

[Croft, C. (2002). The Assessment Resource Banks: From National Testing to a School-Based Resource. *New Zealand Annual Review of Education*, 11, 229-243]

## The Assessment Resource Banks: From National Testing to a School-Based Resource

CEDRIC CROFT

### *Abstract:*

*The Assessment Resource Banks (ARBs) in English, mathematics and science have been designed to reflect current New Zealand curriculum statements and provide schools with an ever-increasing range of valid and reliable assessment material. They are available on the Internet and are now established as the main source of school-based assessment material for New Zealand schools. During their early development phase, national and school-based uses were seen as dual elements, but they now function exclusively within a school-based environment. This article outlines the development and present structure of the ARBs, discusses their consolidation as a school-based resource, and suggests future directions.*

The Assessment Resource Banks in English, Mathematics, and Science (ARBs) developed at the New Zealand Council for Educational Research (NZCER) under contract to the Ministry of Education have been designed to match the structure and terminology of New Zealand curriculum statements in English, mathematics, and science, for levels 2 to 5. Mathematics and science resources first became available via the Internet in small numbers in March and May 1997. English was added in September 1998. As at March 2002, there were in excess of 3000 resources available, made up of 1250 in mathematics, 1300 in science, and 550 in English.

The ARBs have not been developed as alternatives to a school's own assessment procedures, but as a source of complementary material. They provide additional resources to help teachers assess progress within strands or achievement objectives, or "benchmark" the relative

230 Cedric Croft

performance of their students against the "typical" performance of national samples of students at given year levels. Generally, the ARBs are not for the use of students or parents, except that a new range of animated resources requires students to respond to a series of on-screen displays. These are a forerunner to the ARBs having some future interactive capacity.

A strength of the ARBs is that the published assessment materials are prepared co-operatively by assessment specialists and teachers, trialled with a national sample of students, and then selected by teachers to represent their own teaching objectives. This development process ensures acceptable levels of curriculum validity and reliability. The process of selection should ensure classroom validity, as teachers are free to select assessment resources that best match their curriculum objectives and teaching programmes.

Access to the banks is available to enquirers at the NZCER homepage at <http://www.nzcer.org.nz>. To use them it is first necessary to hold a username and password, available from the on-line registration form on the ARB homepage. Staff from registered schools, registered teachers, and staff from teacher support services, are among those eligible for a password. Access is available by arrangement for assessment staff internationally.

All resources are presented in a format that may be printed and photocopied. It is also possible to cut and paste resources electronically, and save ARB files to a word processing package so that the assessment tasks may be adapted to particular needs.

### **Background and Development**

The ARB project began in February 1993, when NZCER was contracted by the Ministry of Education to investigate the feasibility of establishing a bank of materials to assist in the assessment of students at the "transition points" of Year 6/7 and Year 8/9. There was a separate project studying assessment at the "transition point" Year 3/4.

The second stage for the ARBs, beginning in December 1994, was to implement them on a trial basis in 27 schools. The third stage, from April 1996, was to improve a search engine developed during stage two, extend the number of mathematics and science resources in the banks, and make these available to all schools via the Internet. The fourth stage, from July 1997 to June 1999, was to increase the number and range of mathematics and science assessment resources, continue to improve the flexibility of the search strategy, and add English

assessment resources. In the fifth stage, from July 1999 to June 2001 the ARBs continued to grow in size and encompass a more diverse range of assessment material, with strong growth in English being a feature. During this period the ARBs were confirmed as a school-based resource.

The development and growth of the ARBs during the period 1993–2000 have been covered in Croft, Reid and Livingstone (1995), Croft, Gilbert, Boyd, Burgon, Dunn, Burgess and Reid (1996), Croft (1996), Brown and Strafford (1997), and Croft (1999).

