

## Assessment Resource Banks for School-Based and National Assessment

CEDRIC CROFT

### *Abstract:*

*The New Zealand Council for Educational Research (NZCER) is developing assessment resource banks (ARBs) in mathematics and science. The ARBs are broader than traditional item banks, as they contain assessment resources varying from selected-response items to practical tasks. They reflect current New Zealand curriculum statements in mathematics and science, and are presented as computerised databases with classification and retrieval systems which enable users to select assessment material to match teaching objectives. The ARBs are available via the Internet for school-based uses within New Zealand. Their possible role in national assessment policy is also being considered.*

*This article outlines the development and structure of the ARBs, summarises an evaluation of school-based and national uses undertaken in 1995, and discusses the future possibilities of ARBs for school-based and national assessment purposes.*

The development of assessment resource banks (ARBs) in mathematics and science began in February 1993, when the New Zealand Council for Educational Research was contracted by the New Zealand Ministry of Education to undertake a feasibility study and report its findings. The second stage of the ARB project, the implementation trial, began in December 1994, and was designed to trial in 27 schools the resources developed up to that point. The third stage, which began in April 1996, developed more resources for the ARBs, and introduced them into schools. This work, which is incorporated within government policy on "transition-point assessment" continues by contract to the Ministry of Education.

### *Overview of Feasibility Study 1993–94*

The focus of the feasibility study was to prepare a range of assessment resources in mathematics and science, suitable for national and school-based purposes. The investigation of allied measurement issues, plus identification of principles for a model of computer retrieval were other aspects of the contract.

The feasibility study fell into four distinct parts:

- Organisation of the banks
- Content of the banks
- Functions and uses of the banks
- Reporting of performance.

A full overview of this first stage is included in NZCER's final summary report to the Ministry of Education (Reid et al., 1994). Papers by Reid (1994) to the Australasian Conference of Assessment and Certifying Authorities (ACACA), and Croft (1994) to the International Association for Educational Assessment (IAEA), have also commented on aspects of the 1993–94 feasibility study.

### *Overview of Implementation Trial 1994–95*

The implementation trial following the feasibility study also concentrated on mathematics and science at curriculum levels 3–6, with a focus on Years 7 and 9 – two major "transition-points" within the New Zealand education system. Although the general feasibility of ARBs in mathematics and science had been established during 1993–94, apart from developing some 360 assessment resources, no operation of the banks in schools was attempted, and computer technology was not used to store or retrieve the resources.

The central feature of the implementation trial in 27 schools was that the banks were now to be used in schools, both for school-based assessment, and to simulate national assessment. The national assessment trial was designed to compare three different styles of test: in the first, the test was made up solely of selected-response items, which could be machine-scored, and schools had no choice of content; a second style was similar, except that trial schools could choose 50 percent of the items; a third style consisted of material chosen entirely at the school level from the ARBs. The complete trial was set within a balanced research design which would enable subsequent policy decisions to be based on sound evidence.

During the first half of 1995 the priority was to prepare and review assessment material for inclusion in the ARBs. Every resource was referenced back to the appropriate curriculum statement by a classification system reflecting the *learning strands, process skills, achievement objectives*, and *levels* of current New Zealand curriculum statements. The next priority was to conceptualise and plan computer software to represent the classification system, as well as store and deliver the assessment material to schools. Hand-in-hand with the conceptualisation and database architecture that followed, went the incorporation of the material into the database – a task which proved to be quite complex, because of the graphic content of many of the assessment resources written.

A full report of the implementation trial has been published by NZCER (Croft et al., 1996). A paper by Croft (1995) to the ACACA, discusses the implementation trial and other aspects of the project to that date.

#### *Current work*

The focus of work which began in April 1996, was to continue the development of mathematics and science assessment resources, and to make these available to individual schools via the Internet. A Ministry of Education survey (Owens, 1996) projected a rapid growth in the number of schools connecting to the Internet, with 84.3 percent of primary schools and 97.6 percent of secondary indicating that they planned connection by January 1998. How well New Zealand schools will be placed in 1997 to access the ARBs may be problematic, because of limitations of their current computer hardware, and lack of training of teachers in information technology (IT). Hard copy formats of the resources were shown by the implementation trial to be poor substitutes for the electronic versions of the banks, largely because they lacked the searching and retrieval power of the computer, the flexibility to update the resources, and the provisions to access new material easily. However, as an interim measure, some hard copy material will be made available to those schools without Internet access.

