Introduction

Measures of student achievement are regularly used in the development and evaluation of education policy. This can include the relative achievement of sub-groups in the community (such as the Closing the Gap targets for Indigenous students), an assessment of demand (such as university entrance scores) and for international comparisons. This guide provides a brief overview of the major national and international measures of school student achievement that are conducted and used in Australia.

National assessments

- Australian Early Development Index (AEDI)
- National Assessment Program—Literacy and Numeracy (NAPLAN)
- Australian Tertiary Admissions Rank (ATAR)

International assessments

- Programme for International Student Assessment (PISA)
- Trends in International Mathematics and Science Study (TIMSS) and
- Progress in International Reading and Literacy Study (PIRLS).

Australian Early Development Index (AEDI)

What is assessed

The AEDI is designed to measure how young children are developing in different communities. Teachers complete the AEDI checklist of approximately 100 questions for each child in their class. The checklist measures five key areas of development (domains):

- physical health and wellbeing
- social competence
- emotional maturity
- language and cognitive skills (school-based) and
- communication skills and general knowledge.

The responses from the questions in the checklist are compiled to determine up to five AEDI domain scores for each individual child.

To determine whether an individual domain score is ‘on track’, ‘at risk’ or ‘vulnerable’, national AEDI ‘cut-offs’ were established during the first national AEDI data collection in 2009. All the children’s AEDI scores in each
domain were ranked from the lowest to the highest. Those with scores in the lowest 10 per cent were classified as developmentally vulnerable in that domain, while those ranked between 10 per cent and 25 per cent were classified as developmentally at risk. These cut-off scores were used for the 2012 collection and will be used for future collections to enable comparisons across time.

**Who is assessed**

The intention is to complete the checklist for all children in their first year of formal full-time school. Information on children with special needs is not included in the domain results tables, but background information is included to provide communities with information on all children.

In 2009, data was collected on 261,147 children representing 97.5 per cent of children in the relevant cohort, while in 2012 data was collected on 289,973 children (96.5 per cent).

**Frequency**

The initial AEDI collection was undertaken in 2009 and the second collection was undertaken in 2012. The Australian Government has committed to conducting the collection every three years.

**Who runs it**

The AEDI is based on the Canadian Early Development Instrument and is delivered by the Australian Government and state and territory governments working in partnership with the Royal Children’s Hospital Centre for Community Child Health, the Murdoch Children’s Research Institute and the Telethon Institute for Child Health Research, Perth. The data is managed by the Social Research Centre, Melbourne.

**Reporting**

A national report has been published for each of the 2009 and 2012 collections. The reports provide overall results and some information by state and territory. Individual community results are available via the website through a mapping facility.

**Strengths**

- The near universal coverage of the AEDI data collection provides a comprehensive picture of where there are concentrations of children who are developmentally behind their peers. Results are robust both at small geographic levels and for sub-groups such as Indigenous children and those from a non-English speaking background. This information can be used by governments and local communities to determine where to direct efforts to ensure children do not fall behind in the early years of school.

- By using the 2009 results as the benchmark, future collections will also provide a good basis for assessing the effectiveness of early childhood programs implemented since that time on preparing children more effectively for school entry.

- **AEDI research** suggests that the AEDI is a valid predictor of student’s performance later in school; for example, in NAPLAN testing up to Year 7.

**Weaknesses**

- For the 2009 collection, the categorisation of students into ‘developmentally vulnerable’ and ‘at risk’ was based on their relative ranking. This does not provide an absolute measure of students’ development, either in comparison to an agreed benchmark of expected development or to provide a guide of how far behind their peers these children are.

- The AEDI assessment is undertaken by the student’s teacher, who in general will have taught the students for a period of less than six months (the collection is conducted from May to August). There is no validation by another reviewer.

- Although information for the AEDI is collected by teachers, results are reported for the community where children live, not where they go to school. While the majority of children at this age will attend a local school, this means the results are of limited value for schools that draw their students from a range of locations, or where students in a particular location attend a range of different schools.

- Data for small communities must be treated with caution as one or two children may make a significant difference to the proportion reported in each category.
National Assessment Program—Literacy and Numeracy (NAPLAN)

What is assessed
NAPLAN is designed to assess students’ literacy and numeracy skills and determine whether they have the critical skills required for ongoing learning and to contribute effectively in society. Students are assessed in reading, writing, language conventions (spelling, grammar and punctuation) and numeracy. National tests are used which broadly reflect aspects common to the curriculum of all jurisdictions.

