

## Written and Oral Assessment of Secondary Students' Number Strategies: Developing a Written Assessment Tool

Gregor Lomas  
*University of Auckland*  
<[g.lomas@auckland.ac.nz](mailto:g.lomas@auckland.ac.nz)>

Peter Hughes  
*University of Auckland*  
<[pg.hughes@auckland.ac.nz](mailto:pg.hughes@auckland.ac.nz)>

This paper examines the first trial of a Written Strategy Stage Assessment Tool (WSSAT) designed for use within the Secondary Numeracy Project (SNP) in secondary schools. WSSAT was designed to identify students' strategy stages and to provide formative data for teachers to use in their planning and teaching. A year 9 cohort of SNP students took the written assessment, which focuses mainly on number strategy concepts. A small sample of these students was also interviewed by a numeracy expert 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 to establish the stages. Comparison of the written and oral assessment results gave data on the relationship between the two measures. The stages established by the written assessment did not match the stages indicated by the oral assessment. They were generally lower, often by several stages. These results raise issues around the criteria set for establishing stages using the written assessment tool.

### Background

Assessment practices for establishing strategy stages within the Numeracy Development Projects (NDP) have, from their inception, focused on diagnostic interviews for both summative and formative purposes. This has avoided issues of students' reading levels, particularly with very young students, and has encouraged the teasing out of students' understandings via oral interactions that move beyond the assessment script of the NDP diagnostic interview. The NDP oral assessment tools (NumPA [Numeracy Project Assessment] and GloSS [Global Strategy Stage]) establish the stages and indicate "where to next" in terms of teachers' specific planning for strategy (and knowledge) teaching and learning. Both tools have worked satisfactorily in primary school environments, in which the teachers are mainly generalist teachers with a single class. With the establishment of the SNP in secondary schools, involving specialist mathematics teachers who work with multiple classes of students each day, the need for a written assessment tool has been recommended by facilitators and teachers.

It is anticipated that such a tool would:

- reduce the amount of teacher time taken per student
- provide a record of students' work that fulfils the same purposes (formative and summative) as the NumPA diagnostic interview
- provide a standardised "marking" schedule.

Such a schedule would enable the assessment to be marked consistently by someone new to the SNP or even by a person with little knowledge of numeracy, such as a parent or teacher aide. It would also reduce possible individual variability that might occur when conducting diagnostic interviews.

Another benefit would be an enhancement in the accuracy of the assessment, particularly in the multiplicative and proportional domains. These domains take on a significant focus in secondary schools, but Thomas, Tagg, and Ward's (2006) data set the accuracy of various types of numeracy assessments in these domains at 76%.

A written tool was not seen as a replacement for GloSS or other GloSS-type oral assessments or for their use for “in-action” checking of students' progress to inform the teacher's next teaching action. Rather, a written tool was seen as replacing the NumPA as an initial or final assessment of a whole class of secondary-school students.

On this basis, a written assessment tool for the part–whole strategy stages was developed for trialling in secondary schools. Although, in these schools, reading skills were expected to be at a level that would not impact negatively on the student's ability to respond, care was taken in the trial assessment tool to limit the amount and difficulty level of written material. The development took place within the SNP community and was informed by interaction and feedback in a similar way to many other previous developments within the NDP.

## Method

The participants in the study were drawn from the year 9 cohort of a large, Auckland, decile-3 secondary school of mixed ethnic composition. The school had participated in the SNP in 2006 and 2007. The written assessment was given to the majority (278) of the year 9 cohort, with only one small class of special needs students being excluded at the school's request. The oral assessment developed for use in this research was given to a subset (27) of the year 9 cohort, who were drawn from a range of the bands in which the school organised their classes. This banding was based on Assessment Tools for Teaching and Learning (asTTle) measures of the students' general achievement in mathematics and English (see Table 1). The subset sample comprised students whose participation in the study would least inconvenience the school, with the interviewer working around the school's programme to order to access students.

Table 1

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

<b>Class Name</b>	9A1	9A2	9B1	9B2	9B3	9B4	<b>9C1</b>	<b>9C2</b>	<b>9C3</b>	<b>9C4</b>	<b>9D1</b>	<b>9D2</b>	Total
<b>Roll No.</b>	32	32	31	32	31	31	29	29	27	22	26	14	336
<b>Written Ass. No.</b>	27	28	28	29	23	28	25	14	25	16	24	11	278
<b>Oral Ass. No.</b>	10	–	–	–	10	–	–	–	–	–	4	3	27

Note: Students in the bolded classes used the stage 5–7 version of the written assessment; the students in other classes used the 5–8 version (see below).

