ASSESSMENT OF SECONDARY STUDENTS' NUMBER STRATEGIES: THE DEVELOPMENT OF A WRITTEN NUMERACY ASSESSMENT TOOL

GREGOR LOMAS
University of Auckland
g.lomas@auckland.ac.nz

PETER HUGHES
University of Auckland
pg.hughes@auckland.ac.nz

This paper examines the piloting of a *Written Strategy Stage Assessment Tool* designed to identify students' "global" strategy stages and provide formative data for teachers. The Year 9 cohorts from two schools were assessed at the end of the school year. Numeracy experts then interviewed a sample, to identify each student's strategy stage, using an oral assessment. 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. Comparison of the written and oral assessment results showed the stages identified by the two measures to be generally consistent.

Background

The extension of the New Zealand *Numeracy Development Projects* (NDP) from primary into secondary schools, as the *Secondary Numeracy Project* [SNP], led to a call for a written assessment tool to replace the NDP oral assessment tool (NumPA) as the main source of initial information about secondary students' mathematical knowledge and understanding. NumPA information is used to place students initially into stage related teaching groups, as promoted in NDP, and new written assessment information would need to be able to be used in the same way.

A written tool was seen as enabling the initial assessment of whole classes of secondary students to be carried out more "efficiently". The use of oral individual student assessment, as occurred in primary schools, was seen as an unproductive use of teacher time given the larger numbers of students that secondary teachers had to deal with due to their taking multiple classes.

The use of oral assessment negates the challenge of inadequate student reading levels, potentially a major issue with primary school students, whereas reading levels were not seen as problematic with secondary school students. Thus, the larger number of students in secondary schools who need assessing per teacher and the expectation of adequate reading levels were the prime drivers for developing the *Written Strategy Stage Assessment Tool* (WSSAT).

An efficient written assessment tool would:

- reduce the amount of teacher time taken per student;
- allow a standardised (and objective) marking schedule; and
- give a written record of students' answers (without any working).

Most importantly it would give appropriate global strategy stages and useful formative information, reflecting that from NumPA.

The WSSAT, as can be attempted with any written test, was designed to meet the three criteria. The time to complete was aimed at about 30 to 40 minutes, which for classes of 30 (or whole school cohorts) would use significantly less teacher time than individual oral assessments. There was a standardised, straightforward marking schedule (correct/incorrect), enabling quick, consistent and objective marking—particularly by teachers new to SNP or even by a person with little knowledge of numeracy, such as a parent helper. The responses on a carefully formatted answer sheet allowed for quick marking as well as determination of the strategy stage, while providing a clear written record of the students' performance. Experience during the earlier trials and the pilot in this study have indicated that these efficiency aspects of the design have been achieved.

In addition, the availability of the written record for evaluation and moderation purposes reduces the potential for individual variability that can occur when conducting oral assessments. It may also enhance the accuracy of the assessment, particularly in the multiplicative and proportional domains, for which Thomas, Tagg, and Ward (2006) set the accuracy of various types of existing numeracy assessments at 76%. Thomas et al. also found that many secondary 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 if used as intended.

The extent to which the WSSAT achieves a reasonably "useful" determination of students' strategy stages for the initial placement of students into teaching groups is the focus of this paper.

The nature and structure of the written strategy stage assessment tool

The WSSAT is closely aligned with the NDP strategy and knowledge frameworks, which are organised in three and four domains respectively, with a focus on the higher stages (5 to 8) dealing with part-whole thinking (Ministry of Education, 2008). It aims to identify a global or overall strategy stage (as does the NDP Global Strategy Stage [GLoSS] tool) rather than domain-specific strategy stages (as does NumPA), and uses some similar items. Other items consist of short-answer and multi-choice questions on place value and decimals (Brown, Hart & Dietmar, 1984; Hart, 1981) and on number sense (Lomas & Hughes, 2008, 2009; McIntosh, Reys, Reys, Bana, & Farrell, 1997).

