

Written and Oral Assessments of Secondary Students' Number Strategies: Ongoing Development of a Written Assessment Tool

Gregor Lomas
The University of Auckland
<g.lomas@auckland.ac.nz>

Peter Hughes
The University of Auckland
<pg.hughes@auckland.ac.nz>

This paper examines the trialling and piloting of a revised Written Strategy Stage Assessment Tool (WSSAT) designed to help identify students' strategy stages and provide formative data for teachers to use in their planning and teaching of numeracy. A year 9 and a 10 cohort of secondary school students trialled the written assessment mid-year, and then another year 9 cohort, from two other schools, piloted its use at the end of the year. Numeracy experts interviewed a sample of the end-of-year students to identify each student's strategy stage, using GloSS (Global Strategy Stage) oral-type protocols. Results from the written assessment gave relatively consistent measures of stages in terms of the criteria set and a relatively close match to national data sets. Comparison of the written and oral assessment results showed the stages identified by the two measures to be generally consistent. The results of this study suggest that the stages established by the WSSAT are likely to be useful for formative and diagnostic purposes.

Background

The extension of the Numeracy Development Projects (NDP) into secondary schools (as the Secondary Numeracy Project [SNP]) has led to a need for a "more efficient" assessment tool to replace the NDP oral assessment tool NumPA (Numeracy Project Assessment) so that an initial assessment of a whole class of students can be carried out. The use of an oral assessment in the primary school area avoids issues of students' reading levels and encourages the teasing out of students' understandings via oral interactions that move beyond the assessment script. NumPA is therefore relatively time consuming. The larger number of students in secondary schools who need assessing per teacher, due to multiple classes, and the expectation of adequate reading levels was the driver for developing the Written Strategy Stage Assessment Tool (WSSAT).

A written assessment tool would be efficient if it:

- reduced the amount of teacher time taken per student;
- provided a record of students' work that largely fulfils the same purposes (formative and summative) as the NumPA oral assessment;
- provided a standardised "marking" schedule.

A standardised marking schedule would enable consistent marking by someone new to the SNP or even by a person with little knowledge of numeracy, such as a parent helper. It would also reduce the individual variability that can occur, for example, when conducting oral assessments, because the written record would be available for re-evaluation and moderation purposes.

Another benefit might be enhanced accuracy of the assessment, particularly in the multiplicative and proportional domains, for which Thomas, Tagg, and Ward's 2006 data set the accuracy of various types of existing numeracy assessments at 76%. Thomas et al. (2006) also found that many teachers "rated students' strategy stages lower than the rating of the researchers, explaining their decisions in terms of consolidating students' understanding at an existing level" (p. 101), that is, they rated students at a lower stage than NumPA would have assigned.

The Nature and Structure of the Written Strategy Stage Assessment Tool

The WSSAT, which is being developed by Peter Hughes at The University of Auckland, aims to identify a global strategy stage (as does the Global Strategy Stage [GloSS] tool) rather than domain-specific strategy stages (as the NumPA does). The approach taken is closely aligned with, and arises from, the NDP strategy and knowledge sections of the Number Framework but consists of short-answer and multi-choice items that draw upon the approaches used in the NDP assessment tools and research findings in related areas (Lomas & Hughes, 2008).

The focus of the WSSAT is primarily strategy stage identification, with seven items at each stage assessing a range of strategies and further items assessing some knowledge aspects (see Table 1). The knowledge items are a mixture of prerequisite knowledge for, and knowledge directly related to, each stage. To achieve a strategy stage, students need to correctly answer four of the strategy items for that stage. As with the NDP oral assessments, the highest strategy stage achieved is taken as that student's stage. For example, if a student meets the criteria for stages 5, 6, and 7, they are classified as being at stage 7, or if they meet the criteria for stage 5 and not that for stage 6, but do meet the criteria for stage 7, they are classified as being at stage 7. Where the students do not meet the criteria for stage 5 or higher, they are assigned to a category that covers stages 1–4.

