MEASURE FOR MEASURE

A REVIEW OF OUTCOMES OF SCHOOL EDUCATION IN AUSTRALIA

John Ainley
Eveline Gebhardt

Australian Council for Educational Research

ACER

MEASURE FOR MEASURE

A REVIEW OF OUTCOMES OF SCHOOL EDUCATION IN AUSTRALIA

John Ainley
Eveline Gebhardt

ACER

First published 2013
Australian Council for Educational Research Ltd
19 Prospect Hill Road, Camberwell
Victoria, 3124, Australia

www.acer.edu.au

Cover and text design by ACER Creative Services
Printed in Australia by Printgraphics

A record of this publication is available from the National Library of Australia

ISBN: 978-1-74286-224-8

CONTENTS

TABLES

FIGURES

PREFACE

Over the past two decades there has been an increase in the attention given to education policy, practice and improvement. Stories about education abound in news media and other forums to a greater extent than ever seemed possible in previous times. Some of this quickened interest in education has been driven by economic considerations but much has been motivated by an understanding that people need to have a firm grounding in appropriate knowledge, understandings and skills as well as a disposition to continue learning as they grow in order to live as effective and productive citizens in the twenty-first century.

Along with a burgeoning interest in, and rising expectations of, education policy and practice comes a desire to identify and understand best practice at various levels. Such endeavours depend on the accumulation of reliable evidence from multiple sources. Understanding the variations in learning outcomes that exist among systems, schools and students, as well as the factors associated with those variations, can provide a starting point for these endeavours. The information derived from analysing variations in outcomes can guide efforts to ensure that social gradients are minimized and gaps between groups of students are reduced.

In addition it is important to monitor change so as to inform judgements about whether outcomes are improving, to what extent they are improving and whether there is improvement for all students. Judgements about improvement depend upon data that are comparable over time and expert analyses that can identify real change amongst the fluctuations that exist in data gathered over time. In addition, perspectives on change that can be related to changes in policies, practices and contexts provide a stronger basis for understanding than those derived from cross-sectional analyses at any given time.

Large-scale assessment surveys depend on high-level psychometric and technical expertise. That expertise needs to be maintained and developed. It is an area of comparative strength in Australia that has been built on the long involvement of the Australian Council for Educational Research in international achievement studies starting with the first international mathematics study in the 1960s. While research methods have evolved over that period of time, being part of successive international studies has ensured that Australia has been at the forefront of the development of modern methods of design and analysis in this field. Those methods have found application in national, as well as international, achievement surveys.

Governments are increasingly committed to making publicly available as much information about learning outcomes as possible. International achievement studies have a strong tradition of producing comprehensive international and national reports. In addition the organisations conducting international achievement studies make data available for secondary analyses by other scholars. This means that there is expert scrutiny of methods, analyses and conclusions. It is an important aspect of transparency that depends on the independence of those conducting the studies from those managing education systems.

Over past years the reports of national and international achievement studies have provided much valuable information about the achievements of students in Australian schools, differences among jurisdictions and groups of students and differences between the achievement of Australian students and their peers in other countries. The emergence of a national commitment to school improvement makes it timely to provide an integrated appraisal of the results of these large-scale assessment studies.

Judgements about the impact of policies and practices on student outcomes need to be informed by reliable evidence about the ways in which achievement varies with differences in policy and practice and about the extent to which achievement changes over time. This report focuses on changes in achievement over time. It notes that there has been a small decline in reading and mathematics achievement among students in the middle years of secondary school since 2000, stability in science and mathematics achievement among Year 8 students since 1994, a small improvement in mathematics achievement among students in Year 4 since 1994 and a small improvement in reading achievement among students in Year 3 since 2008. The results from future successive assessment cycles in the programs reviewed in this report will be best appraised in relation to the previous trends so that any changes can be related to policy, practice and context.

INTRODUCTION 1

Australia has a history of international and national surveys of educational achievement. About 50 years ago Australia participated in the first international mathematics study (it was conducted among 13-year-olds with data being collected in 1964) (Husen, 1967). It has continued that involvement in international achievement studies to the present day. In 1975 Australia conducted a national study of literacy and numeracy achievement among 10-year-olds and 14-year-olds (Keeves & Bourke, 1976) which was repeated in 1980 (Bourke, Mills, Stanyon, & Holzer, 1981) and followed by a national school English literacy survey among students in Year 3 and Year 5 in 1996 (Masters & Forster, 1997a). The National Assessment Program (NAP) continues the national monitoring of achievement through its annual assessment of literacy and numeracy (of students in Years 3, 5, 7 and 9) since 2008 and a series of triennial sample studies since 2003 in science (2003), civics and citizenship (2004), and information and communication technology literacy (2005).

In addition to national and international surveys of educational achievement there was a number of jurisdictionally-based studies of achievement in literacy and numeracy (as well as some other aspects of learning), beginning with a sample-based survey in Victoria in 1988 (McGaw et al., 1989) and a population-based survey of students in Year 3 and Year 6 in New South Wales in 1989 (Masters et al., 1990) followed by similar programs in other jurisdictions through the 1990s.

PURPOSE

The reports of these studies have provided a plethora of information about the achievements of students in Australian schools and how those achievements differ among jurisdictions and among groups of students. However, each report has often been viewed in isolation from other similar studies. This report is intended to provide an integrated appraisal of the results of the international and national achievement surveys conducted since 1994 but with some references to earlier studies. In addition to limiting the task to manageable proportions, this time span corresponds to the widespread introduction of modern measurement methods so that there is a better basis for the comparison of results across studies and over time. The report examines differences in achievement among groups of students in Australia (including students in different jurisdictions), comparisons of the achievements of students in Australia with their peers in other countries and, where possible, changes in achievement over time for Australian students overall and for groups of students. It is predicated on the assumption that perspectives on the impact of policies and practices on student outcomes can be informed by evidence about the ways in which achievement co-varies with differences in policy and practice and about the extent to which achievement changes over time.

APPROACH

This report is based principally on data that have been published in reports of the studies being considered. Where necessary, standard analytical procedures have been applied to those published data following the principles of meta-analysis. These procedures provide the basis for comparing results from different studies and for determining whether differences are statistically significant (i.e. can be considered to have not arisen by chance). In some cases secondary analyses of original data, where those data are available, have been conducted.

DATA

Table 1.1 contains a summary of the major data sources for this report. These are all large-scale assessment surveys. They focused on a range of assessment domains (but most commonly reading, mathematics and science) and students from different year levels or ages, and were conducted at different times (but with several being linked through time-series data).

The Programme for International Student Assessment (PISA) covers reading, mathematics and science (with some other areas included at various times) among 15-year-olds every three years since 2000. The Trends in International Mathematics and Science Study (TIMSS) have been conducted every four years among students in Year 4 and Year 8 since 1994/5.[1] The IEA Progress in International Reading Literacy Study (PIRLS) has been conducted every five years since 2001 but Australia only began its participation in 2010. The National Assessment Program – Literacy and Numeracy (NAPLAN) has been conducted with the full cohort of students in Years 3, 5, 7 and 9 each year since 2008. This report focuses on data for reading and numeracy.

The report also makes use of other achievement surveys. These include the NAP sample studies conducted every three years since 2003 in science literacy (Year 6), since 2004 in civics and citizenship (Years 6 and 10) and since 2005 in ICT literacy (Years 6 and 10). It also references the IEA Civic Education Study (CIVED) conducted in 1999.

KEY ASPECTS OF THE STUDIES

Psychometric methods

All the projects listed in Table 1.1 use item response theory as a basis for the psychometric analyses of assessment data. Most use the one-parameter version of item response theory known as the Rasch model but the TIMSS and PIRLS projects use the three-parameter version of item response theory. These methods provide a basis for reporting student performance and item difficulty on the same scale thereby facilitating the descriptions of competencies associated with different levels of measured performance. All of the studies report achievement using scale scores and the percentage distribution of scores across performance bands. All of these studies make use of plausible values to generate unbiased population estimates.

1 In 1994 students were assessed in Years adjacent to Year 4 and Year 8 but the focus is on these designated Years.

Table 1.1 Data sources used in this report

Project	Assessment Domain			Population	Year of data collection
	Reading	Mathematics	Other Fields		
PISA	Reading	Mathematics	Science, Digital reading	15-year-olds	2000, 2003, 2006, 2009
TIMSS*		Mathematics	Science	Year 4, Year 8	1994, 1998, 2002, 2006, 2010
PIRLS	Reading			Year 4	2010
IEA CIVED			Civics	Year 9	1999
NAPLAN	Reading	Numeracy	Writing	Year 3, Year 5, Year 7, Year 9	2008, 2009, 2010, 2011, 2012
NAP-SL			Science	Year 6	2003, 2006, 2009, 2012
NAP-CC			Civics and citizenship	Year 6, Year 10	2004, 2007, 2010
NAP-ICTL			ICT Literacy	Year 6, Year 10	2005, 2008, 2011

Notes:
* In this report, the international study is referred to as TIMSS 2011 (because that is how it is referred to in the literature) but the Australian data is referred to as TIMSS 2010 because that is the year of data collection in Australia (and correspondingly for other cycles). See 'Timing of TIMSS and PIRLS' later in this chapter for further explanation.

