International testing: Mathematics and science

Purpose

The media is placing increasing attention on the achievement of Australia and Queensland within international testing programs, such as the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS). In 2013, the Australian Government set a target for Australia to be in the top five countries in these international measures by 2025.

International testing programs exist within a broader social landscape and results play a role in influencing educational and policy debates. Hence, an informed and considered approach is required to interpret data from these programs through lenses that understand and appreciate local contexts.

This paper outlines the current positions around international testing and considers the implications for school leaders and teachers in Queensland. The paper is structured to:

- present the main arguments and interpretations of international testing data (specifically PISA and TIMSS) within the Australian context
- identify considerations that could inform teaching and learning in Queensland schools.

Both PISA and TIMSS consider student performance in mathematics and science; however, the ways this is constructed within the assessments are different. PISA is intended to measure human capital as determined by the Organisation for Economic Cooperation and Development (OECD), whereas TIMSS is intended to measure curriculum content as determined by the International Association for the Evaluation of Educational Achievement (IEA). Therefore, jurisdictions may see improvements in one measure but not the other. Although PISA also tests reading performance, this paper focuses on mathematics and science only as reading is not tested within TIMSS.

It is the position of this paper that Australian student performance in mathematics and science, while higher than the OECD average, presents opportunities for targeted focus on educational actions that improve student outcomes.

Analysing the data

International testing, such as PISA and TIMSS, operates within an emerging global governance of education (Meyer & Benavot 2013) that establishes standardised global benchmarks against which the performance of educational systems can be compared. This presents four broad areas of limitation:

1. *The assessments are point-in-time measures*; while some attempt is made to look for trends over time, each measure only captures a small and finite sample of the entire schooling year and experience.
2. *Data only represents a sample of students*; while attempts are made to make the sample statistically representative, the assessments are administered through selective samples, not across an entire cohort of students. The samples are considered representative of each jurisdiction’s population of 15-year-olds in school, which means that they can be generalised to the population in each jurisdiction with a good degree of accuracy and compared to each other. However, problems arise with how the population is constructed; for example, a city like Shanghai is not the same population as a country like Australia, which has issues of

remoteness. Also, countries that have many 15-year-olds not in school (e.g. Vietnam) are different populations than Australia.

3. *Different forms of sampling are used for different assessments and for different purposes within assessments*; both PISA and TIMSS conduct multiple-matrix sampling (see appendix), which means that an individual student does not complete the entire assessment, but only part of it. Further, the assessments sample for different skills, concepts and abilities.
4. *The assessments assess particular skills*; TIMSS is a curriculum-based assessment that attempts to take into account the broad content and cognitive areas typically represented in the school curriculum. PISA is not curriculum-based but requires students to apply the knowledge and skills they have learnt at school to unfamiliar contexts. The assessments do not reflect the totality of learning or the experiences of students in schools and, while the data does provide some guidance around student learning, it is not intended to be used as an assessment of the full curriculum.

Therefore, although international testing programs provide some insights and interesting comparisons, they have a limited scope and applicability to understanding the effect of practices in the classroom. Sellar, Thompson and Rutkowski (2017) strongly suggested that, given the statistical sampling used in these assessments, it is not possible, and nor should it be intended, to provide analysis at a more fine-grained level than broad jurisdictions. Instead, international testing programs should be viewed as just one source of data that can inform broad educational policy.

Analysis of international testing data, particularly within the media, usually occurs in one or both of two ways, each of which present as problematic.

1. *A simple comparison of country ranks and how they have changed over time*. This presentation is based on two assumptions: (1) that each country has a comparable demographic and sampling process, and therefore a national average is a fair representation of the achievement level of all sub-jurisdictions within the nation, despite potentially significant variations between states and territories, and (2) that the total number of possible ranks have remained stable. Both of these assumptions have been challenged (Gorur & Wu 2015). Additionally, differences in country rankings are presented as being ‘real’, when in fact there are clusters of countries with no statistically significant differences between their average performances.
2. *Comparing average national scores across time*. This presentation assumes that each iteration of the assessment is equally comparable and testing the same constructs and ideas. This can be particularly problematic when considering PISA, which varies the major and minor domains tested in each cycle. This means that comparisons of PISA results should only occur across years where the major domain is the same. For example, in mathematics, only data from 2003 and 2012, when mathematics was the major domain, would be comparable (see the appendix for more detailed explanation).

A preferred option is to consider within-country variations (e.g. comparing Queensland with Western Australia) as there are greater similarities within countries and greater comparability of contexts (Gorur & Wu 2015). Within-country comparisons still need to consider the broader set of factors including gender, geographical location, socio-economic status (SES) and language background. By changing the focus of analysis, more meaningful uses of the data can be applied.

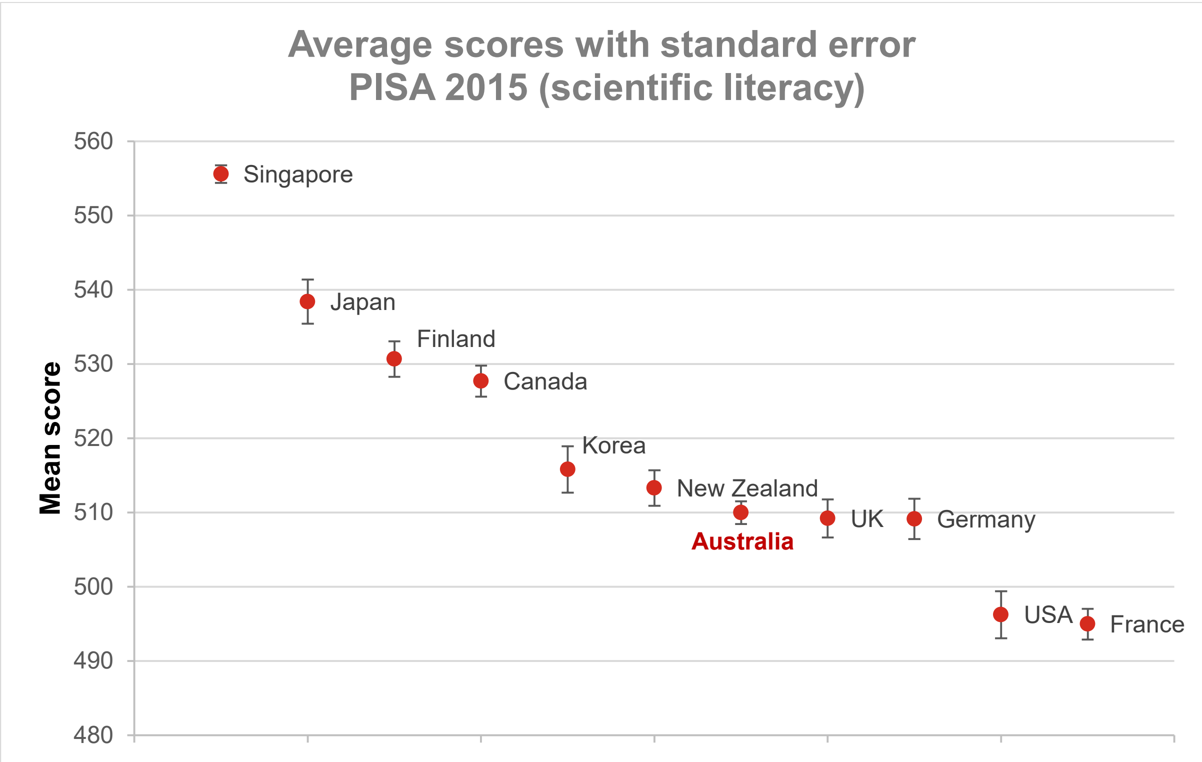
Sources of data

Articles of interest were drawn from published research (i.e. peer-reviewed journals), government reports and analysis, popular media, and independently published reports. Greater value has been placed on published research and government reports. However, inclusion of the popular media and independent reports ensures a contemporary review that highlights some of the political landscape that can influence the translation of the data into public policy. Sources were identified through an internet search using key search terms: ‘PISA’, ‘TIMSS’, ‘Australia results’, ‘international testing’, ‘student outcomes’, and variations of these to ensure a comprehensive review. Initial searches were undertaken through Google News and the Factiva database with regards popular media, and through a university library database for published research. This review is not exhaustive, but attempt has been made to be comprehensive.

Error in results

In determining a result within both PISA or TIMSS statistical uncertainty exists as a consequence of the sampling and processes used to generate the mean scores. This error is presented and discussed in two main ways: mean scores are presented with standard errors (SE) and considered with a confidence interval. Both approaches propose a range within which the true score lies, though they are determined in slightly different ways. Consideration of statistical confidence is important in fully understanding any variation in the mean scores either across time or jurisdictions. For example, a sample of 10 countries from 2015 PISA can be ranked on mean scores. Media reporting on these results claim that Australia ($\bar{x}$ = 510) has fallen behind New Zealand ($\bar{x}$ = 513).[1] These claims are based on a simple presentation of mean scores as though they are absolutes.

A presentation of standard errors (Figure 1) for a sample of countries highlights that for PISA 2015 some results that varied on mean scores are statistically similar. For example, the cluster of Korea, New Zealand, Australia, the United Kingdom and Germany are all statistically similar to each other, despite varying mean scores. Further, the United States and France are similar, but statistically different (lower) to the cluster containing Australia.

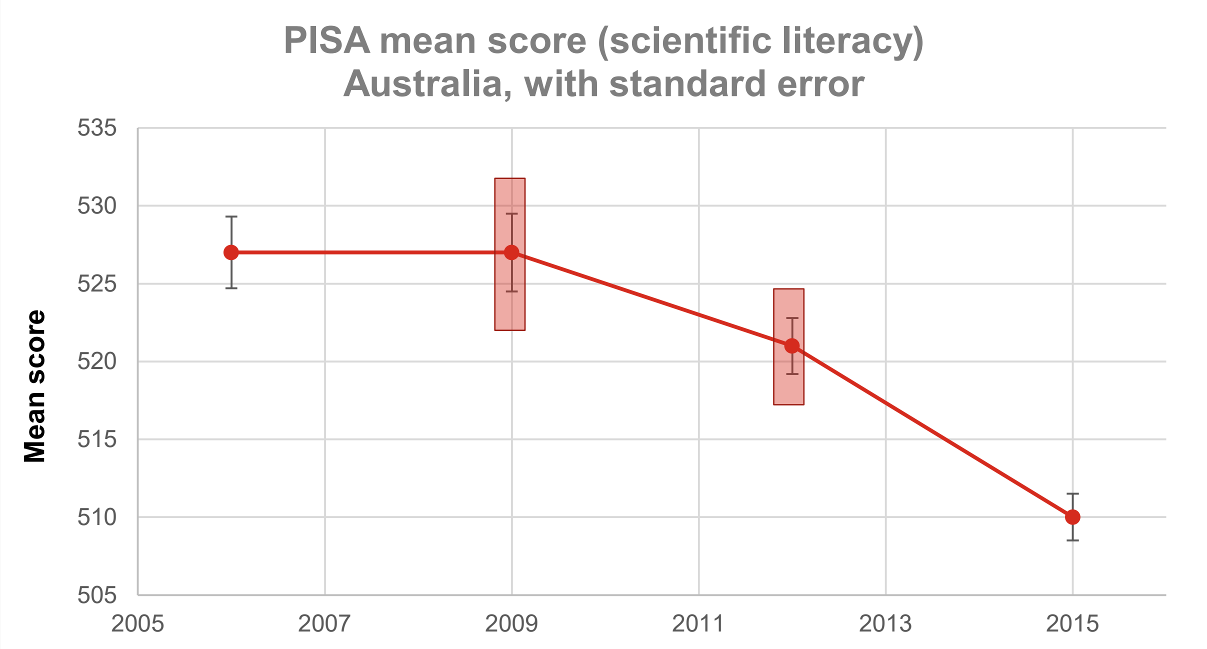


Data from Thomson, S, De Bortoli, L & Underwood, C 2016, *PISA 2015: A first look at Australia’s results*, Australian Council of Educational Research, Melbourne.