Table 1
The Number of Items per Strategy Stage and Criteria for Achieving Each Stage

	Strategy Stages				
	1–4	5	6	7	8
Total number of items (including seven strategy items)	11	11	12	11	10
Number of knowledge items	4	4	5	4	3
Number of correct items to assign to a strategy stage	< 4	≥ 4	≥ 4	≥ 4	≥ 4

Note: The items used to assess stages 1–4 are the same items as for stage 5.

The WSSAT as a written assessment relies on the written answers without any indication of the process used. This is different from NDP oral assessments, in which students give their answer orally and often talk through the process they used to arrive at their answers. Thus, assigning strategy stages by the oral assessments relates largely to process, whereas assigning strategy stages using the WSSAT is based on outcomes (Lomas & Hughes, 2008).

The WSSAT answer sheet has clear directions for the students to follow and was formatted for ease of marking, giving both a quick indication of a student's strategy stage and more detailed formative data for planning and teaching purposes. The answer sheet has room only for the answers because what was sought was the result of mental processes rather than the process itself or written algorithms.

The range of items selected for each stage attempted to isolate and encapsulate some of the conceptual aspects and elements of strategy relating to that stage as per the domains of the strategy section of the Number Framework. For example, item 28 (see Appendix E [pp. 90–93], Part C), "Work out 5 sixths of 42", and item 8 of the oral assessment (see Appendix F [p. 94], Section C), "What is 3 fifths of 35?", relate to stage 7 of the proportions and ratios domain, in which students are expected to use multiplication strategies to solve problems with fractions.

In addition, the nature of the WSSAT items was designed to reflect elements of students' understanding (process) that might be present in an oral dialogue between teacher and student but would not always be evident in written work (which focuses mainly on outcomes). Thus, a key issue in the WSSAT was

to ensure that the items that were used “forced” the student participants to use a particular process and restricted their use of any other approach. That is, the WSSAT minimises the number of items that could be answered procedurally or answered by less sophisticated strategies. For example, some of the written assessment items use combinations of numbers written as words and digits, for example, “Add one-tenth to 4.273” (see Appendix E, Part B, item 17). The aim is to expose the student’s in-depth understanding by stating material in a way that limits the use of procedural methods and requires more understanding of number structure.

This approach is also seen as a way of keeping aspects of “oral” language use within a written format, with, for example, “one-tenth” giving a potentially greater access to the underlying meaning than “zero point one” in symbolic form (0.1), while still requiring the students to be able to connect the underlying meanings.

An example of an item that tries to “force” a particular strategy is item 1 (see Appendix E, Part A), “ $87 + 99 =$ ”. Ninety-nine is close to a tidy number, and the most likely solution consists of “one step”: making the 99 up to 100 and reducing 87 down to 86 at the same time, resulting in 186 as the answer. The choice of the particular numbers lessens the possibility of students using counting on, which could happen, for example, if $48 + 8$ was used, or of them using a variety of different strategies, for example, if $87 + 88$ was used. The extent to which this approach has worked will be determined in part by the extent to which the assignment of stages via the written assessment matches the oral assessment.

The sets of items also have the potential to provide a more detailed and standardised diagnostic map of a student’s learning needs than an oral assessment. This potential is enhanced by the written format, which allows a student to attempt all the items, thus demonstrating any partial understandings that the student might have (with some items answered correctly) beyond the point where an oral assessment would stop.

2007 Trial Findings

The findings from the trial of the initial version of the WSSAT in 2007 showed that the written assessment, while consistent, did not give stage information that reflected the parallel oral assessment, the national curriculum level expectations, or, more specifically, the low-decile data (Lomas & Hughes, 2008). On this basis, a number of changes were made to the organisation of the items within each part of the WSSAT (reflecting a stage) and the suitability/placement of items in each part.

