DEVELOPING PERFORMANCE ASSESSMENT TASKS IN MATHEMATICS: A CASE STUDY

Rosemary A. Callingham
University of Tasmania
<Rosemary.Callingham@Central.tased.edu.au>

Development of assessment tasks that both provide quality insights into students' mathematical understanding and produce rigorous data for measurement purposes is a challenge for mathematics educators. This paper presents a case study of the development of three performance assessment tasks. The tasks addressed several strands of the mathematics curriculum and were designed to match the teaching strategies being used. The tasks were refined following the application of a qualitative developmental model, the SOLO Taxonomy. They were also tested against criteria for good assessment.

Developing assessment tasks that both elicit information about students' deeper understanding of mathematical concepts and provide reliable and valid data for measurement purposes is a challenge for mathematics educators. One way in which this may be achieved is through the use of well designed performance assessment tasks with appropriate scoring rubrics that provide students with the opportunity to draw on a range of mathematics skills and apply these to interesting and motivating situations. Students' approaches to the problems reflect their cognitive processes and mathematical understanding (Harmon *et al* 1997).

Assessment should describe what a student knows and can do in relation to the learning and teaching program within the classroom and should fit within the overall framework of the development of that student (Griffin & Nix 1991). As well, it should "...enhance student motivation ..., provide starting points for further learning..." and "... be fair for all groups of students" (Australian Association of Mathematics Teachers n.d.).

This paper reports the trial of three performance assessment tasks as a case study in preparing suitable tasks that meet the criteria outlined above. These assessments are part of larger study that aims to improve understanding of factors that affect the numeracy development of Indigenous adolescents. The on-going project, Improving Numeracy for Indigenous Students in Secondary Schools (INISSS), targets 17 Tasmanian high schools that had a relatively high enrolment of Indigenous students who were achieving poorly on state-wide measures of numeracy achievement (Callingham 1998). In some of these schools Aboriginal Education Workers worked with Indigenous and non-Indigenous students to provide additional support. Many of the target students were being taught by skilled teachers who had not had training in mathematics teaching; others were taught by highly experienced, well qualified mathematics teachers. The INISSS project provides professional development to teachers and Aboriginal Education Workers involved with the target group of low achieving students in years 7 to 9. Evaluation of the project will focus on year 8 students since most teachers in the project teach students in this grade.

The professional development for teachers is based around the use of Mathematical Problem Solving Task Centres that have been successfully used with Indigenous students (Williams 1997). The tasks come packaged for use with all required materials and manipulatives in reseal plastic bags. Students work on them singly or in small groups. Teachers in the INISSS project schools reported using these tasks in a variety of ways, similar to teachers elsewhere (Pavlou & Clarke 1998). The tasks proved to be motivating to students and, according to the teachers, provided the first real success in mathematics for many students.

RESEARCH QUESTIONS

Because of funding arrangements, the project needed to report on student learning outcomes as part of its evaluation. With this in mind, it was decided to develop a set of assessment tasks that would reflect the nature of the Task Centre material, but which also could be marked rigorously by teachers. The purpose of this initial small-scale trial was to collect qualitative data about the presentation of the tasks, clarity of the instructions and the range of responses elicited from students. The research questions addressed by this trial were then:

- Were the tasks presented in a way that was both understandable and motivating to all students?
- Did the tasks allow for a wide range of responses along a developmental continuum?

Theoretical Model

The theoretical model used as the basis for analysis of the responses was the SOLO Taxonomy (Biggs & Collis, 1982; Collis & Biggs, 1991). This neo-Piagetian model proposes five modes of thinking: sensori-motor, ikonic, concrete symbolic, formal and post-formal. These modes are increasingly abstract and develop as an individual matures from infancy through to adulthood. Although each mode to some extent builds on the previous mode, each mode also continues to develop in its own right. The model can thus provide a description of the development of a student's understanding.