Figure 1: Mean scores (scientific literacy) with standard error (select countries), PISA 2015

[1] For example: Munro, K & Bagshaw E 2016, 'Australian schools two years behind world's best performing systems, The Sydney Morning Herald, December 6. http://www.smh.com.au/national/education/australian-school-students-two-years-behind-worlds-best-performing-systems-20161206-gt4w8p.html

A similar consideration must be given to results across time within a jurisdiction. For example, Figure 2 shows the pattern of results for Australia within the domain of scientific literacy in PISA. In this chart, error bars have been included to demonstrate standard error and shaded boxes to indicate confidence intervals. Although there appears to be a decline between 2009 and 2012 with the mean score decreasing from 527 to 521, these results are not significantly different with overlapping confidence intervals and almost overlapping error ranges. However, the decline in results in 2015 is statistically significantly different to previous results. This highlights that while there is some evidence of a trend of declining results in this element of PISA, a closer examination is needed before a fuller understanding can occur. Educators reflecting on PISA and TIMSS results are cautioned to consider the uncertainties inherent in the development of these figures, rather than just using mean scores as absolutes.



Data compiled from https://www.acer.org/ozpisa/publications-and-data.

Figure 2: Australia PISA results (scientific literacy) with standard error

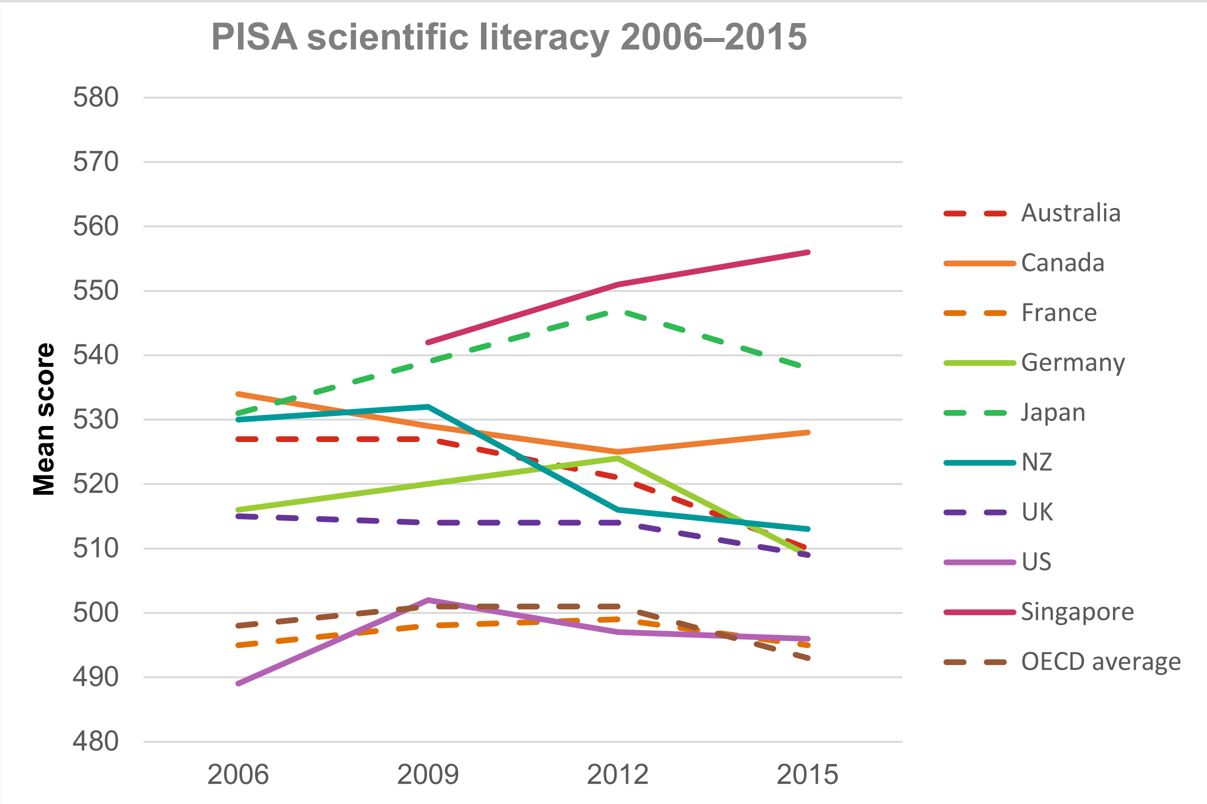
In this paper, for convenience of presentation, data is often presented as simple mean scores. However, where relevant, effort has been made to make statistical significance explicit, either in commentary or in presentation of standard error. The term 'significant' is used to denote statistical significance when describing variations. Where there is interest in interrogating the data more fully, readers are encouraged to examine the additional statistical information provided within both PISA and TIMSS reports (Thomson, De Bortoli & Underwood 2016; Thomson, Wernert, O'Grady & Rodrigues 2016).

Dominant debates

Analysis of data around Australia's performance within international testing programs remains a contested space with two broad positions. One position claims the data indicates declining Australian student performance, and therefore shows a need to focus on replicating what is occurring in successful nations. The other position claims that there are opportunities for performance improvement, but not to the extent of widespread reform across Australian education. This paper begins with an outline of the two positions.

Position 1: Australian performance is declining

The first position claims that Australian educational standards stagnated at the end of last century while other countries changed their trajectories and steadily improved, suggesting that, relative to other countries, Australia is slipping backwards. For example, Masters (2016) asserts the TIMSS Year 8 mathematics gap between Australia and Singapore, the world's highest performing country, widened between 1995 and 2015. In 2015, only 7% of Australia's Year 8 students performed at the Advanced level in mathematics and the same percentage for science, compared with Singapore's 54% and 42% respectively. Figure 3 shows that 2015 PISA data from the OECD demonstrates similar patterns, with an increasing gap between Singapore and Australia in scientific literacy. This figure also shows the complexity in making such a comparison, as the Australian trend is similar to the trend in the OECD average, and significant variations exist in performance across all countries.



Data compiled from https://www.acer.org/ozpisa/publications-and-data.

Figure 3: PISA scientific literacy results 2006–2015 for selected countries

The position that Australian students' ability to apply their knowledge and skills in science and mathematics is not only slipping backwards relative to other countries, but also declining in an absolute sense, appears to be supported by PISA data up to 2015:

- Australia's performance in scientific literacy, the major domain assessed in 2006 and 2015 (see appendix for explanation), declined by the equivalent of approximately half a year of schooling.
- Since 2012, the last time mathematical literacy was the major domain, Australia's performance has declined by the equivalent of about one-third of a year of schooling (Thomson, De Bortoli & Underwood 2016).

Buckingham (2016) argues that the results show continuing declines in the proportion of students at the most advanced level as well as significant increases in the proportion of students below the international standard. Hattie (2016), making reference to Ainley & Gebhardt (2015), asserts that there are too many 'cruising schools' and that the major decline in achievement is among the top

40% of Australian students. Further claims have been made that the gap between the achievement of top performing students in a class and those at the bottom is now as much as seven years (Bagshaw & Smith 2016).

An analysis of PISA data shows that:

- between 2006 and 2015, Australia's proportion of high performers in PISA scientific literacy decreased from 15% to 11%, while the proportion of low performers increased from 13% to 18%
- proficiency (see the appendix for a detailed definition) in PISA mathematics decreased from 20% in 2003 to 11% in 2015, and the percentage of low performers increased from 14% to 22% (Thomson, De Bortoli & Underwood 2016).

Data from TIMSS for the 20 years between 1995 and 2015 shows:

- the percentage of students achieving the Advanced benchmark in Year 8 science declined from 10% to 7%. There was no change in the percentage of students falling below the Low benchmark
- the percentage of students achieving the Advanced benchmark in Year 4 science significantly decreased
- in the majority of countries (10 out of 16) that participated in both TIMSS 1995 and TIMSS 2015, the percentage of Year 8 students achieving at the Advanced benchmark in mathematics significantly increased. Australia was an exception to this with no significant change in the percentage of students achieving the Advanced benchmark (Thomson, Wernert, O'Grady & Rodrigues 2016).

International results also show a large tail of underperformance at each year level in both mathematics and science. In TIMSS, around one-third of Australian Year 4 students and one-third of Year 8 students failed to achieve the nationally agreed proficient standard set by the Australian Curriculum, Assessment and Reporting Authority (ACARA):