The organisation of items within each part was changed to create two sections, with the first section consisting of strategy items and the second of knowledge items. The number of strategy items was standardised for each part, as was the number of correct responses required to achieve a stage. The placement of items in each part was reconsidered in terms of each item’s suitability for assessing that stage from a theoretical viewpoint (alignment with the strategy stages of the strategy section of the Number Framework) and, in some instances, in terms of the number of correct responses in each item’s current placement. For example, “How would you write eleven thousandths as a decimal fraction?”, which only 53 students out of 278 answered correctly in the 2007 trial, was shifted from the stage 6 (Part B) to the stage 8 (part D) section, where it was thought that more students would have the understanding required to function at that stage. It was also rewritten in multi-choice format, “Eleven thousandths equals:” with four choices offered: “A. 0.0011 B. 0.011 C. 0.11 D. 11000”. This was done for two main reasons: to avoid possible confusion about what “decimal fractions” meant and to avoid the answer “11 over (divided by) 1000”, which “side steps” the issue of understanding decimal fraction notation.

Method

The revised WSSAT was trialled mid-year to evaluate the revisions and to identify further aspects for revision on the basis of the data gathered. The further revised WSSAT was then piloted with two schools at the end of 2008, and a parallel oral assessment interview was conducted with a sample from one of the schools. Due to the site-specific nature of the data collected, this research is a form of case study. Thus, the data is unlikely to match the national data sets too closely and care must be taken in generalising any findings.

The participants in the mid-year trial were drawn from the year 9 and 10 cohorts of a large, Auckland, decile 5, girls' secondary school of mixed ethnic composition, with a small number of special needs students being excluded.

The participants in the end-of-year pilot were drawn from two schools: a year 9 cohort of a large, Auckland, decile 3 secondary school of mixed ethnic composition, excluding some special needs students, and the complete year 9 cohort of a medium-sized, Wellington, decile 6 secondary school of mainly New Zealand European students (see Table 2).

Table 2

Percentage Ethnic Make-up of Auckland and Wellington Year 9 Students Included in the Pilot Part of This Study

	NZ European	Māori	Pasifika	Other
Auckland (n = 280) (low decile)	26	18	40	15
Wellington (n = 113) (middle decile)	70	15	7	8

The written assessment was given to both year 9 pilot cohorts, while the oral assessment was given only to a subset (60 students) of the year 9 Auckland cohort. This sample was drawn from several classes from the bands in which the school organised their classes (see Table 3).

Table 3

Auckland School Classes in Bands (High to Low), Showing Student Roll Numbers and the Number of Students Taking the Pilot Written and Oral Assessments

	Pilot Class Name												Total
	P9A1	P9A2	P9B1	P9B2	P9B3	P9B4	P9C1	P9C2	P9C3	P9C4	P9D1	P9D2	
Roll No.	33	32	33	33	33	32	25	20	27	25	28	27	348
Written Ass. No.	26	26	28	24	25	22	25	20	23	16	21	24	280
Oral Ass. No.	10	20	–	–	–	–	–	–	19	11	–	–	60

The interviewer worked around the Auckland school's programme, so the classes accessed were those that least inconvenienced the school. This affected the nature of the sample, which is not representative because it was drawn from upper- and lower-band classes and included no students from the middle-band classes.

The Oral Assessment

The oral assessment research tool (see Appendix F, p. 94) was an expanded form of the GloSS and used some of the GloSS- and NumPA-type items, supplemented by other items that gave increased coverage of higher stages. It also offered the potential for allowing a finer measure of a student's position within a stage, that is, beginning, middle, or later (nearly ready to move to the next stage). As well as the questions being asked orally, a card with the question written on it was placed in front of the student as a reference (as is done with GloSS and NumPA).

Data Collection

For both the trial and the pilot, the WSSAT was generally conducted in each class's usual classroom setting, usually under the supervision of the regular mathematics teacher, in the last few weeks of the second and fourth terms respectively. Standardised instructions were given out explaining how teachers were to conduct the assessments (Lomas & Hughes, 2008), and all the answer sheets were marked by one of the research team to ensure consistency. Copies of the marked answer sheets were returned to the school for potential diagnostic/formative use by the school.