In the current phase from July 2001 to June 2003, a redesigned website is evident, more material with a strong formative basis is included, there has been a strong growth in school-based uses, and criteria have been developed to help determine the optimum size and general shape of the ARBs in future (Croft, 2001). This has been accompanied by a marked increase in web traffic.

### Development and Expansion of the ARBs 1997–2002

Since first opening with just 125 accessible resources, the ARBs have grown to 3100 resources as at March 2002, including 297 resources added from the Third International Mathematics and Science Study (TIMSS). The development and expansion of the ARBs are outlined in the following tables.

In Table 1, the numbers of resources developed and published at NZCER are combined for 12-month periods, 1 July to 30 June. For the three full years from 1 July 1998 to 30 June 2001, the annual average number published to the ARBs has been 613. This number breaks down into 162 for English, 228 for mathematics and 222 for science.

Table 1 Number of NZCER resources published to ARBs in English, mathematics and science by 12 month periods between 1997–2001

12 month periods	English	Mathematics	Science	Total
1 July 1997 – 30 June 1998	–	147	268	415
1 July 1998 – 30 June 1999	140	233	210	583
1 July 1999 – 30 June 2000	163	236	249	648
June 2000 – July 2001	183	216	208	607

It is apparent from Table 1 that the number of resources published in mathematics and science is less for the last year reported than it was for the two preceding years. This resulted from a policy decision agreed to with the Ministry of Education to increase the number of more complex resources designed to engage students in extended responses, with the aim of emphasising the bank's formative uses.

These figures indicate approximate future annual average growth of NZCER resources published to the ARBs, assuming present resourcing and reasonably constant specifications for the styles of assessment resource to be developed and the curriculum levels to which they apply.

Table 2 records the number of TIMSS resources incorporated into the ARBs over a twelve month period. Once they had been classified for their fit to New Zealand curriculum statements, these resources were incorporated into the ARBs, along with their published data on difficulty levels. They were arranged so users could find them by a simple keyword search, providing a timely boost to the numbers in the mathematics and science.

Table 2 Number of TIMSS resources added to the ARB in mathematics and science, by six-monthly periods, 1999–2000

Six month period to:	Mathematics	Science
June 1999	77	62
December 1999	74	66
June 2000	–	18
TOTAL	141	146

### The Structure of the Computerised ARBs

At the heart of the ARBs is the search strategy which allows users to select assessment resources that are most appropriate to their current teaching programmes. Previous experience with item banks (the forerunner of resource banks) had indicated that the quality of the classification system helped significantly in determining the usefulness

of the banks for teachers. Unless the contents of a bank can be retrieved quickly and precisely, the resources it contains may seldom be used. This is exactly where the ARBs' search strategy comes into its own, as it allows a specific search with particular assessment purposes in mind.

Because the structure of the banks and the associated search strategy reflect New Zealand curriculum statements, resources may be found by searching under the following four curriculum categories: *strand*, *achievement objective* or *function*, *process* or *integrating skill*, and *curriculum level*. In addition, two extra categories have been added to assist users find suitable resources: keywords and resource type. Each ARB resource has keywords or phrases that further describe the content and/or predominant skills tapped by the resource. Wherever possible the keywords are drawn directly from New Zealand curriculum statements. The keyword search is very popular with users. Details of these six categories and associated searches are outlined in Croft (1999).

### Resource type

All resources are also classified by the style of response expected from students. The following four types are used in mathematics and science resources:

- **Selected response (SR).** Students *select* a response from a range of options incorporated in the resource. Two or three multiple-choice or matching items may be grouped to form one resource. Examples include multiple-choice items, matching items, true/false and other alternate-choice items.
- **Brief constructed response (BCR).** The student *constructs* the response. Short answers, such as a word or two, a number or two, a phrase, or brief sentence are the essence of a BCR. Correct brief responses will encapsulate a single main idea. *Completing* entries in tables, graphs, or diagrams constitute a BCR. Examples include short-answer questions, completion items for tables, graphs, diagrams, plans, etc., problem-solving tasks requiring brief structured responses.
- **Longer constructed response (LCR).** These have the same general characteristics as a BCR but require a more extended response from students. The LCR resource is generally more open-ended than the BCR, and inferences may be needed to determine relationships

within the task. *Producing* tables, graphs, or diagrams constitute an LCR. Other examples include short essay-type questions (structured or unstructured); a written plan for an experiment, investigation, or practical task, conceptualising and/or producing tables, graphs, diagrams.