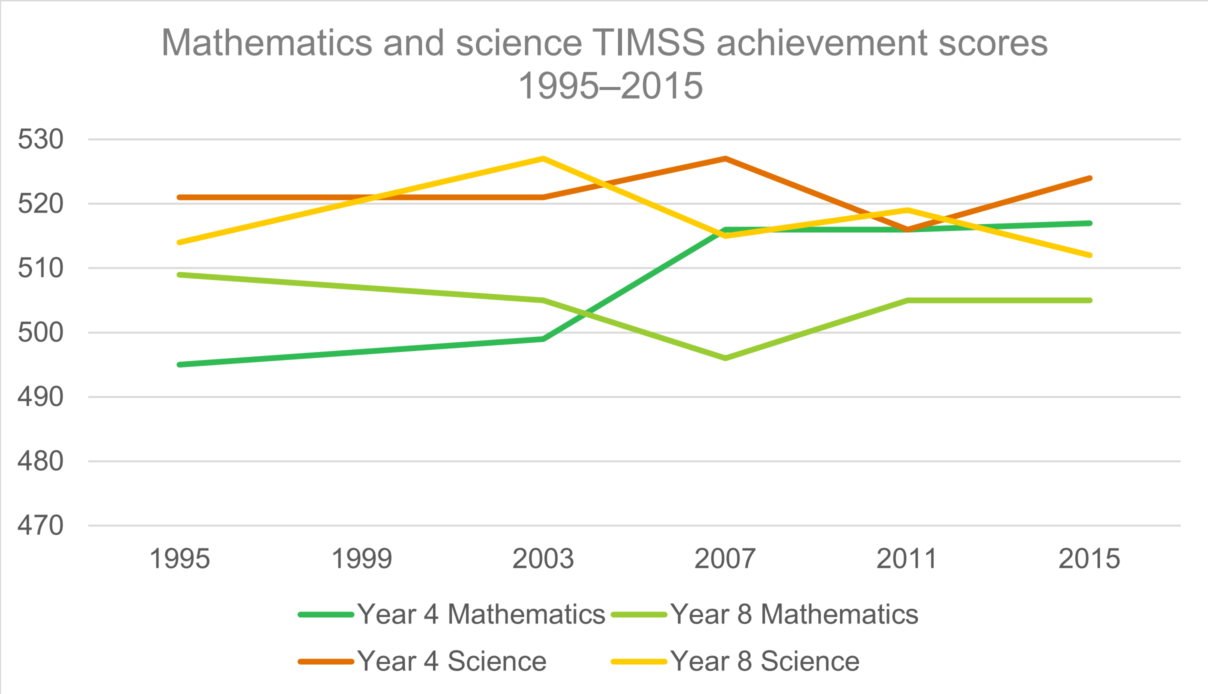
- Year 4 mathematics — 21% performed at the Low benchmark and a further 9% did not reach the Low benchmark
- Year 4 science — 19% performed at the Low benchmark and a further 6% did not reach the Low benchmark
- Year 8 mathematics — 25% performed at the Low benchmark and a further 11% did not reach the Low benchmark
- Year 8 science — 22% performed at the Low benchmark and a further 9% did not reach the Low benchmark (Thomson, Wernet, O'Grady and Rodrigues 2016).

Thomson (cited in Vukovic 2016) argues that such a significant proportion of students below the Australian proficient standard is cause for concern.

Position 2: Australia's performance is stable

The second position claims that Australia still performs well on international measures of student achievement, and fixation on international performance rankings is to our detriment. Australia's performance is higher than the OECD average, which would suggest a successful educational system; however, the OECD average is moveable dependent on 'across the board' performances, with half of results being naturally above the average at all times. International results also show that in the areas of mathematics and science, on average, Australian students outperform students from many other OECD countries. Additionally, Australia's results include learners who are top performers by international standards. PISA 2015 scientific literacy results indicate that Australia's proportion of high performers (11%) was higher than the OECD average (8%) and Australia's proportion of high performers in mathematics (11%) was consistent with the OECD average (Thomson, De Bortoli & Underwood 2016).

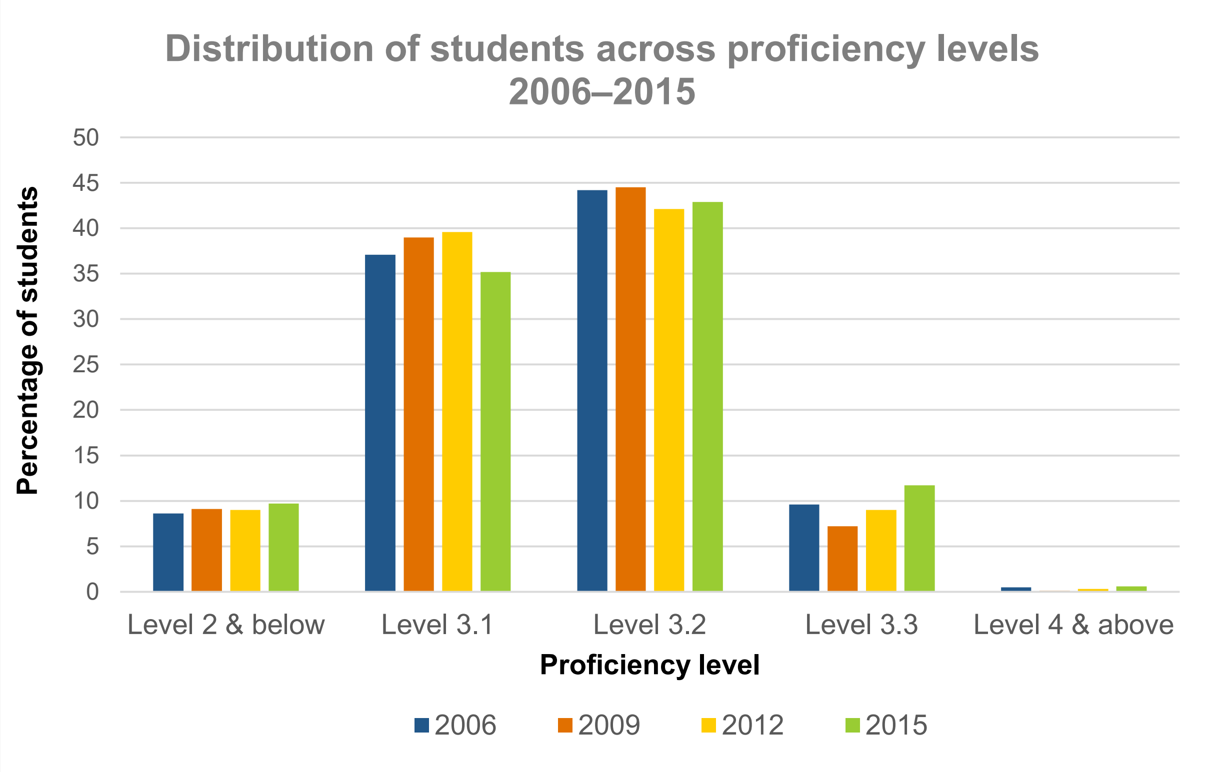
TIMSS data indicates some stability, particularly over the last three cycles. For example, Australia's 2015 Year 4 mathematics score was significantly higher than the corresponding score in 1995; however, this was due to a single increase between 2003 and 2007, with no following decline. For the past three cycles, Australia's scores have been similar. In Year 8 mathematics, there has been no significant change in the percentage of Australian students achieving the Advanced benchmark or falling below the Low benchmark over the past 20 years. In Year 4 science, the overall change since TIMSS 1995 is not significant. Australia's 2015 score in Year 8 science was basically the same as in 1995, with limited fluctuation since 2007 (Thomson, Wernert, O'Grady & Rodrigues 2016); TIMSS scores can be seen in Figure 4.



Data compiled from http://research.acer.edu.au/timss/

Figure 4: Trends in Australia's mathematics and science TIMSS achievement scores, 1995–2015

The position that Australia's results are stable is also supported by a comparison of international test results with national data. For example, National Assessment Program – Science Literacy (NAP-SL) results reflect a relatively consistent Year 6 student science literacy performance from 2006–2015. Figure 5 shows the distribution of student results for Year 6 students across Australia (Level 3.2 and above is considered proficient).



Data from http://www.nap.edu.au/results-and-reports/national-reports#NAP_sample_assessments

Figure 5: NAP-SL – Distribution of students across proficiency levels 2006–2015

Figure 5 indicates that the percentage of students considered proficient in 2015 is similar to 2006 (55.1% in 2015, up from 54.3% in 2006), with the difference not considered statistically significant. There is no indication of a marked decline in the results from 2006 to 2015. In fact, there is evidence of some improvement, with Level 3.1 decreasing and both Level 3.3 and 4+ increasing (ACARA 2017a).

Comparisons to similar countries (e.g. USA, UK, New Zealand, Canada, France and Germany) highlight that Australia is performing as well as, if not better, than these counterparts. Dinham (2013) maintains that Australia's performance on PISA testing has not significantly changed when compared to like nations. In mathematics, Canada and New Zealand were in 10th and 13th, ahead of Australia in 15th and Germany in 16th position. France was in 22nd place, while the UK and USA were below the OECD average at 28th and 31st respectively. In science, New Zealand and Canada were again just ahead of Australia in 7th and 8th position compared to our 10th place. Germany was at 14th, the UK at 16th, with the USA just above the OECD average in 23rd. However, these rankings are based on simple mean scores, and when considered with regards to error ranges, Australia's performance is comparable to culturally similar nations.

Despite the problematic nature of analysing data using country rankings, it does highlight relative achievement, and is often used in popular reporting of results of international testing. Baroutsis and Lingard (2016) propose that using mean scores for global rankings is imperfect because ranks change as the number of participants changes. For example, only 43 nations participated in the 2000 PISA, with participation increasing each assessment year to reach 72 countries and jurisdictions in 2015 (note that in some cases, e.g. China, the entire country does not participate but samples are drawn from city areas or smaller jurisdictions). Baroutsis and Lingard conducted a subsequent analysis of Australia's PISA rank using only the 32 countries that had data across all five assessment years, thereby eliminating this variance. Their results placed Australia 12th in mathematics and 10th in science. While acknowledging the somewhat arbitrary nature of their analysis, they argued the data nonetheless showed that Australia's performance on PISA declined less significantly than simplistic country rankings infer and had been suggested in reporting on the data.

The impact of disadvantage

Analysis of the international testing data, and in particular PISA data, suggests that an important source of variation in performance, particularly for Australia, is reflected in the social and economic circumstances surrounding education. In reporting on PISA, it is consistently highlighted that:

> If a country's scale scores in reading, scientific or mathematical literacy are significantly higher than those in another country, it cannot automatically be inferred that the schools or particular parts of the education system in the first country are more effective than those in the second. However, one can legitimately conclude that the cumulative impact of learning experiences in the first country, starting in early childhood and up to the age of 15, and embracing experiences both in school, home and beyond, have resulted in higher outcomes in the literacy domains that PISA measures. (OECD 2013a, p. 209)

Outcomes in education, it is argued, are only partial indicators around school effectiveness and teacher quality. These outcomes reflect far broader social and economic conditions that both support and hinder academic achievement and success. Evident in results from both PISA and TIMSS is the continuation of the significant inequality in achievement for Australian students. While Australian students are outperforming students from many other OECD countries, the data shows very large variations by region, gender, student SES, language background and Indigenous status as well as widening gaps in achievement as learners progress from stage to stage. Rowe (2006) asserts:

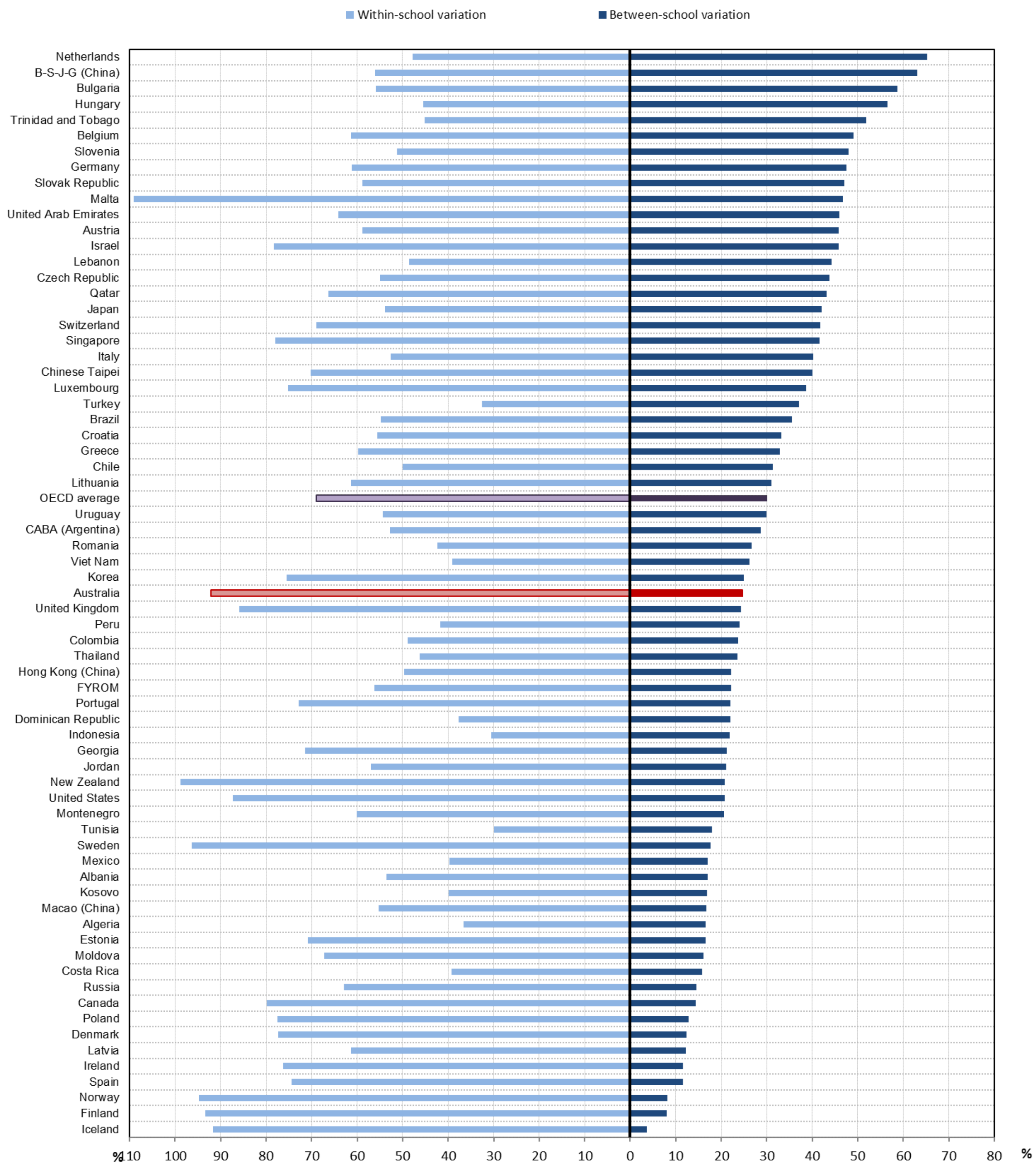
> … the largest source of variation in school performance is typically attributed to differences in what students bring to school: their abilities and attitudes, and family and community wealth and background, [whereas] the research evidence shows that school systems differ in the extent to which students' 'intake' characteristics and socioeconomic (SES) background influences achievement.

Marks (2016) contests the overemphasis on socio-economic status as a determinant for educational outcomes, though acknowledges that it makes some, but arguably small, difference. He instead concludes from an examination of international testing data and NAPLAN results that dominant influences on achievement are early childhood cognitive ability and prior individual achievement. He also proposes that persistence has a moderate effect on student achievement. Similar to Hattie's (2003) research, the argument is made that the most significant impacts on educational achievement are the characteristics and backgrounds students bring the classroom. However, it is the role of teachers to change the trajectories of these students.

Results within PISA 2015 suggest that many socio-economically disadvantaged students do excel. Across OECD countries, 31% of students from disadvantaged backgrounds are 'resilient', meaning that they are among the best performers of all students of similar background internationally (OECD 2011). Among disadvantaged students, learning time in school is one of the strongest predictors of performance. Disadvantaged students tend to spend less time in school studying science and a large proportion of disadvantaged students do not attain the PISA baseline proficiency level in science (OECD 2011). Resilient students were identified to have spent more time studying science, with significant variation occurring where this difference was as little as one hour of additional study per week (OECD 2011). Further:

> Focusing on disadvantaged students, the evidence in PISA reveals that resilient students are engaged and confident learners who enjoy learning Science and display a series of positive attitudes towards learning Science.

Variation in performance due to disadvantage is far more pronounced within schools than between schools (Gaber et al. 2012). Figure 6 highlights the variance in student performance within science as evident in PISA 2015. In this data, Australia (indicated in red) presents with a 'between-school' variance lower than the OECD average (24.7% compared to 30.1%), but within schools the variation was significantly different with Australia (92.1%) being far greater than the OECD average (69.0%). In their analysis of variance through disadvantage, the OECD suggests that a contributing factor may be less supportive households, and that there is a role for teachers and schools to provide the support necessary to enliven student interest in science and mathematics.



Data for graph extracted from PISA 2015, www.oecd.org/pisa, accessed 4 March 2017.

Figure 6: Variation in science performance between and within schools

The conclusions of this position are that school system policies, as well as individual school policies and practices, influence student learning and outcomes. Within Australia, there is evidence of a decline in 'resilient' students, highlighting questions about equity within Australian education and its capacity to meet the needs of all young Australians (Lamb et al. 2015). As education exists within social and economic contexts (Meyer & Schiller 2013), there are myriad influences that need to be considered, particularly in responding to disadvantaged students.

Queensland's performance

Country-to-country comparison of data from international testing programs is problematic due to variations in countries and contexts (Gillis, Polesel & Wu 2016). It is a fairer, more appropriate analysis and use of data to compare similar jurisdictions, and where possible compare data from within countries. In both PISA and TIMSS, Australian data is reported separately across each of the states and territories, thereby allowing comparisons between the different jurisdictions. However, at this level of analysis the data has reduced veracity, particularly within the minor domains, though given the sample and population size of the various jurisdictions, comparisons are still plausible. The following is a summary of data particular to Queensland's performance.

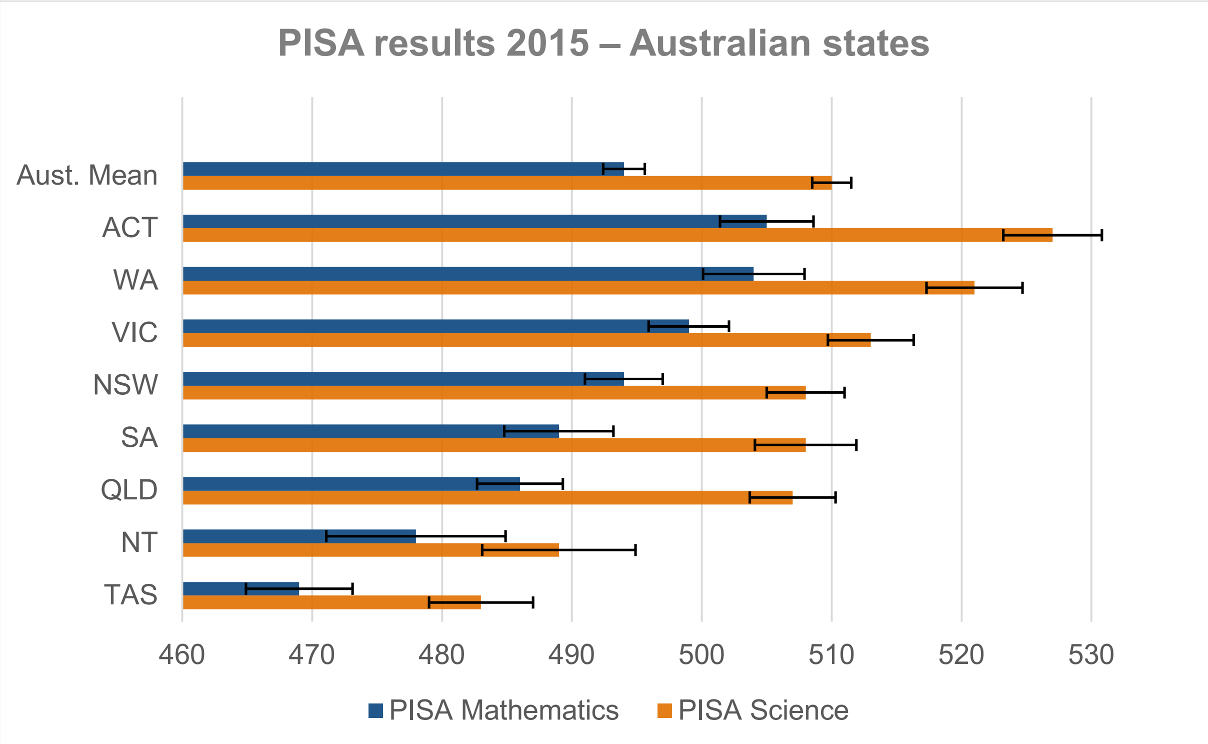
2015 PISA

Thomson, De Bortoli and Underwood (2016) provided a summary of 2015 PISA results.

- Scientific literacy — Queensland (507) performed higher than the OECD average (493) in scientific literacy and had the smallest decline across the states between 2006 and 2015 (15 points). All other Australian jurisdictions declined between 22 and 27 points. The proportion of students in Queensland who reached the National Proficient Standard in 2015 was 50%.
- Mathematical literacy — Queensland's score of 486 was not significantly different to the OECD average (490). Australia's average mathematical literacy scores between 2003 and 2012 declined by between 16 and 46 points across all jurisdictions except for Victoria. Of the seven jurisdictions with declines, Queensland had the smallest (16 points). The proportion of students in Queensland who reached the National Proficient Standard in mathematical literacy was 53%.

Figure 7 presents the achievement scores in mathematics and science for PISA 2015 for each Australian state and territory, and the national mean score, with standard error.

This data shows the statistically similar achievement in science across New South Wales, Victoria, South Australia and Queensland, and the higher achievement in Western Australian and the ACT. In the domain of mathematics there is similar clustering, but with differences between the groups being less clear. Western Australia and the ACT demonstrate similar results at the upper end, and Queensland, South Australia and New South Wales are also similar. Higher achievement in the ACT and Victoria is often attributed to smaller geographical jurisdictions and higher proportions of educated professional parents. Queensland, in terms of these characteristics, is somewhat unique as a large state with a highly dispersed population, and therefore significant rural and regional representation. Even direct comparisons with Western Australia, an equally large state, can be problematic due to more highly centralised population in Western Australia.



Data from Thomson, S, De Bortoli, L & Underwood, C 2016, *PISA 2015: A first look at Australia's results*, ACER, Melbourne.

Figure 7: PISA results 2015 — Australian states

Queensland's PISA performance over time suggests a general decline in mean scores in scientific and mathematical literacies, similar to national trends. Mirroring the national trend, there has also been a decline in the number of high-performing students and an increase in the number of low-performing students (see Tables 1 and 2).

