

QUALITY ASSESSMENT FOR THE NEW CENTURY

Brian Doig

The Australian Council for Educational Research

Abstract

This paper looks at assessment in mathematics and how it may be improved. The basis for improvement is taken to be those aspects of assessment which are the most important and the most amenable to change, namely, the item content, the form of response required, the interpretation of responses, and the interaction between these first three factors. Examples from recent innovative attempts to create improved assessments, with respect to each of these factors, is presented. Whilst not all of the changeable factors are addressed by each assessment tool, the interaction between the innovative and conventional factors does provide better information, which is the fundamental purpose of all assessment.

Introduction

Every teacher would agree that two very important aspects of preparing to teach effectively are to know where one's students are in their conceptual knowledge and how they apply their skills in applying this knowledge. The key question is, therefore, 'How do I get such information?'. Subsidiary questions that one might ask concern the type of information needed, the amount of information required, the effort and time needed to collect it, the use for the information collected and so on. Ideally one would talk to each individual student, reflect on the information gathered and then act accordingly. The reality of the classroom is that a group measure of concepts and skills is the most sensible solution, but this usually means a loss of information. It is my intention to show that this does not have to be the case and that quality assessment is possible.

While acknowledging that the type of information required depends upon the purpose of the assessment, a key point is that the amount of information gained from an assessment is directly related to four factors; these are the content of the questions posed, the form of response required, the interpretation of answers, and the inter-relationship between these factors. It is the interactions between these factors which determine the quality of any assessment, and so particular attention must be paid to these interactions. In this paper I want to concentrate on content and interpretation, two factors with which I have had some experience and where I believe we can seize opportunities for improving assessment in mathematics and producing quality assessment for the near future.

Content

Many different structural approaches have been employed over the years to maximise the information gained from an assessment. A popular method is to form a hierarchy of content, usually defined

by a series of objectives. This approach usually relies on an examination of the mathematical content to be assessed. Such a method was used to construct the Group Review of Algebra Topics (GROAT) (Doig, 1991). In this series of assessments the algebra content has been subdivided into eight separate topics, introductory concepts, algebraic manipulation, factorisation, word problems, linear equations, quadratic equations, inequalities, and sequences and series. Within each topic the questions are ordered by their mathematical complexity, with questions with simpler content coming first. Below are examples of easy (earlier) and harder (later) questions from the GROAT unit on algebraic manipulation.

Insert Figure 1 about here

Through the use of simple grouping techniques for the student

responses it is possible to gain diagnostic information about the students' algebraic abilities in addition to achievement scores. This is demonstrated in the interpretation section below.

An alternative to the mathematical analysis method is to use curriculum experts to suggest questions considered suitable in eliciting information about the students' concepts or skills in the chosen topic. Thus 'experts' would determine the content, form an acceptable context and then order the questions from easiest to most difficult. In this way an assessment instrument is created. I believe that it is better to have the students involved in this procedure. Questions under development are presented to a sample of students (trialled) and from the subsequent analysis of this trial group's responses content is selected and the order of the final content of the instrument established. A good example of this approach can be found in the New South Wales Basic Skills Testing Program (BSTP) (Masters, Lokan, Doig, Khoo, Lindsey, Robinson and Zammit, 1990). In this case, the analysis of the responses of students in the trial group led to the rejection of many questions as the student 'experts' showed (through the analysis of their responses) that in some cases questions were either inappropriate in their content or difficulty. Thus it may be said that the students selected the content of the assessment for their fellows.

Insert Figure 2 about here

It could be argued that this methodology is not new and has been a standard practice for many years. To some extent this is true, but the approach outlined above does not only establish reliability coefficients for the set of questions but also the degree to which the questions form a coherent description of the continuum for the trait being assessed. In the case of the Diagnostic Mathematical Profiles (DMP) for example, a numeracy

continuum is described, on which student achievement can be placed, giving the teacher an idea of where the student is now, and where they should be heading. This is much more useful than providing the teacher with a numerical Cronbach alpha measure of internal consistency. Precisely how this is done is described in the next section on interpretation.