- **Practical (PRA).** These are based around a *performance* component involving responses including, but additional to, paper-and-pencil. An investigation may be undertaken, data may be analysed, a conclusion drawn, or a product may be completed. Examples include simple investigations or experiments in science or mathematics; classifying tangible materials in science; undertaking measurement tasks in mathematics; constructing shapes or figures in mathematics.

In English there are six response types. Of these, Selected response (SR), Short written response (SWR), Longer written response (LWR) resources have characteristics similar to those outlined above for mathematics and science. But there are three additional response types:

- **Oral response (OR).** The predominant response is oral, although a minor written component may be included.
- **Student rating or assessment (SRA).** The essence of these resources is that a rating or assessment is undertaken by the student. This category makes provision for student self-assessment or peer assessment by way of rating scale, observation scale, or checklist.
- **Teacher rating or observation (TRO).** Resources of this type are included to assist teachers' assessments of expressive skills, mostly in the written and oral strands. Multi-level marking guides come within this category, although some multi-level material is also included for LWR resources.

A search may be undertaken by any single classification field, or by combination(s) of the six fields. If a user does not specify particular fields, a broad general search will be carried out, but teachers are always advised to tailor their search to match their teaching, rather than engage in broad, undifferentiated searching. This approach uses the capacity of the search engine better, and increases the likelihood of locating valid resources.

### National Uses of the ARBs

From the outset, national and school-based uses of the ARBs have been dual elements of the project. In the initial stages, and certainly in the politicians' minds at the time, national uses predominated. Indeed, the aim of using the ARBs on a national basis at the "transition points" Year 6/7 and Year 8/9, provided the main impetus for initial funding. This view was not necessarily shared by some policy advisors, as the following quotation indicates. In discussing future national uses of the ARBs, Croft et al. had noted:

The possibility that the banks could provide information on science and mathematics achievement at Years 7 and 9, points which mark the transition primary/intermediate/secondary, and the completion of primary/intermediate schooling, appealed as contributing to the government's Achievement Initiative and its policy on accountability within the school system. Hence the ARBs were to become the major vehicle for implementing transition-point assessment. The general intention of using ARBs to provide some form of national information on Year 7 and Year 9 cohorts had considerable government support. It is of historical interest to note that minutes of the Ministry of Education's Assessment Secretariat Advisory Committee, of 21 July 1994 and 11 November 1994, record concerns about the national uses of ARB data, and the potential for these data to be regarded as "high stakes" information.

... Part of the problem has been that the purposes for collecting national information, apart from contributing to the broad notions of system accountability and targeting resources, have not been made clear at the policy level. (Croft et al., 1996, p. 81)

The requirement for ongoing policy analysis within government, and the need for clearer specification of educational outcomes from national testing had again been noted by Croft when stating:

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. (Croft, 1996, p. 138)

As the banks began to expand and offer more extensive coverage of curriculum statements, alternative possible national uses of the banks had been put forward. These were outlined in Croft (1999, pp. 57–58).

The 1998 Green Paper, *Assessment for Success in Primary Schools*, included a proposal for externally-referenced tests (also referred to as national tests):

Teachers need information to help them to identify whether their judgements about achievement are consistent with national standards.