Table 1: Proportion of high performers

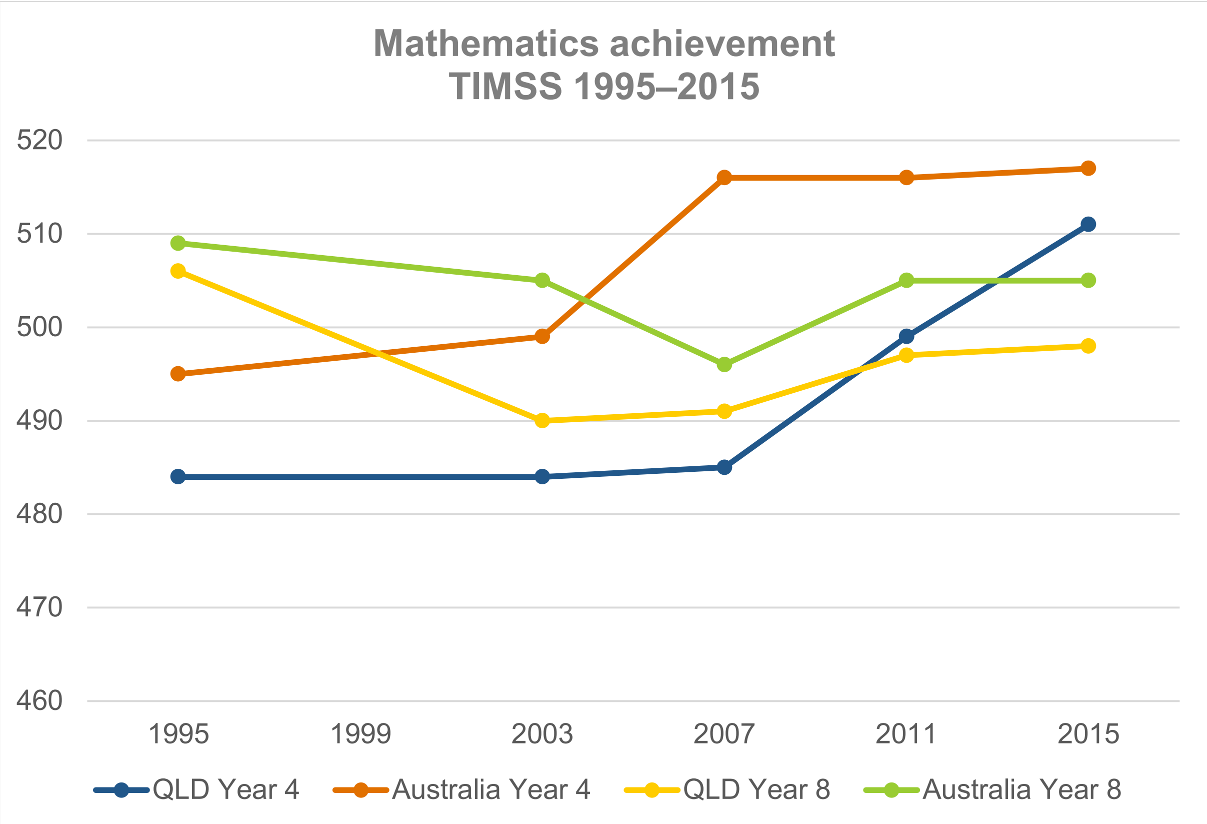
High performers	OECD PISA 2015 average	Qld PISA 2015 attainment	Decline in proportion of high performers (Qld)
Mathematic literacy	11%	9%	9% (since 2003)
Scientific literacy	8%	10%	3% (since 2006)

Table 2: Proportion of low performers

Low performers	OECD PISA 2015 average	Qld PISA 2015 attainment	Increase in proportion of low performers (Qld)
Mathematical literacy	23%	24%	8% (since 2003)
Scientific literacy	21%	18%	5% (since 2006)

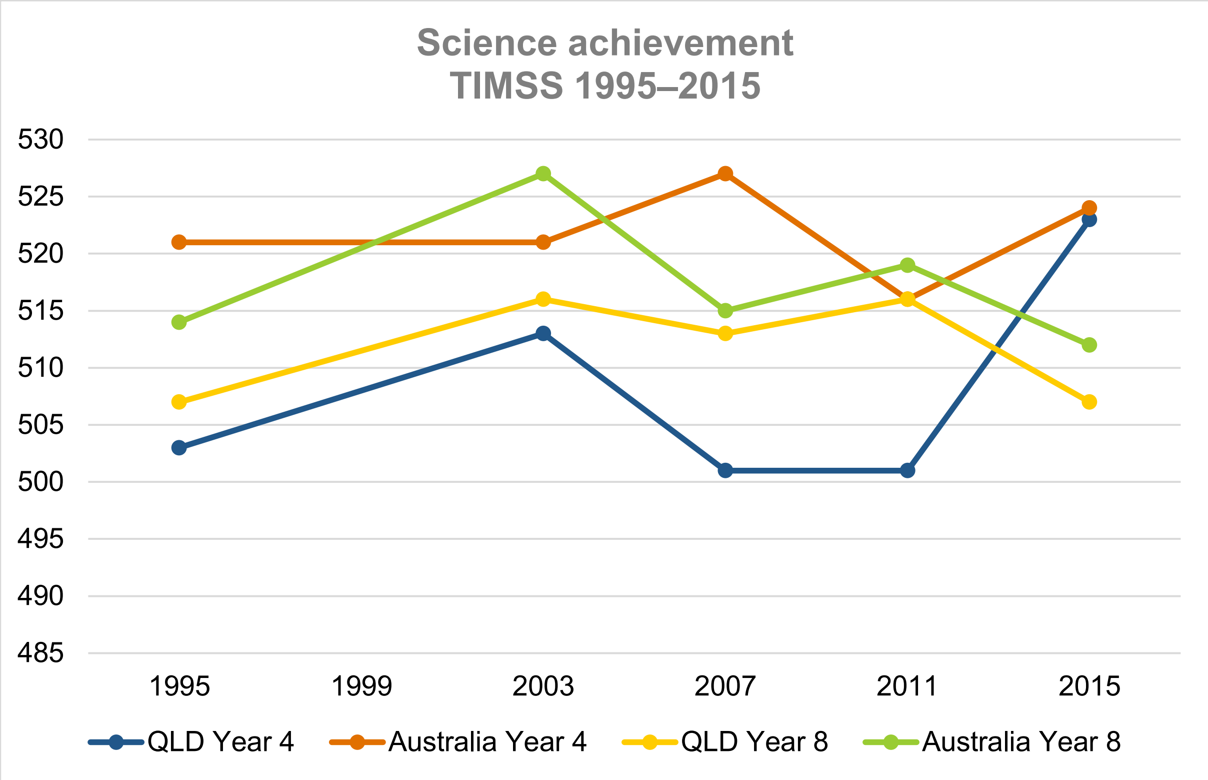
2015 TIMSS

Figures 8 and 9 present achievement in mathematics and science as measured by TIMSS across the period 1995–2015. Each graph shows Queensland and Australian mean scores for both Year 4 and Year 8 students. While there is some stability evident in the results for mathematics, with the suggestion of steady improvement, the outcomes for science appear slightly more erratic, though with a trend approaching the national mean. All results, both state and national, are well above the international average.



Data compiled from http://research.acer.edu.au/timss/

Figure 8: Mathematics achievement in TIMSS 1995–2015 (Note: No data available for 1999.)



Data compiled from http://research.acer.edu.au/timss/

Figure 9: Science achievement in TIMSS 1995–2015 (Note: No data available for 1999.)

Other key points related to Queensland's performance in TIMSS are outlined below.

Mathematics

- In Year 4 mathematics, Queensland's performance showed improvement in 2011 and again in 2015. In 2015, Queensland was the third lowest performing jurisdiction in Year 4 mathematics, although results in Tasmania, Western Australia and South Australia are similar. Queensland had a significantly higher average score in 2015 (511) than in 1995 (484). Queensland has seen a statistically significant reduction in the proportion of low performing students since 1995 as well as an increase in the percentage of students achieving the Advanced international benchmark. In 2015, only 6% of students reached the Advanced benchmark and 10% of students did not reach the Low benchmark.
- Queensland ranked fifth out of the eight jurisdictions in Year 8 mathematics in 2015, equal with South Australia. Only 4% of students achieved the Advanced benchmark and 11% did not reach the Low benchmark. Queensland's performance declined from 1995 to 2003, but has subsequently shown gradual improvements, despite not yet reaching the 1995 level.

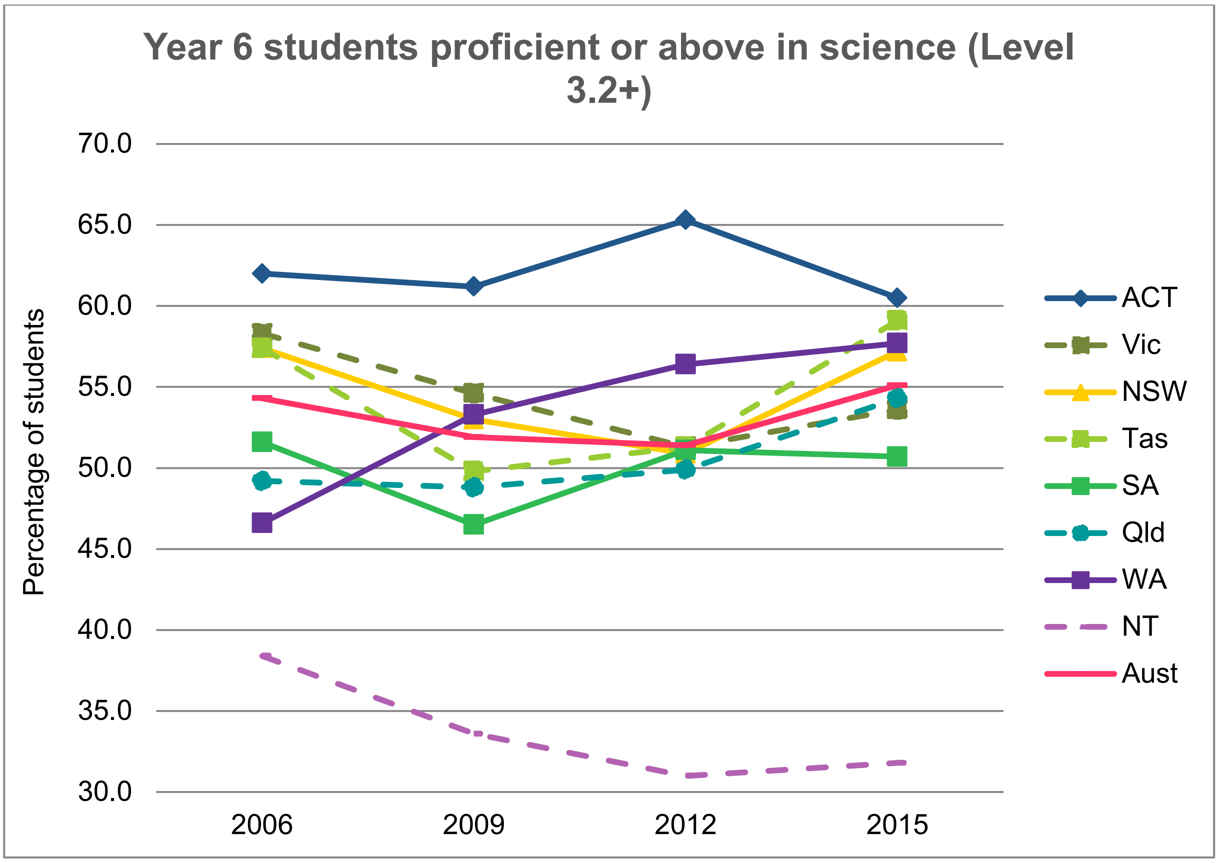
Science

- In Year 4 science, Queensland's performance significantly improved by 23 points and reached its highest ever in 2015. There was also a statistically significant decline in the number of students not achieving the Low benchmark, but this was offset by a slight decrease in the percentage of students achieving the Advanced benchmark. While Queensland was ranked third lowest of the jurisdictions, results are statistically similar to South Australia, New South Wales and Tasmania. Of Queensland's Year 4 students, 7% reached the Advanced level and 7% performed below the Low benchmark.