In general it can be argued that no selection of content can be ideal for every student and that a number of inter-related assessments need to be used. In the long term this means many assessments, which is neither effective time use nor is it good for student morale. One alternative being used experimentally in Victoria is to allow teachers to select the questions on an individual student basis. Teachers select a set of 'worksheets' which contain questions on specific curriculum topics. A complex software program enables any combination of such worksheets to be selected and the results located on a standardized continuum. In this manner teachers are able to tailor the assessment to the needs of their students and themselves. For example a teacher need not use questions on topics that have not been taught, unlike the standard assessment form, where one has to use all the items on the form (this usually means that the assessment cannot be used until the end of the year or term). Examples of this approach to assessment are shown in the example below.

Insert Figure 3 about here

In the example 'Toy Shapes' I have presented the teacher's page which details the indicator or objective of the question sheet, the scoring criteria (note that it is a partial credit scoring), and comments that may be necessary to administer the sheet correctly. In the top left-hand corner is a quarter scale copy of the actual sheet given to the children. Since there are numerous such sheets, the teacher is able to select those whose objectives are appropriate for the children concerned. An interesting side issue is that any number of children may be

assessed at one time – but they may not all be using the same set of sheets. What these sheets represent in fact is an item bank from which the teacher selects her questions!

If the purpose is to investigate the mathematical thinking of students, then assessment instruments (which focus on this thinking rather than simply upon mathematical content) will obviously have different content, possibly some that is unfamiliar to students. This is a radically different notion. Two instruments of this nature are the Collis-Romberg Mathematical Problem Solving Profiles by Kevin Collis and Tom Romberg (1992) and the Profiles of Problem Solving (POPS) (Stacey, Groves, Bourke and Doig, in press).

The Collis-Romberg assessment is based on the SOLO taxonomy for classifying student thinking and examines five mathematics

content areas. The problems posed are such that they form challenging problem-solving situations for students, and the analysis of the responses lead to specific follow-up teaching ideas. This can be seen quite clearly in the case of Mary, whose score sheet is reproduced below.

Insert Figure 4 about here

Details of the interpretation of Mary's responses using this approach are in the interpretation section below.

The POPS assessment of problem solving focuses upon the strategies and processes that student's use in solving problems. The categories of problem solving abilities described by POPS include the method used, accuracy, extracting information and quality of explanation (of the solution process). Students are not only assessed on whether they can obtain correct solutions, but also on the strategies they use to solve problems. The POPS questions are such that although they are unfamiliar, children who use strategies such as being systematic, visualising, and estimating can solve these problems. One such question is shown in Figure 5.

Insert Figure 5 about here

Through an analysis of student's responses and their problem-solving strategies specific teaching activities (included in the POPS manual) can be selected which match student strengths and weaknesses.

Both the POPS and the Collis-Romberg Mathematical Problem Solving Profiles use sophisticated analytic techniques and high quality visual presentations to assist users in gaining the maximum information with the least effort. Just how they do this is demonstrated in the interpretation section following.

In some instances it may be neither practicable nor particularly desirable to assess a student in a single instance. In cases like this it would appear that teachers need assistance in interpreting student's day-to-day class work. Those who are correct in their application of the standard algorithms are not the issue here of course, but those who are having difficulty are. The idea behind Stop! Look and Lesson is to assist teachers interpret the errors made by students in order to facilitate the remediation process. Not only does Stop! Look and Lesson guide diagnosis, but it also contains remedial activities and advice directly linked to each error identified. These errors have been collected from the actual work of some 80 000 students in Australia. This is not theory but practice on a large scale. The use of a guide like Stop! Look and Lesson means that students who are making errors in their daily work can be helped without immediate recourse to special assessments. Their daily work is sufficient data for their teacher to make diagnostic assessments. Below is an example of a Stop! Look and Lesson error description and one of Stop! Look and Lesson's remedial activities.

Insert Figure 6 about here

In all of the assessment instruments described above, part of their claim to quality rests upon the theoretical basis upon which they are founded. However an even stronger claim can be made on the basis of the techniques and approaches used for analysis and interpretation of the data on student learning. Although the analytic techniques may vary, in every instance the assessment has been designed with particular analysis in mind thus ensuring a coherence over the entire assessment process.