Within these modes, cycles of response to different learning situations can be found. The five categories of response depend on the Structure of the Observed Learning Outcome (SOLO) produced by the student from the use of information available to the individual when giving an answer. The categories are of increasing complexity and are described as follows:

- Pre-structural - no use or irrelevant use of the information available;
- Uni-structural (U) - use of only one piece of the relevant information;
- Multi-structural (M) - a number of pieces of information are strung together, usually sequentially;
- Relational (R) - the various pieces of information are integrated into an understanding of the relationships involved in the situation;
- Extended Abstract - over-arching principles are called upon, and the thinking is thus moved to a new level. This response provides the uni-structural response for the next higher mode of thinking.

In the school context, cycles that occur in the ikonic and concrete symbolic modes are particularly important. These progress through the uni-structural, multi-structural, relational cycle and are termed U-M-R cycles. Several studies have identified at least one and sometimes two U-M-R cycles of response in the concrete symbolic mode in different learning situations (e.g. Watson, Collis & Moritz 1997). From a measurement perspective, the use of U-M-R cycles potentially provides a basis for the development of a scoring rubric, which can then be translated onto a developmental scale addressing the underlying concept. For this reason, it was important to ensure that the tasks allowed for at least one cycle of U-M-R responses.

THE TASKS

Of five tasks trialed, one was discarded and one completely rewritten following comments from teachers and students. The three assessment tasks reported here addressed aspects of several strands of the mathematics curriculum. They were presented in a similar format to the Task Centre materials. Students were provided with a record sheet for each task. Each task had between 8 and 10 individual questions, each of which was provided with a scoring rubric based on anticipated responses. This provided a standard format for collecting the assessment information, but did make the tasks more structured than those of the Task Centres. Overall, however, the tasks were more open than traditional assessment materials and had the 'flavour' of the teaching tasks. An initial scoring rubric was developed for every question on each task, based on experience and information from the initial sources where this was available. It was hoped that this approach would be easier for teachers to use rather than a holistic score on the whole task.

Bean Counter was adapted from a similar assessment task used in New Zealand (Flockton & Crooks 1998). It required students to use number patterns to complete magic squares. The name came from the beans that were provided as part of the task package. Students used the beans to 'juggle' the numbers in the empty squares. The underlying concept was recognition and use of number patterns. *Come in Spinner*, as its name implies, addressed aspects of the students' understanding of chance. It was developed from an idea in the *Mathematics Work Samples* (Curriculum Corporation 1994). Students made spinners and tested them to see if they met the specified criteria.

Street Party was an original task addressing pattern and algebra concepts. Students developed patterns by building small tables into a 'long table' for a street party, and used the patterns to find the numbers of small tables needed to seat a given number of people. Both direct and inverse relationships were targeted.

THE TRIAL

The tasks were trialed in three year 8 classrooms in one school in early December 1998. The school enrolled students from a very wide area and its students came from a range of socio-economic backgrounds. The students in the grade were divided broadly into upper and lower ability classes. The trial classes included all levels of ability and were taught by two very experienced mathematics teachers. Altogether 32 students produced 45 completed record sheets covering these three tasks.

Students worked through the tasks, usually in pairs, but providing an individual response on the record sheet provided. They were also asked what they thought about the tasks they attempted and these written comments were collected separately. No pro-forma was supplied for this—students were free to respond in any way that they wished.

The two teachers had both previously used Task Centre materials in a variety of settings. They marked the tasks using the scoring rubrics provided based on anticipated responses, and provided feedback about these and practical aspects of using the materials. Marked responses and students' and teachers' written comments were collected for analysis.

FINDINGS

Motivation

Teachers reported that the students enjoyed the tasks and were motivated to complete them. In general the students were on task throughout the lesson—a first for some of these students. One teacher reported that he had never seen one student work so well. This student was of Polynesian descent and usually struggled in the classroom, becoming at times a behaviour problem. At the end of the task the student wrote "Thanks to [the teacher] and I really like the Tasks".