During 1996, work was guided by results and feedback obtained during the 1995 implementation trial. In particular, the search engine used for the trial was modified to allow a more flexible approach to identifying and retrieving assessment resources. The style of assessment resources within the banks was broadened, following a decision to increase the numbers of resources requesting answers in a constructed-

response format, rather than a selected-response format such as multiple-choice.

#### **Item Banks and the New Zealand Assessment Resource Banks**

At this point, some comment on the types of resources contained in the New Zealand trial banks may be in order. Item banks, also known as “item pools”, “question banks”, “item files”, or “test item libraries”, have been in existence for 30 years or more. Essentially, they comprise a group of items arranged and classified in a manner that makes them accessible to users “on demand”. However, experience has shown that it is the quality of the classification system as much as the quality of the items which often determines the popularity of item banks with teachers. “Classification is the key that unlocks the item bank” (Millman & Arter, 1984). Unless their contents can be retrieved quickly and precisely, teachers will leave item banks on their shelves, after the first exploratory trial.

Hitherto, multiple-choice items have usually outnumbered other types of assessment material in established banks, as they are easy and quick to mark objectively, and calibration data can be more readily calculated. The experience of most New Zealand teachers with item banks has been largely of hard-copy collections of mathematics items. But item banks of this type have been regarded as less than “user friendly”, mainly because the identification and retrieval of suitable items has been tedious, the business of cutting and pasting material has proved cumbersome, and the regular updates necessary are difficult and expensive to obtain.

#### **The New Zealand Assessment Resource Banks**

For the New Zealand ARBs, a resource which is broader than the traditional item bank is now being extended, with a major aim of helping teachers assess important aspects of national curricula in mathematics and science. The emphasis is on making available a range of tasks which reduce emphasis on a simple objective approach to assessing educational achievement, and encourage extended student responses in problem-solving contexts. This approach poses significant challenges for those who contribute material to the banks, and for those who use the material for assessment.

From the outset there was considerable debate on the dimensions of ARBs which would make them suitable for national use, and at the

same time, sufficiently flexible to be valid within a context of school-based assessment. This is the classic dilemma of trying to have one assessment strategy serve multiple purposes. Nevertheless, in the case of the ARBs, it is anticipated that by providing a diversity of resources, plus guidance for teachers, some of these difficulties will be overcome. The eventual scope of school-based uses may need to be modified by the requirements of national uses of the banks, within the government's policy of transition-point assessment. At this time these requirements remain unknown.

The ARBs are not seen as alternatives to a school's own assessment material, but as a source of additional material. It is also clear that to incorporate the potentially wide range of assessment material, and allow teachers to utilise ARBs fully, a time horizon beyond 1997 is in view.

Figure 1 Structure of Assessment Resource Banks

MATHEMATICS AND SCIENCE CURRICULUM STATEMENTS		
Achievement and Diagnostic Tasks and Items		
<i>Selected Response</i>	<i>Brief- and Longer-constructed Response</i>	<i>Practical Tasks</i>
Multiple-choice	Completion	Construction and assembly
Matching	Short-answer	"Fair test" experiments
Alternate-choice	Enhanced multiple-choice	Constructing geometric shapes
	Essay-type	Investigations
	Problem-solving	
	Interpreting tables and diagrams	
	Drawing geometric shapes	
	Planning investigations	
Attitude and Interest Scales		
	Rating scales	
	Checklists	
	Interview schedules	
	Questionnaires	

Figure 1 gives some idea of the way in which the resource banks are structured. At present they contain mostly achievement and diagnostic material, requiring a written response from students. But other tasks of a practical nature are envisaged in future; for example, constructing geometric figures, drawing plans, graphing data, carrying out simple experiments, all with scales or checklists to help assess the process and the outcome.

