AN INTRODUCTION TO THE ASSESSMENT RESOURCE BANKS (ARBS) AND THEIR DIAGNOSTIC POTENTIAL.

Paper presented at Technology in Mathematics Education (TIME 2000)
December 2000, Auckland, New Zealand

Alex Neill

Researcher, NZCER,
P.O. Box 3237,
Wellington,
New Zealand.

The Assessment Resource Banks (ARBs) are computerised banks of assessment material that are available on the Internet. They are linked to the current New Zealand curriculum statements in mathematics, science and English. This workshop will introduce and demonstrate the ARBs. This will be followed by a description of the diagnostic potential of the ARBs. The final part will be an interactive discussion of the diagnostic dimension of a selection of resources from the mathematics ARB.

Introduction

The Assessment Resource Banks (ARBs) are computerised banks of assessment material which are based on the New Zealand curriculum statements in mathematics, science and English. They are all at Levels 2 to 5 of the curriculum, except for a few science resources at Level 6. The ARBs first opened in 1997. As at the end of June 2000, there were 951 mathematics resources, 953 science and 304 English resources. They can be used not only for summative assessment, to identify what a student has achieved, but also for formative or diagnostic assessment, to identify and remedy student problems.

One of the main purposes of the ARBs is to give teachers a wide range of assessment material. This includes formative and diagnostic assessment as well as summative assessment. The ARBs are similar to the traditional item banks. They are distinctive in that they are delivered electronically and employ a customised search engine that allows teachers to access relevant, valid and reliable assessment materials. The questions cover a range of styles rather than just multiple choice. This allows the ARBs to be used for a range of styles including both summative and formative assessment. The ARBs are the only such collection available.

The ARBs can supplement the assessment tools that schools currently use. The ARBs are at the more formal end of assessment, but they do incorporate a variety of question types. They include multiple-choice questions, as do most item banks. They also include questions where the students have to construct their own responses or perform practical tasks (Croft, 1998). This paper looks at the formative and diagnostic dimensions in more detail.

The ARBs are available to all New Zealand schools free of charge. The ARBs are accessible on the Internet at http:\\www.nzcer.org.nz. The ARBs let teachers find relevant resources using the search capability of the Internet. They incorporate a classification system based on the New Zealand curriculum statement. This allows teachers to match their own teaching and curriculum objectives. This is analogous with locating a library book using a computer. Teachers can then design valid assessments specific to their own classroom and school requirements.

Development of the ARB

The ARB project began in February 1993 when NZCER was contracted to undertake a feasibility study for the Ministry of Education. This was followed in December 1994 with an implementation trial in 27 schools. The focus was on both large-scale national and school-based assessment in mathematics and science. This was especially focused at the transition points between primary and intermediate, and intermediate and secondary. Since 1996 the banks have become a reality with rapidly increasing numbers of both resources and registered user. The English ARB came on-line in September 1998. School-based uses have predominated, and now have become the sole focus.

Resources are written in a variety of ways, by panels of teachers, NZCER staff (which include primary and secondary teachers) and other consultants. Before a resource is trialled it goes through an extensive process of reviewing to ensure maximum validity and usefulness. Once this is completed, resources are grouped together. Trials involve about 200 students from about eight schools. The schools are chosen to get a representative sample to ensure a good distribution geographically and by SES.

The results are then marked, and scoring rubrics are modified if necessary. The results are subjected to a classical analysis to ensure reliability. This analysis also allows each resource to be given a level of difficulty. Individual schools can then use the level of difficulty as a benchmark to compare their own students performance with a national sample.

Resources are then reviewed again. Acceptable resources are put onto the NZCER website using HTML. Some resources are amended and retrialled, while a small minority are rejected at this point.

The structure of the on-line ARBs

The ARBs can be accessed through NZCER's Homepage (http://www.nzcer.org.nz) by clicking on the Assessment Resource Button. From the ARB homepage there is access to:

1. the mathematics, science, and English banks,
2. the online registration page,
3. information about various aspects of the ARBs, and
4. NZCER's Homepage.

