Learning from international studies: PISA, TIMSS and PIRLS from an Australian research perspective

(シンポジウム開催日: 2010年12月10日)

Dr. Sue Thomson [Australian Council for Educational Research]

Education

1998, Doctor of Philosophy, Monash University
1995, Master of Educational Studies, Monash University
1993, Graduate Diploma in Mathematics Education, Deakin University
1983, Diploma in Education (Mathematics, Computer Science), Monash University
1982, B.App. Sci (Mathematics), Royal Melbourne Institute of Technology

Experience

2011 - Director, Educational Monitoring and Research
Australian Council for Educational Research

2005 – 2010. Principal Research Fellow
Australian Council for Educational Research

2002 – 2004. Senior Research Fellow
Australian Council for Educational Research

Australian Council for Educational Research

1998, Statistical and Research Consultant
Department of Community Medicine and General Practice
Monash University
Lecturer (Educational Statistics)
Faculty of Education, Monash University
Project Officer
Council of Australian Postgraduate Associations

1997 – 1998. Lecturer (Statistical Methods)
Swinburne University of Technology

1996 – 1997. Lecturer (Statistical Methods)
Department of Psychology, Australian Catholic University

Monash University

1993 Lecturer (Mathematics)
Swinburne University of Technology

Victorian Department of Education

Selected Publications


Dr. Sue Thomson is the Head of Educational Monitoring and Research and the Director of the National Surveys research program at the Australian Council for Educational Research.

Dr. Thomson is also the National Research Coordinator for Australia in the Trends in International Mathematics and Science Study (TIMSS), which measures achievement in mathematics and science for students in grades 4 and 8, and the Progress in International Reading Literacy Study (PIRLS), which measures reading literacy of grade 4 students, and is the National Project Manager for Australia for the OECD Programme for International Student Assessment (PISA), which examines reading, mathematical and scientific literacy of 15-year-old students.
Introduction

For Australia, participation in international studies enables policy makers and educators to compare the performance of Australian students with students in countries which may be our near neighbours, our trading partners or our competitors. Domestically, international studies provide information on how well Australian schools are providing for the educational needs of various sub-groups such as students with language backgrounds other than English, Indigenous students, geographically remote students and students from different socio-economic backgrounds. International studies also provide information on the strengths and weaknesses of Australian students in particular subject domains such as reading, mathematics and science, and more generic areas such as problem solving. This information is extremely valuable in identifying areas that require further attention in school curriculum or classrooms.

Background

Australia has a federal system of government, with the states and territories each responsible for their own educational administrations, albeit with similar overall structure. Collaboration on matters of policy takes place in the Ministerial Council of Education, Early Childhood Development and Youth Affairs (MCEECDYA), which is comprised of education ministers from all jurisdictions and the Commonwealth.
In Australia, PISA, TIMSS and PIRLS are managed by staff in the National Surveys Program of the Australian Council for Educational Research (ACER), on behalf of the Commonwealth and State governments. ACER plays a dual role in PISA, as it also the lead partner in the consortium that manages PISA internationally on behalf of the OECD. ACER is an independent research organisation with more than 300 staff working in offices across Australia, as well as in Dubai and Delhi.

The importance of strong literacy and numeracy skills in the population cannot be understated. Student achievement in school is a key determinant of later individual success in education and work. Recent research has also suggested that improving levels of literacy and numeracy in the workforce is associated with better economic outcomes for the nation as a whole. These skills are fundamental for improvements in productivity, innovation and competitiveness, and it is important that we are able to monitor and benchmark the effectiveness of our schools in developing these capabilities in our children not only across Australia but also internationally.

As a basis for reporting ongoing progress towards the goals, by drawing on agreed definitions of Key Performance Measures, MCEECDYA developed a Measurement Framework for National Key Performance Measures, which sets out the National Assessment Program.

In the *Melbourne Declaration on Educational Goals for Young Australians*, MCEECDYA commits “to supporting all young Australians to become successful learners, confident and creative individuals, and active and informed citizens”, and to promoting equity and excellence in education.
The National Assessment Program encompasses tests such as the national literacy and numeracy tests (National Assessment Program – Literacy and Numeracy, NAPLAN), three-yearly sample assessments in science literacy, civics and citizenship, and ICT literacy, and Australia’s participation in the international assessments Programme for International Student Assessment (PISA), managed by the OECD, and the Trends in International Mathematics and Science Study (TIMSS), managed by the International Association for the Evaluation of Educational Achievement (IEA). Australia has participated in international tests of mathematics and science since the First International Mathematics Study in 1964, and has participated in all cycles of TIMSS (since it was the Third International Mathematics and Science Study in 1985) and all cycles of PISA. In 2010, Australia participated for the first time in the IEA Progress in International Reading Literacy Study (PIRLS).

