 17). Auckland, NZ: University of Auckland, Project asTTle.

Fennema, E., Carpenter, T., et al. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27*(4), 403–434.

Fuson, K. C., Smith, S. T., LoCicero. (1997). Supporting Latino First Graders' Ten-structured Thinking in Urban Classrooms. *Journal for Research in Mathematics Education, 28*(6), 738–766.

Higgins, J. (2003). *An evaluation of the Advanced Numeracy Project 2002.* Wellington, NZ: Ministry of Education.

Irwin, K. C. (2003). *An evaluation of the Numeracy Project for Years 7—10, 2002.* Wellington, NZ: Ministry of Education.

Jones, G. A., Thornton, C. A. Putt, Hill, Mogill, Rich & Van Zoest. (1996). Multidigit number sense: A framework for instruction and assessment. *Journal for Research in Mathematics Education, 27*(3), 310–336.

Ministry of Education. (2003a). *The Number Framework*. Numeracy Professional Development Projects 2003 Book 1. Ministry of Education: Wellington.

Ministry of Education. (2003b). Assessment Tools for Teaching and Learning (asTTle) V2, 2003 [Computer software]. Wellington, NZ: Learning Media.

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation, 9*(4).

Thomas, G. & Ward, J. (2001). *An Evaluation of the Count Me In Too Pilot Project*. Wellington, NZ: Learning Media.

Thomas, G., Holton, D., Tagg, A., & Brown, G. T. L. (2002). *Numeracy item signature study: A theoretically derived basis.* (Technical Report No. 25). Auckland, NZ: University of Auckland, Project asTTle.

Thomas, G., Tagg, A., & Ward, J. (2003). *An evaluation of the Early Numeracy Project 2002.* Wellington, NZ: Ministry of Education.

van Gardaren, D. (2002). *Paper & pencil strategy assessment items for the New Zealand Number Framework: Pilot study.* (Technical Report No. 18). Auckland, NZ: University of Auckland, Project asTTle.