The positive comments by teachers were borne out by other students' comments.

> I thought it was a good worksheet and fun to do. Comment on *Bean Counter* RL
>
> I found this activity more interesting. Although it was a bit easy. Comment on *Street Party* CH

Comments about the apparent lack of difficulty were frequent, even from students who produced a low-level response:

> The spinner unit was pretty good because it made you think a lot and was pretty easy.
> Comment on *Come in Spinner* BC

The comments also produced some interesting insights into students' perceptions of problem solving:

> ...also question 4, 5, and 6 are a bit obious[*sic*] and aren't problem solving just looking problems. It's a good activity but a bit short! Comment on *Bean Counter* CJ

This student had produced one of the lowest level responses to the task.

It was clear from these comments that students enjoyed the tasks, found them understandable and non-threatening and were prepared to work at them—even those who were unable to produce high level responses.

Developmental Progression

By considering the nature and complexity of students' responses, the development of the targeted concept could be evaluated. Findings are summarised for each task.

Bean Counter

This task provided students with packets of beans and some blank grids to help them find their answers. This allowed a range of strategies to be used—some students enjoyed using the beans while others preferred to use blank grids and write the different solutions.

Placing beans on one square and then arranging others to fit was the commonest problem solving strategy. Where this was unsystematic, it was coded as an uni-structural response in SOLO terms—each response was unique and independent of the others. Where a similar but systematic strategy was applied, this was coded as a multi-structural response because the student was using one bit of information sequentially. One student appeared to be visualising her response:

> Well I went through the numbers in my mind to come up to the answer. I had to go all through the numbers. (McC)

Since the final problem had eleven different solutions, not surprisingly this student had been unable to find all of these. This was coded uni-structural since the student was only using one number at a time, and appeared unable to systematically record her findings.

No student explicitly referred to patterns in the solutions although this was implied in some responses. Probably the most sophisticated response came from the Polynesian student referred to earlier, who had clearly recognised the patterns but lacked the English language skills to express this clearly. Having described his process for the easy initial questions he went on to write:

> We get the answer easyier(*sic*) because we just have to get the answer from number 1, 2, 3 and 4. (JR)

The numbers referred to were those used to label his first few solutions of the eleven-solution problem. This response was also coded multi-structural but with better language skills the student might have reached a relational response. Interestingly, his partner, while participating in the same process had not reached the same insights:

> Me and [JR] worked our methods out of jubiling (*juggling*) our numbers around and trying to work them out and add the number up with the correct answer. (HM)

This response was coded uni-structural as an unsystematic strategy. The common criticism of the weakest student's performance being over-inflated when group tasks are used for assessment was not upheld in this instance. A summary of the highest responses to this task achieved by students is presented in Table 1.

Table 1
Summary of Highest Responses to Bean Counter

Uni-structural N=7	Multi-structural N=8	Relational N=0
Use of unsystematic strategies based on random trials of numbers	Use of strategies using progressive, systematic movement of beans from one cell to another.	Use of systematic strategies based on recognising the patterns and constraints of the numbers in the squares.

Come in Spinner

All students' responses demonstrated an understanding of equal likelihood but their understanding of variability was less certain. They were asked to make two spinners—one with an equal likelihood of landing on each of three colours, and one with four colours where one colour would come up most, one least and the other two were about equal. They had to test these and explain their tests and results. The spinners were correctly constructed in all but one case. Students were able to set up a test and record the findings but they were satisfied with one short run. Some students gave no explanation of their test findings other than a comment unsupported by data, or recorded the data without comment. These kinds of response were classified as uni-structural since the students appeared to focus on only one aspect of the test. Others suggested an emerging understanding of variability. For example:

> I found out that you don't get equal amount of colours even though you have equal spaces. (BC)

These responses were classified as multi-structural since the students were attempting to read their data in relation to the design of the spinner.