For the Auckland pilot school, a GloSS-type oral assessment was given during the two days after the written assessment to a sample of 60 students, drawn roughly, and at the school's convenience, from across the "ability" bands into which the year 9 cohort was organised (see Table 3). The oral interviews were conducted by external interviewers who had expert knowledge of NDP.

Analysis

The results of the trial and pilot written assessments were first analysed for the internal consistency of the tool in identifying a student's stage, that is, whether a student assigned as being at stage 6 had also been assigned as being at stage 5, and so on. Then they were analysed against three other measures of achievement, one school-based and two based on nationally collected data from the NDP, which give measures of global, rather than domain-specific, strategy stages. The stages that the students achieved were compared with:

- the banding (where applicable) of the class they were in, to see whether this reflected the school's placement of students;
- the national, year 9, low- or middle-decile stage distribution data from the SNP;
- the national, year 9, stage distribution data (the numeracy curriculum expectations).

Issues identified in the mid-year trial were taken into consideration in a further minor revision of the WSSAT prior to this revised version (see Appendix E) being piloted at the end of the year.

The results of the written and oral assessments from the Auckland pilot school were compared to establish a relationship between these two forms of assessment. The oral assessment was assumed to be the more accurate and was taken as the baseline for the comparison due to its alignment with national data collection methods. This assumption was based on two main factors. Firstly, the oral assessment was an extension of the NumPA tool and thus was collecting some of the same data. Secondly, the extra questions were provided by a numeracy expert with an intimate knowledge of the development and use of both GloSS and NumPA. This connection to these existing and "proven" NDP assessment tools provided a sound basis for comparison with the database of student results arising from the WSSAT's use.

Results

The Trial Data

The trial data was collected from both year 9 and year 10 students to meet the school's requirements. Although this was a mid-year rather than an end-of-year assessment, the groups fall either side of the national end-of-year data and so can be related to it, although not directly. The year 9 and the 10 trial data was analysed separately to allow direct comparison of the year 9 data with that of the 2007 trial and of both year 9 and 10 data with the national data.

Internal consistency

Of the 74 year 9 students who could be assigned as being at stage 6 on the basis of the written assessment, only two had not also achieved the criteria (missing by one correct response) for being assigned as being at stage 5. For the 62 students assigned as being at stage 7, only six (almost one-tenth) had not achieved the criteria for being assigned as being at stage 6. Of these six, five had missed the criteria for stage 6 by one correct response and one by two correct responses. For the 56 students who could be assigned as being at stage 8, 18 (almost one-third) had not achieved the criteria for earlier stages. Of the 14 who missed only stage 7, five had missed by one correct response and five by more than one. The other four had not been assigned either stage 6 or stage 7.

Of the 57 year 10 students who could be assigned as being at stage 6 on the basis of the written assessment, only three had not also achieved the criteria (missing by one correct response) for being assigned as being at stage 5. For the 55 students assigned as being at stage 7, only ten (just over one-fifth) had not achieved the criteria for being assigned as being at stage 6. Of these ten, five had missed the criteria for stage 6 by one correct response and five by two correct responses. For the 52 students who could be assigned as being at stage 8, five (almost one-tenth) had not achieved the criteria for earlier stages. Of the four who missed only stage 7, two had missed by one correct response and two missed by two; the fifth student had not been assigned as being at any of stages 5, 6, or 7.

All this data suggests that the WSSAT has a high level of internal consistency in assigning students at stages except at stages 7 and 8, in which a greater level of variation was evident. This indicates that further revision of these stages should be investigated.

Conformity of assigned stages with students banding into classes

The stages assigned by the trial WSSAT generally conformed to the banding of the classes: students in classes in higher bands achieved more of the higher stages, and students in classes in lower bands achieved fewer of the higher stages. Additionally, in line with the internal consistency of the WSSAT, as discussed above, the meeting of the criteria for particular stages also aligned with the banding of the classes, with fewer students from lower-band classes meeting the criteria for each stage. For example, in year 9, class 9A2 had 16 students meeting the stage 8 criteria, 23 the stage 7 criteria, and all 26 the stages 6 and 5 criteria, compared with class 9B2, in which one student met the stage 8 criteria, 10 the stage 7 criteria, 18 the stage 6 criteria, and 24 the stage 5 criteria, while class 9D1 had only one student meet the stage 8 criteria, none the stage 7 criteria, three the stage 6 criteria, and 14 the stage 5 criteria (see Table 4). Similar patterns can be seen in the year 10 data of classes 10A2, 10B2, and 10D1, although more students from these classes achieved higher stages.