- In Year 8 science, Queensland and South Australia ranked equal fifth in jurisdictional rankings; 5% of Queensland Year 8 students achieved the Advanced benchmark and 9% of students performed below the Low benchmark. Queensland's results in 2015 are equal to the 1995 level, which are the lowest across the 20 years.

2015 NAP-SL

2015 NAP-SL data indicates there may be small improvements in the percentage scores of Year 6 students in similar states over time (ACARA 2017a). ACARA reports that the change is not statistically significant in any state except Western Australia. In the context of this discussion, the data does not indicate a decline in the results as could be inferred from PISA and TIMSS reports.



Data from http://www.nap.edu.au/results-and-reports/national-reports#NAP_sample_assessments

Figure 10: Percentage of Year 6 students achieving proficient level or above (Level 3.2+), NAP-SL 2006–2015

Values, context and aspirations

Interpretation of international testing data must be undertaken with a clear understanding of the values, context and aspirations of the educational jurisdiction. The Melbourne Declaration on Educational Goals for Young Australians (MCEETYA 2008) signposts academic, personal and social development and achievement as being important for students in Australian schools. Therefore, while international testing data provides one indication of performance in a limited area of education, they do not, nor do they claim to, represent the totality of an educational system. The value assigned to these scores in determining the success of education within Queensland or Australia must be determined by considering the broad set of values and aspirations for education.

The 'top-ranking' PISA and TIMSS jurisdictions are generally very different to Australia socially, culturally, demographically, geographically and linguistically, and any comparison of student performance should be viewed in this context. The conditions for learning within these jurisdictions, including long school hours, extra tutoring outside of school including on weekends and school holidays, cramming and intense test preparation can be extreme (Dinham 2013) and it is debatable that this is what Australia would desire for its young people. Likewise, attention and effort with respect to performance on international tests is substantially different, with reports, for example, of South Korea approaching the tests as indications of national pride and commitment, as opposed to Australia's broader lassiez-faire approach (Gorur & Wu 2015).

Raising performance on international testing requires a more considered discussion based on recognising and developing our own strengths rather than attempting to emulate practices from top-performing countries with different approaches and values to us. Gorur and Wu (2015) argue that, before looking overseas, Australian jurisdictions should look to other states for improvement strategies. For example, Western Australia, which shares many similarities with Queensland, has achieved results that compare favourably with other PISA participants:

- Western Australia's 2015 PISA mathematical literacy scores are significantly lower than only 10 countries and not significantly different to 12 countries including Poland, Ireland and Germany
- In 2015 PISA scientific literacy, Western Australia performed significantly lower than only six countries and not significantly different to six others including Hong Kong, Vietnam and Canada (Thomson, De Bortoli & Underwood 2016).

2015 NAP-SL data (see Figure 10) is further evidence of Western Australia's high performance. This means there are within-country role models Queensland could look to before considering distant and culturally radically different systems such as Singapore or Shanghai.

Despite the espoused concerns about Australia's educational performance, particularly in science and mathematics, as a nation we still produce high quality scientists and contribute to international scientific research far beyond what would be expected of a nation of our size and economy (Office of the Chief Scientist 2013). Although we continue to do well in these fields, it remains important that investment (both monetary and effort) be made into continuing to build the scientific and mathematical skills of young students in preparation for a society in which the vast majority of new jobs will be generated in the STEM (science, technology, engineering and mathematics) fields (Office of the Chief Scientist 2012). The question, therefore, is not whether we should aim to continue to improve teaching and outcomes in mathematics and science, but how this is best done and what measures will properly indicate success.

Considerations for Queensland schools

International testing is not designed to capture everything taught in Australian schools. This means that results do not reflect our whole education system nor the interests or abilities of all our students. While external measurements can provide valuable data, they are not the sole source of evidence of the degree of a school's or system's success, with much variation in student achievement explained through many non-educational factors (Meyer & Schiller 2013). PISA, for example, does not measure what is taught in schools, only what has been determined by the OECD that students should be able to do by 15 years of age.

It is proposed that Queensland schools already effectively provide students with skills and experiences to prepare them for active citizenship and to enable them to continue as lifelong learners. Any reforms must prioritise this value of equipping students with '21st century skills' to meet their future challenges ahead of performance on international testing. Students need to be able to think deeply and critically about issues, solve problems creatively, work in teams, show initiative and leadership, and possess intercultural understandings. To measure success, Queensland schools should continue to collect data and feedback from within their schools, including aspects that relate to the formal curriculum, such as increased retention, attendance, sense of belonging and integration of students with disadvantage. This will provide a more valuable understanding of the success (or otherwise) of education (Starr 2014).

In reviewing the data from international testing programs alongside a range of commentary and broader analysis, the following points are identified as opportunities to positively impact student performance, particularly in mathematics and science:

- develop and improve teaching and learning
- encourage student engagement in mathematics and science
- allocate resources to address individual student needs.

Develop and improve teaching and learning

What happens inside the classroom is crucial for students' learning (OECD 2016). This means that schools must maintain the focus on continuous improvement in teacher quality by providing ongoing, high-quality and evidence-based professional learning for teachers, especially in terms of high-level skills in assessment and analysis of data. This allows an emphasis on each student's learning and the best next teaching steps to advance that learning (Thomson 2013).

Hattie (2016) asserts that school improvement requires:

- identifying and valuing expertise
- supporting teachers to work collaboratively
- targeting resources at need
- accepting evidence and evaluating progress over time.

The OECD Teaching and Learning International Survey (Jensen 2012) identifies the following aspects as necessary for improved learning:

- teachers' content knowledge
- teachers' pedagogical knowledge, both general and subject-specific
- clear, well-structured lessons supported by effective classroom management resulting in a high proportion of time used for effective learning and teaching
- individualised instruction
- commitment to higher-order problem-solving, deep analysis of content, and activities requiring advanced thinking skills and deductive reasoning
- active professional collaboration that has a direct impact on learning and teaching. Key elements include classroom observations, team teaching and constructive feedback.

To improve learning and address inequity across Australian schools, schools must ensure there is a quality teacher in every classroom (Dinham 2013).

Encourage student engagement in mathematics and science

Engagement, retention and attendance are important indicators of success at school (Starr 2014). PISA surveys provide valuable data on students' dispositions towards schooling and lifelong learning. In 2000, 21% of Australian students who participated in PISA assessment felt a low sense of belonging at school, slightly better than the OECD average of almost a quarter. 2012 PISA data reflected a weakening of students' sense of belonging by around ten percentage points since 2003 (OECD 2013b). The same report states that:

> Socio-economically disadvantaged students are less likely than advantaged students to feel like they belong at school, are more likely to feel like outsiders, and are less likely to feel happy and satisfied with their school (p. 54).

To further develop students' post-schooling outcomes, schools need to identify students who are at risk of disengagement and work with them individually to re-engage them as learners.

Improving engagement in science is vital for disadvantaged students. The OECD (2011) reported that time spent learning science is one of the correlates of better performance that benefits the most disadvantaged students. Their findings showed that, in Australia, the odds of disadvantaged students with significant exposure to science learning being resilient (i.e. beating the odds and succeeding at school) are four times greater than those who do not have the opportunity to learn science at school, after accounting for a host of student and school background factors, approaches to learning and school policies. Schools have an important role to play in fostering resilience, self-confidence, motivation and opportunities to learn. Policies geared towards improving disadvantaged students' access to, and learning time in, science should be considered to improve equity outcomes and boost average performance.

NAP-SL results show that more than 80% of Year 6 students appear interested in learning new things in science, learning about science and doing science-based activities (ACARA 2017a). This provides a strong foundation to build student awareness of and confidence in science and inspire continued engagement and the pursuit of excellence. Unfortunately, TIMSS evidence seems to suggest that student interest in science declines in later years. In Year 4, just 11% of female students and 13% of male students said that they do not like learning science. By Year 8, this had increased to 32% of female students and 26% of male students, indicating the challenge for secondary schools is to build engagement in science.

In a survey commissioned by the Office of the Chief Scientist (2012), senior secondary school students and university students nominated teachers as the most influential factor in determining their interest in and attitudes toward science. Students identified the most stimulating styles of teaching and learning as student-led research, practical activities and the study of real-world examples within the student's sphere of experience.

In almost all education systems in PISA 2015, students scored higher in science when they reported their science teachers 'explain scientific ideas', 'discuss their ideas' or 'demonstrate an idea'. Students also scored more highly when their teachers adapted lessons to their needs or provided individual help when students had difficulty understanding a topic (OECD 2016).

Goos (2016) asserts that different types of teaching and teacher qualities also affect student achievement in mathematics. Referring to the work of Mewborn (2003), she states that research demonstrates that students' mathematics learning and their dispositions towards mathematics are influenced — for better or for worse — by the teaching they experience at school. To bring mathematics to life, teachers need to use creativity and vitality to communicate a belief that mathematics is a tool for thinking with, a unique and concise language, a way of investigating patterns and relationships, and a part of everyday life.

Some of the strongest correlations with achievement and school completion are based on how students view mathematics. The correlations between achievement and anxiety, and confidence and mathematics self-efficacy, are moderately strong across the OECD, but they are even stronger in Australia. How well Australian students achieve in mathematics is linked more strongly to their beliefs and views about themselves as mathematics students than for learners across the OECD (Lamb et al. 2015).