The focus of the WSSAT is primarily global strategy stage identification rather than the identification of specific strategies. There are seven domain specific strategy items at each of four stages (5 to 8) assessing a range of strategies and further items assessing some knowledge aspects. The knowledge items are a mixture of prerequisite knowledge for, and knowledge directly related to, each stage. To achieve a global strategy stage, students need to correctly answer four of the domain specific strategy items for that stage. Four correct items are used to assign an overall (global) stage as they indicate the mastery of strategies, at a particular stage, in some but not necessarily all of the three strategy domains (addition and subtraction, multiplication and division, and proportions and ratios). That is, there is a range of evidence about strategy mastery that underpins the assignment of an overall strategy stage.

As with the NDP oral assessments, the highest strategy stage achieved is taken as that student's stage for teaching purposes. 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 those 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 WSSAT criteria for stage 5 or higher, they are assigned to a category that covers stages 1–4.

The WSSAT as a written assessment relies on the written answers without any indication of the process (or strategy) used. This is different from NDP oral assessments, in which students give their answers orally and can be prompted to talk through the processes they used to arrive at their answers thus revealing the particular strategies used. Therefore, using oral assessment to assign strategy stages can relate more to process and particular strategies, whereas assigning strategy stages using the WSSAT is based solely on outcomes (Lomas & Hughes, 2008).

The WSSAT answer sheet has clear directions for the students to follow with space for answers only. It 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. Calculator use was not allowed, mental working was promoted, and writing ("working"), other than for recording answers, discouraged.

The range of items selected for each stage attempted to isolate and encapsulate some of the conceptual aspects and strategy elements relating to that stage as per the domains in the strategy section of the NDP Number Framework. For example, item 36, 'Work out 5 sixtieths of three thousand six hundred' relates to stage 8 of the proportion and ratio 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 that might be present in a dialogue between teacher and student but would not always be evident in written work with an outcome focus. Thus, a key issue in the WSSAT was to ensure that items used "forced" the student participants to use a particular process and restricted their use of any other approach. That is, the WSSAT attempts to minimise the number of items that could be answered procedurally or answered by less sophisticated strategies. For example, some of the written assessment items such as item 3, '7 tens + 20 =' and item 28, 'Work out 5 sixths of 42' use combinations of numbers written as words and digits. Another example is 'Eleven thousandths equals:' with four choices offered: 'A. 0.0011 B. 0.011 C. 0.11 D. 11000'. This format avoids the answer '11 over (divided by) 1000', which "side steps" the issue of understanding decimal fraction notation. The aim is to expose the student's in-depth understanding by stating material in a way that both limits the use of procedural methods and requires more understanding of number structure. This approach was also seen as a way of keeping aspects of oral language use within a written format.

An example of an item that tries to force a particular strategy is item 14, '341 - 'what' = 299'. Here, 299 is close to a tidy number, and the most likely NDP solution consists of making the 299 up to 300 and increasing 341 to 342, resulting in 42 as the answer. The choice of these particular numbers lessens the possibility of students using a strategy such as doubling if numbers such as '51 - what = 26' were used. The extent to which this approach has worked will be determined in part by the alignment of the stages assigned by the written assessment to the national data and oral assessment.

The sets of items also have the potential to provide a more detailed and standardised diagnostic map of students' 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 pieces of knowledge and strategy understandings that the student might have beyond the point where an oral assessment would stop.

Initial development of the WSSAT

The WSSAT was trialled twice during the initial development phase giving rise to the version examined here. The initial trial demonstrated sufficient internal consistency but the assigned stages did not match sufficiently the parallel oral assessment, the national curriculum level expectations, or, more specifically, the low-decile[1] 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), the positioning of items in each part, and the style of some items (Lomas & Hughes, 2008). The trial of the second version gave rise to further minor revisions to enhance consistency for parts C and D, that is, Stages 7 and 8 (Lomas & Hughes, 2009).

Method

The revised WSSAT was piloted with two schools, 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—particularly the national aspirational expectations—too closely, and care must be taken in generalising any findings.

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

The written assessment was only given to the students present on the particular day for each Year 9 cohort, while the oral assessment was later given to a subset (60 students) of the Year 9 Auckland students who had responded to the WSSAT. This sample was drawn from several classes from the four bands (strands/streams) into which the school organised their classes (see Table 1) at the schools convenience. This affected the nature of the sample, which is not representative because it was drawn equally from upper- and lower-band classes and included no students from the middle-band classes.