Table 4

Classes in Band Order, Showing the Number of Year 9 and 10 Students Taking the Trial WSSAT and the Number of Students Meeting the Criteria for Achieving a Particular Stage

Year 9	Trial Classes													Total
	A1	A2	B1	B2	B3	B4	B5	B6	B7	B8	B9	D1	D2	
No. of students	26	26	25	24	23	17	20	27	12	16	26	18	15	275
No. meeting stage 5	26	26	23	24	23	17	19	27	10	15	25	14	15	264
No. meeting stage 6	26	26	12	18	15	14	13	16	3	11	17	3	3	177
No. meeting stage 7	21	23	6	10	7	6	4	6	1	3	9	0	4	100
No. meeting stage 8	15	16	0	1	4	2	4	5	3	2	3	1	0	56
Year 10	A1	A2	B1	B2	B3	B4	B5	B6	B7	B8	B9	D1	D2	Total
No. of students	29	26	21	26	19	25	21	22	-	-	-	13	15	217
No. meeting stage 5	29	26	20	24	19	24	19	21	-	-	-	11	14	207
No. meeting stage 6	29	25	14	17	14	9	16	18	-	-	-	7	2	151
No. meeting stage 7	25	18	9	15	9	5	9	11	-	-	-	1	0	102
No. meeting stage 8	14	15	6	2	2	2	4	7	-	-	-	0	0	52

This data suggests that the banding of classes is reflected in the trial WSSAT measurement of the students' numeracy achievement and the assigning of stages.

Comparison with national numeracy curriculum expectations

The assigning of stages from the written assessment gave rise to a distribution reasonably similar to both the low-decile schools' data and the national expectations (see Table 5) for year 9. While the year 10 percentage data shows a small shift to higher stages (23% at year 9 compared with 25% at year 10 for stage 7 and 20% compared with 24% for stage 8), as expected, the increases are not large. This may well reflect the greater difficulty of the stage 7 and 8 material and the longer time needed to "master" that material.

The area of greatest disparity (around a 50% difference or more) with the middle-decile data for year 9 students at the end of the year is the higher number of middle-decile year 9 trial students assigned as being at stage 8 (20% compared with 10%). The areas of greatest disparity for the year 9 (mid-year) trial students compared with the year 9 national expectations related to the lower number of year 9 trial students assigned as being at stage 7 (23% compared with 40%) and the higher number of year 9 trial students assigned as being at stage 5 (27% compared with 12%). Similar disparities are also evident in the year 10 data. The disparities may partly reflect a difference between students from middle-decile schools and the national expectations, or they may be a feature of the particular school's population.

Table 5

The Percentage of Stages Assigned to Trial Students by the WSSAT and the Middle-decile and National Numeracy Curriculum Expectation Data for Year 9 Students (End of Year)

	Stages				
	1-4	5	6	7	8
Percentage of year 9 students (n = 275)	4	27	27	23	20
Percentage of year 10 students (n = 217)	3	22	27	25	24
Middle-decile ¹ (average ²) percentage end year 9 (Tagg & Thomas, 2008)	6	22	32	30	10
National numeracy curriculum expectation percentage end year 9 (Tagg & Thomas, 2008)	2	12	25	40	21

The patterns evident in this data are considerably different from those found in the 2007 trial, where few students were assigned as being at stage 7 (7%) and none at stage 8, while 46% were assigned as being at stages 1–4, 13% at stage 5, and 34% at stage 6 (Lomas & Hughes, 2008). This suggests that the revision of the 2007 version of the WSSAT has at least partially addressed the major concern identified in the 2007 trial, namely the mismatch between stages assigned by the WSSAT and those evident in the middle-decile and national expectations data, while maintaining high levels of internal consistency.

