

Inappropriate uses of NAPLAN results

Margaret Wu and David Hornsby

NAPLAN tests are done once a year, using only around 40 questions per test. As a consequence, NAPLAN tests are limited in their coverage of the wide range of skills in literacy and numeracy, and in their capability to measure the achievement levels of individual students. Yet this very limited assessment system is being used as if it is capable of much more. The following summarises specific issues regarding the NAPLAN tests.

1. Content coverage of NAPLAN tests

With around 40 test questions per test, NAPLAN only measures fragments of student achievement. Testing a small bit of a curriculum does not indicate a student's learning in the whole curriculum area. Students' results on NAPLAN tests show the percentages of questions they can answer on those tests, but the results do not necessarily reflect students' achievement in the whole numeracy domain and literacy domain.

Bureaucrats may refer to an achievement gap between students (and between schools) but what they mean is a test score gap. Since the test assesses very limited aspects of learning, the results can not be used to make claims about overall achievement.

Further, student achievement should not be narrowly confined to achievement in numeracy and literacy only. Achievement should include creativity, critical thinking, ability to follow an inquiry, compassion, motivation and resilience – important skills, strategies and behaviours that are not assessed with NAPLAN pencil and paper tests. In contrast, teachers do know about students' wider abilities beyond numeracy and literacy.

2. Accuracy in identifying students' overall levels in numeracy and literacy

A test instrument with only 40 questions cannot accurately separate students into finely graded levels of achievement. This is because a student's results on short tests can vary quite widely. If we know nothing about a student, a NAPLAN test can give us a rough idea of whether a student is struggling, on target, or achieving well above the expected level, but no finer grading than that. However, teachers do know

their students well, so NAPLAN rarely provides information that a teacher does not already know. In order to locate student levels more accurately through testing, we would need many tests and longer tests, which would not be in the best interests of students – or the taxpayers!

3. Matching assessment with curriculum

For assessments to be relevant to teaching and learning, what is being assessed should match what is being taught. What curriculum is NAPLAN testing? ACARA claims the test items are 'informed by' the National Standards of Learning for English, but that document is unknown in most schools. It is inappropriate to base NAPLAN on the new Australian Curriculum, as some States have not yet adopted it, and even after adoption it will take years for the new curriculum to be fully implemented in schools. Since student learning is cumulative, it will take a long time before students' learning completely reflects the new curriculum. It will be a long time before NAPLAN truly matches what is taught.

4. Providing diagnostic information for teaching and learning

The NAPLAN tests are not diagnostic tests. They are standardised tests which are designed to assess and compare the overall achievement of very large groups, not individual students or schools. Because there are very few questions testing very few areas of literacy or numeracy, NAPLAN tests do not provide sufficient diagnostic information to identify areas of weakness or strength to support classroom learning. Despite this, schools are being required to use the results as if they are diagnostic and to identify 'weaknesses' to be 'fixed'. (Even if the NAPLAN tests were diagnostic, the five-month delay in providing the results would make them useless for informing teaching.)

Good diagnostic tests are generally constructed for focused curriculum units or particular areas of learning with questions specifically written to identify common misconceptions and errors. NAPLAN questions are not designed to uncover particular learning problems.

Further, evaluation is a continuous process, not an event. A single event, once a year, will miss many of the strengths or weaknesses individual students may have.

5. Making inferences about school effectiveness

NAPLAN tests are tests for students. Using students' test results to judge teachers and schools requires making an inference. Such an inference is not valid. Test scores can not tell us whether a teacher or a school is good or bad because many other factors influence test scores (such as poverty, parental support, personality, interests, aspiration, motivation and peer pressure). Attributing low school performance to poor teaching is not only invalid but insulting since this implies that teachers are not doing their job. Given the nature of the comparisons in NAPLAN where schools are compared with each other, half of the schools will always be described as below average or under-achieving. There is an assumption that the staff in below-average schools are not doing the best they can. We know this

assumption can not be made, but the government applies pressure to make these schools meet certain targets of improvement. Such target setting is often unrealistic, to the point of being ludicrous.

6. Monitoring trends

The current NAPLAN test format has severe limitations for monitoring trends. This is because each NAPLAN test is short and there is an insufficient number of test items to provide robust links between tests from one year to another.

At the system level, NAPLAN data can be used to provide useful information for comparing large groups. For example, the results can help us make generalisations about the performance of girls and boys, about rural and urban students, and about students from non-English-speaking backgrounds. Even so, it should be noted that tracking large groups of students over time can be done just as effectively by testing a sample of students every three or so years. It is unnecessary, and a waste of public money, to test every student every year for the purpose of monitoring trends.



Margaret Wu has a background in statistics and mathematics. Her main research interests are in educational measurement which deals with measuring students' performance levels in various subject domains. Margaret has worked as a psychometrician for many years, and has been closely involved with large-scale international assessments including PISA and TIMSS.



David Hornsby taught at every year level and was then an education department consultant working in both primary and secondary schools for 4 years. After further years of teaching, he became a primary school principal. He then lectured in teacher education and is now working as a consultant in schools. He has completed many lecture tours of the US and the UK, and worked with teachers in Costa Rica, Hong Kong, Indonesia and Singapore. This year, he will be working with teachers in Shanghai. He has written or co-authored many popular teacher reference books.