

Education Research Brief

Have Kids Stopped Trying at PISA and NAPLAN?

Trevor Cobbold

September 2018

Save Our Schools

<http://www.saveourschools.com.au>

<https://twitter.com/SOSAust>

saveourschools690@gmail.com

Summary

A much-ignored aspect of school results in Australia is the sharp contrast between significant improvements in Year 12 results compared to declining PISA results for Year 10 and stagnating Year 9 NAPLAN scores. How is it that trends in school outcomes only two or three Year levels apart are so different?

Many research studies show that students make less effort in low stakes tests such as PISA and it leads to lower results. Students tend to put in less effort in low stakes tests because they have no personal consequences. Students put more effort into Year 12 assessments because they affect their future careers and lives.

There is credible evidence indicating that Australian students do not try as hard on the PISA and NAPLAN tests as they do in Year 12 assessments.

A new study shows that nearly a quarter of Australian students do not treat the PISA test seriously. It may at least partially account for the contrasting trends between PISA and Year 12 results. There is some evidence of declining student effort in PISA. There was a significant decline in student engagement which may have led to lower motivation and effort.

Student motivation and effort should also be considered in assessing NAPLAN results. There is research evidence of low effort in NAPLAN by some students and it may contribute to the contrast between stagnating Year 9 NAPLAN results and improving Year 12 results.

The studies also call into question the worth of international and national league tables. Country rankings are affected by differing proportions of students in different countries who do not take tests seriously. The new study estimated that Australia's PISA ranking would move up several places with greater student effort.

Similarly, league tables based on NAPLAN results are likely to be affected by varying degrees of student effort in different schools. School rankings can move up and down depending on student effort.

The studies show that caution is needed in interpreting results from PISA and NAPLAN and drawing policy conclusions. The results could be as much a measure of student effort as a measure of student learning. More research is needed on student motivation and effort in PISA and NAPLAN.

One of the most puzzling features about student results in Australia over the past decade or more is the sharp contrast between declining scores on the OECD's Programme for International Student Assessments (PISA) and improvements in Year 12 results. Surprisingly, the contrasting outcomes have attracted little comment but deserve examination.

A new study published by the US National Bureau of Economic Research (NBER) points to low student motivation as a factor because a large proportion of Australian students do not take the PISA tests seriously. It complements evidence from the OECD and many overseas studies that low student effort in low stakes tests which leads to lower results.

Australia's reading, mathematics and science scores in PISA have declined significantly over the last 10-15 years and are amongst the largest declines in the OECD. The declines are equivalent to 6-12 months of learning. The PISA tests are for 15-year-old students, of whom 75% are in Year 10 in Australia. In addition, Year 9 NAPLAN results in reading, writing and numeracy have stagnated since the tests began in 2008 (reading and numeracy) and 2011 (writing).

In contrast, Year 12 results have improved significantly over the past 10-15 years. The percentage of the estimated Year 12 population that completed Year 12 increased from 68% in 2001 to 76% in 2016. In 2016, 90% cent of 20 to 24-year-olds had attained Year 12 or Certificate II, up from 79% in 2001, while 89% attained Year 12 or Certificate III compared to 77% in 2001. The percentage of the estimated Year 12 population achieving an ATAR score of 50 or above increased significantly from 25% in 2006 to 42% in 2015.

In addition, a larger proportion of disadvantaged students now complete Year 12. For example, the percentage of low socio-economic status (SES) students who completed Year 12 increased from 62% in 2001 to 73% in 2016 and the rate for remote area students increased from 66% to 78%. The Year 7/8 to Year 12 retention rate for Indigenous students increased from just under 40% in 2001 to 78% in 2016.

These are indicators of an improving education system, not a deteriorating one. They are difficult to reconcile with the decline in PISA results and the stagnating Year 9 NAPLAN results. How is it that the trends in school outcomes only two or three Year levels apart can be so different?

Of course, one possible explanation is that Year 12 standards have declined, but there does not appear to be any evidence that this is the case.

Save Our Schools has previously speculated that the contrasting results may partly reflect a difference in student attitudes to standardised tests which have no, or few personal consequences attached to them compared to the importance of Year 12 assessment [Cobbold 2017]. PISA is regarded as a low stakes test; students are not even told their individual PISA results. In contrast, the results of Year 12 assessments are much more important for students as they have a major influence on their future careers and lives.