Interpretation

The second factor of any assessment that may be varied to improve quality is the form of analysis that is applied to the data gathered. While most users of standard assessments are aware of means, stanines and the like, most are unused to other techniques. An example of a technique which is different from the usual, yet simple in its use and interpretation is that used in the Group Review of Algebra Topics (GROAT). For each question in a GROAT unit there are four possible answers: one of these is correct, the other three are incorrect. Analysing these distractors (incorrect answer options) gives the most detailed diagnostic information, since each distractor has been designed to elicit particular student misconceptions or weaknesses. The GROAT analysis chart not only facilitates the correction of student answers, but also allows teachers to assess the extent to which particular errors exist in the group of students assessed. In addition teachers may arrange the students' answers to facilitate further diagnosis. For example, by ordering the students' answers in overall performance order (that is, in order of best to worst total scores) group difficulties with specific question types can be assessed by visual inspection. Other arrangements to suit the needs of the teacher are also possible. Such analyses are of course variations on what many of us already do, but there are other more sophisticated analyses now available. These newer methods of analysis are attempting to maximise the information available to teachers while minimising the time and work required. I would like now to discuss some of these techniques.

The first technique relates to the Collis-Romberg Mathematical Problem Solving Profiles. The analysis used here is based on the Structure of Observed Learning Outcomes (SOLO) taxonomy (Biggs and Collis, 1982). The SOLO taxonomy describes the types of mental manipulations available to students who are functioning in the sensori-motor, ikonic, concrete-symbolic and formal modes (the first four Piagetian modes). In brief, the elements of the SOLO taxonomy used in the Profiles are:

Unistructural where the student uses one obvious piece of information coming directly from the stem of the question;

Multistructural where the student uses two or more pieces of information from the question, but treats them as separate;
Relational where two or more pieces of information are combined to show an integrated understanding;
Extended abstract where the student uses an abstract general principle derived from or suggested by the information in the question.

While it is often possible to infer these levels of thinking from a student's response to a question, in the Collis-Romberg Profiles questions have been structured to elicit these levels of thinking. This approach maximises the opportunity to observe the levels of thinking. Thus in the example given earlier, the response required by part A is merely a restatement of the stem (a unistructural response), the response to part B requires all the information to be used in sequence, in a number of discrete steps, like a set of instructions to be followed. This requires multistructural thinking. The response to part C involves being able to understand the principle involved in order to 'reverse' the instructions. This indicates a relational thinking mode. Finally, a correct response to part D requires the extraction of

the general principle, writing it in abstract form while dismissing any distracting cues. This is the extended abstract thinking mode. In the same way all the Profiles questions have been carefully structured.

The student score sheet shown in Figure 4 shows how Collis-Romberg Profiles questions are corrected. A tick or cross as appropriate is given for each part of each of the five Profiles questions and the 'last' (highest thinking level) tick for each question shaded. By totalling the U, M, R and E columns the overall highest thinking level can be estimated. In Mary's case this is M or multistructural. When the number of ticks in the question rows are totalled, areas of mathematical weakness become apparent. In the sample given, both algebra and measurement are problems. The last, but certainly not the least, step is to look at the intersection of the SOLO level row and the area of mathematical weakness column. (These intersections are numbered 35 to 52, which correspond to page numbers in the manual). The intersection number indicates the page where learning activities are described, suitable for a student at that level of thinking and with those particular weaknesses. Sample activities for Mary in algebra are given on page 40 of the manual, and some of these are shown in Figure 7.

Insert Figure 7 about here

In the Collis-Romberg Profiles the careful structure of the questions, combined with analysis of responses based on a strong theory-driven framework provide much more than a standard assessment which gives at best normative information only. The

use of the SOLO taxonomy gives access to the level of the thinking processes being used by the student, while the range of the mathematical topics provides a diagnosis of weaknesses (and strengths) in mathematics. The combination of these two insights provides a powerful tool for the classroom teacher.

The Profiles of Problem Solving (POPS) focuses not on the student's levels of thinking, but on the skills and strategies used by the students. POPS questions require a variety of problem solving strategies for their solution

The categories of problem solving abilities assessed by POPS include one not usually found in mathematics assessments, quality of explanation (of the solution process). The authors of POPS believe that explaining one's thinking processes is a very important, but often neglected, part of mathematics. This became very clear during the development phase of the POPS instrument, where students often corrected their erroneous working when faced with having to communicate their solution strategies. The development of each of the five problem solving processes has been expressed on continua based on the scoring criteria for the POPS questions. Development on each scale has been described verbally, based on responses from sample students. To further assist teachers these scales have been sub-divided into levels of problem solving proficiency (Beginning, Developing and Advanced). A sample POPS profile is shown in Figure 8.