The samples drawn for each of the studies included as part of the National Assessment Program reflect the purposes of the testing and reporting. NAPLAN tests full cohorts at Year 3, Year 5, Year 7 and Year 9, and reports student achievement based on a common test to students, parents, teachers and schools. TIMSS tests a sample of students in a sample of schools at Year 4 and Year 8, PIRLS a sample of students in a sample of schools at Year 4, and PISA a sample of 15 year old students in a sample of schools. The difference in the samples between the national and international assessments reflects the differing purposes of the assessments – NAPLAN assesses the achievement and progress of individual students, PISA, TIMSS and PIRLS measure the overall performance of the educational system in a number of areas at a number of levels and in different ways. Accurate results from the international studies are not routinely provided for students or schools for reasons that will become clear.
The first stage in drawing samples for PISA, TIMSS and PIRLS is defining the sample. To be acceptable internationally, national sample designs must result in probability samples that give accurate weighted estimates of population parameters such as means and percentages, and for which estimates of sampling variance could be computed. Sampling procedures are strictly monitored by the respective expert bodies – WESTAT for PISA and Statistics Canada and the IEA Data Processing Center for the IEA studies. The basic sample design is defined for each study, and while countries were able to adjust the sample design and increase the complexity to suit their own situations, all adaptations have to be approved by the sampling experts.

Basic across studies is that the selection of schools and students is kept as inclusive as possible, so that the sample of students comes from a broad range of backgrounds and abilities. In each country a small number of exclusions is allowed, for example for very small schools in geographically remote areas or for students with intellectual or severe functional disabilities or limited proficiency in English (or the mother tongue). In Australia, a small number of students are excluded from other populations such as prison populations and students from correspondence schools. In all cases, the detail of schools and students excluded is the subject of scrutiny and only a small percentage is allowed. Students are not able to be excluded for behavioural reasons, nor simply because they are of lower ability.

The first stage of sampling in all studies consists of a sample of schools, which may be stratified. Stratifying the sample ensures that not only the overall population, but also key subgroups of the population, will be represented in the appropriate proportions and so able to be reported effectively. In Australia the TIMSS, PIRLS and PISA samples are stratified by state, by type of school (government, Catholic or independent), and by geographic region (metropolitan, regional, rural and remote). Selection is probability proportional to size within each strata.

Australia takes a much larger sample for PISA, in order to derive more accurate measures for the smaller states, and also because the PISA sample becomes the beginning cohort for the next wave of the Longitudinal Surveys of Australian Youth. Students participating in the Longitudinal Surveys of Australian Youth are contacted in each subsequent year by telephone survey until they are 25 years of age, and have, in most cases, made the transition from education to work. Many reports
examining the school to work or further study have been produced.

**Value of participation**

PISA and TIMSS are highly regarded international studies of student achievement. By participating in these studies Australia becomes part of the international education community that, through research, aims to improve student outcomes globally. Student results from PISA and TIMSS, and the analysis of these results to identify factors impacting on the outcomes, are reported in publications that are distributed internationally and discussed widely by people with an interest in school education.

---

**Value of participation**

- enables policymakers and educators to compare the performance of Australian students with students in countries which may be our near neighbours, our trading partners or our competitors.
- provides information on how well Australian schools are providing for the educational needs of various subgroups such as:
  - students with language backgrounds other than English,
  - Aboriginal and Torres Strait Island students,
  - geographically remote students and
  - students from different socio-economic backgrounds.

---

**The assessments**

In order to cover the entire PISA, PIRLS or TIMSS frameworks, the pool of items developed is extensive, and in each case would require a great deal more testing time than is possible. To accommodate this, all studies use a matrix-sampling technique that involves dividing the entire assessment pool into a set of unique item blocks, distributing these blocks across a set of booklets, and rotating the booklets among the students. This is the reason that students and schools are not provided with scores in TIMSS, PIRLS or PISA – each student only sits a portion of the entire testing materials and even over a whole school, not all items are completed by all students. Item Response Theory (IRT) scaling is used to compute scores across the entire item pool. IRT also allows scaling to be conducted so that comparisons can be made across time, which is one of the main benefits of participating in the international studies.