Results are benchmarked against standards which are based on the Statements of Learning for English and Mathematics, which have been developed collaboratively by state and territory education authorities and other representatives. For each of the assessments, results are reported on an achievement scale from Band 1 to Band 10, representing increasing levels of skills and understanding, with a national minimum standard defined at each year level.

Who is assessed
Nearly all students in years 3, 5, 7 and 9 are assessed. The exceptions are ‘exempt’ students: those with a significant disability and those from a non-English speaking background who have been in the school system for less than a year; those whose parents withdraw them from testing; and those who are absent on the day of testing or unable to participate due to a mishap. In 2012, about 95 per cent of students in each of years 3, 5 and 7 either sat the tests or were exempt (exempt students are included in the results and deemed not to have met the national minimum standard). The participation rate for Year 9 students was 91.5 per cent.

Frequency
NAPLAN was first conducted in 2008 and is held every year in May.

Who runs it
The Australian Curriculum Assessment and Reporting Authority (ACARA) runs NAPLAN under the direction of the Council of Australian Governments’ Standing Council on School Education and Early Childhood.

Reporting
All students receive an individual report in mid-September showing how they rated against the achievement scale, how the majority of students in their year performed, and the national average for their year for each test. The school average is also included in some states and territories.

School results are published on the My School website, and comparisons are available with ‘similar schools’ (that is, those with a similar socio-economic profile) and all schools.

A national summary report is released each year in September, with a more detailed report released in December that provides results broken down by a wide range of demographic characteristics for each year level and each area of testing.

NAPLAN has also been the subject of two inquiries by the Senate Standing Committee on Education, Employment and Workplace Relations: The Effectiveness of the National Assessment Program - Literacy and Numeracy in 2013 and the Administration and Reporting of NAPLAN Testing in 2010.

Strengths
- NAPLAN’s near universal coverage provides good data on the skill levels of Australian students across a significant proportion of their school education. The large numbers enable analysis of the performance of sub-groups with some confidence.
- It is possible to track the performance of cohorts through the school system (for example, students in Year 3 in 2008 were also tested in Year 5 in 2010 and Year 7 in 2012). This provides information not just on achievement levels for each year tested, but also the different rates of progress for particular groups.
- While small variations in each year should be treated with caution, annual testing provides regular data which enables early identification of trends.

Weaknesses
- The participation rate in the testing has declined consistently over the period from 2008 to 2012 (albeit only one to two per cent in total for each year level). There is some concern that this reflects a tendency for poorly performing students to not participate in the assessments. This is a particular concern at Year 9,
where participation is now down to 91.5 per cent, meaning a significant number of students are not included in the results.

- NAPLAN’s effectiveness as a diagnostic tool is limited by the four-month delay in the release of the results, meaning that teachers and schools cannot take action on any identified weaknesses until late in the academic year. While NAPLAN should only be a supplement to ongoing monitoring of student progress, nonetheless this does limit NAPLAN’s value at the individual level.

- While ACARA states that NAPLAN is just one part of school assessment and specific coaching is not recommended, there are commercial websites and publications offering practice tests and it is likely that some students receive extra coaching on how to perform well in these tests. Use of the results for purposes such as determining class dux, allocation of teacher performance pay and ranking of schools is likely to increase the use of additional coaching. There are also concerns that teachers are distorting their teaching to focus on the NAPLAN tests to the detriment of other aspects of the curriculum. It is not clear the extent to which coaching specifically for NAPLAN testing actually increases a student’s skill level and understanding, and hence reflects improved educational outcomes.

**Australian Tertiary Admissions Rank (ATAR)**

*What is assessed*

The primary purpose of the ATAR is to provide a consistent score for consideration in university admissions processes, and is calculated on the basis of Year 12 certificate results. It is used in all states except Queensland (which has a similar ranking called Overall Position [OP]). Students are ranked from a high of 99.95, decreasing in 0.05 increments, theoretically reflecting their Year 12 achievement relative to their Year 7 age cohort. That is, an ATAR of 80 means a student is in the top 20 per cent of their age cohort. Complex scaling processes exist to attempt to ensure consistent treatment across subjects and schools.