PISA: Programme for International Student Assessment
TIMSS: Trends in International Mathematics and Science Study
PIRLS: Progress in International Reading Literacy Study
CIVED: IEA Civic Education Study
NAPLAN: National Assessment Program – Literacy and Numeracy
NAP-SL: National Assessment Program: Science Literacy
NAP-CC: National Assessment Program: Civics and Citizenship
NAP-ICTL: National Assessment Program: ICT Literacy

Test design

All the studies in Table 1.1 except NAPLAN make use of rotated block designs in constructing the assessments. This means that items are organised in blocks, blocks are assigned to booklets and booklets are then randomly assigned to students. This design enables the assessment to cover (or assess) more material than any individual student could possibly complete. For example, PISA 2009 included approximately 450 minutes of test material (covering 270 minutes of reading, 90 minutes of mathematics and 90 minutes of science) but each student was assigned one of 13 booklets and spent 120 minutes completing the assessment. In TIMSS 2006/7 there were 429 items (215 in mathematics and 214 in science) for Year 8 which represented approximately 630 minutes of assessment time. These items were arranged in 14 blocks of mathematics items and 14 blocks of science items, with the blocks being assembled in 14 booklets. Each booklet contained two maths and two science items. Each student was randomly assigned one booklet which took 90 minutes to complete.

Equating

The studies listed in Table 1.1 equate scores over time so that scores in any given cycle are represented on the original scale. For most of the large-scale assessments in this table this is achieved through common-item equating in which a number of items in any cycle are kept secure and included in the next cycle. The common items that behave in a consistent manner in both cycles are then used as link items to equate scores to the original scale. For example, PISA 2009 Reading contained 37 of the total 131 items that had been included in previous cycles. TIMSS 2006/7 included 189 items that had been used in previous cycles of TIMSS in its total of 429 items. NAPLAN equates scores through common-person equating in which samples of students complete equating tests together with the current NAPLAN tests.

The assessments conducted as part of the NAP (including NAPLAN) equate scores over Year levels using common-item methods so that any given score denotes the same achievement level regardless of the Year level of the student. In practice this means, for example, that some items are common between the adjacent levels that are tested. In other words, in NAPLAN the Year 5 assessment contains some items that also appear on the Year 3 assessment and some items that also appear on the Year 7 assessment.

Proficiency levels, proficiency bands or benchmarks

PISA and NAPLAN all report results in terms of the percentages of students in score ranges called variously proficiency levels or proficiency bands. The methods used to do this are broadly similar. The studies also each define points on the achievement scales that represent what is considered to be a 'satisfactory' performance for a student although the terms used, the meanings associated with those terms and the cut-points on the distributions of scores differ greatly across the studies.

Each of the studies uses item response theory to establish a scale on which items are ordered according to their relative difficulty (and the same scale represents the distribution of student performance). The ranges of difficulties are then divided into a set of levels each covering an equal range of difficulty. This involves consideration of an appropriate range of difficulty (or width) and the probability of a student in the middle of the range successfully answering the easiest items in the level.[2] Item descriptions (or descriptors) are written that describe content and processes that are assessed by each item. Summary descriptions of the levels or bands are then generated from syntheses of the constituent item descriptions. Standards, or the performance considered to be satisfactory, are then established through consultation with panels of experts considering each item near the boundary and judging whether or not it represented a satisfactory performance: a process referred to as an empirical judgement technique (McLarty, Way, Porter, Beimers & Miles, 2013).

PISA 2009 reported reading achievement in terms of seven proficiency levels.[3] Level 2 has been defined as 'a baseline level of proficiency at which students begin to demonstrate the reading literacy competencies that will enable them to participate effectively and productively in life' (OECD, 2009: 52).

In NAPLAN achievement is represented in terms of the location of a score in a range called a proficiency band. Ten proficiency bands along each scale are defined using an approach similar to that developed for PISA. The proficiency bands are based on the difficulties of the items and encompass a range of difficulties so that all bands are of equal width. This ensures that the notion of being at a level can be interpreted consistently and in line with the fact that the achievement scale is a continuum. It takes account of the expected success of a student of a given achievement level on items at that level and the probability that a student in the middle of a level would correctly answer an item of average difficulty for that level. The 10 proficiency bands encompass the full range of the scale. For each year level, six of these bands encompass most of the range of student achievement. And for each year level, a national minimum standard is defined and located on the NAPLAN scale. For Year 3, Band 2 is the national minimum standard; for Year 5, Band 4; Year 7, Band 5; and Year 9, Band 6.

In NAP sample studies proficiency levels are established at equally-spaced intervals across the relevant scale following similar procedures to those adopted in PISA and NAPLAN. Each level description

2 In PISA the response probability for the analysis of data is set at p = 0.62 and the width of the proficiency levels is set at 1.25 logits.

3 In 2000 there were six levels but in 2009 the bottom level was subdivided into two levels. The seven levels were labelled as 1b (scores from 262.0 to 334.8 points), 1a (334.9 to 407.5 points), 2 (407.6 to 480.2 points), 3 (480.3 to 552.9 points), 4 (553.0 to 625.6 points), 5 (625.7 to 698.3 points) and 6 (698.4 and above). In addition the percentage of students who scored below level 1b is reported.

provides a synthesised overview of the knowledge, skills and understandings that a student working within the level is able to demonstrate. However, in the sample studies Proficient Standards were established to represent a 'challenging but reasonable' expectation of student achievement at a year level. Proficient Standards provide reference points of reasonable expectation of student achievement at that Year in the area.

TIMSS and PIRLS use different procedures to identify four equally-spaced points along the achievement scales to use as international benchmarks (Advanced, High, Intermediate and Low) as well as an unbounded range below the 'Low' benchmark. Expert Groups (Mathematics, Reading and Science) of the IEA conduct scale anchoring analyses to describe student competencies 'at the benchmarks' (Martin & Mullis, 2013). Scale anchoring is described as identifying items that students at the benchmarks answered correctly, examining item content to determine the competencies demonstrated and generating a content-referenced description of achievement at each international benchmark (Martin & Mullis, 2013: 2). In practice, students scoring within five score points above or below each benchmark are selected and the percentages of those students answering each item correctly are computed. The items correctly answered by 65 per cent or more of these students, but fewer than 50 per cent of students at the next benchmark below, are used as the basis for describing the benchmark. Thus the TIMSS and PIRLS international benchmarks reflect a concept of progression in the field[4] in an analogous manner to the performance levels of PISA.

Student background

All studies collect information about student background although the method of collecting those data varies between student reports (in the international studies) and school records from parent-supplied information (in the case of the NAP). Those data typically involve sex, socioeconomic background (using either parental occupation or education or both), Indigenous status, language background (and/or country of birth of parents)[5] and geographic location. For the international, and some national, surveys these data are based on student responses to questions included in a questionnaire. For the studies conducted as part of the NAP since 2008 these data are based on school records of information supplied by parents and sometimes organised as files across school systems.[6]

Indigenous status refers to whether a student is Aboriginal and/or a Torres Strait Islander. The detailed data that are obtained at the point of data collection are reported as Indigenous or non-Indigenous because the numbers in some of the detailed categories are typically very small.

Indicators of socioeconomic background are based on information about parental education, parental occupation and the presence of designated home educational resources, cultural possessions and wealth. Parental education represents the highest level of parental school or non-school education that either parent/guardian has completed. Parental occupation represents the occupation group that characterises the main work undertaken by each parent or guardian. If a parent or guardian has more than one job, the occupation group which reflects their main job is reported. The higher occupational group of either parent or guardian is reported. In PISA, and the IEA International Civic

4 The report of PIRLS summarise this progression as follows:
Students at the Advanced International Benchmark take the entire text into account to provide text-based support for their interpretations and explanations. Students at the High International Benchmark were able to distinguish significant actions and information, make inferences and interpretations with text-based support, evaluate content and textual elements, and recognize some language features. At the Intermediate International Benchmark, students could retrieve information, make straightforward inferences, use some presentational features, and begin to recognize language features. Lastly, students at the Low International Benchmark demonstrated the ability to retrieve information from a text when it is explicitly stated or easy to locate. (Mullis, Martin, Foy & Drucker, 2012: 64)

5 PIRLS also collects information from a parent questionnaire about the language spoken by the child at the time they started school in the country.

6 Earlier NAP sample studies derived student background data from a student questionnaire.

and Citizenship Education Study (ICCS), parental occupations are coded on a continuous scale of occupational status. In PISA scores on this scale are combined with parental education and home possessions to form an index of economic, social and cultural status (ESCS) (OECD, 2010a: 170). For the NAP sample studies, data on parental education and occupation are coded in a set of ordinal categories. TIMSS reports against the categories of parental education (but does not use parental occupation) and the numbers of books in the home. PIRLS only uses number of books in the home (recorded in three categories) as an indicator of home background resources. In the case of PISA, breakdowns of mean scores by quarters of the distribution of the ESCS scale are reported as well as relationships between values of the ESCS scale and achievement measures. In the case of NAPLAN, NAP sample studies and TIMSS, mean achievement scores for each of the categories of parental education and occupation are reported.