Although the response to the materials in schools has been very positive, the trials have revealed some problems in integrating the ARBs into school-based assessment programmes. Points needing further attention are:

- the difficulty of constructing valid and reliable assessment tasks, which are still manageable within classroom constraints;
- the time taken to extract useful diagnostic information from tasks requiring open-ended responses;
- the length of time students take to complete tasks requiring multiple responses; the time needed to assess certain curriculum objectives can expand to an unreasonable degree, or if an attempt is made to restrict it, inadequate sampling within objectives will limit the validity of the assessment.
- the influence of English skills on an assessment, when extended written or oral responses are needed;
- the context-bound nature of "authentic" (realistic, true-to-life) problem solving; attempting to create realistic problem-solving situations may help improve the validity of the assessment for some students, but may also disadvantage a significant minority who may not have experienced the particular "realistic" situation created for the problem.
- the lack of a strong "tradition" in New Zealand for the formal assessment of science at primary/intermediate school level;
- the discovery of simple but effective ways of reporting performance from complex assessment procedures.

**What Did Teachers Think about School-based Uses of ARBs?**

An important aspect of the feasibility study was to tap teachers' views of the proposed ARBs. Accordingly, samples of teachers of mathematics and science were surveyed in 1993. Teachers at this early stage were responding to information supplied by NZCER about the principles of ARBs, as the ARBs themselves were not then in existence.

For school-based uses, respondents supported a broad range of assessment resources and wished to have access to a variety of individual student and school-based results. Practical individual tasks, co-operative tasks, and checklists of observable skills were seen as priorities by both science and mathematics teachers. Respondents commented most often that they would favour resources that were:

- easy to use
- consistent with new curricula
- valid for assessing a broad range of skills, including individual and group skills
- applicable to both formative and summative assessments
- accompanied by statistics and “benchmark” information on each assessment resource
- available to teachers in hard copy and electronic forms.

Various aspects of the 1993 survey were replicated during the implementation trial in 1995. In addition, 60 teachers from the 27 trial schools were interviewed twice, asked to maintain logs and records, and requested to provide additional information about the trial. A further 58 teachers from these schools gave additional information on the trial of national uses that formed part of the implementation trial.

A summary of 1995 findings regarding the school-based uses of the ARBs, briefly stated, is as follows:

- There was strongest support for continued school-based use of the banks, with 90 percent of teachers involved in the trial wishing to see more assessment resources included, other areas of the mathematics and science curricula added, and the assessment resource bank concept extended to other curricula.
- Forty-nine percent of teachers nominated English as the next curriculum area for incorporation into the assessment resource banks model.
- Twice as many teachers preferred the computer versions of the banks to the hard-copy versions.
- A minority of users reported problems with the computerised versions; searching and selecting items, and lack of items in particular curriculum areas were the difficulties cited most often.
- Thirty-six percent of teachers reported that physical access to school computers was a barrier to use.

- Fifty-three percent nominated on-going help line support and 12 percent indicated more computer training as priorities.

### Future Possibilities for School-Based Assessment

The NZCER team that conducted the implementation trial was in no doubt that the proposed school-based uses of the assessment material represented an innovative approach in its contribution to quality school assessment practices. At the time the implementation trial was reported, the combination of electronic delivery, the broadening of the item-bank concept, and the development of the classification and retrieval system representing the main elements in the New Zealand Curriculum Framework, together represented a unique development in school-based assessment (Croft et al., 1996, March).

Strong support has been evident from New Zealand teachers for the ARB concept and its implementation, but this support was not unanimous. Most reservations came from the lack of coverage of some curriculum areas by the ARBs, and the formal and traditional nature of some of the material. Coverage of curricula will be improved with any continued work, as will the range of assessment materials developed. However, the resources included in the New Zealand ARBs will continue to be predominantly formal, as there is little alternative if satisfactory validity and reliability are to be achieved across a range of school sites. Resources that are embedded in the curriculum and which are essentially the same from school to school have increased chances of remaining valid measures. Reliability of scoring needs to be achieved if the performance data derived from national trial samples and included in every resource are to provide a meaningful benchmark for each school.

Some trial scoring guides developed have been regarded as too detailed by a minority of users. This suggests a lack of appreciation of the general concept of reliability, and inter-marker reliability in particular, where answers to constructed-response assessment tasks are concerned. A lack of reliability has been one of the problems encountered elsewhere with attempts to move towards performance assessment (Linn, 1994). The aim has been to eliminate as far as possible the error that might arise from vague, subjective scoring guides. The nature of score reliability, and how best to ensure consistency, must be addressed in more detail in future.

