

First national literacy & numeracy tests introduced



More than one million students in years 3, 5, 7 and 9 participated in the National Assessment Program – Literacy and Numeracy (NAPLAN) in May 2008. **Chris Freeman** describes ACER's contribution to NAPLAN through test development, data analysis and marking.



Chris Freeman is Research Director of ACER's Systemwide Testing program

NAPLAN 2008 measured students' performances in literacy, including reading, writing, spelling, grammar and punctuation, and numeracy, including mathematical concepts such as algebra, functions, patterns, space, measurement, chance and data.

This inaugural sitting of NAPLAN, conducted by the Ministerial Council for Education, Employment, Training and Youth Affairs (MCEETYA), was the first time all students in Australia in Years 3, 5, 7 and 9 have been assessed in literacy and numeracy using the same year level tests.

The national tests, which replaced a raft of tests administered by Australian states and territories, have reduced the level of duplication and improved the comparability of students' results across states and territories. In the future, NAPLAN results may be useful in evaluating the effectiveness of educational policies and programs; in identifying the need for targeted interventions for individual students and groups of students; and in developing improved measures of school performance and greater transparency in national reporting.

Benefits for schools and students from NAPLAN depend on literacy and

numeracy tests that are thorough, accurate and objective. In 2008 and 2009, NAPLAN test development, administration and marking are being conducted for MCEETYA by a range of educational organisations coordinated by Curriculum Corporation.

ACER is contributing to test development, data analysis and the marking of NAPLAN tests in some jurisdictions.

Test development

NAPLAN test development is a collaborative process involving literacy, numeracy and assessment experts from state and territory education departments and the Catholic and independent school sectors. It relies on the goodwill and support of schools and teachers around the country to assist in the test development process.

The development of NAPLAN tests involves a specific set of steps designed to ensure that they are of the highest standard. The tests are constructed to assess knowledge, skills and understandings appropriate to each year level; to be interesting and engaging to students throughout Australia; and to challenge students at all levels of ability.



The questions for the tests – the test items – are created through a process that draws on the creativity, experience and expertise of teachers. Teams of literacy and numeracy test developers – themselves teachers – conduct this specialised work which is underpinned by a NAPLAN assessment framework and by the national Statements of Learning which describe the key literacy and numeracy skills that all students in Australia should have.

Test items are reviewed at several stages in their development, first by other test developers and then by curriculum and measurement specialists from the Commonwealth, state and territory departments and Catholic and independent school sectors. Reviewers ask questions of the following kind:

- Do the items adequately address the literacy or numeracy domain that is the focus of the test?
- Is the item format – multiple choice, short answer, or extended response (in the case of writing) – the best format for assessing the particular knowledge, skills or understandings being tested?

- Is the item at the right level, in terms of the expected ability level and year level?
- Is the test material interesting?
- When a unit includes more than one item, are there dependencies between the items? Does one item give a clue to the next one? Would a different order of items within a unit make a difference? If a response to one item is incorrect, does this affect possible responses to other items in the unit?
- Is the item likely to be biased; is it likely to be easier or harder for certain subgroups in the test population for reasons other than differences in the ability being measured?
- Are there any ‘tricks’ in the item that should be removed?
- How will the students perceive the test material? To answer this, the panel members must imagine the cognitive, cultural and response format demands of the items from the point of view of test-takers.
- If the item is multiple-choice, are the incorrect choices – the distractors – indisputably incorrect, while being reasonable and plausible? Is the correct answer – the key

– indisputably correct, and not simply the best of the alternatives presented? Are the alternatives of similar length and written in a similar style to the key?

- Do test materials reflect cultural sensitivity and avoid potential biases?
- Do items relate to the national assessment framework?
- Are items appropriate from a local perspective; do they assess knowledge, skills and understandings that local curricula would expect students to have developed by the May test date?

Following each review, revisions are made and national stakeholders recheck the revised items. The test developers may then ‘pilot’ specific items with groups of students to explore how students interact with them.

Item trialling

ACER had the contract to trial test items for NAPLAN 2008 and has recently conducted a trial of items for the 2009 tests in 360 schools across Australia.

In this process, at least twice the number of items required for the final tests are trialled in schools. Multiple-choice items are then marked by computer. Other items which require single word answers also are computer marked using optical character recognition, and any response that the computer interprets as incorrect is checked by a human marker. The writing tasks are marked by teachers experienced in the marking of student writing.

A team of ACER psychometricians analyses the data from the trial, establishes the difficulties of the trialled items, compares these difficulties with the abilities of the students in the trial and produces statistical reports on the items. Items are eliminated from the item pool if they have unacceptable psychometric properties.

ACER test developers then identify the best items and assemble a set of proposed test papers which are submitted to the national stakeholders for approval and subsequent sign off. For NAPLAN 2009, ACER is coordinating the national review of all test items.

Data analysis and reporting

Following the administration of NAPLAN across Australia in May 2008, ACER was contracted to conduct the central analysis of students' test performances. ACER psychometricians analysed the test results, constructed the NAPLAN scales on which student results were to be reported, and supplied the states and territories with the information they required to prepare reports for individual students and schools. ACER also is preparing results in a form appropriate for public reporting and inclusion in a national report on students' literacy and numeracy levels.

The NAPLAN tests are the result of an iterative process of development by teachers who are expert in test construction; scrutiny by curriculum and measurement experts; piloting and trialling with students; expert marking; and psychometric data analysis and reporting. ACER's involvement has contributed to a high-quality, nationally-comparable snapshot of how Australian students are performing in these key areas of the school curriculum.

More information can be found at www.naplan.edu.au ■