Before teachers can use the banks they must currently obtain a password from NZCER, which requires filling out the online registration form. NZCER monitors the profile of ARB users over time. The total number of schools registered as at July 2000 was 1825 compared with 1100 in May 1999. Most of the current registrations are schools, but some are other educational institutions, including educational consultants and colleges of education.

The mathematics search page

The mathematics search page is where the user can define the parameters they wish to search by. Each of these has a drop-down menu of options to specify. The first four of these search parameters are explicitly modeled on *Mathematics in the New Zealand curriculum*. The search parameters are:

1. Strand *Number, Measurement, Geometry, Algebra* and *Statistics*.

2. Level — *Levels 2 to 5* of the curriculum statement.
3. Objective — Specific to the strand chosen e.g. *Exploring number*.
4. Process Skill — *Problem Solving; Logic and Reasoning; Communicating Mathematical Ideas.*
5. Resource Type — *Selected Response (including multiple choice) Brief Constructed Response, Longer Constructed Response,* and *Practical*.
6. Keyword

Keywords are brief descriptors of the main elements of the task. A thesaurus of the keywords that are used is available online, and allows a different strategy of searching to the more usual strand-level-objective route. Keywords can be concatenated with the usual boolean *and, or, not* type constructs.

A search is carried out by selecting specific search fields. For example:
Strand: *Number* Level: *five* Objective*: Exploring number*
This will list out all the resources that meet all of these criteria. The user can then browse from the resources listed (by clicking on the resource number) and chose the ones they find relevant. If the user knows the resource number, then they may also search by this.

The structure of a resource
Each resource is presented in a standard format as follows:
Page 1: Assessment task(s) or question(s).
A *scoring* button to go to page 2.
Page 2: A scoring guide which gives:
a model answer,
a recommended mark allocation,
a degree of difficulty for each part of the question.
Diagnostic information – available on some resources only.

The level of difficulty shows how students performed on this resource during trial testing of about 200 students in the clustered, representative sample of schools. The year level and date of the trial are also given. Table 1 shows the percentage bands associated with each difficulty level that have been used to indicate students' performance. The bands are deliberately broad, reflecting the margin of error associated with a sample of about 200 students.

Table 1: Difficulty levels and percentage bands of student performance.

Difficulty Level	**Percentage Band**
Very Easy	80% and above
Easy	60% to 79.99%
Easy	60% to 79.99%
Moderate	40 to 59.99%
Difficult	20 to 39.99%
Very Difficult	19.99% and below

Use of resources.

Because the ARBs have a wide range of question types they can be used for a variety of purposes. These can be briefly summarised as follows:

a) *Formative assessment*. This helps indicate how well a student is progressing with the intention of giving a guide to further teaching and learning.
b) *Diagnostic assessment*. This analyses a student's response to indicate the misconceptions which the student has with the aim to rectify these.
c) *Summative assessment*. This aims to assess what a student has achieved at the end of a period of teaching or for reporting purposes.
d) *Pre- and post-tests.* The administration of the same of similar resources before and after a period of teaching to assess their prior knowledge before and their understanding after teaching.
e) *Monitoring school-wide performance*. This could either be over time by a repeated administration of comparable collections of resources under uniform conditions to build a picture of student performance or as a one-off to compare different groups of students across the school.
f) *Confirming and reporting levels* of students' performance against national standards.
g) *Exemplars of assessment material*. The ARB resources reflect good assessment styles that teachers may wish to adapt or follow when writing their own assessment material.
For a discussion on the seven different types of assessment the ARBs can be used for, see Croft (1999).

Diagnostics and the ARB

Analysis of the ARB trials initially centred on the correct answers. This left a large part of the data untapped, namely the pattern of incorrect responses. During work on the ARB mathematics project, it became increasingly clear that the incorrect answers that students came up with provided valuable insights into their thinking, the common misconceptions that they held, and the incorrect processes that they were applying. These wrong responses are now being turned into diagnostic information.