The precise nature of the calculation varies from state to state, in part because of the different nature of each jurisdiction’s Year 12 accreditation. For example, results in English are only included in the ATAR in South Australia (SA), Western Australia (WA) and the Australian Capital Territory (ACT) where it is one of the student’s four highest scoring subjects (or best five in Tasmania). However, an English score is compulsorily included in the ATAR for both New South Wales (NSW) and Victoria. In some states results from Year 11 can form part of the score, while others are based purely on Year 12 performance, and different rules apply to students who repeat Year 12 or do an additional year of study. Vocational courses are typically not included in the scoring formula, or only form a small proportion. Students undertaking other programs, such as the International Baccalaureate, receive a notional ATAR based on a national conversion table.

*Who is assessed*

All students who complete a ‘tertiary’ package of study in Year 12 will receive an ATAR.

*Frequency*

The ATAR was introduced in 2010. Prior to that time different (but similar) measures were used in NSW and the ACT (University Admissions Index (UAI)); Victoria (Equivalent National Tertiary Entrance Rank (ENTER)); and in WA, SA, Tasmania, and the Northern Territory (Tertiary Entrance Rank (TER)). It is issued every year.

*Strengths*

- The ATAR provides a consistent basis on which to compare university admission cut-off scores across most Australian universities. It provides an indicator of relative demand for courses both within a year and across years.

*Weaknesses*

- ATAR scores form only part of the entry requirements for university courses. Non-school leavers have other entry requirements and additional criteria may also be considered for school leavers (for example, principals’ recommendations). Some courses also use aptitude tests or interviews.

- Different universities allocate bonus points to students for a wide range of characteristics (for example, attending a regional high school, Indigenous status, sporting excellence) which are added to the student’s ATAR for the purposes of determining cut-off admission scores. Hence, the actual minimum achieved ATAR for a course may be somewhat less than the published cut-off score.
• The different basis used in each state and territory to calculate ATARs means that they are not strictly comparable between states. Each year is scaled relative to the current cohort of students so scores are also not directly comparable across years.

• ATAR is only a weak predictor of academic performance at tertiary level. While students with an ATAR of above 80 typically perform well, so do many with middle and lower level scores.

Programme for International Student Assessment (PISA)

What is assessed
PISA attempts to assess the extent to which students have acquired the skills for lifelong learning and to participate fully in society. As such, it does not assess how well students have learned a specific curriculum, but rather how they apply their learning to everyday problems and situations.

The three domains of reading, mathematics and science are included in each cycle, with one being chosen as the major domain each time, providing greater depth on that subject on a rotating basis. Problem solving was also included as a separate domain in 2003.

Who is assessed
Each cycle of PISA assesses a random sample of 15-year-old students from a representative sample of schools in each participating country. In 2009, 65 countries participated (compared to 32 countries in 2000), with almost 500,000 students participating—353 schools and 14,251 students participated in Australia.

Frequency
PISA has been conducted every three years since 2000. Australia has participated in all five cycles to date.

Who runs it
PISA is a program of the Organisation for Economic Development and Co-operation (OECD), although non-OECD countries and economies (such as Hong Kong and Shanghai) also participate. The Australian PISA National Centre is located at the Australian Council for Educational Research (ACER).

Reporting
ACER has published several reports and a number of statistical tables on the Australian results from each cycle of PISA. These provide international comparisons as well as state and system level data, and for different demographic groups.

The OECD produces a wide range of reports and analysis at the country level as well as detailed statistical tables.

Strengths
• PISA is the largest of the international student assessments, with both the greatest representation of countries (providing a greater range of external comparison) and the largest Australian representation (providing increased scope for internal analysis).

• The program has now been conducted five times (although results have not yet been released for 2012). This provides for an assessment of trends as well as providing greater confidence in the results for each cycle.

• In addition to the subject matter testing and standard demographic characteristics, PISA collects a wide range of data from the student (for example, educational resources in the home, attitudes to study, relationships with teachers, occupational aspirations) and principals (school context and resources). This allows for a much wider range of factors affecting student achievement to be analysed.