Teachers use assessment information to help them to evaluate the effectiveness of their teaching and learning programmes. (p. 23)

In its *Response to the Green Paper*, NZCER (p. 6) had noted:

A consolidated statement of the purposes of the national tests is not evident (apart from the helpful diagram on p. 19), but various elements are presented throughout the text. For example:

- to provide primary schools with a source of comparative information (pp. 2, 19, 23, 26)<sup>1</sup>
- to provide information on educational outcomes to the Government (pp. 2, 6, 15, 26)
- to provide information to improve primary school programmes (pp. 2, 4, 13, 23)
- to provide information for parents (pp. 8, 19, 24)
- to follow international trends (pp. 9, 27)

Again, a lack of clearly communicated and specific goals had been noted for national testing in respect of valid educational outcomes.

NZCER had recommended against national testing in the report of the Stage Two implementation trial for the ARBs (Croft et al., 1996) and it did so again in its Green Paper Response, concluding (p. 12):

We are unaware of research findings or reputable literature which supports the validity of a single test for large-scale testing. We are aware of reputable literature which notes the invalidity of a single test for large-scale testing, e.g., Brown, McCallum, Taggart, & Gipps (1997); Jones (1997); Davey & Neill (1991).

and also (p. 16):

The case against the GP version of national testing is multi-faceted and educationally compelling. We conclude that national testing as outlined in the GP will not fulfil its purposes. Our view is that it should not proceed in the form it is outlined.

We recommend that our suggestion for Intact tests within the ARBs be developed to the point where they may be field tested.

The Intact tests referred to above were to be assembled from the existing resources within the ARBs. Intact tests were seen as subsidiary to the primary purposes of the ARBs, which were to make available an expanding collection of individual assessment resources that schools could match to their own teaching programmes.

The advantages of Intact tests created from the ARBs, compared with the national tests proposed in the Green Paper, were that the basic test development had already been completed and a range of tests would result, which assessed more diverse curriculum outcomes than would be possible in a single national test. Initiating large-scale paper and pencil testing could be avoided at a time when many countries were preparing to use computer technology for large-scale assessment. The flexibility of electronic delivery would make it more cost effective to produce new annual versions of Intact tests, if these were seen as necessary. The whole notion of compulsory “national testing” could be set aside and a more useful educational model adopted.

There was no deep conviction within NZCER that Intact tests would necessarily be of high educational value, or that they were consistent with the notion that the ARBs should be searched to find material that met a teacher’s objectives. But they were a pragmatic response to the continuing notion of national testing, which they regarded as somewhat flawed.

Subsequently, and with a change of government in 1999, the impetus to promote compulsory national testing, in the short term at least, was lost. NZCER did again suggest the Intact test approach to the present Labour administration, as this approach was consistent with its view of voluntary national testing. To the long term benefits of the ARBs, an alternative approach known as Assessment Tasks for Teaching and Learning (AsTTle), was embarked on, within the Government’s literacy and numeracy strategy. The AsTTle will provide literacy and numeracy tests for Years 5, 6 and 7 in English and Maori. In concept they are little different from the Intact tests proposed for the ARBs. AsTTle provides some choice, but at the whole test level only, as schools are unable to determine the selection of items that make up a test. The development of AsTTle was undertaken by the University of Auckland and the Education Testing Centre of the University of New South

Wales. However, the Centre subsequently discontinued its work, delaying the production of the AsTTle, and the first is yet to be published. However, it appears that, although their proposed structure will not allow national reporting of results, they may address the mandatory reporting requirements for schools contained in the Education Standards Act 2001.

### **School-based Uses of the ARBs**

Although the school-based uses of the ARBs were dominated by their potential national uses during the period 1996–1999, there was no doubt in the minds of the developers that the school-based aspects would pay better dividends in the long term. It had been noted in Croft et al.:

We have no doubt that the proposed school-based uses of ARBs represent an innovative approach to improving the quality of school assessment practices. When combined with electronic delivery, a broadening of the item-bank concept, and development of a classification system representing the New Zealand Curriculum Framework, school-based uses of ARBs constitute a unique development. We find no references in assessment literature to any other system which combines these features in a similar manner.