As a consequence of the higher levels of inconsistency at stages 7 and 8, some minor changes were made to some of the strategy items in parts C and D of the WSSAT (see Appendix E). Only three items in each of parts C and D were changed, while one further item (30) in Part C was reformatted (fraction answers were replaced by pairs of whole numbers) to reduce the operational demand.

In an attempt to enhance consistency, items covering the same strategy but at different levels were created. For example, trial item 35, which deals with a multiplication estimation, became pilot item 24; the new pilot item 35 deals with a division estimation. The pilot item 25, “98 × 5”, had easier numbers that replaced the trial numbers “999 × 9”, and trial item 37, dealing with fraction addition, was replaced with an item similar to item 25 but with harder numbers, “9998 × 5”. Trial item 36, dealing with fractions, was replaced with a harder version of item 28, and trial item 29, dealing with fractions but requiring two answers, was replaced with a multiplication item requiring only one answer.

The Pilot Data

The pilot data was collected from a year 9 cohort in a low-decile school in Auckland and another from a middle-decile school in Wellington. The data for each cohort is analysed separately in this paper to allow direct comparison with the year 9 national data.

Internal consistency

All the Auckland students assigned as being at stage 6 had also been assigned as being at stage 5, while of the 84 students who could be assigned as being at stage 7 based on the written assessment, 21 (one-quarter) had not achieved the criteria for stage 6. Of these 21, 15 had missed the criteria by only one correct response. A further two students who achieved the criteria for being assigned as

¹ The middle-decile and national curriculum expectations data in tables 5, 9, and 10 come from Secondary Numeracy Project data collected by researchers. The middle-decile percentages are an average figure derived from the percentage data from the additive, multiplicative, and proportional domains.

² The average figure (as in i.) gives a more global strategy stage rather than being specific to domains.

being at stage 7 had not achieved the criteria for either stages 5 or 6. For the 46 students assigned as being at stage 8, only ten (over one-quarter) had not achieved the criteria for stage 7, and one further student had not achieved the criteria for either stages 6 or 7. Of the ten not achieving the criteria for stage 7, four had missed by only one correct response.

All the Wellington students assigned as being at stage 6 had also been assigned as being at stage 5, while of the 39 students who could be assigned as being at stage 7 based on the written assessment, eight (almost one-fifth) had not achieved the criteria for stage 6. Of the eight, three had missed the criteria by only one correct response. For the 21 students assigned as being at stage 8, only one had not achieved the criteria for stage 7.

This data suggests that the WSSAT was largely internally consistent in assigning stages except at stages 7 and 8, where a greater level of variation was evident (although less variation was evident in the middle-decile Wellington school data).

Conformity of assigned stages with students banding into classes

The stages assigned by the WSSAT generally conformed to the banding of classes in the Auckland school used in the pilot: classes in higher bands achieved more of the higher stages, and classes in lower bands achieved fewer of the higher stages. Additionally, in line with the internal consistency of the WSSAT, as discussed above, the meeting of the criteria for particular stages also aligned with the banding of the classes, with fewer lower-band students meeting the criteria for each stage. For the Auckland students, for example, class P9A2 had 11 students meet the stage 8 criteria, 20 students the stage 7 criteria, 25 the stage 6 criteria, and 26 the stage 5 criteria, compared with class P9B2, in which six students achieved the stage 8 criteria, 13 the stage 7 criteria, 17 the stage 6 criteria, and 23 the stage 5 criteria, while class P9D1 had only two students meet the stage 7 criteria, five the stage 6 criteria, and 19 the stage 5 criteria (see Table 6).