Geographic location is recorded according to the MCEETYA Schools Geographic Location Classification System (Jones, 2004). The MCEETYA Schools Geographic Location Classification draws on the work of the Accessibility/Remoteness Index of Australia (ARIA) project (ABS, 2001) and provides a means of identifying a geographic location on the basis of postcode and place of residence. The classification distinguishes between locations in terms of population size and rurality and outlines eight broad geographic location classifications ranging from mainland state capital city regions to very remote areas. A geographic location can be classified using home location (suburb/town) and postcode or by using the location and postcode of the student's school. For reasons associated with the numbers in each category most of the studies cited in this report analyse geographic location in three categories: metropolitan, provincial, and remote.[7] Typically, the distribution of students is approximately 70 per cent of students from metropolitan locations, 28 per cent from provincial locations and two per cent from remote locations (which includes very remote locations).

Language background is captured in two indicators. The first is in terms of the main language spoken at home as either English or a language other than English. Typically the distribution is such that approximately 10 per cent of students have a language other than English as the main language spoken at home. The second indicator is based on the country of birth of the students and their parents.

POPULATIONS AND SAMPLES

Populations

Table 1.1 records the target population for each of the listed large-scale assessments. The target population refer to all students enrolled in schools, across all sectors, for the designated year or age level. Most define the population in terms of a year level. However, PISA refers to 15-year-old students who may be spread across several year levels and that spread can be different in different jurisdictions. For that reason, any comparison of relative results on PISA with results from other large-scale assessments needs to take account of differences in the age–grade distributions. For example, in PISA 2009, 53 per cent of 15-year-olds in Western Australia were in Year 11 compared to five per cent in New South Wales and none in Tasmania.

7 The metropolitan category includes mainland State capital city regions and major urban statistical divisions of 100,000 people or more. The provincial category includes provincial city statistical districts (plus Darwin) as well as other provincial towns and areas. The remote and very remote categories include locations with an ARIA index greater than or equal to 5.92.

Samples

NAPLAN surveys the population of students in each of the year levels with which it is concerned. Participation in NAPLAN is high, with participation rates around 96 per cent in Years 3, 5 and 7, and 92 to 93 per cent in Year 9,[8] so that the effects of bias from non-participation are likely to be small.

The other large scale assessments listed in Table 1.1 make use of samples of schools and students. Sampling is typically conducted as part of a two-stage design in which schools are sampled with a probability proportional to enrolment followed by a random selection of a sample of students in each school. In the cases of PISA and NAP-ICTL this is a random selection of a fixed number of students from all students in the defined population within each school. In the cases of TIMSS and PIRLS as well as NAP-CC and NAP-SL an intact class is selected at random within each school. Criteria govern population coverage, sampling procedures and participation rates so that participation bias is not introduced.[9]

Sample size

The studies listed in Table 1.1 are typically referred to as large-scale assessments so it is appropriate to review what this means in terms of numbers of participants. NAPLAN aims to assess the population in each of the designated year levels, except for certain specified exemptions, so that it involves between 240,000 to 260,000 students at each of the four year levels that are assessed.

In Australia, PISA 2009 involved 14,251 students from 353 schools. TIMSS 2010/11 in Year 8 involved 7556 students from 275 schools and in Year 4 (along with PIRLS) involved 6146 students from 280 schools.[10] NAP-SL in 2009 involved a sample of 13,162 Year 6 students from 618 schools; NAP-CC in 2010 involved a total sample of 13,655 Year 6 and Year 10 students from 647 schools; and NAP-ICTL in 2011 involved 11,023 Year 6 and Year 10 students from 649 schools.

One of the reasons for basing these studies on large samples is to ensure greater precision in the population estimates. In fact samples are typically designed so that the confidence intervals[11] associated with the mean estimates are less than one-tenth of a standard deviation for Australia as a whole and a little greater (0.15 to 0.2 of a standard deviation) for the mainland states.

A shift in the age–grade distribution over time (for example arising from changes in the school starting age) could contribute to changes in test scores for 15-year-olds simply because a higher percentage of 15-year-olds were in Year 11.

THE PISA SEQUENCE OF ASSESSMENTS

PISA focuses on three domains (reading, mathematics and science literacy) over a three-year assessment cycle. A different domain is chosen to be the major domain in each assessment cycle. This means that more assessment items are included from the major domain than from the two minor domains, and consequently more assessment time is allocated to the major domain than the two

8 In addition there are small percentages of exempt students (less than two per cent) at each year level.

9 In PISA no more than five per cent of the target population can be excluded from the sampling frame. A minimum participation rate of 85 per cent of schools (this could be achieved with sampled replacement schools if the initial response rate was between 65% and 85%) was required as well as a minimum average participation rate of 80 per cent of sampled students within schools (schools with a participation rate of less than 50 per cent were considered to have not participated). Similar criteria apply in TIMSS and PIRLS as well as in NAP-CC and NAP-ICTL.

10 There were just 20 fewer students in PIRLS than TIMSS.

11 The confidence interval is a measure of the precision of the estimate. The 95 per cent confidence interval is the range within which the estimate of the statistic based on repeated sampling would be expected to fall for 95 per cent of samples that might have been drawn. Its value is 1.96 times the 'standard error' of the estimate. In keeping with international practice this report uses the standard error to indicate precision.

minor domains. Reading literacy was the major domain in PISA 2000 and PISA 2009. Mathematical literacy was the major domain in PISA 2003, and scientific literacy was the major domain in PISA 2006.

There are several consequences that follow from this design. One is that more precise assessments are possible for a major domain than for minor domains. A second is that measures relating to subscales are possible when an area is a major domain but not when it is a minor domain. A third consequence is that more accurate measures of trends or changes are possible between cycles that involve a common major domain than between cycles involving minor domains. Specifically, trends in reading literacy are measured more accurately between PISA 2000 and PISA 2009 than between any other pair of cycles.

The assessment frameworks for each domain become established when it is a major domain. Hence, although it is possible to measure trends from 2000 onwards for reading literacy, it only became possible to measure trends from 2003 onwards for mathematical literacy and from 2006 onwards for science literacy.

TIMING OF TIMSS AND PIRLS

The populations for TIMSS and PIRLS studies are defined in terms of Years (or Grades). The schedule is designed so that students are assessed at the same stage of the school year in northern and southern hemispheres. For the northern hemisphere countries the surveys are conducted in the period from March to July and for the southern hemisphere in the period from September to December of the preceding calendar year. The most recent cycle of TIMSS was conducted in 2010 in the southern hemisphere and 2011 in the northern hemisphere so that students were assessed at the same point in the relevant school year. The convention is to refer to the study as TIMSS 2011 or PIRLS 2011. In Australia these assessments were conducted in 2010. In this report the whole study is referred to as TIMSS 2011 but Australian data are referenced as 2010 when only the national data are being described.

THIS REPORT

Chapter 2, which follows this introduction, is concerned with reading and draws on data from PISA, NAPLAN and PIRLS to examine patterns and trends in achievement in reading. It compares data for Australia with other countries (mainly with OECD countries) and explores in more detail the distribution of reading achievement within Australia. Chapter 3 focuses on mathematics and numeracy using data from PISA, NAPLAN and TIMSS. Data from TIMSS enable a longer time perspective on mathematics than was possible for reading. Again the focus is on broad comparisons with other countries combined with more detailed investigation of patterns of mathematics achievement within Australia. Chapter 4 explores patterns and trends in achievement in other fields: science, digital literacy, and civics and citizenship. The consideration of science achievement uses data from TIMSS since 1994/5 as well as the more recent data from PISA and the NAP sample studies of science literacy in Year 6. Achievement in other fields makes use of the NAP sample studies covering civics and citizenship (conducted in 2004, 2007 and 2010) and digital or ICT literacy (conducted in 2005, 2008 and 2011) as well as the PISA investigation of digital reading in 2009. Chapter 6 provides a summary with some links to major policy developments in school education.

2 READING

Given that the development of reading proficiency is universally seen as a central purpose of schooling and as essential to functioning in modern society, it is not surprising that most national assessment programs include reading as a central focus (e.g. National Assessment of Educational Progress (NAEP) in the United States). Reading achievement developed by mid-secondary school is closely related to reading proficiency at 24 years of age (OECD, 2010d). Higher reading achievement in schools is associated with higher levels of participation in post-school education and training and through that participation to employment and career outcomes (OECD, 2010d).

This chapter focuses on data from three large-scale assessment surveys that include reading as an assessment domain: the OECD's Programme for International Student Assessment (PISA), the National Assessment Program – Literacy and Numeracy (NAPLAN) in Australia and the Progress in International Reading Literacy Study (PIRLS). The chapter devotes most attention to change in achievement. The premise is that much can be learned about influences on achievement by examining change and the policies and practices associated with change (see also Mourshed, Chijioke & Barber, 2010). However, the discussion also explores cross-national differences at particular times for those differences can also inform judgements regarding the correlates of achievement. In the discussion of results from PISA the focus is on change through four assessment surveys over the nine-year period from 2000 to 2009. For NAPLAN the time scale is shorter: the five assessments took place between 2008 and 2012. In the case of PIRLS the time series began in 2001 and continued through a survey in 2006 to 2011. However, Australia began its participation only in the 2011 cycle (which was in 2010 in Australia).