As indicated in chapter 7, there is strong support from teachers for the ARB concept and its implementation, but this support is not unanimous. Most reservations stem 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 continuation of the work, as will the range of assessment materials developed. However, the resources included in the ARBs will continue to be predominantly formal, and hence future supporting literature should emphasise that ARBs are designed to complement teachers’ own assessments, and should not be seen as the sole source of assessment material for classrooms. (1996, pp. 85–86)

With the consolidation of the ARBs as a school-based resource, and as a reflection of better and speedier internet access and a broader range of resources becoming accessible, the emphasis has moved to supporting teachers’ use of the ARBs for formative purposes. Steadily increasing numbers of resources in mathematics and science are reporting diagnostic information in the form of common errors made by

students, or common misconceptions that may be inferred from their responses (Neill, 2001). It has been noted by Mendelovits et al. (2000) that the former summative purposes of the ARBs have probably influenced the nature of some material published prior to 2000.

The stance adopted within the ARB project has always been that giving teachers the opportunity to select assessment material for their students would be a major contribution to enhancing validity, as an implicit aspect of current policy is that the individual teacher is best placed to exercise this choice. Although this places a major responsibility on all teachers using the ARBs, feedback in the early stages of the project (Reid et al., 1994) and later from Gilmore and Hattie (2000), indicates that most teachers appreciate the opportunity to exercise general control over the assessment material administered in their classrooms. However, it would be false to give the impression that selective searching is embraced totally by all users. Gilmore and Hattie (2000) noted that a number of users were printing screeds of material from the ARBs, and using them more in the manner of hard copy material.

Since the majority of primary school assessments are of an informal nature, i.e., short tests, conferencing, observation, running records (Croft, Strafford, & Mapa, 2001), data from national samples provide a complementary external measure to assess progress in defined tasks linked to curriculum objectives. However, the nature of the national data is such that it needs wise and careful evaluation as to its relevance to a particular school context. It is one thing to know that nationally, a certain ARB task is of moderate difficulty (answered correctly by between 40% – 59% of students nationally), but quite another thing to be able to judge how applicable these norms are to the standards or expectations of a particular school.

For secondary schools the value of the performance data may be restricted to helping choose tasks of appropriate difficulty, as much secondary assessment involving ARB resources seems still to be for summative purposes within schools. The chief value of the resource in these circumstances is probably its close articulation with the curriculum.

Earlier, the growth in the number of assessment resources that constitute the ARBs was described. There has been a corresponding growth in use of the ARBs, as this is indicated by the regular monitoring

of “hits” on the website. A constant two-week period at the start of the school year has been monitored from 1998 to 2002. This shows that during that time hits progressed from 250 to 6500 (Marston, personal communication, February 2002). There has been strong growth on a monthly basis too. Gilmore and Hattie (2000) reported a little under 9000 hits for August 1999. For August 2000, hits were recorded as 21,000 and for the same month in 2001 they were 25,000. It is not possible to interpret these rates precisely in terms of “user sessions” or “school activity”, as hits are counted on only four pages in a single session, i.e., main page for site and first search pages for English, mathematics and science. Nevertheless it indicates continuing strong use of the ARB website.

Gilmore and Hattie (2000) also provide the most recent data on how the ARB resources are used after they have been downloaded. About 50 percent of their respondents were using the ARBs as part of a pre-test or post-test, or for summative (end of unit) assessment, as distinct from mid- or end-of-year exams. About 30 percent were using them to compare students’ achievement with objectives, to seek diagnostic information on students’ achievement, or as a source of ideas on how to assess in curriculum areas, i.e., as exemplars of assessment. A quarter reported they used resources for formative assessment during a unit of work, or as a teaching resource.