Insert Figure 8 about here

The descriptors for POPS proficiency levels are shown in Figure 9. These descriptions are extremely valuable for reporting to parents and colleagues, since they give much more information about a student than simple numerical scores.

Insert Figure 9 about here

The students' responses to POPS questions are also assessed by considering the strategies used to obtain solutions. These strategies include being systematic, visualising, and estimating. By assessing these strategies, deficiencies in a student's problem solving repertoire can be gauged, and suitable teaching activities (suggested in the manual) brought to bear on the

problem. Figure 10 describes some of these suggestions.

Insert Figure 10 about here

POPS has the advantage over other problem solving assessments in that it offers information on both skills and strategies, makes links between these, and offers suggestions for fostering improvement. This is quality assessment!

Another technique for analysing student responses, which has been used for several of the assessments described above, is that involving probabilistic models of student responses. The model

employed in these assessments is the model developed by Georg Rasch and which bears his name. The Rasch model has one very important feature which makes it eminently preferable for educational use. This feature is that it measures the difficulty of questions and the ability of students on the same scale. Because of this feature it is possible to estimate the likelihood of a student of a particular ability correctly answering a question of a given level of difficulty. This information can be expressed visually through what has been termed a 'kidmap' (Wright, Mead and Ludlow, 1980). Figure 11 shows a simple kidmap describing a student's performance on part of a primary mathematics curriculum. The questions are arranged vertically in difficulty order and the student's ability estimate is recorded on the central part of the scale.

Insert Figure 11 about here

How does one interpret a kidmap? It is useful to think of a kidmap as being divided into four quadrants by the intersection of the central scale and an imaginary line drawn through the estimated student ability (in the case of Tony in figure 11 this is about 62). The two quadrants to the left of the central ability scale record the numbers of those questions that Tony had correct and those to the right of this scale those he had wrong. Questions below Tony's ability estimate he is expected to get correct, and the kidmap clearly shows those questions for which this is true. On the 'incorrect' side of the scale are those questions that Tony had wrong but was expected to have correct. These indicate aspects for remedial instruction. Those questions above Tony's ability estimate we expect him to have wrong as they are expected to be too difficult for him. Thus those questions on the 'wrong' side are of no surprise or worry. However, those questions that Tony had correct which are above his ability estimate are an indication of, perhaps, a particular strength in his mathematical skills. In the case of this kidmap the teacher has written comments to further focus attention on Tony's needs. The kidmap provides an excellent visual interpretation of data usually conveyed less clearly by tables. In practice kidmaps ease the task of identifying students with specific strengths and weaknesses. Of course the software used to build kidmaps can do so for any number of students or indeed even produce similar maps for groups of students. In the case of the BSTP some 300 000 kidmaps have been produced in the last four years giving teachers considerable detail of their students' strengths and weaknesses. Unfortunately as you will have realised, to produce kidmaps one needs sophisticated software and hardware which are not always available to schools or even school systems. There is a solution to this problem. If it is possible to establish question difficulties via a trial group of students, then a version of kidmaps can be employed. This alternative is the diagnostic map (DIAMAP) and it allows the classroom teacher to produce kidmaps

with nothing but a pencil! The idea of the DIAMAP is quite simple; a sample of students is used to calibrate the questions from the assessment instrument. Next, questions are printed in their difficulty order on both sides of the central ability scale. The central scale is marked with raw score values in the correct position to give ability estimates. To construct a

DIAMAP for a student one simply circles those questions the student had correct on the 'correct' side of the DIAMAP and those which were incorrect on the 'incorrect' side of the DIAMAP. Finally the student's ability estimate (raw score) is circled on the central ability scale. The interpretation of the DIAMAP now proceeds exactly as for a kidmap. Figure 12 shows a completed DIAMAP from the Diagnostic Mathematics Profiles (addition). DIAMAPs have been used in mathematics, music, English, Indonesian and other languages (Doig, 1992). Their use is only restricted by our imagination.