CROSS-NATIONAL PERSPECTIVES ON READING IN SECONDARY EDUCATION FROM PISA

The PISA concept of reading literacy is broader than traditional 'notions of the ability to read' that centre on decoding information and literal comprehension. The PISA concept extends to interpretation and reflection as well as the ability to use written information in situations that students may encounter in their life at and beyond school. In the assessment framework for PISA (OECD, 2009), reading literacy is defined as 'understanding, using, reflecting on and engaging with written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society'.

Reading literacy in PISA is conceptualised around three dimensions: formats, aspects and situations. The PISA reading literacy assessment framework refers to continuous and non-continuous text formats (and provides measures in relation to these two formats). Continuous texts involve sentences organised into paragraphs and larger structures. Non-continuous texts involve matrix formats such as lists or combinations of lists in tables, diagrams and forms. PISA 2009 also incorporated combinations of these two forms in mixed and multiple formats. The framework also refers to the competencies (called aspects) involved in reading as three broad categories: access and retrieve, integrate and interpret, and reflect and evaluate. It also recognises that there are 'complex' tasks that combine all three of the competencies. In addition the PISA reading framework refers to situations (or contexts) as 'personal, public, educational or occupational' depending on the intended audience and purpose.

Cross-national comparisons of reading literacy in 2009

Summary statistics for reading literacy achievement from PISA 2000 and PISA 2009 for Australia and other selected OECD countries are recorded in Table 2.1. From the PISA reading literacy results for 2009 it can be inferred that Australian 15-year-olds performed similarly to their peers from New Zealand, Japan and Netherlands but significantly less well than 15-year-olds from Korea, Finland and Canada[12] (Thomson, De Bortoli, Nicholas, Hillman & Buckley, 2011: 52). The average score for Australian students was 515 scale points compared to the average for the 26 OECD countries in Table 2.1 of 496 points (OECD, 2010b). The average standard deviation for these countries in 2009 was 94 scale points.

One indicator of the spread of student scores is the standard deviation.[13] The standard deviation for Australia in 2009 was 99 points compared to the OECD average of 94 points. Australia is one of the countries with a relatively wide spread in reading performance. In other words the Australian variance in reading is larger than the average variance of the other 25 OECD countries in both 2000 and 2009.

More detailed reporting shows that, although there is a difference between the Australian means for continuous texts and non-continuous texts, there is little difference in the relative achievements of Australian students in relation to students in other countries on the continuous and non-continuous texts scales (OECD, 2010a: Tables 1.2.3, 1.2.16 & 1.2.19). In terms of continuous texts, Australian students performed less well than a group of three countries (Korea, Finland and Canada) and not significantly different from three other countries (Japan, New Zealand and the Netherlands). In terms of non-continuous texts, Australian students performed less well than a group of three countries (Korea, Finland and New Zealand) and not significantly different from three other countries (Canada, Japan, and the Netherlands).

Changes in reading literacy achievement from PISA 2000 to PISA 2009

Table 2.1 shows that, between 2000 and 2009, the average achievement in reading literacy for Australia declined from 528 to 515; a difference that is small (about one-eighth of a standard deviation) but statistically significant. Other countries to record a significant decline in average reading scores included Ireland, Sweden and the Czech Republic. Seven countries (Chile, Israel, Poland, Portugal, Korea, Hungary and Germany) recorded significant improvements (with gains of 13 to 40 scale points) in mean reading scores (OECD, 2010b).

---

12 Table 2.1 focuses on the OECD countries for which data are available for both 2000 and 2009. Other countries in which 15-year-olds performed better than Australia on PISA reading in 2009 were Hong Kong and Singapore (as well as the city of Shanghai).

13 Approximately 67 per cent of student scores are expected to fall between minus one and plus one standard deviation around the mean. A low standard deviation indicates that the scores are not spread out widely, whereas high standard deviation indicates that the scores are more spread out.

Table 2.1 Mean and range of reading performance for OECD countries: PISA 2000 and 2009

Country	PISA Reading 2000 Mean score[1]	PISA Reading 2000 Standard deviation	PISA Reading 2009 Mean score[1]	PISA Reading 2009 Standard deviation	Difference Mean Difference	Difference Ratio of variances (09/00)[2]
Korea	525 (2.4)	70	539 (3.5)	79	**14**	**1.30**
Finland	546 (2.6)	89	536 (2.3)	86	**-10**	0.93
Canada	534 (1.6)	95	524 (1.5)	90	-10	**0.91**
New Zealand	529 (2.8)	108	521 (2.4)	103	-8	**0.90**
Japan	522 (5.2)	86	520 (3.5)	100	-2	**1.37**
Australia	528 (3.5)	102	515 (2.3)	99	**-13**	0.94
Belgium	507 (3.6)	107	506 (2.3)	102	-1	0.90
Norway	505 (2.8)	104	503 (2.6)	91	-2	**0.77**
Iceland	507 (1.5)	92	500 (1.4)	96	-7	1.08
Switzerland	494 (4.2)	102	501 (2.4)	93	7	**0.84**
Poland	479 (4.5)	100	500 (2.6)	89	**21**	**0.80**
United States	504 (7.0)	105	500 (3.7)	97	-4	**0.85**
Germany	484 (2.5)	111	497 (2.7)	95	**13**	**0.73**
Sweden	516 (2.2)	92	497 (2.9)	99	**-19**	**1.15**
Ireland	527 (3.2)	94	496 (3.0)	95	**-31**	1.03
Denmark	497 (2.4)	98	495 (2.1)	84	-2	**0.73**
France	505 (2.7)	92	496 (3.4)	106	**-9**	**1.32**
Hungary	480 (4.0)	94	494 (3.2)	90	**14**	0.92
Portugal	470 (4.5)	97	489 (3.1)	87	**19**	**0.80**
Italy	487 (2.9)	91	486 (1.6)	96	-1	1.10
Spain	493 (2.7)	85	481 (2.0)	88	**-12**	1.07
Greece	474 (5.0)	97	483 (4.3)	95	9	0.96
Czech Republic	492 (2.4)	96	478 (2.9)	92	**-14**	0.92
Israel	452 (8.5)	109	474 (3.6)	112	**22**	1.04
Chile	410 (3.6)	90	449 (3.1)	83	**40**	**0.85**
Mexico	422 (3.3)	86	425 (2.0)	85	3	0.97
Average	496 (0.7)	96	496 (0.5)	94	1	0.97

Notes:
1 Standard errors are shown in parentheses.
2 This is the ratio of the squared standard deviations with values for 2009 divided by 2000. Ratios less than one mean a decrease in spread and ratios more than one reflect an increase. The test is whether this ratio is significantly different from one. Computations were performed for this report.
3 Differences and variance ratios that are statistically significant are shown in bold.
4 Countries listed in order of mean scores for 2009. Countries shown with shading are not significantly different from Australia in 2009.
5 The average in this table refers to the 26 countries listed for which data were available for 2000 and 2009.
Data Source: OECD (2010b) *PISA 2009 Results: Learning Trends*

In Australia there was no significant change in the spread of reading scores (as shown by the standard deviation) between 2000 and 2009. There were reductions in the spread of scores in Canada, Chile, Belgium, Germany, Denmark, Norway, Poland and Portugal. In three countries there was a significant increase in the spread of scores: Korea, France and Japan. Out of the seven countries with significant gains in performance, Germany, Chile, Poland and Portugal recorded significant reductions in the spread of scores. However, Korea recorded an improvement in mean reading score as well as an increase in the spread of scores.

OECD countries with improved reading achievement and reduced dispersion

It is of interest that in the four countries that had an increase in the mean score accompanied by a reduced spread of scores there had been identifiable reforms in school education in the period prior to 2009.

- Following the PISA 2000 assessment German education authorities invested strongly in research-based curriculum development as well as support for immigrant students from low socioeconomic backgrounds (OECD, 2010b; Lohmar & Eckhardt, 2011). There has also been greater collaboration among the states and the establishment of federal standards on the quality of teaching.
- In Poland there were structural changes in the school systems from 1999 which resulted in delayed specialisation into academic or vocational programs and the development of a core curriculum together with a reform of the examination system (OECD, 2010b: 33). Other reforms introduced in 1999 included decentralising the management of educational institutions to local municipal authorities and the heads of schools and a new system of teacher development and evaluation (Jakubowski, Patrinos, Porta & Wisniewski, 2010).
- From 2005 onwards Portugal introduced reforms focused on the first cycle (primary school) of education by consolidating small schools, grouping schools in 'clusters' (headed by a director) so as to provide better facilities and in-service education for teachers and strengthening evaluation processes (Matthews, Klaver, Lannert, Ó Conluain, & Ventura, 2008). National assessments were introduced in language and mathematics for students in Grades 4, 6 and 9. In 2006 a national plan for reading was introduced to improve reading proficiency among students and to encourage wider reading (OECD, 2010b: 68-69).
- Chile substantially increased its spending on education from the 1990s onwards, doubling between 1995 and 2007 (OECD, 2010b: 85). There were also increases in teacher salaries introduced between 2000 and 2006. More specifically, there had been concern over issues of quality and equity including the coverage of pre-school education. During the 1990s programs to support low-performing and disadvantaged students were introduced (Cox, 2004). These involved subsidies based on students from underprivileged socioeconomic backgrounds. One of the programs targeted the 10 per cent of lowest performing primary schools (OECD, 2010b: 85). Subsidised and municipal schools could apply for this if they committed to improvements, especially with regard to lifting their mean scores in the national assessments. Schools could design improvement plans and use the subsidies to employ specialist assistance. Following PISA 2000 there were curriculum changes in language courses to emphasise to a greater extent reading comprehension and communication (Cox, 2004).