### *The Nature and Structure of the Written Strategy Stage Assessment Tool*

The Written Strategy Stage Assessment Tool (WSSAT) is closely aligned with, and arises from, the NDP strategy section of the Number Framework (and to a lesser extent, given the tool's strategy focus, the knowledge section of the Framework). The items (short answers and multi-choice) draw upon the material of the NumPA and GloSS assessment tools as well as on the work of Hart (1981) and Brown, Hart and Dietmar (1984) on place value and decimals and on that of McIntosh, Reys, Reys, et al. (1997) on number sense. The focus of the WSSAT is primarily strategy, with two overlapping versions: one covering stages 5–7 for less advanced students and the other stages 5–8 for more advanced students (this version was the stage 5–7 version with an additional stage 8 component).

Each stage has a set of items, with the students required to correctly answering a specified number of items to signify their achievement of that stage (see Table 2). As with the SNP diagnostic interviews,

the highest strategy stage achieved in the WSSAT trial was taken as the stage for the student. For example, if a student met the criteria for stages 5, 6, and 7, they were classified as being at stage 7, or if they met the criteria for stage 5 and not that for stage 6, but did meet the criteria for stage 7, they were classified as stage 7. Where the students did not meet the criteria for stage 5 or higher, they were assigned to a category covering stages 1–4.

**Table 2**  
*The Number of Items per Stage and Criteria for Achieving Each Stage*

Stages	1–4	5	6	7	8
Number of items	12	12	11	9	10
Number correct to assign to the given stage	< 9	> 9	> 8	> 7	> 9

Although the WSSAT is a written assessment in the sense that the students read the problems on their own and write their answers down, they do not write down the process by which they arrive at their answers. This is different from a diagnostic interview, in which the questions are communicated orally (read) to the students, printed material is sometimes given as a prompt/reminder, and the students give their answer orally and often talk through the process they used to arrive at their answers. Further prompts can then be given by the interviewer to elicit explanation and clarify points about a student's understandings. Thus, assigning stages by oral assessment relates largely to process, whereas assigning stages in the WSSAT assessment is based on outcomes.

A key issue in the WSSAT is to ensure that the items used force the student participants to use a particular process and restrict their use of any other approach, that is, items that could be answered procedurally are not included. For example, some of the written assessment items use combinations of numbers written as words and in digits, for example, "2 fifths is equivalent to 0.4". 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. It is also seen as a way of keeping aspects of "oral" language use within a written format, with "two-fifths" giving a potentially greater access to the underlying meaning than "zero point four" (or possibly "four tenths") in symbolic form, while still requiring the students to be able to connect the underlying meanings. Whether this has worked will be determined in part by the extent to which the assignment of stages via the written assessment matches the oral assessment used in the trial.

For the use of the WSSAT in this study, an answer sheet was provided that was laid out with clear directions for the students to follow. The format ensured ease of marking, giving both a quick impression of a student's stage and more detailed formative data for planning and teaching purposes. The answer sheet had room only for the answers because what was required was the result of mental processes rather than the process itself or procedures such as written algorithms.

A range of items for each stage needed to isolate and encapsulate some of the conceptual aspects and elements of strategy that might be present in an oral exchange. These sets of items have the potential to provide a more detailed and standardised diagnostic map of a student's learning needs than an oral assessment because they allowed the 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.

## *The Oral Assessment*

The oral assessment tool used in this trial was developed as a research tool rather than as a tool for general use with students. The tool used some of the GloSS items, but these were supplemented by a few items that 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).

## *Data Collection*

The WSSAT was conducted during term 4 in each class's usual classroom setting, mostly under the supervision of the regular mathematics teacher, on the Monday of the week before the students' end-of-year exams. Both the teachers and the students received a set of instructions on how to conduct the assessment, a key point of which was that the students could stop if they felt the items had become too difficult for them. This instruction was an attempt to mirror the practice used in the NDP diagnostic interviews of the assessor only asking questions until the student started to give incorrect responses or showed other evidence of failing to understand what the questions required. At that point, the assessor would bring the interview to a close and assign a stage to the student on the basis of the student's responses up to that point.

Although the research team did not brief the teachers before giving them the written assessment, the liaison teacher who worked with the teachers had been briefed. After the assessment, the teachers collected all the answer sheets and gave them to the researchers. To ensure consistency, only one person marked the assessments. Copies of the marked answer sheets were returned to the school for potential diagnostic use by the school.