There is some anecdotal evidence to indicate that low student motivation is a factor in Australia's PISA results. For example, one 15-year-old who participated in PISA 2015 said:

My peers who took part in this test were unanimous in that they did not, to the best of their ability, attempt the exam. [Jackson-Vaughan 2016]

The possibility of low student effort in NAPLAN and low stakes international tests such as PISA and Trends in International Mathematics and Science Study (TIMSS) cannot be ignored as it can affect

the results, how they are interpreted and how they are used. As one study noted, the “results could be a measure of motivation and effort as much as a measure of student knowledge” [Eklof et.al. 2014: 32].

Higher stakes for students are associated with NAPLAN than PISA and TIMSS in the sense that students may get to know their results through the reports to parents. Also, many schools make the tests a major event in the school year, including devoting considerable time to preparing for the tests, because school results are posted on the My School website and affect school reputations. Feedback on their NAPLAN results may make it more personally meaningful to students.

However, the personal stakes associated with NAPLAN are less than for many standardised tests overseas where students are required to pass before advancing to the next grade or graduating from school. It may be that the slightly higher personal stakes associated with NAPLAN compared to PISA is a factor in the difference between the declining results in PISA and the stagnation in Year 9 NAPLAN results.

There appears to be little research on student motivation in standardised tests in Australia. Data from PISA 2012 shows that the effort made in the test by low performing Australian students in mathematics was amongst the lowest of 64 participating countries/economies [OECD 2016: 113]. Their effort was ranked equal 9th lowest on the PISA “effort thermometer” (out of 30 rank scores for the participating countries/economies, that is, several countries had the same score) [online Table 3.7]. The difference in effort made by low performing students and those above baseline proficiency was the equal 5th largest out of 17 score differences for all participating countries/economies.

A study prepared for the Australian Curriculum, Assessment and Reporting Authority (ACARA) for its project to move NAPLAN online found that students with more negative motivation for mathematics were less motivated in the NAPLAN numeracy test and had lower results [Lifelong Achievement Group & Martin 2015]. An incentive program to improve outcomes in a low SES Queensland public high school in the 2016 NAPLAN tests elicited greater efforts by students and improved performance in most tests in Years 7 and 9 compared to a baseline test with no incentives [Sarkar et.al. 2018]. A study of the year 9 cohort of a large regional North Queensland high school found that students who saw themselves as unable or with poor ability to achieve in NAPLAN had little desire to perform well [Belcastro & Boon 2012].

Many overseas studies over the past 20 years have found that students make lower effort in low stakes tests and it leads to lower results [Finn 2015]. As one of the earlier studies stated:

Without consequences for performance, many students will not give their best effort to such “low-stakes” tests; as a result, their assessment test scores may not serve as valid indicators of what they know and can do.

....the research discussed in this article indicates that when low-stakes assessment tests are used, the underestimation of student proficiency can be substantial. All low-stakes assessment programs are vulnerable to this threat, [Wise & DeMars 2005: 1, 15]

The new NBER study and other recent studies provide additional evidence of this. The NBER study is one of very few that attempt to estimate how many students do not take low stakes tests seriously [Akyol et.al. 2018]. It is the first study to provide estimates for many countries.

The study shows that a relatively high proportion of Australian students do not take the PISA tests seriously. Using data from PISA 2015 supplied by the OECD, the researchers identified the proportion of non-serious students taking the science test as those who skipped questions, spent insufficient

time answering them, or quit the test early. It found that the proportion of non-serious students varies enormously by country from 13.6% in Korea to 67% in Brazil. It estimated 22.5% of Australian students as non-serious compared to 15-18% in high achieving countries such as Finland (15.8%), Japan (18.1%), Korea (13.6%), Taiwan (14.2%) and Singapore (17%).

High SES, low skilled and male students tended to take the test less seriously than other students. Countries in which students reported sitting for more “high stakes” exams had a higher proportion of non-serious students.

The study estimated the impact of low student effort by using a statistical technique to plug in missing answers based on how each student likely would have performed given their other answers and those of similarly skilled students. It found that the number of correct answers would increase.