Changes in the relative variation of achievement between and within schools

In this section of the report attention is focused on the variance in achievement scores. Students vary in reading achievement within Australia and the extent of variation is indicated as the variance in achievement scores. Variance is a measure of dispersion calculated as the mean of the squared deviations of observed values from a mean. The total variance in student achievement can be envisaged as made up of two sources: the variance within schools and the variance between schools' mean scores. The percentage of the total variance that is between schools provides an indication of the extent to which schools differ in their average achievement scores.

Table 2.2 Variance in reading scores in 2000 and 2009, total, within and between schools

	Variance in performance							
	PISA 2000				PISA 2009			
	Total	Within	Between	Percentage between	Total[1]	Within[1]	Between[1]	Percentage between
Australia[2]	10,171	8,294	1,805	18%	9,783	**(-) 7,707**	2,440	24%
Canada	8,954	7,632	1,934	20%	**(-) 8,163**	**(-) 6,780**	1,877	22%
Chile	8,074	3,981	4,081	51%	**(-) 6,833**	4,005	4,893	55%
Finland	7,994	7,117	591	8%	7,467	6,993	665	9%
Germany	12,367	4,717	6,667	59%	**(-) 8,978**	**(-) 3,890**	5,890	60%
New Zealand	11,700	9,765	1,867	16%	**(-) 10,575**	**(-) 8,228**	2,622	24%
Sweden	8,495	7,729	786	9%	**(+) 9,729**	8,290	**(+) 1,877**	19%
United States	10,979	7,846	3,306	30%	**(-) 9,330**	**(-) 6,476**	3,638	36%
OECD Average[3]	9,260	5,922	3,324	36%	**(-) 8,793**	5,875	3,420	37%

Notes:
1 Figures for 2009 are shown in bold if they are significantly different from 2000.
2 Including extra students in 2000 that were not part of the sample for the international study.
3 The OECD average refers to 26 OECD countries with data for 2000 and 2009.
Data source: Based on data from OECD (2010b: 161) *PISA 2009 Results: Learning Trends*

Table 2.2 records the percentage of the variance that is between schools for Australia and selected OECD countries in 2000 and 2009. The percentage of variance that is between schools indicates the extent to which schools are differentiated in terms of this achievement outcome. The highest level of differentiation is found in tracked education systems where entry to secondary school is based on measured performance (e.g. Germany). The lowest level of differentiation is found in fully comprehensive school systems where there is little social stratification by location (e.g. Finland). The extent of differentiation is influenced by factors such as explicit selectivity in entry to types of secondary school, the extent of enrolment in private schools and the extent to which residential location is differentiated on the basis of socioeconomic background. Between 2000 and 2009 there was a significant increase in the differentiation of achievement by school in Australia, New Zealand Sweden and the United States.

Table 2.3 records measures of the effect of socioeconomic background on student performance. The effect, or the slope, is estimated overall, within schools and between schools. In Australia, although there was no change in the effect overall or within schools (on average), there was a significant increase in the between-school effect. That is, for a given difference in average socioeconomic status between schools, the difference in performance has increased. In other words, the gap in performance between schools with a low and high socioeconomic status has increased. Australia was the only country in Table 2.3 (or in the full range of OECD countries) where this increase was observed. There were several countries, including Canada, where the school level effect of index of economic, social and cultural status (ESCS) on performance decreased.

In summary, from 2000 to 2009 Australian secondary schools became more differentiated in reading achievement (as shown in Table 2.2) and that differentiation became more strongly linked to the average socioeconomic context of the school (as shown in Table 2.3).

Table 2.3 Effect of socioeconomic background on reading performance, overall, within and between schools

	Relationship between reading achievement and ESCS (index of economic, social and cultural status)					
	PISA 2000			PISA 2009		
	Overall	Within-school	Between-school	Overall	Within-school	Between-school
Australia	47 (2.7)	32 (3.1)	47 (7.0)	46 (1.8)	30 (1.9)	**66 (6.2)**
Canada	38 (1.3)	29 (0.7)	49 (3.4)	**32 (1.4)**	**21 (1.4)**	**32 (6.7)**
Chile	39 (1.7)	11 (2.1)	62 (6.0)	**31 (1.7)**	8 (1.8)	50 (4.3)
Finland	25 (2.3)	22 (1.7)	65 (55.3)	**31 (1.7)**	**28 (2.0)**	19 (10.3)
Germany	52 (2.6)	14 (2.3)	142 (17.7)	**44 (1.9)**	10 (1.6)	122 (8.4)
New Zealand	47 (2.7)	33 (2.9)	57 (10.4)	52 (1.9)	36 (2.9)	61 (9.3)
Sweden	36 (1.8)	27 (2.2)	43 (9.6)	**43 (2.2)**	**34 (2.2)**	52 (10.1)
United States	52 (3.0)	30 (4.6)	90 (10.9)	**42 (2.3)**	23 (2.9)	63 (12.1)
OECD Average-26	39 (0.5)	18 (0.5)	66 (3.2)	38 (0.4)	**19 (0.4)**	**61 (1.9)**

Notes:
1 Standard errors are shown in parentheses.
2 Bold if significantly different from 2000.
3 Overall effect of ESCS is based on the slope resulting from a single-level bivariate regression of reading on ESCS.
4 Within-school effect is the result of a two-level regression of reading on ESCS: average within-school effect (slope) of student level ESCS on performance.
5 Between-school effect is the result of a two-level regression of reading on ESCS: the effect (slope) of school average ESCS on student performance.
6 The OECD average refers to 26 OECD countries with data for 2000 and 2009.
Data Source: OECD (2010b: 161 and 163) *PISA 2009 Results: Learning Trends*

NATIONAL PERSPECTIVES ON READING IN SECONDARY SCHOOL FROM PISA

Over the four PISA assessment cycles the change in reading literacy scores for Australia was from 528 in 2000, through 525 in 2003 and 513 in 2006, to 515 in 2009. The data for PISA 2012 will inform judgements about whether the small decline has continued or whether the data for 2006 and 2009 indicate a flattening off of the trend. In this section the focus is on change in reading achievement for sections of the Australian population between 2000 and 2009.

Table 2.4 Distribution of PISA reading achievement for Australia in 2000 and 2009

	PISA 2000	PISA 2009		
	Mean	Mean	Difference	Significance of difference
Distribution of scores				
95th percentile	685 (4.5)	668 (3.9)	-17	Yes
90th percentile	656 (4.2)	638 (3.2)	-18	Yes
75th percentile	602 (4.6)	584 (2.7)	-18	Yes
50th percentile	534 (4.2)	521 (2.4)	-13	No
25th percentile	458 (4.4)	450 (2.9)	-8	No
10th percentile	394 (4.4)	384 (3.1)	-10	No
5th percentile	354 (4.8)	343 (3.8)	-11	No
Percentage of students in proficiency levels				
Level 5 & above	17.6 (1.2)	12.6 (0.8)	-4.9	Yes
Below level 2	12.5 (0.9)	14.2 (0.6)	1.8	No

Note: Standard errors are shown in parentheses.
Sources: OECD (2010b); Thomson et al. (2011); Lokan, Greenwood & Cresswell (2001)

Changes in the distribution of achievement

Data regarding the distributions of student scores in reading in 2000 and 2009 are recorded in Table 2.4. An examination of differences in percentiles suggests that, superimposed on the overall decline in reading scores, there has been a greater decline at the 75th, 90th and 95th percentiles than at the 5th, 10th and 25th percentiles. The decline of the 90th percentiles was 18 scale points which was statistically significant whereas the decline for the 10th percentile was 10 points and not statistically significant (OECD, 2010b: 147). Figure 2.1 is a percentile-plot that shows the shift in the upper part of the distribution. Despite the decrease of these percentiles, the total variance did not change significantly between 2000 and 2009 (see Table 2.2). It is worth noting that it is not only the top section of distribution that has declined. If we examine all four cycles the results for the lower half are very similar in 2000 and 2003 and the 25th and 50th percentiles decline significantly between 2003 and 2009.[14]

The change in the top three percentiles can also be described in terms of the percentages in the described proficiency levels used in PISA. There was a decline in the percentage of students at proficiency level 5 and above (18% in 2000 compared to 13% in 2009) but no significant change in the percentage of students below level 2 (13% in 2000 compared to 14% in 2009) (OECD, 2010b). The pattern of a decline in the percentage of students in level 5 and above combined with no significant change in the percentage below level 2 was also evident in data for Canada, Finland, New Zealand and Norway (OECD, 2010b: 147).