The GloSS-type oral assessment was given the next day, Tuesday, to a sample of 27 students drawn roughly, and at the school's convenience, from across the "ability" bands into which the year 9 cohort was organised. The oral interviews were conducted by a person external to the school who had expert knowledge of the NDP and was well experienced in conducting and interpreting oral assessments.

## *Analysis*

The results of the written assessment 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. The stages that the students achieved were compared with:

- the banding of the class they were in to see whether this reflected the school's placement of students on the basis of asTTle results
- the 2007 national, end-of-year, year 9, low-decile stage distribution data from the SNP (Tagg & Thomas, this volume)
- the 2006 national, end-of-year, year 9, stage distribution data from the SNP (Tagg & Thomas, 2007).

The results of the written and oral assessments were compared to establish a relationship between the two forms of assessment. For the purpose of comparison in this study, the oral assessment was assumed to be accurate and thus provided a baseline for the comparison. This assumption was used in order to establish the connection to the existing and "proven" NDP assessment tools and the database of student results arising from their use.

## Results

### *Internal Consistency*

Of the 109 students who could be assigned as being at stage 6 on the written assessment, only 14 had not also achieved the criteria for being assigned as being at stage 5. Of these 14, eight had missed the criteria by only one correct response and three by two correct responses. For the 19 students assigned as being at stage 7, only four had not achieved the criteria for one or both of stages 5 and 6. Of these four, two had missed the criteria for stage 6 by only one correct response. These data suggest that the written assessment was largely internally consistent in assigning stages.

A further analysis, conducted for the situation where the criteria for assigning stages were adjusted by reducing one of each component, resolved the ten cases mentioned above and created only seven more. Of the 136 students who could be assigned as being at stage 6 on the adjusted criteria, only nine had not also achieved the criteria for being assigned as being at stage 5, and of the 49 assigned as being at stage 7, only six had not achieved the criteria for one or both of stages 5 and 6. This analysis also supports the internal consistency of the written assessment in assigning stages 5–7. No stage 8 was assigned under either set of criteria levels, so consistency across stages 7 and 8 could not be determined.

### *Conformity of Assigned Stages with Students Banding into Classes*

The stages assigned by both the written and oral assessment generally conformed to the banding of the classes: students in higher band classes achieved more of the higher stages, and students in lower band classes achieved fewer of the higher stages. Additionally, in line with the internal consistency of the written assessment tool, as discussed above, the meeting of the criteria for particular stages also aligned with the banding of the classes, with fewer students meeting the criteria for each stage. For example, of the students in class 9A2, 10 students met the stage 7 criteria, 23 the stage 6 criteria, and 22 the stage 5 criteria, compared with class 9B2, in which no students met the stage 7 criteria, 17 met the stage 6 criteria, and 23 met the stage 5 criteria, while in class 9C2, only one student met the stage 6 criteria and two met the stage 5 criteria. No one in 9D2 met any of the criteria for stages 5–7. (See Table 3 for the written assessment data.)

Table 3  
*Classes in Band Order, Showing the Number of Students Taking the Written Assessment and the Number of Students Meeting the Criteria for Achieving a Particular Stage*

Class	9A1	9A2	9B1	9B2	9B3	9B4	9C1	9C2	9C3	9C4	9D1	9D2	Total
<b>Number of students</b>	27	28	28	29	23	28	25	14	25	16	24	11	278
<b>Number meeting stage 5</b>	24	22	19	23	8	17	4	2	8	5	3	0	135
<b>Number meeting stage 6</b>	23	23	10	17	5	18	5	1	4	0	3	0	109
<b>Number meeting stage 7</b>	4	10	0	2	0	3	0	0	1	0	0	0	20

Note: Students in the bolded classes used the stage 5–7 version of the written assessment; the students in other classes used the 5–8 version.

This data suggests that the asTTle measure of general achievement in mathematics and English that was used to place the students in bands, and classes within those bands, is reflected in the WSSAT measurement of their numeracy achievement and assigning of stages.

However, the results for class 9B3 do not match the results for other classes in that band, being lower for all stages. This may reflect some variation in: the composition of this class compared with others in the same band; the way in which the assessment was actually conducted for this class despite the instructions that were distributed to each teacher; or a teacher/teaching practice difference.

The oral assessment also appeared consistent with the placement of the students in their bands. Of the ten students in 9A1, eight were assigned as being at stage 8 and two at stage 7; of the ten students in 9B3, five were assigned as being at stage 7, four at stage 6, and one at stage 5; and the three students in 9C2 were assigned as being at stage 5. The students in 9C1 did not fit as well, with three being assigned as being at stage 7 and one at stage 5.