• Australian students participating in PISA are invited to participate in the Longitudinal Surveys of Australian Youth (LSAY), which then survey the students for each of the following 10 years as they move to further study, employment and other activities. This means that student outcomes can be linked to a measure of academic ability, as well as a range of other factors.

Weaknesses
• As PISA is a sample survey, results have to be treated with some caution. In particular small differences in scores between countries or across different survey cycles may well be the result of statistical variability rather than real differences.
• More countries are participating in PISA in each cycle and hence ‘rankings’ from one survey to the next are not comparable.

• While every effort is made to ensure the assessment tools are robust across culture and language, school systems and students are not directly comparable internationally. For example, non-school based tuition (such as private classes after school and personal coaching) is a dominant feature of some education systems, particularly in East Asia. In these cases, therefore, the results do not just reflect the performance of the relevant school system.

**Trends in International Mathematics and Science Study (TIMSS) and Program in International Reading Literacy Study (PIRLS)**

**What is assessed**
TIMSS assesses student achievement in mathematics and science in order to improve the teaching of these subjects. It is curriculum based in that it attempts to consider the intended curriculum, the implemented curriculum and attained curriculum. PIRLS assesses reading achievement on a similar basis. In addition to assessing student achievement, the studies also collect a range of data on the education system, school and classroom environment. This includes asking parents to complete a survey on the home environment.

**Who is assessed**
TIMSS assess students in years 4 and 8. In most countries the study is conducted as a sample survey. In Australia, a stratified sample of schools is selected and then a random mathematics classroom is chosen from each school, with all students in the class participating in the assessment. In the 2011 cycle, 52 countries and economic regions participated in the Year 4 TIMSS assessment and 45 countries in the Year 8 assessment, representing more than 600,000 students in total. In Australia, 280 primary schools and 290 secondary schools participated, with around 6,100 Year 4 students and 7,600 students in Year 8.

PIRLS assesses students in Year 4. In the 2011 cycle, 48 countries and regions participated. However, some countries tested older or younger students, so only 45 countries or regions have results comparable with Australia. In total, approximately 325,000 students participated in PIRLS 2011, including countries assessing students at more than one grade, benchmarking participants, and pre-PIRLS participants (a simpler version of the PIRLS test designed to prepare countries to move to full PIRLS participation).

In Australia, PIRLS was conducted jointly with TIMSS in 2011, so the same Year 4 schools and students participated in both studies.

**Frequency**
TIMSS is conducted every four years. The first survey was in 1995 and Australia has participated in all five cycles to date.

PIRLS is conducted every five years, with the first study in 2001. Australia participated in PIRLS for the first time in 2011.

**Who runs it**
Both TIMSS and PIRLS are projects of the International Association for the Evaluation of Educational Achievement (IEA) and are directed by the TIMSS International Study Center at Boston College, United States of America. ACER is responsible for undertaking the data collection in Australian schools for both studies.

**Reporting**
ACER has published a range of reports and statistical data for each cycle of TIMSS and PIRLS. More information on international performance is available through the International Study Center website.

**Strengths**
• TIMSS is the oldest of the major international comparative studies and hence provides longer term trends than PISA.

• As TIMSS is conducted every four years in both Year 4 and Year 8, cohort effects can be monitored (for example, the Year 4 results in 1995 is the same cohort as the Year 8 results in 1999).

• Both TIMSS and PIRLS assess expected understanding of core curricula at Year 4, thus providing relatively early feedback on the effectiveness of the school system. Benchmarks of ‘expected standards’ are reported...
against, from ‘low’ (which is similar to the NAPLAN national minimum standard measure) through to ‘advanced’. This provides greater information about the range of achievements in each school system than score averages.

**Weaknesses**

- As with all sample surveys, results are subject to sampling variability and hence small differences should be treated with caution.
- More countries are participating in these studies in each cycle, and hence ‘rankings’ from one cycle to the next are not comparable.
- TIMSS and PIRLS assess knowledge rather than problem-solving skills (whereas PISA assesses the application of knowledge to solve problems). This may be of less value in assessing the ability of the school system to prepare students for life. Comparison of the results from TIMSS and PISA suggest there is limited correlation in the relative performance of countries across the two studies.
- While both TIMSS and PIRLS attempt to assess students against measures common across all countries’ curricula, international comparisons must be treated with caution due to different environments and cultural conditions.