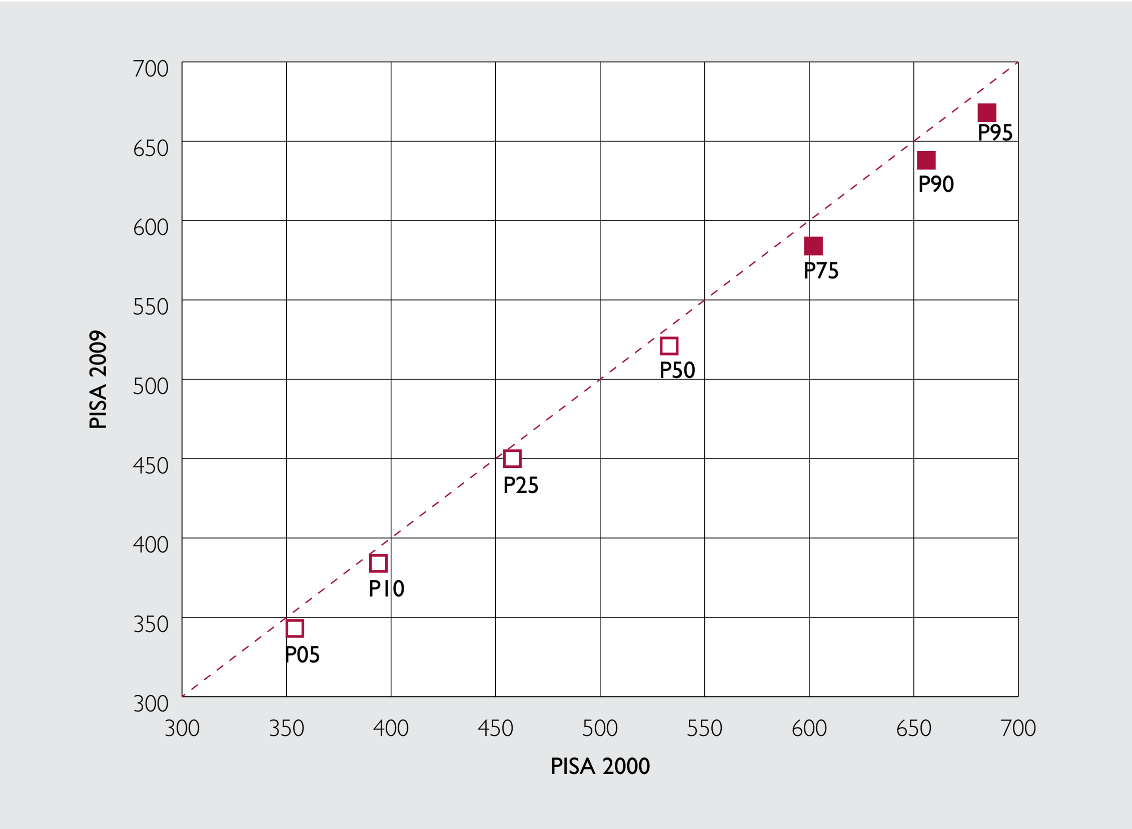


Figure 2.1 P-P Plot for PISA reading achievement in 2000 and 2009

Note: Points shown in red indicate that the differences in the percentiles between 2000 and 2009 are statistically significant.

14 The differences are not significant compared with the 2000 cycle because the sample size is smaller, and the standard errors are larger, in 2000.

Differences among jurisdictions

Table 2.5 indicates that there were differences among jurisdictions in the change in mean reading scores between 2000 and 2009. In Tasmania (31 points), South Australia (31 points), New South Wales (23 points) and the ACT (21 points) there were significant declines. The effect of each of these states on the national negative trend in reading varies due to the population size. Percentages of the 15-year-old population that live in these states are: about three per cent in Tasmania, eight per cent in South Australia, 30 per cent in New South Wales and two per cent in the ACT. Therefore, the negative change in New South Wales probably affected the national decline the most. There were no significant changes in Western Australia, the Northern Territory, Victoria or Queensland (Thomson et al., 2011). The variations among Australian jurisdictions in the extents of the declines suggest that there may be some systemic factors associated with curricula, the availability of qualified teachers or school organisation that may be linked to the declines in achievement in the lower secondary years.

Table 2.5 Jurisdictional mean reading performance in Australia: PISA 2000 and PISA 2009

Jurisdiction	PISA 2000	PISA 2009	Difference (PISA 2009 – PISA 2000)
Tasmania	514 (9.7)	483 (5.8)	**-31**
South Australia	537 (7.7)	506 4.8)	**-31**
New South Wales	539 (6.3)	516 (5.6)	**-23**
ACT	552 (4.6)	531 (6.0)	**-21**
Western Australia	538 (8.0)	522 (6.3)	-16
Northern Territory	489 (5.6)	481 (5.6)	-8
Victoria	516 (7.6)	513 (4.7)	-3
Queensland	521 (8.6)	519 (7.0)	-2

Notes:
1 Standard errors are shown in parentheses.
2 Differences between 2000 and 2009 that are statistically significant are shown in bold.
Source: Thomson et al. (2011: 127)

Changes in differences associated with social and demographic characteristics of students

PISA relates achievement scores to a number of student characteristics including sex, Indigenous status, socioeconomic background, language background, immigrant background, and geographic location. Most of these are categorical variables but socioeconomic background is measured by a continuous scale: economic, social and cultural status (ESCS). Data are shown in Table 2.6.

- *Sex*. In both PISA 2000 and PISA 2009 females recorded higher average reading literacy scores than males and over this period there was no significant change in the difference between females and males. The difference was between 34 and 37 scale points. The decline was statistically significant for males and not statistically significant for females but there was no change in the difference.
- *Indigenous status*. In both PISA 2000 and PISA 2009 non-Indigenous students recorded higher average reading literacy scores than Indigenous students and over the period from 2000 to 2009 there was no significant change in the difference between these two groups of students. The difference was constant at 82 to 83 scale points.
- *Language background*. In PISA 2000, but not in PISA 2009, students whose language spoken at home was mainly English recorded higher average reading literacy scores than students whose language spoken at home was mainly a language other than English. Over the period there was a decline in the scores of the former group but not the latter.

- *Immigrant background*. In PISA 2009 but not in PISA 2000 students of an immigrant background recorded higher average reading literacy scores than students not of an immigrant background. There was a change in the difference between these two groups. There was no change in the average scores of students of an immigrant background but there was a decline in the scores of students that were born in Australia.

Table 2.6 PISA reading statistics for groups of Australian students in 2000 and 2009

	PISA 2000 Mean		PISA 2009 Mean		Significance of difference between cycles
Gender					
Females	546	(4.7)	533	(2.6)	
Males	513	(4.0)	496	(2.9)	*
Difference	**34**	(5.4)	**37**	(3.1)	
Indigenous status					
Non-Indigenous	531	(3.4)	518	(2.2)	*
Indigenous	448	(5.8)	436	(6.3)	
Difference	**83**	(6.7)	**82**	(6.7)	
Language background					
English language at home	535	(3.6)	518	(2.0)	*
LBOTE	504	(7.5)	509	(8.9)	
Difference	**31**	(7.4)	10	(8.3)	
Immigrant status					
Australian born	532	(3.6)	515	(2.1)	*
Immigrant background	520	(6.7)	524	(5.8)	
Difference	12	(6.6)	**-10**	(5.8)	
Location					
Metropolitan	535	(4.8)	521	(2.9)	*
Non-metropolitan	518	(7.0)	496	(4.0)	*
Difference (metro–non-metro)	**17**	(8.8)	**25**	(5.1)	
Economic, social and cultural status (ESCS)					
Top quarter	587	(4.9)	562	(1.7)	*
Upper quarter	538	(4.5)	532	(1.5)	
Lower quarter	516	(3.8)	504	(1.9)	*
Bottom quarter	476	(3.6)	471	(2.1)	
Difference (Top–Bottom)	**112**	(6.1)	**91**	(2.7)	*

Notes:
1 Standard errors are shown in parentheses.
2 Differences between groups that are significant are shown in bold.
3 Differences across cycles that are significant are designated with an asterisk *.
Sources: OECD (2010b); Thomson et al. (2011)

- *Location*. In both PISA 2000 and PISA 2009 students in metropolitan locations recorded higher average reading literacy scores than students in non-metropolitan locations and over the period from 2000 to 2009 there was no significant change in the difference between the two groups. The difference was between 17 and 25 scale points.
- *Socioeconomic status*. The association of reading achievement with the socioeconomic background scale ESCS is shown in the differences in mean scores for the quarters of the distribution of ESCS in Table 2.6. Those data show that the difference in achievement between the top and bottom quarters is substantial but appears to have reduced a little

between 2000 and 2009 mainly as a consequence of the drop in the average score of those in the top quarter. In the discussion of Table 2.3 it was noted that the overall association between reading achievement and socioeconomic status in Australia had not changed over the decade but there was an increase in the between-school association of average school achievement and average school ESCS scores.

Changes by social and demographic characteristics of students within jurisdictions

Since only four of the jurisdictions showed a significant negative trend between 2000 and 2009, it is worthwhile to have a closer look at changes in performance by social and demographic characteristics within each of these states and territories. Given the national change in variation in performance between schools and the school-level effect of socioeconomic status on performance, the first goal was to decompose the variance in performance for each state and territory and to decompose the effect of socioeconomic background on performance.