High results on international assessments do not necessarily correlate to engagement with science and mathematics. While students from many Asian countries attained top ranks in mathematics and science, they scored lowest in 'enjoyment' of these subjects. For example, Korea scored lower than the international average on learning attitudes and confidence, and this was replicated in Hong Kong, Japan, Taiwan and Singapore (Seong 2012). This is more likely to influence whether they pursue careers in mathematics and science after school than scores on international assessments. Hattie (2016) claims that schools are driving down mathematics and science participation and success by the way they teach and attract students. He suggests mathematics and science should not be promoted as the domain of the academically talented or those who see themselves as future scientists. Instead, he suggests schools need to reconsider methods of teaching mathematics and science, how students are enticed to enjoy the learning in these subjects, and how we promote them as relevant and exciting. This is supported by the OECD (2016), and the Australian Chief Scientist, who assert that promoting a positive and inclusive image of science is important. As knowledge and understanding of science are useful well beyond the work of scientists, school science should be promoted more positively as a source of interest and enjoyment.

Allocate resources to address individual student needs

Successful education systems understand that learning should not be hindered by disadvantage and therefore find ways to allocate resources to address inequality. 2015 PISA results indicate:

- Australia's equity profile was not significantly different to the OECD average, except in scientific literacy where the effect of socio-economic background was higher than average
- Indigenous students were on average about two-and-a-half years behind non-Indigenous students
- students from metropolitan schools were about one year of schooling ahead of students in provincial schools and one-and-a-half years ahead of those in remote schools (Karp 2017).

The OECD (2016, 2017) raises a number of considerations that can address within-school disadvantage, with varying degrees of resource and public policy demands. They suggest the following:

Public policy responses

- developing teacher quality to ensure that all students, particularly those facing disadvantage, receive quality teaching and learning experiences
- as much as possible, ensuring mathematics and science classes are taught by qualified teachers
- targeting resources at disadvantaged students or those who struggle with science, which can make a difference in helping students acquire a baseline level of science literacy and develop a lifelong interest in the subject
- adequately diagnosing and addressing every child's needs

School and sector responses

- timetabling sufficient time for the teaching of science; PISA results show that students score five points higher in science for every additional hour spent per week in regular science lessons, after accounting for socio-economic status
- requiring students to attend science classes; limiting the exposure of low-performing students to science only widens the gap with better-performing students
- improving school attendance and retention rates
- extending the range of enriching extracurricular activities (e.g. competitions and clubs) to make science more engaging, relevant and interesting

Teacher responses

- encouraging the beneficial effects of peer influences by valuing students' achievements and efforts
- treating all students with the same level of attention and respect
- showing interest in the various cultural traditions represented in the student body
- having high expectations for all students.

Conclusion

Data from international testing programs is extensive. Hundreds of pages of reports and data analysis have been developed by the OECD, ACER and other organisations. However, the value of any analysis must be aligned to the intended outcomes and aspirations of an educational system. Achieving within the top five countries on PISA or TIMSS is only of value where such an achievement reflects the values and aims of education in Queensland and Australia. Otherwise, this data is just one part of a more complex puzzle that helps teachers understand what may work in their classrooms and supports evidence-led decision-making in education.

This paper has only highlighted a small amount of the available analysis of international testing programs, but, through this, it has identified some possible focus areas for schools. A commitment to continuous improvements in teaching and learning, more effort to engage students in mathematics and science, and allocating resources within the school to meet individual student needs emerge as three considerations to support in-school improvements. These are not the only areas of focus that schools will benefit from, but they provide an evidence-based starting point for schools to target improvements in student outcomes in mathematics and science.

Acknowledgments

The QCAA acknowledges the input and feedback provided by Professor Merrilyn Goos, University of Queensland, and Associate Professor Greg Thompson, Queensland University of Technology, as part of the review process.

References

Australian Council for Educational Research 2016, 'PISA 2015: Australian students' achievement' Research Developments, https://rd.acer.org/article/PISA-2015-australian-students-achievement 6 December.

Ainley, J & Gebhardt, E 2013, 'Measure for measure : a review of outcomes of school education in Australia', Australian Council for Educational Research, Melbourne.

Australian Curriculum Assessment and Reporting Authority 2017a, *NAP Sample Assessment Science Literacy* http://www.nap.edu.au/docs/default-source/default-document-library/20170309-nap_sl_final.pdf?sfvrsn=2 accessed 31 July 2017

—— 2017b, 'Student performance in science remains static', http://www.acara.edu.au/docs/default-source/Media-Releases/20170315-nap-science-literacy-media-release.pdf?sfvrsn=2, accessed 15 Mar 2017

Bagshaw, E & Smith, A 2016, 'Education policy not adding up: OECD asks what's wrong with Australia's schools', *Sydney Morning Herald*, 26 March.

Baroutsis, A & Lingard, B 2016, 'Counting and comparing school performance: an analysis of media coverage of PISA in Australia, 2000–2014', *Journal of Education Policy*, pp. 1–18.

Buckingham, J 2013, 'Don't panic about PISA', ABC News, viewed 22 March 2017, http://www.abc.net.au/news/2013-12-04/buckingham-pisa-panic/5133364.

—— 2016, 'Why Australia's PISA results are a catastrophe', *The Financial Review*, 7 December.

Buckingham, J 2016 'New PISA reading, maths and science results confirm Australia's educational decline', The Centre for International Studies, 6 December, https://www.cis.org.au/commentary/articles/new-pisa-reading-maths-and-science-results-confirm-australias-educational-decline/

Centre for International Research on Education Systems 2015, 'Fact sheet 4: Senior school years: School completion uneven across Australia', *Educational opportunities in Australia 2015,* Mitchell Institute for Health and Education Policy, Melbourne.

Dinham, S 2013, 'The quality teaching movement in Australia encounters difficult terrain: A personal perspective', *Australian Journal of Education*, vol. 57, no. 2, pp. 91–106.

Donnelly, K 2014, 'Social class affects school achievement less than you think', *The Conversation*, 7 March.

Gaber, S, Cankar, G, Marjanovic Umek, L & Tasner, V 2012, 'The danger of inadequate conceptualisation in PISA for education policy', *Compare: A journal of comparative and international education*, vol. 42, no. 4, pp. 647–63.

Gillis, S, Polesel, J & Wu, M 2016, 'PISA data: Raising concerns with its use in policy settings', *The Australian Educational Researcher*, vol. 43, no. 1, pp. 131–46.

Goos, M 2006 'Why teachers matter', *Australian Mathematics Teacher,* vol. 62 no. 4, pp. 8–13.

Gorur, R & Wu, M 2015, 'Leaning too far? PISA, policy and Australia's 'top five' ambitions', *Discourse: Studies in the Cultural Politics of Education*, vol. 36, no. 5, pp. 647–64.

Hattie, J 2003, 'Teachers make a difference: What is the research evidence?' Paper presented at the Building Teacher Quality: What does the research tell us? ACER Research Conference, Melbourne, Australia. Retrieved from http://research.acer.edu.au/research_conference_2003/4/

Hattie, J 2016, 'Shifting away from distractions to improve Australia's schools: Time for a reboot', paper presented to ACEL 2016 annual conference, Melbourne, 28–30 September', paper presented to ACEL Annual conference, Melbourne, 28–30 September.

Jensen, B 2012, *Catching Up: Learning from the best school systems in East Asia,* Grattan Institute, February 2012

Karp, P 2017, 'Socioeconomic disadvantage puts students three years behind, report finds', https://www.theguardian.com/australia-news/2017/mar/15/socioeconomic-disadvantage-students-three-years-behind-report-finds, accessed 16 March 2017.

Lamb, S, Jackson, J, Walstab, A & Huo, S 2015, 'Educational opportunity in Australia 2015: Who succeeds and who misses out', Centre for International Research on Education Systems, Victoria University, for the Mitchell Institute, Melbourne: Mitchell Institute.

Marks, GN, 2016, 'The relative effects of socio-economic, demographic, non-cognitive and cognitive influences on student achievement in Australia', *Learning and Individual Differences* 49, pp.1–10.

Masters, G 2016, '20-year slide in maths and science learning', https://www.teachermagazine.com.au/geoff-masters, accessed 30 January 2017.

Meyer, H-D & Benavot, A 2013, 'PISA and the globalisation of education governance: Some puzzles and problems', in H-D Meyer & A Benavot (eds), *PISA, power and policy: the emergence of global*

educational governance, Oxford Studies in Comparative Education, Symposium Books, Southhampton, pp. 9–26.
Meyer, H-D & Schiller, K 2013, ‘Gauging the role of non-educational effects on large-scale assessments: Socio-economics, culture and PISA outcomes’, in H-D Meyer & A Benavot (eds), *PISA, power and policy: the emergence of global educational governance, Oxford Studies in Comparative Education*, Symposium Books, Southhampton, pp. 207–24.
Ministerial Council on Education, Employment, Training and Youth Affairs 2008, *Melbourne Declaration on Educational Goals for Young Australians*, December 2008.
OECD 1999 *Measuring Student Knowledge and Skills: A new framework for assessment*. Paris: OECD.
OECD 2011, ‘PISA in focus’ http://www.oecd.org/pisa/pisaproducts/pisainfocus/48165173.pdf
OECD 2013a, ‘PISA 2012 results: What makes schools successful?’ *Resources, policies and practices* (Volume IV), PISA: OECD Publishing.
OECD 2013b, PISA 2012 Results: Ready to Learn: Students’ Engagement, Drive and Self-Beliefs (Volume III), PISA, OECD Publishing.
OECD 2016, ‘PISA 2015: results in focus’, https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf
OECD 2017, ‘PISA 2015 results: Volume III. Students’ well-being’, PISA, OECD Publishing, Paris.
Office of the Chief Scientist 2012, *Health of Australian Science*, Australian Government, Canberra.
—— 2013, *Science, Technology, Engineering and Mathematics in the National Interest: A strategic approach,* Australian Government, Canberra.
Riddle, S & Lingard, B 2016, ‘PISA results don’t look good, but before we panic let’s look at what we can learn from the latest test’, *The Conversation*, 6 December.
Rowe, K 2006, *School Performance: Australian state/territory comparisons of student achievements in national and international studies*, Australian Council of Educational Research, Melbourne.
Sellar, S, Thompson, G & Rutkowski, D 2017, *The Global Education Race: Taking the measure of international testing*, Brush Education Inc., Calgary.
Starr, K 2014, ‘The influences and implications of PISA: An Australian perspective’, *AASA Journal of Scholarship & Practice*, Winter 2014, vol 10. no. 4. pp. 19–30.
Thomson, S 2013, ‘Declining PISA outcomes: Time to stop the slide’, https://rd.acer.org/article/declining-pisa-outcomes-time-to-stop-the-slide, accessed 17 Mar 2017.
Thomson, S 2010, *Mathematics learning: What TIMSS and PISA can tell us about what counts for all Australian students*. Camberwell: ACER.
Thomson, S, De Bortoli, L & Underwood, C 2016, *PISA 2015 : A first look at Australia’s results*, Australian Council of Educational Research, Melbourne.
Thomson, S, Wernert, N, O'Grady, E & Rodrigues, S 2016, *TIMSS 2015 : A first look at Australia's results*, Australian Council of Educational Research, Melbourne.
Vukovic, R 2016, ‘Australian student performance flatlining’ www.teachermagazine.com.au/?ACT=23&ad_id=273 accessed 24th January 2017.