**Who participates?**

A growing number of countries participate in international studies. Sixty-seven countries or economies participated in PISA 2009, including all 32 OECD member participants and 35 partner economy participants. A further 9 countries/economies will complete PISA 2009+, the same assessment carried out a year later. While it is impossible to characterise the non-OECD or partner countries that participate, they range across most continents other than Africa. There are a number of developing countries participating in PISA, however there are problems with the sample populations when a large proportion of 15-year-olds have already left school and thus the sample is skewed. Eighty countries or economies will or have participated in TIMSS 2011, including some African countries (Botswana, Ghana, South Africa). It is more appropriate for many developing countries to participate in TIMSS and PIRLS because of the grade levels at which they are conducted.

The goal of participating in educational studies is to find out how well your countries’ education system is performing. For developing countries the aim of participating is to monitor change and progress.
Results from PISA 2009

So, what have we learned about Australian students' performance in reading from PISA 2009? Australian students achieved at a level significantly higher than the OECD average, and our scores were similar to those of students in New Zealand, Japan and the Netherlands. Students in these four countries scored at a level significantly lower than the countries shown in slide 14, namely Shanghai, Korea, Finland, Hong-Kong, Singapore and Canada. Of course there are some issues with comparability when making comparisons between an entire country such as Australia and an economy such as Shanghai in which the full range of schooling alternatives is not necessarily present. In PISA these are referred to as 'partner economies'.

Scaling the reading literacy tasks

The reading literacy scale was constructed using Item Response Theory, with reading literacy items ranked by difficulty and linked to student proficiency. Using such methods means that the relative ability of students taking a particular test can be estimated by considering the proportion of test items they answer correctly, while the relative difficulty of items in a test can be estimated by considering the proportion of students getting each item correct. On this scale, it is possible to estimate the location of individual students, and to describe the degree of reading literacy that they possess. The inclusion of a broader range of reading literacy tasks in PISA 2009 has enabled the reading literacy proficiency scale to be expanded from the five levels identified in 2000, to seven levels. The new levels describe the reading literacy skills at each end of the proficiency scale – those students with very high or very low reading proficiency. A new level (Level 6) located above Level 5 describes the reading literacy skills of students with very high levels of reading proficiency. At the other end of the proficiency scale, Level 1 has been relabelled as Level 1a and a new level (Level 1b) has been introduced to describe the skills of those students who previously were simply described as not having achieved Level 1. The unbounded level below Level 1b does not provide a description about these students as there is an insufficient number of items on which to base a description of these students’ reading proficiency. The other Levels (2, 3, 4 and 5) remain the same in PISA 2009 as in 2000. Expanding the reading literacy proficiency scale in PISA 2009 allows for almost all PISA students to be accurately described.

As has been the case in previous PISA cycles, Level 2 has been defined internationally as a “baseline” proficiency level. This level does not separate reading literacy and illiteracy; rather it defines the level of

---

1 For the purposes of this discussion, not all countries’ results are reported. Readers should refer to the PISA International Reports for the full list of achievement results.
achievement on the PISA scale at which students begin to demonstrate the reading literacy competencies that will enable them to actively participate in life situations. Students performing below this baseline are considered to be at serious risk of not achieving at levels sufficient to allow them to adequately participate in the 21st century workforce and contribute as productive citizens.

Slide 15 shows the proportions of students achieving at each of the proficiency levels in reading literacy. In Shanghai, around two per cent of students achieved the highest proficiency level, a similar proportion to Finland, Canada, Singapore, Japan, New Zealand and Australia. However around 19 per cent of students in Shanghai achieved in the highest two levels, compared with 16 per cent in Singapore, and 13 per cent in Australia and Japan. More telling is that just four per cent of students in Shanghai failed to achieve the minimum proficiency level 2, compared with 13 per cent of students in Japan and 14 per cent of Australian students. A focus of education in Australia over the past several years has been to lift the achievement of students at the lower levels of achievement – this provides some indication that this has not been entirely successful.