Conclusion

As the century draws to a close it is imperative that we address the question, 'How do we produce the best assessment?' The answer is of course through striving for quality in our assessment programs, and this quality assessment in mathematics is within our grasp in a variety of ways. Whether it is through the construction of instruments using student as well as expert input or carefully selected questions, or whether we choose to place our emphasis on the best possible analytic and reporting techniques, the choice is ours. In the foregoing many different instruments have been described, all different but having in common the stamp of quality. It is now up to us as educators to take these tools and use them in the best interests of our students.

References

- Biggs, J. B. and Collis, K. F. (1982). *Evaluating the Quality of Learning: The SOLO Taxonomy (Structure of Observed Learning Outcomes)*. New York: Academic Press.
- Collis, K. and Romberg, T. (1992). *Collis-Romberg Mathematical Problem Solving Profiles*. Hawthorn: Australian Council for Educational Research.
- Doig, B. (1990). *Diagnostic Mathematical Profiles*. Hawthorn: Australian Council for Educational Research.
- Doig, B. (ed) (in press). *Stop! Look and Lesson*. Hawthorn: Australian Council for Educational Research.
- Doig, B. A. (1991). *Group Review of Algebra Topics*. Hawthorn: Australian Council for Educational Research.
- Doig, B. A. (1992). *DIAMAPs – Self-correcting Kidmaps*. Paper presented at the American Educational Research Association annual conference, San Francisco, April 1992.
- Masters, G. N., Lokan, J., Doig, B., Khoo, S-T., Lindsey, J., Robinson, L. and Zammit, S. (1990). *Profiles of Learning*.

Hawthorn: Australian Council for Educational Research.
Stacey, K., Groves, S., Bourke, S. and Doig, B. (in press).
Profiles of Problem Solving. Hawthorn: Australian Council for
Educational Research.

Wright, B. D., Mead, R. J. and Ludlow, L. H. (1980). Kidmap.
Research Memorandum Number 29. Chicago: Statistical Laboratory,
Department of Education, University of Chicago.

Figure 1: Easy (early) and harder (later) questions from the
Group Review of Algebra Topics (Manipulation).

Figure 2: Questions ordered (by sample students) for the Basic
Skills Testing Program.

Figure 3: Sample teacher's page from the Victorian Standard
Achievement Tasks (Space collection).

Figure 4: Sample diagnostic profile from the Collis-Romberg
Mathematical Problem Solving Profiles.

Figure 5: Sample item from the Profiles of Problem Solving
(POPS).

Figure 6: Sample error from the Stop! Look and Lesson and a
suggested remedial activity.

Figure 7: Sample activities for Mary in algebra from the Collis-
Romberg Mathematical Problem Solving Profiles.

Figure 8: Sample student profile from the Profiles of Problem
Solving (POPS).

Figure 9: Level descriptors from the Profiles of Problem Solving
(POPS).

Figure 10: Suggestions for further learning from the Profiles
of Problem Solving (POPS).

Examples of suitable activities to help develop estimation
strategies are 'Hundreds and Thousands' from Teacher Tactics for
Problem Solving (Stacey and Southwell, 1983) and the 'Map of
Australia' activity from Mathemastics Curriculum and Teaching
Program Activity Bank (Lovitt and Clarke, 1988).

Suitable resources for teachers wishing to develop this skill [visualisation] among their students include the materials Tangram Sets , Tangrams for the Overhead Projector, Tangram Template and the books Tangrams (Millington, 1986) and Tangram Treasury (Fair, 1987), all of which use tangram puzzles as the medium for strengthening visual skills. Other suitable resources include the books Symmetries with Pattern Block Designs (Davidson and Willcott, 1987), Visual Thinking Cards (Seymour, 1983) Geometric Playthings (Pederson and Pederson, 1973), Build Your Own Polyhedra (Hilton and Pederson, 1988), Exploring with Polydrons (Komarc and Clay, 1991), Puzzles in Space (Stonerod, 1982) and the DIME materials.

Figure 11: Kidmap for part of a primary mathematics curriculum.

Figure 12: DIAMAP for addition from the Diagnostic Mathematics Profiles.