The idea of diagnostics emerged only part way through the life of the project. This means that there are many resources in the ARBs that were not analysed for diagnostic purposes. Virtually all the resources that are currently being trialled and entered into the banks are being analysed to see if they give useful diagnostic information.

This diagnostic information can be used by teachers in their planning, as it gives an indication of how difficult or easy the area they plan to teach is. It also gives them information on the misconceptions they are likely to encounter among their students. It can be used diagnostically to help identify possible causes of particular errors. This can be fed back into teaching to help rectify students' misconceptions or errors in performing the mathematical activities.

For classroom use, teachers may wish to have a face-to-face conference with an individual student to confirm the cause of error. They may also have a small group session, a whole class discussion, or a whole class teaching session. The diagnostic information can be used in these sessions as starting points to help teachers establish the true causes of error.

Analysing Incorrect Responses to Create Diagnostics

To produce diagnostic information on a question, the pattern of incorrect answers must be analysed. This is a routine task for multiple-choice questions. If students construct their own responses, then the likely underlying causes for the incorrect responses must be analysed.

In mathematics the answers are often numerical. There is a need to unwrap the likely cause of the wrong numerical answer. The likely incorrect calculation is given, with an accompanying likely underlying cause. Sometimes a wide range of wrong numerical answers are given. Neill (1997) quotes an example where 60 different wrong answers were given by students, 30 of which had clear diagnostic potential. Not all of these could be quoted. For many incorrect responses, however, no likely underlying cause could be identified. If the answer is not numerical in nature, only a likely underlying cause is given.

A two-pronged approach is used to analyse constructed responses. Firstly, our markers look for common incorrect responses as they mark. The markers can sometimes infer the misconceptions from clues in the students' scripts (their working, etc.). A sub-sample of approximately two to three classes of students have their scripts analysed in a more rigorous way.

The causes of the errors are then inferred from the wrong answers. It would be more satisfactory, of course, to be able to ask the students what their thinking processes were. However, this would be an enormous task to perform on hundreds of questions given to scores of students. The approach of inferring the misconception is ideal for numerical responses.

It is highly probable that a numerical answer has been arrived at via the anticipated method, and there is only a slight chance that an alternative logic has been applied. Sometimes the erroneous calculation has been written in the margin of the test script, increasing the marker's confidence in the misconception being the one identified.

On the other hand, in a multiple-choice situation each distracter has a diagnostic explanation, but we do not know if the student engaged in this anticipated logic, or used alternative approaches such as language clues or other test-wise strategies. They may well have been guessing, in which case the diagnostic potential of the distracters drops to zero. There is usually less evidence of working associated with multiple-choice responses.

Producing diagnostics

Two criteria are used for establishing diagnostics:

1. Is the error common to many students?
2. Is there a clearly definable misconception or incorrect process that the response indicates?

The diagnostics are then written onto each resource, with the common errors and the likely misconception identified. Often this is accompanied by the likely calculation that the student has used. These diagnostics are specific to the actual questions in an individual resource.

Other Diagnostic Approaches

Several authors have addressed the issue of diagnostic information in mathematics. Two distinct styles of diagnosis are made: those that focus on the mathematical errors made, and those that look at a wider range of error sources. The first style is typified by Palmer (1994) and Booker (1995), and the second style by Newman (1977, 1983) and Kirsch and Mosenthal (1993). *Mathematics in the New Zealand curriculum* (1992) espouses a model of mathematics in a real world context. This makes the diagnostic styles of the latter more relevant. However, the more clearly delineated mathematical misconceptions of the former make them highly useful as well. Newman looks at five dimensions where the error may occur:

1. *Reading* the problem.
2. *Comprehending* what is read.
3. *Transforming* the problem into the appropriate mathematical form.
4. Applying the *process skills* selected correctly. This means correctly performing the mathematics involved.
5. *Encoding* the answer in an acceptable written form.