Another study based on PISA results examined the effects of differences in student effort on differences in country scores in PISA 2009 [Zamarro et al. 2016; see also Borgonovi et.al. 2018]. Its measure of student effort was based on the rate of decline in performance over the course of the test, item non-response rates and a measure of careless answering patterns. It found that between 32 and 38% of the cross-country variation in PISA scores can be explained by different levels of student effort across countries. The gap between the highest and lowest performing countries in the sample was driven partly by student effort and the range of PISA scores between countries narrowed considerably after making adjustments for student effort.

An earlier study examined effort and mathematics performance in three (Norway, Sweden and Slovenia) of the ten countries participating in the TIMSS Advanced test in mathematics in 2008 for students in the last year of secondary school [Eklof et.al. 2014]. As with PISA, the stakes of the test are low for the participating students. Students do not get any feedback on their performance, individual test results are not returned and test results do not count toward student grades.

Only 35% of Swedish students, 43% of Slovenian students and 48% of Norwegian students reported that they did their best on the test. It found that the low level of reported effort was associated with a lower level of performance. This is all the more remarkable because it would be expected that students taking advanced courses in mathematics would be a select group of students and have high levels of motivation. The study concluded:

The above findings raise the question of whether the achievement test scores really are valid indicators of what the students in these samples knew and could do, and highlights the need to assess effort and motivation in low-stakes assessment contexts where results are used by stakeholders to make inferences about student knowledge. [Eklof et.al. 2014: 42]

A Swedish study found that where students perceive a test to be unimportant they put in less effort and achieved lower results [Knetka 2017]. It compared results from a field trial in Sweden’s national test in science for 9th grade students with those in the actual national test. Test-taking motivation was measured by a combined scale derived from student self-reporting on effort, expectancies, importance, interest, and test anxiety. The largest difference between the two test situations was in reported effort, importance, and test anxiety.

Another recent study published by the NBER compared the results of a mathematics test of US and Shanghai students who were offered financial incentives to increase their effort in the tests with those of students in each country who were not offered the incentives [Gneezy et al. 2017]. The performance among Shanghai students did not change while the scores of the US students increased substantially. With the incentives, US students attempted more questions (particularly towards the

end of the test) and were more likely to answer those questions correctly. The effects on test scores were concentrated among students whose baseline performance is near the US average.

Thus, the effort made by students in low stakes tests appears to be a significant factor influencing results. If nearly one-quarter of Australia's 15-year-old students do not take the PISA tests seriously, it is likely to account for at least some part of the difference with the Year 12 results.

There is little data available to determine whether student effort in PISA has declined over the years and contributed to the decline in PISA results since 2000. However, there is some indicative evidence that effort may have declined. A study using data from the PISA effort thermometer found that effort made by Australian students in reading declined from PISA 2000 to PISA 2003 [Butler & Adams 2007: 296]. In PISA 2003, Australia was ranked equal 5th in the OECD on the effort thermometer in the reading test and in PISA 2012 it was ranked equal 10th in the effort made by low performing students in mathematics and equal 9th in the effort made by other students [Butler & Adams 2007: 285; OECD 2016: online Table 3.7].

There may be indirect influences on student attitudes to low stakes tests over time. Student engagement at school amongst 15-year-olds has declined significantly since 2003 [De Bortoli 2018]. Between PISA 2003 and 2015, there was a large decline from 88 to 72% of students who felt they did not belong at school. There were declines across all demographic groups with the largest declines recorded for female, low SES, Indigenous and remote area students. This decline in student engagement at school may have led to lower motivation and effort in PISA and NAPLAN over time.

Some students may have become more resistant to putting effort into PISA and NAPLAN because of the pressure many schools put on students each year to perform in the NAPLAN tests. NAPLAN has become a big event in the school year and many schools devote considerable time and resources to spur students to achieve good results. There is a range of student experiences and responses to the tests some of which involves high levels of anxiety and stress [Howell 2017; Howell 2016; Rice et.al. 2016]. Also, teachers are in general highly critical of NAPLAN and a significant proportion of parents are opposed to the tests [Rogers et.al. 2018; Thompson & Harbaugh 2013; Thompson 2013; Dulfer et.al. 2012]. Such attitudes may have filtered through to an increasing proportion of students and influenced their approach to NAPLAN and PISA. The difference in trend between the declining PISA results and the stagnating NAPLAN results in Year 9 may be partly due to the slightly higher stakes attached to NAPLAN by students, parents and schools.