Materials to assess the achievement of practical outcomes in mathematics and science remain among the more demanding to

develop; clearly such material is most suited to school-based uses. The development of assessment material that requires students to recognise and select a possible answer in some shape or form is relatively straightforward to produce (although the continual challenge is to find novel ways of asking the question and presenting the stimulus material, while still maintaining a sharp focus on validity). The development of assessment material that has as its end point

- a product, for example, geometric figure;
- a process, for example, a science experiment;
- or a combination of process and product (e.g., a statistics project)

poses additional challenges for the specification of the task and its subsequent assessment. These types of assessment materials and activities take time and skill to develop. They are also the ones that are implicated in numerous New Zealand achievement objectives and valued highly by many teachers.

The strength of the ARB approach to school-based assessment is that materials may be prepared, trialled and calibrated by assessment specialists, then chosen by individual teachers to represent their teaching objectives. Curricular validity and high levels of reliability should be assured, and a good match between teaching and assessment should result. But small-scale trials have shown that issues of comparability between and within schools are difficulties which still need to be addressed if ARBs are to realise their full potential.

### What Did Teachers Think about National Uses of ARBs?

For national uses, most respondents in 1993 did not support information from ARBs being used to improve the allocation of resources to schools. Very few teachers supported national reporting of results, Ministry control of the banks, or Ministry access to the results for individual students. Though respondents supported *most* Ministry uses of this information, they were wary of the possible misuse of statistical information by commentators. Respondents supported the general conception of ARBs, providing:

- control of the resource banks remained within the school;
- results were available to schools for the purposes of *enhancing learning and teaching*;
- the banks would provide teachers with resources applicable to current assessment policy.

As far as the national uses of ARBs are concerned, the main 1995 findings may be summarised as follows:

- In comparison with 1993, there was *weakening* support for any results to be reported to, and/or used by the Ministry, or anyone outside the immediate school. This appears to reflect a general lack of support for national testing.
- In comparison with teachers' responses in 1993, there is *significantly increased* support for the Ministry to use ARB information to improve the allocation of resources nationally, and *significantly less* support for the Ministry to use ARB information to make public statements about national strengths and weaknesses in student achievement.
- Results provided to teachers whose students sat computer-marked tests, received *greater* support than the other forms of testing used. The more detailed information provided on students' performance appears to be of greater interest and use to teachers.
- Overall, teachers in schools where computer-marked tests were administered appeared to be *least* satisfied with the whole concept of national testing. This could be partly attributed to their lack of input into the whole process. Teachers in schools where other options for national testing were used, had some input into item/task/objective selection, and therefore had some influence over the test content. As a consequence, teachers may have helped assemble a test which was more valid.
- Since teachers have now had working contact with the banks, support for central marking has *increased* significantly since the feasibility study. This suggests that the benefits of central marking, and the nature of the analyses and score-reports possible under this option, are now better appreciated. It may also mean that in view of other workload pressures, teachers are less keen to be involved in marking national tests.

Papers by Croft (1996) to IAEA, and Croft, Boyd, Dunn, and Neill (1996) at the annual conference of the New Zealand Association for Research in Education have discussed teachers' views and other aspects of the project to date.

Overall, developing the resource banks for school-based uses was the strongest message in NZCER's findings. Teachers generally saw the

potential for the banks to add to their classroom assessment, planning, and teaching.

### Future Possibilities for National Assessment

From the outset, and within the context of the policy for transition-point assessment, ARBs were to be used for both school-based and national testing purposes. This has remained the official position since the beginning of this development. Three approaches to national testing were investigated during the implementation trial, but each was found wanting on a number of counts, particularly in terms of low validity. Accordingly, and in view of these validity issues, three alternative approaches to gaining national information were outlined in Croft et al., (1996). These are:

1. **Systematic review of performance data from banks.** Users would have an immediate benchmark against which to judge the performance of students if performance data (such as calibration data or difficulty indices) are regularly updated.
2. **Administration by schools of selected materials.** Schools could be requested to administer to their students a predetermined set of resources from the banks, or select resources from a predetermined range, or select resources from predetermined strands and/or objectives.
3. **Reporting by schools of the levels achieved by students.** The resource banks could be used by schools as the basis for reporting the achievement of all students by level and learning strand. These could be collated nationally.