Appendix

This appendix provides more details about the assessments considered in this report.

PISA[2]

What is PISA?

The Programme for International Student Assessment (PISA) is a triennial international study that measures how well 15-year-olds (which means students could be in Years 9, 10 or 11) use their knowledge and skills in mathematics, science and reading. PISA does not measure a specific curriculum, but it is used to evaluate education systems. Australia has been involved in PISA since its inception in 2000.

What does PISA assess?

> **Scientific literacy** is the ability to engage with science-related issues, and with the ideas of science, as a reflective citizen. A scientifically literate person is willing to engage in reasoned discourse about science and technology, which requires the competencies to explain phenomena scientifically, evaluate and design scientific enquiry, and interpret data and evidence scientifically.
>
> **Reading literacy** is an individual's capacity to understand, use, reflect on and engage with written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society.
>
> **Mathematical literacy** is an individual's capacity to formulate, employ and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. It assists individuals to recognise the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens. (OECD 2016, p. 13)

How is PISA administered?

PISA is conducted every three years. The dominant assessment domain rotates between scientific, reading and mathematical literacy as reflected in the table below from Thomson, De Bortoli and Underwood (2016).

PISA 2000	PISA 2003	PISA 2006	PISA 2009	PISA 2012	PISA 2015
Reading literacy	Reading literacy	Reading literacy	Reading literacy	Reading literacy	Reading literacy
Mathematical literacy	Mathematical literacy	Mathematical literacy	Mathematical literacy	Mathematical literacy	Mathematical literacy
Scientific literacy	Scientific literacy	Scientific literacy	Scientific literacy	Scientific literacy	Scientific literacy

Major domain (coloured cells: Reading literacy in PISA 2000 and 2009; Mathematical literacy in PISA 2003 and 2012; Scientific literacy in PISA 2006 and 2015)   Minor domain (all other cells)

This has implications for comparing data across major and minor domain cycles. In 2009, the major domain of reading was given 60% of the testing time, meaning it was more thoroughly assessed than mathematics and science (OECD 2012, p. 28). By allocating a domain additional testing time, more 'linking items' can be included. This allows developers to ensure that the assessment is comparable from cycle to cycle and lessens the link error in score comparison

---

[2] Descriptions from:

Thomson, S, De Bortoli, L and Underwood, C 2016, *PISA 2015: A first look at Australia's results*, Australian Council for Educational Research Ltd, Camberwell.

OECD 2016, 'PISA 2015 Assessment and analytical framework: Science, reading, mathematics and financial literacy', PISA, OECD Publishing, Paris. http://www.mecd.gob.es/dctm/inee/internacional/pisa-2015-frameworks.pdf?documentId=0901e72b820fee48 between cycles. In PISA, the link error is statistically accounted for and documented in the PISA technical reports (e.g. OECD 2012). Link error estimates provide information about how trends were estimated. Along with sampling errors, link errors manifest themselves in the measures of uncertainty that surround achievement estimates across time.

Also relevant to the major/minor domain distinction is the fact that the frameworks for PISA were not well developed until the point at which they served as the major domain: 2000 for reading, 2003 for mathematics and 2006 for science. Consequently, mathematics achievement is comparable only back to 2003, and science achievement is directly comparable across cycles only back to 2006 (OECD 2012). In addition, given the small number of overlapping linking items between major and minor, or minor and minor domain years, it is advisable to only compare performance differences between major domains (e.g. reading in 2000 and 2009). In other words, because the minor domains include only a small number of questions, there are not enough linking items that are similar between the previous assessment and the current assessment to be confident that the changes in scores are due to changes in student ability (Sellar, Thompson & Rutkowski 2017, pp. 51–2).

The number of changes in administration for the 2015 assessment also mean that direct comparisons with previous cycles should be treated with caution. For example, the assessment mode was changed, with the assessment being delivered online for the majority of countries for the first time. Non-attempted items are referred to as 'non-reached items' in 2015, whereas previously they were referred to as 'incorrect' when estimating proficiency.

All Australian states and jurisdictions participated in PISA 2015, with about 758 schools and 14 500 students completing the assessment.

Multi-matrix sampling

Using what is called multiple-matrix sampling, assessment material is divided up into overlapping item clusters that are assembled into partially overlapping assessment booklets. Approximately 10 hours of testable material is packaged into 120-minute booklets, meaning an individual student takes a part of the assessment, not the entire test. Testing organisations use their background information along with how students performed on a portion of the assessment to calculate five likely scores. When this is aggregated to the population level, the scores are stable; however, it is not possible for the testing organisation to calculate accurate individual and school scores. As such, these five scores should never be reported as individual results, and results from PISA should not be reported at the school level (Sellar, Thompson & Rutkowski 2017, pp. 45–46).

How are PISA results reported?

PISA results are reported on a set of scales, originally constructed to have an average score of 500 and a standard deviation of 100. Results are reported as average scores, providing a summary of student performance and allowing for comparisons of the relative standing between different countries and subgroups. The OECD average is the average of the data values across all OECD countries and can be used to compare a country on a given indicator with a typical OECD country.

In scientific and reading literacies, there are seven levels of proficiency, ranging from Level 1b (the lowest proficiency level) to Level 6 (the highest proficiency level). A difference of 75 score points represents one proficiency level on the PISA scientific literacy scale, while a difference of 73 score points represents one proficiency level on the PISA reading literacy scale. For mathematical literacy, there are six levels, ranging from Level 1 (the lowest proficiency level) to Level 6 (the highest proficiency level), with 62 score points representing one proficiency level.

TIMSS[3]

What is TIMSS?

The Trends in International Mathematics and Science Study (TIMSS) is conducted every four years and assesses students in Years 4 and 8. TIMSS has a curriculum focus with its main aim to provide nations with information so that they can monitor and evaluate mathematics and science teaching and learning. Australia has been involved in all TIMSS cycles since inception in 1995.

What does TIMSS assess?

TIMSS is organised around two dimensions: a content dimension, which specifies the domains or subject matter to be assessed in mathematics and science, and a cognitive dimension, which specifies the thinking processes and sets of behaviours required by students.

How is TIMSS administered?

In Australia, 287 primary schools and 285 secondary schools participated in TIMSS 2015, representing Australian Year 4 and 8 populations. At each school at least one intact class from the relevant year level — along with all Indigenous students in that year level — was selected to participate, resulting in a sample of 6057 Year 4 students and 10 338 Year 8 students.

How are TIMSS results reported?

The TIMSS 2015 mathematics and science results are represented as average scores, with a mean of 500 and a standard deviation of 100.

Changes in mean performance of students from one cycle of an assessment can provide evidence of improvement in the quality of schools and education systems. However, mean scores can mask significant variation within an individual class, school or education system. Information on internal disparities in performance is reported by examining the difference between the 5th and 95th percentiles. Countries are generally shown in decreasing order of achievement; however, this should not be interpreted as a simple ranking.

When comparing groups of students across and within countries, summary statistics such as mean scores are used, but this does not provide information as to what types of tasks the students were able to undertake successfully. Instead, TIMSS uses points on the scale as international benchmarks.

These four levels summarise the achievement reached:

- the Advanced international benchmark, which was set at 625
- the High international benchmark, which was set at 550
- the Intermediate international benchmark, which was set at 475
- the Low international benchmark, which was set at 400.

The descriptions of the levels are cumulative, so that a student who reached the High benchmark can typically demonstrate the knowledge and skills for both the Intermediate and the Low benchmarks as well.

[3] Descriptions from: Thomson, S, Wernert, N, O'Grady, E & Rodrigues, S 2016, *TIMSS 2015: A first look at Australia's results,* Australian Council for Educational Research Ltd, Camberwell.

Mullis, IVS & Martin, MO (Eds) 2013, *TIMSS 2015 Assessment Frameworks.* Retrieved from Boston College, TIMSS & PIRLS International Study Center, http://timssandpirls.bc.edu/timss2015/frameworks.html

NAP-SL[4]

What is NAP-SL?

NAP-SL is the National Assessment Program – Science Literacy and is one of three sample assessments conducted by ACARA on a triennial basis. One of its main objectives is to monitor trends in science literacy performance in Year 6 students over time. NAP-SL is the only sample assessment that focuses entirely on Year 6 students as the Education Council (formerly MCEECDYA), agreed to use the PISA as the national measure of performance for science literacy among secondary students. The other NAP sample programs are: civics and citizenship, and information and communication technologies (ICT) literacy and sample students at Year 6 and Year 10.

What does NAP-SL assess?

NAP-SL assesses students' ability to apply broad conceptual understandings of science to make sense of the world, to understand natural phenomena and to interpret media reports about scientific issues. Science literacy is defined using the definition from the original PISA framework:

The capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity. (OECD 1999, p. 60)

2015 represents a transition to the Australian Curriculum: Science, which will be complete for NAP-SL 2018.

The Australian Curriculum: Science has three interrelated strands, designed to be taught in an integrated way:

- Science understanding
- Science as a human endeavour
- Science inquiry skills.

Together, these three strands provide students with understanding, knowledge and skills through which they can develop a scientific view of the world. These three strands are then further divided into sub-strands, year levels and specific content descriptors.

How is NAP-SL administered?

In previous NAP-SL cycles, the assessment was conducted through printed booklets. Assessment included practical tasks where students conducted an investigation in groups and then responded individually to a set of items about the investigation. This provided opportunities for students to demonstrate their science inquiry skills in context.

In 2015, NAP-SL was delivered online making the assessment of inquiry skills in context no longer viable. Instead, students were presented with a simulated science investigation using video and other stimuli. Students answered items as they progressed through the stages of the investigation.

All states and jurisdictions participated in NAP-SL 2015 with about 600 schools and 12 000 students completing the assessment.

How are NAP-SL results reported?

The results of NAP-SL are reported as mean scores and distributions of scores across proficiency levels. They are also described in terms of the understandings and skills that students demonstrated in the assessment.

[4] Descriptions from:

ACARA, 2017, http://www.nap.edu.au/nap-sample-assessments/Science-literacy

ACARA 2017, *NAP Sample Assessment Science Literacy, 2015*, Australian Curriculum Assessment and Reporting Authority

Five levels of proficiency are defined and described for science literacy. Typically, students at a particular proficiency level can demonstrate the understandings and skills associated with that level as well as the understandings and skills of lower proficiency levels.

Initially, three proficiency levels, corresponding with Levels 2, 3 and 4 of the assessment domain, were identified. However, as 90% of students' scores fell within Level 3 in the 2003 assessment, three further proficiency levels within Level 3 were created, providing five levels for reporting student performance. Students who perform at or above Level 3.2 are considered proficient. For NAP-SL 2015, 55.1% of Australian students were proficient, up from 54.3% in 2006, but the difference is not statistically significant.