The new NBER study also shows that country rankings on the PISA league tables are affected by differing proportions of students in different countries who do not take the tests seriously. The paper estimates the effect of differential student effort on the cross-country rankings and found large changes in rankings for several countries if students took the test seriously. For example, the ranking for Portugal changed 15 places from 31st to 16th; Sweden's ranking changed 11 places from 33rd to 22nd and Norway's changed 9 places from 25th to equal 16th. Australia's ranking increased four places from 16th to 12th.

In contrast, there was little to no change for the high achieving countries. There was no change for Singapore and Taiwan. Japan's ranking increased by two places and that of Finland and Korea by three places.

However, if all countries were able to ensure that all students took the tests seriously there would be much less change in the rankings with some increasing and others falling. This is obvious, but it is unlikely that all countries can ensure that all students take the tests seriously and differences in student effort would remain.

Based on its findings, the study concludes:

Using PISA scores and rankings as done currently paints a distorted picture of where countries stand in both absolute and relative terms. [p. 23]

The earlier NBER study simulated the impact on US performance in PISA 2012 mathematics and found that the increased effort would improve the US results by 22-24 points, which was equivalent to moving the US from 36th to 19th in the 2012 rankings. [Gneezy et al. 2017]. A study based on data from PISA 2006 also found significant changes in the rankings of countries when motivational factors are taken out, with some countries moving up the league table while other countries move down in the rankings [Borghans & Schils 2013]. Australia moved from 13th on a combined reading, mathematics and science scale to 7th.

Given there is a very wide range in the proportion of students who exhibit low motivation between countries there is therefore considerable need for caution in interpreting country rankings on standardised tests such as PISA and TIMSS. They cannot be considered as an accurate measure of the comparative quality of different education systems as often interpreted.

...when comparing student performance across countries, the extent to which student performance on international tests might be influenced by the effort that students in different countries invest in the assessment must be considered. [OECD 2010: 36; also OECD 2013: 36]

...many low-stakes tests are used for accountability and comparison measures, underscoring the importance of considering motivation when interpreting low-stakes test performance. Low motivation poses a serious problem for determining the validity of the test, lowers the credibility of the assessment if it is seen as an invalid measure, and leaves institutions potentially drawing questionable conclusions about the status of their programs and the proficiency of their teaching faculty. [Finn 2015: 1]

As one study observed “...the conclusions drawn from them should be more modest than current practice” and “...policy reforms that ignore the role of intrinsic motivation to perform well on the test may be misguided and have unintended consequences” [Gneezy et.al. 2017: 21].

In conclusion, there is credible evidence indicating that Australian students do not try as hard on PISA and NAPLAN tests as they do in Year 12 assessments. There is strong evidence that low student effort on the PISA tests is a factor influencing Australia’s results and position in international league tables and part of the difference between its PISA results and Year 12 outcomes. Whether low student effort has increased over the years and contributed to the decline in Australia’s PISA results since 2000 is difficult to say; student effort may have declined but further research is needed to be definite about this. The important point from the new NBER study and other recent studies is that national and international test results are affected by the degree of student effort, among other factors, and therefore caution is needed in interpreting the results and drawing policy conclusions from them.