These are all untried approaches to national testing, but they have the potential to utilise a wider range of resources, beyond multiple-choice items, developed for ARBs, and address validity issues by better sampling of the range of achievement within current curriculum statements in mathematics and science.

Tests consisting of selected-response items to be computer-marked are the cheapest option for national testing and make the least demand on teachers' time. All data obtained from this style of testing are identical in format, allowing for the greatest range of possible reports to be generated from machine scoring. By careful choice of common items between years and over time, growth and change may be measured on a full-cohort basis. Data of this nature could prove invaluable for

national planning purposes, as in traditional terms these objective tests achieve higher levels of reliability than newer forms of testing. Additionally, administration of objective tests is familiar to teachers and students, and responses to the items are not influenced by students' writing skills.

A major disadvantage is that a single national test made up of 50 selected-response items is unlikely to be a valid measure of many achievement objectives, or be in keeping with major assessment principles of the *New Zealand Curriculum Framework*. A flow-on effect is that the lowered levels of validity associated with selected-response items may reduce some of the benefits of higher levels of reliability of a traditional selected-response tests. Indeed, the lowered levels of validity must reduce the utility of the data for planning and other purposes.

The single issue of reduced validity of a selected-response test to assess current curricula is paramount. Considering the range of constructed-response material being developed for the ARBs, using selected-response items only as a source of national data, would not make optimum use of the resource banks.

On balance, the disadvantages were regarded as outweighing the advantages for those approaches to national testing investigated during the implementation trial. Accordingly, there was no recommendation to the New Zealand Ministry of Education to proceed with any form of national testing on the basis of the investigation so far.

Caution is still required when considering national testing and national information. Much clearer and more precise government policy is needed, including the purposes of national testing, the nature of the data required, and the uses to which the data will be put.

### References

- Croft, C. (1994, October). *Assessment resource banks programme*. Paper presented at twentieth annual IAEA conference, Wellington, New Zealand.
- Croft, C. (1995, August). *Implementing and evaluating assessment resource banks in mathematics and science*. Paper presented at ACACA conference, Fiji.
- Croft, C., Reid, N., & Livingstone, I. (1995). Assessment resource banks in mathematics and science. *Best of set Assessment*, 2, item 14. Wellington: New Zealand Council for Educational Research.

- Croft, C., Gilbert, A., Boyd, S., Burgon, J., Dunn, K., Burgess, L., Reid, N. (1996, March). Assessment Resource Banks in Mathematics and Science – Implementation Trial. Wellington: New Zealand Council for Educational Research.
- Croft, C. (1996, September). *Resource banks in mathematics and science for school-based assessment*. Paper presented at annual IAEA conference, Beijing, China.
- Croft, C., Boyd, S., Dunn, K., & Neill, A. (1996, December). *Symposium: resource banks in mathematics and science for school-based assessment*. Symposium presented at NZARE annual conference, Nelson.
- Linn, R. (1994). Performance assessment: Policy, promises and technical measurement standards. *Educational Researcher*, 23(9), 4–14.
- Millman, J., & Arter, J. (1984). Issues in item banking. *Journal of Educational Measurement*, 21, 4, 315–330.
- Owens, J. (1996, November). A survey of computer use in New Zealand schools. *The Research Bulletin*, 7, 1–10. Wellington: Research and International Section, Ministry of Education.
- Reid, N. (1994, July-August). *The development of assessment resource banks in mathematics and science: Problems and prospects*. Paper presented at the ACACA conference, Adelaide, Australia.
- Reid, N., Armstrong, L., Atmore, D., Boyd, S., Croft, C., & Livingstone, I. (1994). *Assessment resource banks feasibility study: Summary report*. Wellington: New Zealand Council for Educational Research.

### The author

Cedric Croft is currently Chief Research Officer at the New Zealand Council for Educational Research, where he has management responsibility for the Council's Learning, Curriculum and Assessment research programme. He has experience as a teacher, tutor, lecturer, school psychologist, member and chairperson of a School Committee and Board of Trustees, council member and secretary of the New Zealand Association for Research in Education. He has led the Assessment Resource Banks development since 1995.