A recent survey at NZCER (yet to be published), has found that the predominant uses of the ARBs, as recorded from a random sample of teachers in a representative sample of schools, are for teaching and learning purposes, monitoring students’ progress and providing information on achievement to parents and caregivers. In comparison with the findings of Gilmore and Hattie (2000), this information suggests a marked increase in formative uses of the ARBs.

### **Looking Ahead**

Croft (2001) has noted that the database which incorporates ARB trial data now has information on the performance of more than 60,000 students, with analyses possible on many combinations of curriculum outcomes, school type and individual student variables.

When data are required for curriculum review purposes, information from the performance of diverse national samples as embodied in this database is potentially more useful than data from a

single form of national testing. This is because data combined from numerous ARB resources enable a broader range of curriculum outcomes to be assessed than is likely under the restrictions of a single national test.

Marston and Croft (1999) have noted that:

ARBs are now established as the major collection of nationally developed material in science, mathematics, and English, published for classroom assessment purposes. As the ARBs accumulate data on national achievement in these three learning areas, they will become valuable sources of information on national achievement trends. They will provide an objective basis for reviewing curriculum statements and focusing teachers' professional development.

Although broader reporting of curriculum outcomes is not a major purpose of the ARBs at present, its potential contribution in this area is worth noting and planning for now, as a possible future development which could be very useful.

#### Notes

1. All specific quotes as noted are contained in an Appendix.

#### References

- Brown, G., & Strafford, E. (1997, November). *The feasibility of expanding the assessment resource banks to incorporate English*. Unpublished project report. Wellington: New Zealand Council for Educational Research.
- Croft, C. (1997). Assessment resource banks for school-based and national assessment. *New Zealand Annual Review of Education*, 6, 127–139.
- Croft, C. (1999). *School wide assessment: Using the assessment resource banks*. Wellington: New Zealand Council for Educational Research.
- Croft, C. (2001). *Assessment resource banks in English, mathematics and science: Criteria for determining the optimum size and nature of the ARBs*. Report to the Ministry of Education. 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., Reid, N., & Livingstone, I. (1995). Assessment resource banks in mathematics and science. *set: Research Information for Teachers, Best of set Assessment*, 2, Item 14.
- Croft, C., Strafford, E., & Mapa, L. (2001). *Stocktake/evaluation of existing diagnostic tools in literacy and numeracy, in English*. Final report to the Ministry of Education. Wellington: New Zealand Council for Educational Research.
- Gilmore, A., & Hattie J. (2000). *Evaluation of the assessment resource banks in schools: Final report*. Christchurch: University of Canterbury, Department of Education, Unit for Studies in Educational Evaluation, and Auckland: University of Auckland, School of Education.
- Marston, C., & Croft, C. (1999). What do students know in science? Analysis of data from the assessment resource banks. *set: Research Information for Teachers*, 2, Item 12.
- Mendelovits, J., Farkota, R., & Lindsey, J. (2000). *Evaluation of the New Zealand assessment resource banks' methodology for Ministry of Education, New Zealand*. Camberwell, Victoria: Australian Council for Educational Research.
- Ministry of Education. (1998). *Assessment for success in primary schools* (Green paper). Wellington: Ministry of Education, Learning and Evaluation Policy.
- Neill, A. (2001). An introduction to the assessment resource banks (ARBs) and their diagnostic potential. *New Zealand Mathematics Magazine*, 38(1), 29–38.
- New Zealand Council for Educational Research. (1998). *Assessment for success in primary schools: Green paper. A response from the New Zealand Council for Educational Research*. Wellington: NZCER.
- 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.

**Author**

Cedric Croft is Chief Researcher and Research Director, Learning, Curriculum and Assessment at the New Zealand Council for Educational Research. He has written widely on the Assessment Resource Banks and school assessment more generally. Recently he was a member of the Literacy Experts Group. He has studied large-scale testing programmes in the United States and United Kingdom as a Visiting Researcher at the Educational Testing Service, National Centre for Research on Evaluation, Standards, and Student Testing, and at the Institute of Education, University of London.