Some students seemed to think that the number of spins used to test the spinner should be related to the number of colours on the spinner. Many students tested their 3-colour spinner by spinning it 30 times and their 4-colour spinner by spinning it 40 times. This may be a reaction to a common teaching strategy of asking students to undertake short runs and then combining the results to get a larger sample size. Highest level responses to *Come In Spinner* are summarised in Table 2.

Table 2
Summary of Highest Responses to Come In Spinner

Uni-structural N=7	Multi-structural N=8	Relational N=0
Spinners correctly constructed and tested but test limited to one short run and no comment on the variability of the result.	Spinners correctly constructed and tested but test limited to one short run, test results recorded and explained by reference to the spinner construction only.	Spinners correctly constructed and tested. Long runs suggested to improve predictions of outcomes and possibility of spinner bias acknowledged.

Street Party

Responses to this task indicated students' developing understanding of relationships. The first two questions, which all students successfully completed by continuing the pattern, only required an uni-structural response in SOLO terms.

When asked to record their findings most students simply copied and extended the diagrams that had been provided. Some, however, drew up a table showing the number of small tables used to make the big table and the number of people seated, and two went straight to the underlying rule. This appears to suggest a cycle of understanding of the relationship the students were investigating. In continuing the given pattern in the same form, students demonstrated only a uni-structural response, while those who translated this into a table required at least multi-structural understanding since they had to recognise two variables and abstract these into a different symbolic form.

The two students who used the rule appeared to demonstrate a relational understanding of the information. Their work, however, showed some interesting differences.

One of these, a girl, had described the rule in words:

> To get the results you can times the number of tables by 2 and add 2 for the 2 at the end. (HC)

This student, in later parts of the task, clearly indicated that she had a good understanding of this direct relationship, arising from a concrete situation. She correctly solved several more difficult problems, including those where a different pattern was used, to find the number of people that could be seated. She could not, however, solve any of the inverse problems where she was asked to find the number of tables needed to seat a given number of people. For this reason, her response was coded as multi-structural rather than relational.

The other student, a boy, used a more mathematical notation:

> People at tables = 2 x Tables + 2 (RL)

This student completed the whole worksheet correctly, including the inverse problems. He appeared, however, to need some work on refining his mathematical notation. He correctly solved the problem of finding the number of tables needed to seat 200 people but when asked to explain how he got his answer wrote $200 - 2 \div 2$, rather than $(220 - 2) \div 2$. Despite the poor notation, the response was coded as relational with respect to the underlying concept of understanding relationships.

Few other students correctly solved the inverse problems. One who did solve the problem requiring the number of tables for 200 people, used a laborious method of drawing long lines of tables and counting the people seated. Not surprisingly, this student did not complete the rest of the task. His uni-structural method was successful but very inefficient. The most successful students were clearly able to articulate the relationship using mathematical language. For example, this relational response:

> I subtracted 2 from 200 because of the 2 people at the end and then I divided 198 by 2. This gave me 99. (SJ)

Another response that had not quite reached the same level of understanding was the following, coded as multi-structural because of the apparent inability to resolve it completely:

> There is[*sic*] three people at each end of the tables and two for each of the middle table so you work out a number to times to add to the end. (BG)

The task appeared to be successful in drawing out the range of understanding of relationships. Highest level responses to Street Party are summarised in Table 3.

Table 3
Summary of Highest Responses to Street Party

Uni-structural N=4	Multi-structural N=7	Relational N=4
Correctly solved simple relational problems using counting strategies only.	Recognised a straightforward relationship between two variables and used this to solve direct relationship problems only in a practical situation.	Developed a straightforward rule relating two variables in a practical situation and used this to solve direct and inverse problems.

CONCLUSIONS

The trial met its goal of providing utilitarian information about the tasks and data concerning the range of responses in relation to the underlying concepts being addressed. In answer to the first question, the tasks were clearly motivating and accessible to all students. Both teachers and students commented favourably on this aspect, and it was also clear from student work that they had understood what was asked, even if they produced a low-level response. Teachers also reported that the scoring rubrics were provided in an easy-to-use format and they could match their students' performances to the descriptions of anticipated response provided.