Table 6

Auckland School Classes in Band Order, Showing the Number of Students Taking the Pilot WSSAT and the Number of Students Meeting the Criteria for Achieving a Particular Stage

	Auckland Pilot Classes												Total
	P9A1	P9A2	P9B1	P9B2	P9B3	P9B4	P9C1	P9C2	P9C3	P9C4	P9D1	P9D2	
No. of students	26	26	28	24	25	22	25	20	23	16	21	24	280
No. meeting stage 5	26	26	28	23	25	22	25	15	22	15	19	22	268
No. meeting stage 6	25	25	21	17	16	16	15	2	15	6	5	4	167
No. meeting stage 7	23	20	11	13	14	5	5	5	11	7	2	4	120
No. meeting stage 8	16	11	2	6	1	4	4	0	1	0	0	1	46

In the case of the Wellington school, the classes were not banded and included the entire cohort (see Table 7). There was no evidence of any particular factors that offered an explanation for the variation of class W9-2, which had higher numbers and proportion of students assigned at stage 8 than the other classes.

Table 7

Wellington Classes, Showing the Number of Students Taking the Pilot WSSAT and the Number of Students Meeting the Criteria for Achieving a Particular Stage

	Wellington Pilot Classes						Total
	W9-1	W9-2	W9-3	W9-4	W9-5	W9-6	
No. of students	17	22	18	19	21	16	113
No. meeting stage 5	17	22	18	19	21	16	113
No. meeting stage 6	12	21	8	11	12	9	73
No. meeting stage 7	7	21	7	8	11	5	59
No. meeting stage 8	3	14	1	2	0	1	21

Comparison between oral and written assessments

The stages determined by the pilot WSATT closely matched the stage determined by the oral assessment of students at stage 8 but less so at other stages (see Table 8). A third of students achieving stage 5 on the oral assessment achieved stage 6 on the WSSAT, and two-thirds of students achieving stage 6 on the oral assessment achieved stage 7 on the WSSAT. However, there were no stage differences of more than one stage, unlike the 2007 trial data (Lomas & Hughes, 2008). This may reflect the more even spread across stages achieved by the 2008 revised WSSAT versions.

Table 8

Comparison of Stages Assigned to Auckland Students by the Pilot WSSAT Compared with the Oral Assessment

	Number of students (n = 60)												
	2	1	-	9	5	-	5	9	-	7	1	3	18
Oral assessment stage	1-4	1-4	5	5	5	6	6	6	7	7	7	8	8
WSSAT stage	1-4	5	1-4	5	6	5	6	7	6	7	8	7	8

Comparison with national numeracy curriculum expectations

The assigning of stages from the WSSAT for the pilot schools gave rise to a distribution reasonably similar to the data for both the middle- and low-decile schools respectively and to the national expectations (see tables 9 and 10).

For the middle-decile year 9 cohort used in the pilot, the areas of greatest disparity with the middle-decile data for end-of-year year 9 students were the higher number of students achieving at stage 8 (19% compared with 10%) and the lower number of students achieving at stages 1-4 (0% compared with 6%) in the trial data. The greatest area of disparity for the middle-decile year 9 students compared with the year 9 national expectations was the higher number of students achieving at stage 5 (28% compared with 12%) in the trial data.

Table 9

The Percentage of Stages Assigned to Wellington Students from the WSSAT and the Middle-decile and National Numeracy Curriculum Expectation Data for Year 9 Students (End of Year)

	Stages				
	1–4	5	6	7	8
WSSAT: percentage of students (n = 113)	0	28	19	34	19
Middle-decile ³ (averaged) percentage at end year 9 (Tagg & Thomas, 2008)	6	22	32	30	10
National numeracy curriculum expectations percentage year 9 (Tagg & Thomas, 2008)	2	12	25	40	21

For the low-decile year 9 cohort, the areas of greatest disparity (around a 50% difference or more) with the low-decile data for end-of-year year 9 students were the higher number of students achieving at stage 8 (16% compared with 5%) and the lower number of students achieving at stages 1–4 (4% compared with 11%) in the trial data (see Table 10). This may partly be explained by the exclusion of a group of lowest-performing year 9 students from the data collection process.