To begin to identify where particular problems might lay, the data were then examined by gender. As is clear in Slide 17, boys do not achieve at the same levels as girls in reading literacy. Whilst the data are comparable to the OECD averages for males and females, the discrepancies are large and notable. In Australia, 10 per cent of males but 16 per cent of females achieve at least proficiency level 5, however 20 per cent of boys compared to just nine per cent of girls failed to achieve proficiency level 2. These data identify for Australia that a particular target of programs to improve overall reading must be boys.

Slide 16

As this is now the fourth administration of PISA, we can begin to look at trends. For Australia in reading literacy the trends are not promising. As seen in Slide 17, Australia’s scores in reading literacy declined significantly from PISA 2000 to PISA 2006, and have not improved significantly since then.
A more worrying trend, which can be seen in Slide 18, is that the decline in reading achievement is a combination of more boys failing to achieve at the lower levels and fewer students of either sex achieving at the higher achievement levels.

One of the important considerations for the Australian Commonwealth and State governments is the principles of social justice. Part of the Melbourne Declaration states that an overarching aim of education in Australia is to improve outcomes for Indigenous youth, disadvantaged students and those from a low socioeconomic background. We are able to use PISA data to provide national outcomes on these measures.

Slide 20 shows the gap in achievement scores between Indigenous and non-Indigenous students in PISA 2000 and PISA 2009. Using these data it is clear that little progress has been made over nine years, with Indigenous students still achieving at a level that is about two and a half years lower than that of non-Indigenous students in Australia.
Further analysis found that this is compounded by low socioeconomic background, with almost half of the Indigenous students in the sample in the lowest quartile of socioeconomic background as defined for the overall Australian population.

Examining the relationship between socioeconomic background and achievement on PISA, a very similar pattern can be seen. Students with lower levels of socioeconomic background generally perform at a level about two and a half years lower than students from high levels of socioeconomic background. How much of the low achievement of Indigenous students is a result of low socioeconomic background is a question to be explored further.

Slide 23 further examines the strength of the relationship between socioeconomic background and achievement in Australia. This scatterplot shows that while socioeconomic background is not deterministic, in that there are some instances in which students from low-level socioeconomic backgrounds achieve high scores and students from high-level socioeconomic backgrounds achieve at quite high levels, for the bulk of students there is a direct positive linear relationship between socioeconomic background and achievement in PISA.

Using linear regression we can see the levels of equity in countries participating in PISA. The proportion of variance in student achievement explained by the socioeconomic measure in PISA, ESCS (the students’ educational, social and cultural status) is plotted against the mean score in reading literacy for the country. This is then compared to the averages for the OECD.

Countries can be thought of as high equity if the proportion of variance explained by ESCS is significantly lower than that for the OECD on average, and high quality if their average score on reading literacy is significantly higher than that of the OECD on average.

As can be seen in Slide 24, both Japan and Australia fall into the high-quality, high-equity quadrant of the graph,
however the level of equity for Australia is not significantly different to the OECD average, so Australia is high-quality but only average equity. We have some way to progress to achieve what we want to achieve given the statements made by our Ministers of Education.

The graphs plotted on slide 25 show the ESCS measure of social background plotted against reading literacy for a selection of countries, to further investigate the relationship between social background and reading literacy. These graphs show the general tendency is most countries for there to be a linear relationship between social advantage and achievement. In Hong Kong, the pattern is slightly different, with a decreasing payoff for increasing social advantage. In Australia, at low levels of social advantage, the difference in scores between Australian students and the OECD average is about 1/3 of a school year. Social advantage brings with it many benefits in Australia, with the achievement of those students at high levels of socioeconomic background having an advantage of about 2/3 of a school year over the OECD average. The aim for a country is to have a line as high and flat as possible, so that social advantage does not define students’ scores. On this measure of social equity, Australia (and Japan) could be placed in the high-quality, low-equity quadrant.

I mentioned earlier that the PISA sample becomes the beginning cohort for the Longitudinal Surveys of Australian Youth. As an example of research utilizing these longitudinal data, a recent publication examines the links between success and outcomes in later life. In this analysis, the authors wanted to investigate whether the argument that students who did not achieve at proficiency level 2 would find it difficult to achieve success in their post-school outcomes was investigated. “Success” in this instance was defined as being occupied full-time (either with study, work or a combination of these) and being happy with their lives. Students’ pathways were reported, along with an analysis of the likelihood of the student achieving a successful outcome.