Table 1. Auckland school classes in bands (high to low), showing student roll numbers and the number of students participating in the WSSAT and oral assessment.

	Class Name (Auckland)												
	P9A1	P9A2	P9B1	P9B2	P9B3	P9B4	P9C1	P9C2	P9C3	P9C4	P9D1	P9D2	Total
Roll	33	32	33	33	33	32	25	20	27	25	28	27	348
WSSAT	26	26	28	24	25	22	25	20	23	16	21	24	280
Oral	10	20	–	–	–	–	–	–	19	11	–	–	60

[1] Deciles are measures of socio-economic status.

The oral assessment

The oral assessment research tool was an expanded form of the GLoSS (Lomas & Hughes, 2009) and used some of the GLoSS- and NumPA-type items, supplemented by other items that gave increased coverage of higher stages. 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). The oral assessment was conducted by external interviewers who had expert knowledge of NDP, NumPA, and GLoSS.

Data collection

The WSSAT was conducted in each class's usual classroom setting, mostly under the supervision of the regular mathematics teacher, in the last few weeks of the fourth term. Standardised instructions were given 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 but what, if any, use was made of these is unknown.

The GLoSS-type oral assessment was conducted in the two days following the written assessment at the Auckland school.

Analysis

The results of the 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 medium-decile stage distribution data from the NDP; and
- the national, Year 9, stage distribution data (the aspirational expectations).

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 GLoSS and NumPA tool and thus was collecting some of the same data, although by experts rather than classroom teachers. 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 basis for comparison of student results arising from the oral tools' use with the national data sets.

Results

The data for each cohort is analysed separately to allow direct comparison with the appropriate decile (socio-economic) level of Year 9 national data.

Internal consistency of WSSAT

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, 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 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, 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. However, 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, 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.

These data suggest 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 medium-decile Wellington school data.

Conformity of WSSAT assigned stages with students banding into classes

The Wellington classes were not banded but the banding of classes in the Auckland school generally reflected the stages assigned by the WSSAT: classes in higher bands achieved more of the higher stages, and classes in lower bands achieved fewer of the higher stages (see Table 2). Additionally, in line with the internal consistency of the WSSAT 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.

Table 2. Auckland school classes in band order, showing the number of students participating and the number of students meeting the criteria for achieving a particular stage.

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

Comparison between oral and written assessments

The stages determined by the oral assessment of students closely matched the stage determined by the WSATT at Stages 7 and 8, but less so at other stages (see Table 3). 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 differences of more than one stage, unlike the initial trial data (Lomas & Hughes, 2008). This may reflect the more even spread across stages achieved by the revisions of the WSSAT.

Table 3. Stages assigned to Auckland students from 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

The level of alignment between the two sets of data is 61% with 41 students having the same stage for both assessment tools. This is not too dissimilar to the 75% accuracy given by Thomas et al. (2006) for secondary teachers' assigning of stages. Of the 32% of students with different stages, 16 (27% of the total) have higher WSSAT stages assigned and 3 (5% of the total) have lower. The higher WSSAT stages could be problematic with students being placed in inappropriate teaching groups—particularly in light of secondary teachers' assigning students lower groups for consolidation purposes (Thomas et al., 2006).

Comparison with New Zealand national aspirational expectations

The assigning of stages from the WSSAT for the two schools gave rise to a distribution reasonably similar to the data for both the medium- and low-decile schools respectively and to the national aspirational expectations (see Tables 4 and 5).

For the medium-decile Year 9 cohort used in the pilot, the areas of greatest disparity (around a 50% difference or more) with the medium-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%) (see Table 4).

Table 4. The percentages of stages assigned to Wellington students from the WSSAT, the medium-decile data, and national aspirational 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
Medium-decile (averaged) percentage (Tagg & Thomas, 2008)[2]	6	22	32	30	10
National aspirational expectations percentage (Tagg & Thomas, 2007)	2	14	27	39	18

[2] The low- and medium-decile percentages are average figures derived from the respective percentage data for the additive, multiplicative, and proportional strategy domain percentage data.