Figure 2.2 and Table 2.7 give the results of decomposing the variance in reading within and between schools. Only New South Wales showed a significant increase in between-school variance from 2000 to 2009. The proportion of the total variance in reading that lies between schools – the intra-class correlation – increased from 0.11 to 0.29 (compared to 0.18 and 0.24 at a national level). No causality can be established, but this change could be related to the negative trend in reading performance in New South Wales. Another observation that may be noteworthy is the relative large between-school variance in Tasmania.

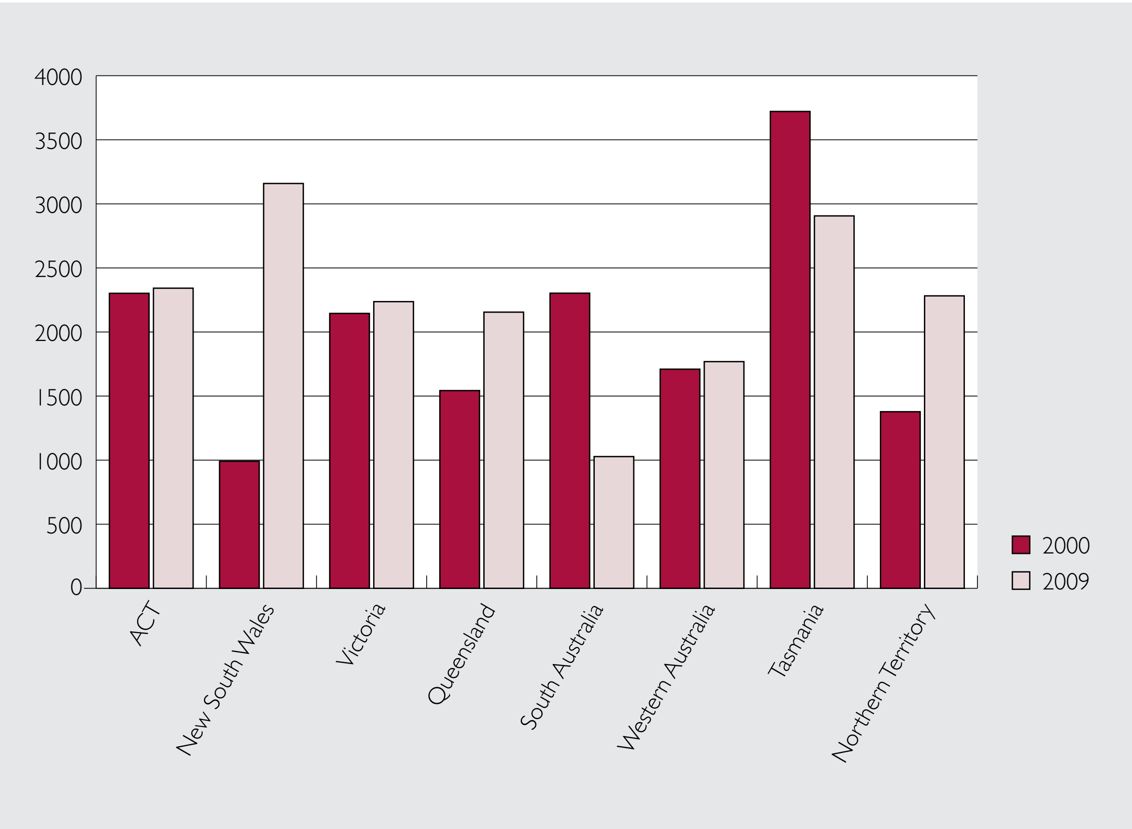


Figure 2.2 Between-school variance within each jurisdiction in 2000 and 2009

Table 2.7 Variance decomposition of reading performance within jurisdiction

	Variance	Variance in 2000	Variance in 2009	Difference (t-value)
ACT	Within	8772 (757.0)	8637 (541.0)	-0.14
	Between	2302 (821.4)	2342 (603.4)	0.04
	Intraclass correlation	0.21	0.21	
New South Wales	Within	8092 (489.9)	7564 (288.0)	-0.93
	Between	991 (290.7)	3159 (934.2)	**2.22**
	Intraclass correlation	0.11	0.29	
Victoria	Within	8329 (462.4)	7432 (318.2)	-1.60
	Between	2145 (678.7)	2237 (406.6	0.12
	Intraclass correlation	0.20	0.23	
Queensland	Within	8499 (487.1)	7965 (345.7)	-0.89
	Between	1543 (389.2)	2155 (521.4)	0.94
	Intraclass correlation	0.15	0.21	
South Australia	Within	7077 (540.9)	6851 (333.0)	-0.36
	Between	2303 (683.9)	1028 (367.1)	-1.64
	Intraclass correlation	0.25	0.13	
Western Australia	Within	8206 (606.0)	8124 (495.6)	-0.10
	Between	1710 (855.8)	1769 (411.0)	0.06
	Intraclass correlation	0.17	0.18	
Tasmania	Within	9497 (1193.9)	8087 (478.5)	-1.10
	Between	3721 (1002.9)	2906 (911.4)	-0.60
	Intraclass correlation	0.28	0.26	
Northern Territory	Within	11566 (1005.0)	12109 (1488.2)	0.30
	Between	1378 (737.5)	2282 (1600.0)	0.51
	Intraclass correlation	0.11	0.16	

Notes:
1 Standard errors are shown in parentheses.
2 Significant differences are shown in bold.
3 States and territories with a negative trend in average performance on reading are underlined.

Consistent with the analysis at the national level the effect of socioeconomic status on reading performances was estimated between and within schools. Figure 2.3 and Table 2.8 present the results of these analyses. The only significant change was in New South Wales, where the between-school effect increased between 2000 and 2009, similar to the results at the national level. In other words, the gap in performance between schools with a given difference in average socioeconomic level in New South Wales was larger in 2009 than in 2000. Generally, the school-level effect is larger than the student-level effect, except for in Queensland, where the difference in performance for a given difference in socioeconomic status is larger within school than between schools.

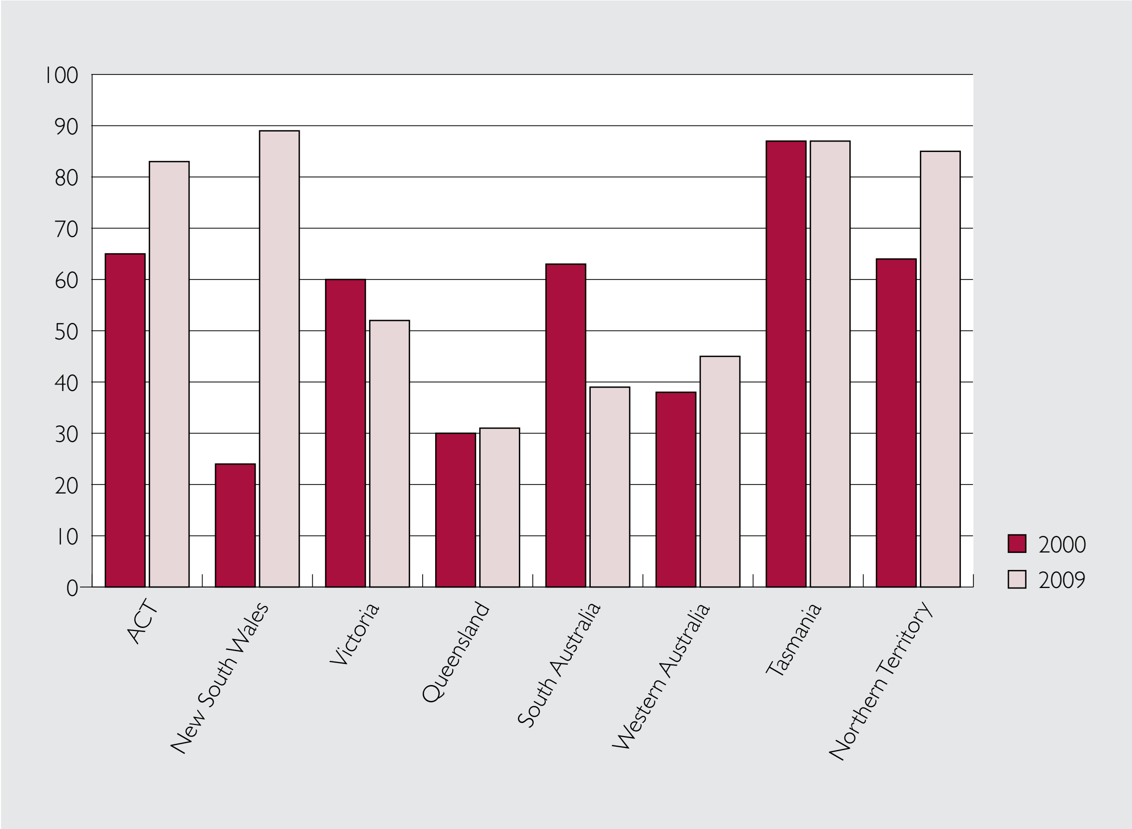


Figure 2.3 Between-school effect of socioeconomic status on performance by state or territory in 2000 and 2009