This model is similar to that of Kirsch and Mosenthal, who describe three dimensions:

1. I*dentification*: broadly speaking this encompasses the reading and comprehension dimensions.
2. P*roblem formulation*: this is the same as Newman's transformation dimension.
3. *Computation:* this has a somewhat narrower connotation than does process skills.

Newman's encoding dimension is a vital extra dimension that completes the model. Two distinct styles exist for deciding what category of error has been made. Both Newman and Booker use one-to-one interviewing of children. While this is ideal, it is often not practical in a classroom setting, nor even for producing large collections of questions such as the ARBs.

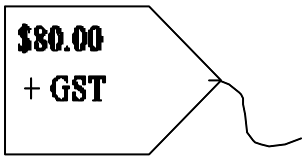
The alternative approach is to perform a test. Something like the PAT test outlined by Reid (1993), or the Diagnostic Mathematics profile of Doig (1990), or even a school or teacher produced pre-test are the typical options here. In the ARBs, however, we were not able to exploit the cumulative diagnostic evidence directly, but had to infer it from students' answers, or occasionally from their working.

To demonstrate the different classes of Newman errors, Example 1 has been analysed. This example comes from the Number strand:

Example 1

Number - Level 5; Computation. Part a) of resource number **NM1082**

Jenny saw a jacket that she wanted to buy. The price tag looked like this:



a) If the rate of GST is 12.5%, what was the ***total cost*** *of the jacket?* $__________

Reading/Comprehension	If an answer of $10 is given, the likely explanation is that the student did not *read* carefully. They gave the GST component rather than the total cost.
Transformation	If the answer given is $92.50 it is highly likely the student added $80.00 and 12.50. This means they did not *transform* the problem into the appropriate mathematics.
Processing	If the answer given was $86.40, it is highly likely that the student made an error in turning 12.5% into 0.08 (i.e. they confused one eighth and 0.08). This is an error in the *processing* of percentages and decimal fractions.
Encoding	If the answer given was $9000, the student has wrongly *encoded* 9000 cents as $9000.

If these errors were commonly seen, then they would be included in the resource as a diagnostic. This would be specific to the actual resource, rather than being described in more general terms.

Although detailed discussion of errors is beyond the scope of this paper, it is interesting to note that many errors stem from reading or comprehension errors. This is particularly true in the Number strand. This finding persists right through to Level 5, which we have typically tested on Year 10 students. Many other errors indicate that the student cannot turn the word problem into the correct mathematics (i.e. a transformation error). Less than half of the diagnostics concern incorrectly performing the appropriate mathematics. The reading component is probably underestimated in mathematics, because students who cannot adequately read the questions will either not respond at all, or come up with idiosyncratic answers that will not lead to diagnostic information.

Each question in the ARB focuses on a particular objective and skill (i.e. mathematical tool). It is useful to distinguish what is being tested. One useful model for this is that put forward by Clark (1997) when he discussed the hierarchy of assessment. He differentiated between tool possession, tool understanding, tool application and tool selection. Clearly these will lead to different emphases of diagnostics. Tool possession and tool understanding will rely more on process skills, while tool application and tool selection will span the range of Newman's dimensions far more extensively.

Other Potential Uses of the ARBs

With the exception of papers on the diagnostic uses of the mathematics banks (Neill, 1997, 1998; Gilbert & Neill, 1999; Croft, 1998), most other articles on the ARBs have concentrated on their development and organisation, or on how teachers can access the resources and use them for classroom assessment.

Clearly, there is broader information available from the ARBs than sets of resources containing items and tasks for classroom assessment purposes. As indicated in the examples covered earlier, the ARBs provide solid empirical data to underpin formative assessment. This is the fundamental purpose of the ARBs in classrooms (Croft, 1999, p. 32). These data may be used to alert teachers *in advance of teaching a topic* to:

- possible misconceptions their students might demonstrate;
- common errors they might make;
- aspects of their study that they are likely to master readily;
- understandings they may generalise to other aspects of their learning.