### *Comparison between Oral and Written*

The stages determined by the written assessment generally did not match the stages determined by the oral assessment (see Table 4). Only one student at stage 6 was the same. The stages assigned by the written assessment were generally lower, with six students lower by more than two stages (see the bolded figures in Table 4).

Table 4

*Comparison of Stages Assigned to Students by the Written Assessment Compared with the Oral*

Oral assessment stage	5	6	6	6	7	7	7	8	8
Written assessment stage	1–4	1–4	5	6	1–4	5	6	6	7
Number of students	5	<b>2</b>	1	1	<b>4</b>	2	4	6	2

The assignment of generally lower stages using the written assessment can also be seen in Table 5, where 46% of the students were assigned as being at stages 1–4 using the written assessment, while none of the students assessed orally were assigned as being at less than at stage 5. At the upper stages, only 7% were assigned as being at stage 7 on the written assessment, much fewer than the 41% assigned as being at stage 7 by the oral assessment. Also, no students were assigned as being at stage 8 using the written assessment, while 26% of the students assessed orally were assigned as being at stage 8.

### *Comparison with 2007 Low-decile Data and 2006 Year 9 Data*

The assigning of stages from the written assessment gave rise to a considerably different distribution (see Table 5) from both the low-decile-schools data (averaged across the strategy domains; see Tagg & Thomas, this volume) and the 2006 year 9 data (Tagg & Thomas, 2007). The most significant differences are evident in the percentage of students in the stage 1–4 category, with the written assessment at 46%, compared with 11% and 2% respectively for the oral assessment, and in the comparison of the top two stages, in which the percentages of stages 7 and 8 assigned by the written assessment are very much less than those for the oral assessments. The school in which the written assessment was conducted is decile 3, which might indicate an achievement level below the 2006 year 9 data, but this does not account for the variation from the low-decile-schools data.

Table 5

*The Percentage of Stages Assigned to Students for Each Assessment Tool and for the End-of-year, Year 9, Low-decile (2007) and the Year 9 as a Whole (2006)*

Stages	0–4	5	6	7	8
<b>Written assessment</b>					
% of students (n = 278)	46	13	34	7	0
<b>Oral assessment</b>					
% of students (n = 27)	0	18	15	41	26
<b>Low-decile data</b>					
% of students (n = 412) (Tagg & Thomas, this volume)	11	29	33	22	5
<b>2006 year 9 data</b>					
% of students (n = 5807) (Ministry of Education, n.d.)	2	14	27	39	18

However, the oral assessment's assigning of stages to students is reasonably close to the 2006 year 9 data percentages, particularly when taking into account the small sample size. This would support the assumption that the oral assessment is accurate and provides a useful baseline for calibrating the written tool.

## Discussion

The initial trial of the written assessment tool has highlighted a number of issues. Although the WSSAT is internally consistent for stages 5–7 and can be used to assign students a stage, there is a significant divergence of these stages compared with the data for low-decile-schools at the end of their first year in the SNP, the 2006 year 9 data, and the oral assessment assignment of NDP stages. Thus, the written assessment tool in its current form does not determine a student's numeracy strategy stages with any apparent degree of accuracy. Although these results clearly indicate that further development work is definitely needed, the internal consistency suggests that there is a solid basis from which to conduct such a development.

Further steps are needed to examine the criteria for assigning stages and to see if adjusting these criteria could assist in creating a closer match between the stages assigned by the WSSAT and the national data sets for the current data. This could be done by repeating the above analysis. Following this, and any changes to items or format suggested by teachers from the study school or prompted by expert feedback, such a refined tool would need to be further trialled.

## 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.
- McIntosh, A., Reys, B., Reys, R., Bana, J., & Farrell, B. (1997). *Number sense in school mathematics: Student performance in four countries*. Perth, WA: MASTE.
- Ministry of Education (n.d.) Curriculum expectations. Retrieved December 2007 from: [www.nzmaths.co.nz/Numeracy/Lead\\_Teacher/PLC/Expectations/index.aspx](http://www.nzmaths.co.nz/Numeracy/Lead_Teacher/PLC/Expectations/index.aspx)
- Thomas, G., Tagg, A., & Ward, J. (2006). Numeracy assessment: How reliable are teachers' judgments? In *Findings from the New Zealand Numeracy Projects 2005* (pp. 91–102). Wellington: Learning Media.
- Tagg, A., & Thomas, G. (this volume). Performance of SNP students on the Number Framework. In *Findings of the Secondary Numeracy Project 2007*. Wellington: Learning Media.