In terms of the developmental continuum, SOLO analysis suggested that some tasks did not attract relational level responses. Modifications were subsequently made to improve the range of responses possible. Additional questions were added to the spinner task to elicit responses more clearly about variability and the 4-colour spinner was made more open by asking students to design their own spinner having a different chance of stopping on each colour. Minor changes were made to the wording of *Bean Counter* to encourage students to look for patterns. Changes to *Street Party* asked students to design their own tables and determine relationships based on these designs to allow a greater degree of freedom of response. The U-M-R cycles identified were also used to inform the scoring rubrics of the rewritten tasks before large-scale use, while maintaining the format. Samples of student work were also used to provide exemplars for the teachers' manual to accompany the final tasks.

There were also some unexpected outcomes from the trial. Both of the teachers involved indicated that these tasks had provided them with insights into their students' thinking that they would not otherwise have had, such as the response from the Polynesian student. These insights also applied to teaching strategies. For example, following discussion about the spinner responses, one teacher said that he would only ask students to complete a limited number of trials in the future if he specifically wanted to compare what happened with different small samples. If he wanted large sample sizes he would instead give students five minutes to complete as many trials as possible so that not all students produced the same amount of data. This teacher has also spent some time providing opportunities for his low-ability students to translate patterns from diagrams into other summary forms such as lists and tables. In these ways, the tasks affected the teaching programs in the classrooms more than might have been expected from a small-scale trial at the end of the school year.

Finally, the trial provided some evidence of the use of a theoretical developmental model to underpin assessment tasks to be used in large-scale quantitative studies. The SOLO model applied to student work provided a means of identifying the developmental sequence of the concept being addressed by a task. In addition the underlying U-M-R cycles offered a basis for developing improved scoring rubrics for individual questions in each task. Zero was used for a pre-structural response throughout. Other questions were coded according to the highest level of the U-M-R cycle possible. Thus a question that could get a relational response was coded 0 - 3, where unistructural (1) and multi-structural (2) responses provided partial credit. Further work is needed in this area, however, particularly in relation to large-scale testing.

REFERENCES

Australian Association of Mathematics Teachers (AAMT). (n.d.). *AAMT discussion paper on assessment and reporting in school mathematics*. Adelaide: Author.

Biggs, J.B. & Collis, K.F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.

Callingham, R. A. (1998) *Year 9 numeracy assessment and monitoring program 1997: Summary report.* Hobart, Tas.: Department of Education, Community and Cultural Development.

Collis, K.F. & Biggs, J.B. (1991). Developmental determinants of qualitative aspects of school learning. In G.T.Evans (Ed.) *Learning and Teaching Cognitive Skills.* Melbourne: ACER.

Curriculum Corporation (1994). *Mathematics-Work Samples* Carlton: Author.

Flockton, L. & Crooks, T. (1998). *Mathematics assessment results 1997.* Otago, NZ: Ministry of Education.

Griffin, P. & Nix, P. (1991). *Educational assessment and reporting*. Sydney: Harcourt Brace Jovanovich.

Harmon, M., Smith, T.A., Martin, M.O., Kelly, D.L., Beaton, A.E., Mullis, I.V.S., Gonzalez, E.J. and Orpwood, G. (1997). *Performance assessment in IEA's Third International Mathematics and Science Study.* Boston: IEA.

Pavlou, M. & Clarke, D. (1998). *Problem solving task centres and the professional development of mathematics teachers.* Available: <http://www.srl.rmit.edu.au/mav/PSTC/refer/abstract.htm>

Watson, J.M., Collis, K.F., & Moritz, J.B. (1997). The development of chance measurement. *Mathematics Education Research Journal*, 9, 60-82.

Williams, D. *Mathematics task centre project for Aboriginal and Torres Strait Islander Students.* Carlton: Curriculum Corporation.