References

- Akyol, S. Pelin; Krishna, Kala & Wang, Jinwen (2018). Taking PISA seriously: How accurate are low-stakes exams? Working Paper No. 24930, National Bureau of Economic Research, Cambridge, MA, August.
- Belcastro, Lauren & Boon, Helen (2012). Student motivation for NAPLAN tests, *Australian and International Journal of Rural Education*, 22(2): 1-19.
- Borghans, Lex & Schils, Trudie (2012). The leaning tower of pisa: decomposing achievement test scores into cognitive and noncognitive components. Working paper, Labor Studies Group, National Bureau of Economic Research, Cambridge, MA.
- Borgonovi, Francesca; Hitt, Collin; Livingston, Jeffrey A., Sadoff, Sally & Gema Zamarro (2018). The effort students put into standardised tests varies widely by country, *Vox*, 16 January.
- Butler, Jayne & Adams, Raymond J. (2007). The impact of differential investment of student effort on the outcomes of international studies, *Journal of Applied Measurement*, 8(3): 279-304.
- Cobbold, Trevor (2017). What's behind Australia's tottering PISA results: A review of Australia's PISA 2015 results. Education Research Paper, Save Our Schools, February.
- De Bortoli, Lisa (2018). Sense of belonging at school, PISA Australia in Focus No.1, Australian Council for Educational Research, Camberwell, May.
- Dulfer, N., Polesel, J., & Rice, S. (2012). The experience of education: The impacts of high stakes testing on school students and their families: An educator's perspective. The Whitlam Institute, University of Western Sydney.
- Eklof, Hanna; Pavesic, Barbara Japelj & Gronmo, Liv Sissel (2014). A cross-national comparison of reported effort and mathematics performance in TIMSS Advanced, *Applied Measurement in Education*, 27:1, 31-45
- Finn, Bridgid (2015). Measuring motivation in low-stakes assessments, *ETS Research Report*, No. RR-15-19: 1-17.
- Gneezy, Uri; List, John A; Livingston, Jeffrey A.; Sadoff, Sally; Qin, Xiangdong & Xu, Yang, Jinwen (2017). Measuring success in education: the role of effort on the test itself. Working Paper No. 24004, National Bureau of Economic Research, Cambridge, MA, November.
- Howell, Angelique (2017). 'Because then you could never ever get a job!': children's constructions of NAPLAN as high-stakes, *Journal of Education Policy*, 32(5): 564-587.
- Howell, Angelique (2016), Exploring children's lived experiences of NAPLAN. In Bob Lingard, Greg Thompson & Sam Sellar (eds.) *National Testing in Schools: An Australian assessment*, Routledge, Abingdon, 164-180.
- Jackson-Vaughan, Paloma (2016). Why 15-year-olds don't care about PISA rankings, *Sydney Morning Herald*, 7 December.

Knetka, Eva (2017). Are all pupils equally motivated to do their best on all tests? Differences in reported test-taking motivation within and between tests with different stakes, *Scandinavian Journal of Education Research*, 61(1): 95-111.

Lifelong Achievement Group & Martin, Andrew J. (2015). Online NAPLAN Testing and Student Motivation: Exploring Adaptive and Fixed Test Formats. Report prepared for the Australian Curriculum, Assessment and Reporting Authority, July.

OECD (2010). *PISA 2009 Results: What Students Know and Can Do – Student Performance in Reading, Mathematics and Science (Volume I)*, PISA, OECD Publishing, Paris.

OECD (2013). *PISA 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science (Volume I)*, PISA, OECD Publishing, Paris.

OECD (2016). *Low-Performing Students: Why They Fall Behind and How to Help Them Succeed*, PISA, OECD Publishing, Paris.

Rice, Suzanne; Dulfer, Nicky; Polesel, John & O’Hanlon, Clare (2016). NAPLAN and student wellbeing: teacher perceptions of the impact of NAPLAN on students. In Bob Lingard, Greg Thompson & Sam Sellar (eds.) *National Testing in Schools: An Australian assessment*, Routledge, Abingdon, 72-85.

Rogers, S.; Barblett, L. & Robinson, K. (2018). Parent and teacher perceptions of NAPLAN in a sample of Independent schools in Western Australia, *Australian Educational Researcher*, 45:493–513.

Sarkar, Jayanta; Dulleck, Uwe; Kocher, Martin & Sarkar, Dipanwita (2018). Do incentives improve test scores? New evidence from a field experiment. Paper presented to the International Conference in Public Economics, Aix-en-Provence, 25-26 June.

Thompson, G. (2013). NAPLAN, MySchool and accountability: Teacher perceptions of the effects of testing, *The International Education Journal: Comparative Perspectives*, 12(2), 62–84.

Thompson, G. & Harbaugh, A. G. (2013). A preliminary analysis of teacher perceptions of the effects of NAPLAN on pedagogy and curriculum. *Australian Educational Researcher*, 40, 299–314

Wise, Steven L., & DeMars, Christine. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, 10(1): 1–17

Zamarro, Gema; Hitt, Collin and Mendez, I. (2016). When students don’t care: Reexamining international differences in achievement and non-cognitive skills. EDRE Working Paper 2016-18, University of Arkansas.