The information on common errors and misconceptions may also play a role in helping teachers and children create realistic expectations of what children might know about particular aspects of mathematics. Knowing, for example, that stem-and leaf graphs are often not known at Levels 2 - 4 (see Gilbert & Neill, 1999) will help put individuals' accomplishments related to these graphs into a firmer framework of expectations.

Beyond individual schools or classrooms, accumulated ARB data could help indicate areas of the mathematics curriculum where national samples of children are performing strongly or weakly. Insights of this nature would be invaluable if an empirical review of the curriculum were to be undertaken. They would also be a major improvement on simply relying on teachers' unsupported anecdotal judgements about the areas of the curriculum that 'work' and those that 'do not work'. Information based on the performance of diverse national samples on ARB resources is potentially more useful than data from mandatory national testing, when the data are required for curriculum review purposes. This is basically because data accumulated from ARB resources administered in many classrooms enables a broader range of curriculum outcomes to be assessed than is possible under the restrictions of national tests.

The ARBs are now established as the major collection of nationally developed material, in science, mathematics and English, published for classroom assessment purposes. They will be able to provide an objective basis for reviewing curriculum statements and focussing teachers' professional development, in order to strengthen students' achievement in identified areas of these curriculum statements.

A key point about the diagnostics part of this paper is that it is worth reflecting on the processes that a student may have used to come up with an incorrect answer. Often the answer alone will provide enough clues. This is especially so when the answer is a constructed response and is numerical. Incorporating diagnostic information about common incorrect answers, and likely explanations, gives teachers a powerful tool, not only to test students understanding, but also to remedy their misconceptions. Roadblocks to student understanding and learning can be identified, and appropriate strategies to remedy them can be developed.

References

Booker, G. (1994) *Booker profiles in mathematics numeration and computation*. Melbourne: Australian Council for Educational Research.

Clark, C (1997) *Teaching learning and assessment: an essential symbiosis.* Paper presented at NZAMT conference, Palmerston North, New Zealand.

Croft, C. (1998) Testing on the net: Assessment Resource Banks in mathematics and science. ***set:*** *Research information for teachers, No. 1*. Wellington: New Zealand Council for Educational Research.

Croft, C. (1999) *School-wide assessment: Using the Assessment Resource Banks.* Wellington: New Zealand Council for Educational Research.

Doig, B. (1990) *Diagnostic mathematical profiles*. Melbourne:Australian Council for Educational Research.

Dunkels, A. (1991) Interplay of the number concept and statistics using attitudes and techniques of EDA., *Proceedings of the third international conference on teaching statistics pp. 129-139*. Dunedin, New Zealand.

Gilbert, A. & Neill, A. (1999) *Assessment Resource Banks in mathematics and their diagnostic potential*. Paper presented at NZAMT conference, Dunedin, New Zealand.

Kirsch, I. & Mosenthal, P. (1993) Strategies for solving math document problems, *Journal of Reading* pp. 498-501.

Mathematics in the New Zealand curriculum. (1992) Wellington, New Zealand : Ministry of Education.

Newman, M. (1977) An analysis of sixth-grade pupils’ errors on written mathematical tasks, *Victoria Institute for educational research bulletin, 39,* pp. 31-43.

Newman, M. (1983) *Strategies for diagnosis and remediation*. Sydney: Harcourt, Brace, Jovanovich.

Neill, A. (1997) *Time or place? Diagnostic information in mathematics*. Paper presented at 19th NZARE conference, Auckland, New Zealand.

Neill, A. (1998) Assessment Resource Banks in mathematics: How can we help teaching and learning? *Mathematics teaching: Where are we at? Seminar proceedings.* Wellington: New Zealand Council for Educational Research.

Palmer, D., Kays, M., Smith, A., & Doig, B. (1994) *Stop, look and lesson: Manual* Melbourne: Australian Council for Educational Research.

Reid, N. (1993) *Progressive Achievement Test of mathematics*. Wellington: New Zealand Council for Educational Research