The area of greatest disparity for the low-decile year 9 students with the year 9 national expectations was the lower number of students achieving at stage 5 (28% compared with 12%) in the trial data. This may partly reflect a difference between low-decile students and a national expectation, although a similar disparity was evident in the comparison for the middle-decile data (see above).

Table 10

The Percentage of Stages Assigned to Auckland Students for Each Assessment Tool and the Low-decile and National Numeracy Curriculum Expectation Data for Year 9 Students (End of Year)

	Stages				
	1–4	5	6	7	8
WSSAT: percentage of students (n = 280)	4	28	23	30	16
Oral assessment: percentage of students (n = 60)	5	23	23	13	35
Low-decile ⁴ (averaged) percentage at end year 9 (Tagg & Thomas, 2008)	11	29	33	22	5
National numeracy curriculum expectations percentage year 9 (Tagg & Thomas, 2008)	2	12	25	40	21

The oral assessment's assigning of stages to students is reasonably close to the national expectation percentages for all stages except those achieving at stages 6 and 7. However, the sample was skewed in that it took half from the upper band and half from the lower band without sampling the middle band, where more students achieved at stages 6 and 7. If the consistency of the stages assigned with the banding of classes (see Table 6) had been taken into account, a less skewed sample would have shown a closer fit overall.

³ The middle-decile percentages are average figures derived from the respective percentage data for the additive, multiplicative, and proportional domains.

⁴ The low-decile percentages are average figures derived from the respective percentage data for the additive, multiplicative, and proportional domains.

Discussion

A significant factor to consider in comparing the WSSAT and oral assessment results with the national data sets is the degree to which they accurately represent the stages that the students at year 9 have achieved. This is highlighted by Thomas et al.'s (2006) research findings on the accuracy of secondary teachers' assessments (76%) and teachers' tendency to assess at lower levels based on their perception of students' needs rather than actual performance on a numeracy assessment tool. This would suggest an underestimation of student performance overall, but possibly more so at stages 7 and 8, in which the learning demands are greater. This may be evident in the trial data, where there is nearly twice the percentage of students achieving at stage 8 compared to the national middle-decile data, although the percentage of students achieving at stage 7 in the national middle-decile data is higher. The possibility of such a trend is more apparent in both the medium- and low-decile pilot data. For example, compared with the national middle-decile data, there are three times more students achieving at stage 8 and one and a half times more students achieving at stage 7, but only about one-third fewer students achieving at stage 6.

The trial of the revised WSSAT gave a better spread of assigned stages and highlighted a number of issues relating to consistency across stages 7 and 8. With the pilot version, there appears to have been an improvement in consistency across stages 7 and 8 due to the minor revisions to stage 7 and 8 items but less consistency across stages 6 and 7. However, overall, the WSSAT has reasonably high levels of internal consistency for stages 5–8 and can be used to assign students a numeracy (global) strategy stage. In addition, there is a reasonable congruence of the stages assigned with both the low- and middle-decile school data and the national numeracy curriculum expectations and with the oral assessment assignment of NDP stages. Thus, the WSSAT, in its pilot form, determines a student's numeracy strategy stage with a reasonable degree of accuracy (and consistency) for formative and diagnostic purposes.

References

- Lomas, G., & Hughes, P. (2008). Written and oral assessment of secondary students' number strategies: Developing a written assessment tool. In *Findings from the New Zealand Secondary Numeracy Project 2007* (pp. 32–38). Wellington: Learning Media.
- Tagg, A., & Thomas, G. (2007). Do they continue to improve? Tracking the progress of a cohort of longitudinal students. In *Findings from the New Zealand Numeracy Development Projects 2006* (pp. 8–15). Wellington: Learning Media.
- Tagg, A., & Thomas, G. (2008). Performance of SNP students on the Number Framework. In *Findings from the New Zealand Secondary Numeracy Project 2007* (pp. 5–16). Wellington: Learning Media.
- Thomas, G., Tagg, A., & Ward, J. (2006). Numeracy assessments: How reliable are teachers' judgments? In *Findings from the New Zealand Numeracy Development Projects 2005* (pp. 91–102). Wellington: Learning Media.