The greatest area of disparity for the medium-decile Year 9 students compared with the Year 9 national aspirational expectations was the higher number of students achieving at Stage 5 (28% compared with 14%).

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%) (see Table 5). 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 aspirational expectations was the lower number of students achieving at Stage 5 (28% compared with 14%). This may partly reflect a difference between low-decile students and a national norm, although a similar disparity was evident in the comparison for the medium-decile data (see above).

Table 5. The percentage of stages assigned to Auckland students for each assessment tool, the low-decile results, and national aspirational 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 (averaged) percentage (Tagg & Thomas, 2008)[2]	11	29	33	22	5
National aspirational expectations percentage (Tagg & Thomas, 2007)	2	14	27	39	18

The oral assessment's assigning of stages to students is reasonably close to the national aspirational expectation percentages for all stages except those achieving at Stages 6 and 7. However, if the sample had been less skewed and included middle band students where more students achieved at Stages 6 and 7 there may have been a closer fit overall.

Discussion

A factor to consider in comparing the WSSAT and oral assessment stages with the national data sets is the degree to which the national data sets accurately represent the stages that the students at Year 9 can achieve. The national data is based primarily on aggregated teacher gathered data and its' accuracy may be variable. For example, secondary teachers' accuracy of 76% and their assigning of lower stages (Thomas et al., 2006) would suggest an underestimation of student performance overall, but possibly more so at Stages 7 and 8, in which the learning demands are greater. The possibility of such a trend is apparent in both the medium- and low-decile cohort data. For example, compared with the medium-decile data, there are three times the percentage of students achieving at Stage 8 and 50% more students achieving at Stage 7, but only about two-thirds the percentage figure of students achieving at Stage 6.

The WSSAT numeracy stages achieved by the Auckland students reflected their placement in ability banded class groups indicating that the WSSAT results paralleled other school based measures of students' mathematical performance used for student

placement. Similarly, the WSSAT assigned stages for the two Year 9 cohorts stage distribution were reasonably close to the national aspiration expectation distribution, allowing for their decile levels, and to the low- and medium-decile distributions. These data indicate that the WSSAT items are measuring strategy (or possibly something that gives a parallel measure) to the extent of being able to assign students' global strategy stages with some accuracy. Thus, the written items appear to access elements of student strategy, with the written record allowing later access to explore responses to items designed to elicit specific strategies.

Conclusion

Overall, the WSSAT has reasonably high levels of internal consistency for Stages 5 -8 and could be used to assign students a global (numeracy) strategy stage. In addition, there is a reasonable congruence of the stages assigned by WSSAT, with both the low- and medium-decile school data, the national aspirational expectations, and with the oral assignment of NDP stages. Thus, the WSSAT could determine a student's global numeracy strategy stage with a sufficient degree of accuracy to allow for students' initial placement into stage appropriate teaching groups. In this sense, WSSAT appears to be a potentially useful tool in secondary schools, given its time efficiency. However, its potential and usefulness for teachers as a diagnostic and formative planning aid in working with their students needs to be explored.

WSSAT may also have uses with other groups where the reading level is adequate such as Year 7 and 8 (11 and 12 year old) primary students for the assigning of an initial global strategy stage, and with pre-service teacher education students and in-service teachers for professional development purposes in identifying and addressing deficiencies in their mathematical knowledge relevant to teaching.

References

Brown, M., Hart, K., & Dietmar, K. (1984). *Chelsea mathematics tests: Place-value and decimals*. Berkshire, UK: NFER-Nelson.

Hart, K. M. (Gen. Ed.) (1981). *Children's understanding of mathematics: 11–16*. London: John Murray.

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.

Lomas, G., & Hughes, P. (2009). Written and oral assessment of secondary students' number strategies: Ongoing development of a written assessment tool. In *Findings from the New Zealand Secondary Numeracy Project 2008* (pp. 17–28). Wellington: Learning Media.

McIntosh, A., Reys, B., Reys, R., Bana, J., & Farrell, B. (1997). *Number sense in school mathematics: Student performance in four countries*. Perth, WA: MASTEC.

Ministry of Education. (2008). *Book 1:The Number Framework: Revised Edition 2007*. Wellington, NZ: Ministry of Education.

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.