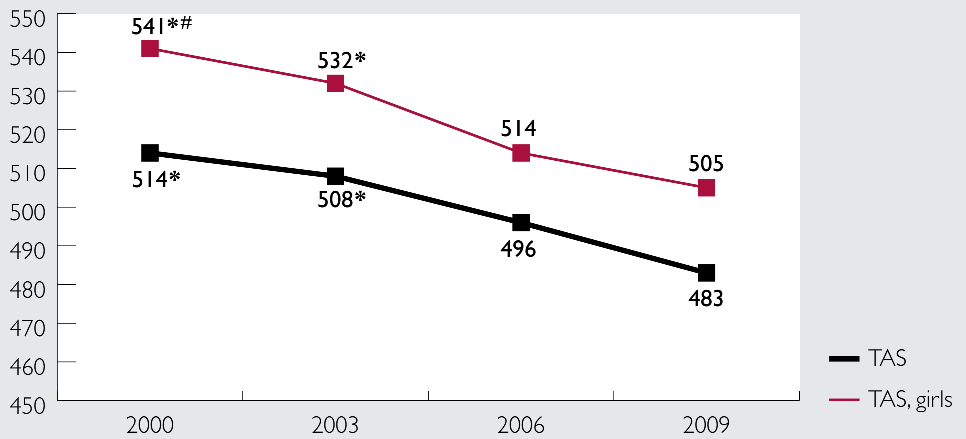
Table 2.8 The effect of socioeconomic status on reading performance within and between schools in 2000 and 2009

			2000		2009		Difference (t-value)
ACT	ESCS slope	Within	43	(6.7)	41	(5.6)	-0.22
		Between	65	(13.4)	83	(20.8)	0.72
	Residual variance	Within	7400	(571.1)	7781	(491.9)	0.51
		Between	678	(596.4)	739	(182.7)	0.10
New South Wales	ESCS slope	Within	33	(4.6)	26	(3.5)	-1.22
		Between	24	(11.4)	**89**	(14.8)	**3.52**
	Residual variance	Within	7309	(436.1)	7175	(261.7)	-0.26
		Between	355	(160.7)	1018	(343.3)	1.75
Victoria	ESCS slope	Within	26	(4.4)	29	(3.7)	0.39
		Between	60	(8.6)	52	(7.0)	-0.70
	Residual variance	Within	7974	(432.4)	6966	(292.3)	-1.93
		Between	375	(240.2)	781	(413.7)	0.85
Queensland	ESCS slope	Within	46	(10.0)	73	(11.8)	1.75
		Between	30	(6.1)	31	(3.0)	0.23
	Residual variance	Within	7997	(534.6)	7184	(277.8)	-1.35
		Between	371	(303.3)	634	(281.6)	0.63
South Australia	ESCS slope	Within	31	(4.5)	26	(5.3)	-0.84
		Between	63	(13.8)	39	(11.2)	-1.32
	Residual variance	Within	6471	(555.6)	6383	(275.5)	-0.14
		Between	424	(172.8)	323	(109.7)	-0.49
Western Australia	ESCS slope	Within	35	(6.1)	40	(6.2)	0.55
		Between	38	(17.1)	45	(17.8)	0.30
	Residual variance	Within	7508	(585.7)	7451	(477.2)	-0.08
		Between	525	(202.1)	663	(381.8)	0.32
Tasmania	ESCS slope	Within	21	(7.3)	34	(3.6)	1.63
		Between	87	(13.4)	87	(14.6)	0.03
	Residual variance	Within	8885	(1101.5)	7461	(490.0)	-1.18
		Between	601	(196.1)	473	(199.4)	-0.45
Northern Territory	ESCS slope	Within	30	(8.5)	51	(7.5)	1.90
		Between	64	(18.3)	85	(30.7	0.58
	Residual variance	Within	9864	(762.7)	10573	(1185.9)	0.50
		Between	142	(248.5)	728	(340.3)	1.39

Notes:
Standard errors are shown in parentheses.
Significant differences are shown in bold.
States and territories for which there was a decline in average achievement between 2000 and 2009 (see Table 2.5) are underlined.

In addition to the decomposition of variance and effect of socioeconomic status between and within schools, trends were estimated for subgroups of students and tested for significance for each state or territory. Following are the main results for each state and territory in order of the size of the negative trend.

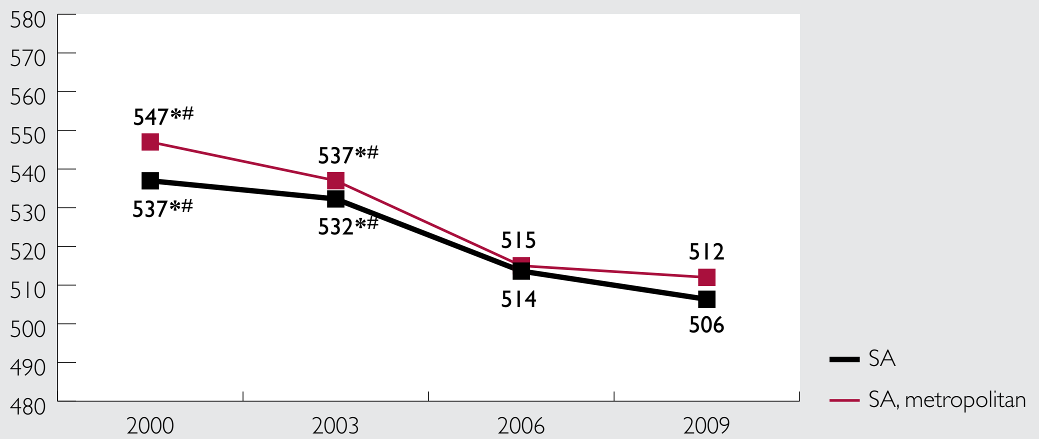
Tasmania: There was a negative trend overall, in particular for the girls (see Figure 2.4). The overall and girls' performance is significantly higher in 2000 and 2003 than in 2009. In addition, the girls' performance is higher in 2000 than in 2006.



Note: * indicates significantly different from 2009
# indicates significantly different from 2006

Figure 2.4 Average reading performance since 2000 in Tasmania

South Australia: There was a negative trend overall, particularly in metropolitan areas (see Figure 2.5). Performance in 2006 and 2009 is significantly lower than in 2000 and 2003 for both groups.



Note: * indicates significantly different from 2009
# indicates significantly different from 2006

Figure 2.5 Average reading performance since 2000 in South Australia

New South Wales: There were negative trends overall and in particular in provincial areas (see Figure 2.6). Student performance in 2006 and 2009 was significantly lower than in 2000 for both groups.

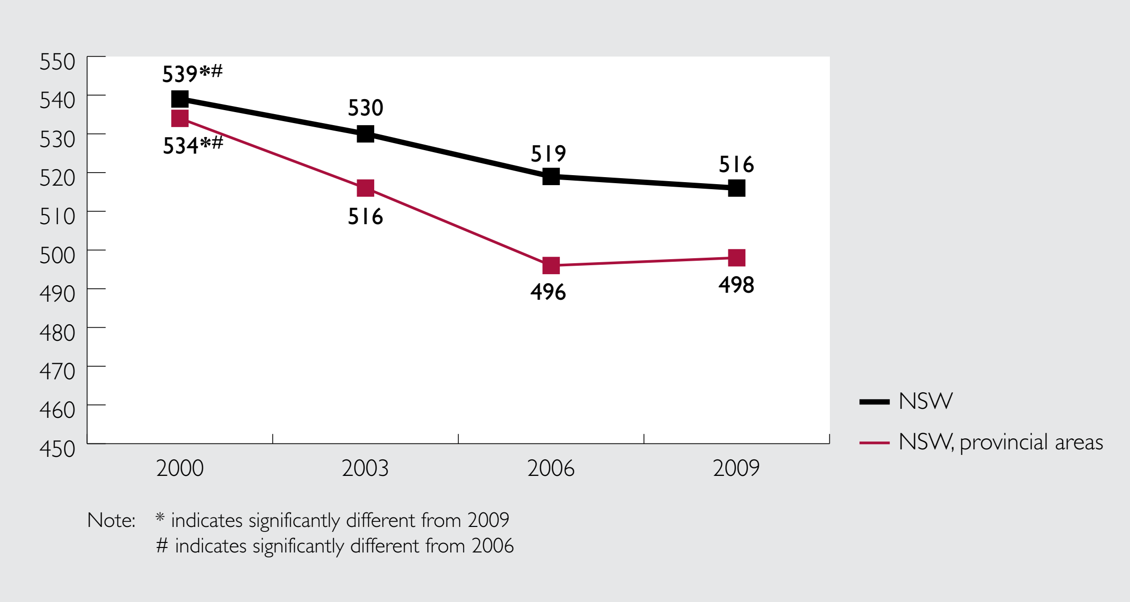


Figure 2.6 Average reading performance since 2000 in New South Wales

ACT: There was a negative trend overall, in particular for the majority group (born in Australia, non-Indigenous, English speaking background). Overall performance, and the performance of the majority group, was higher in 2000 than in 2009 (see Figure 2.7). In addition, the performance of the majority group was higher in 2000 than in 2006.

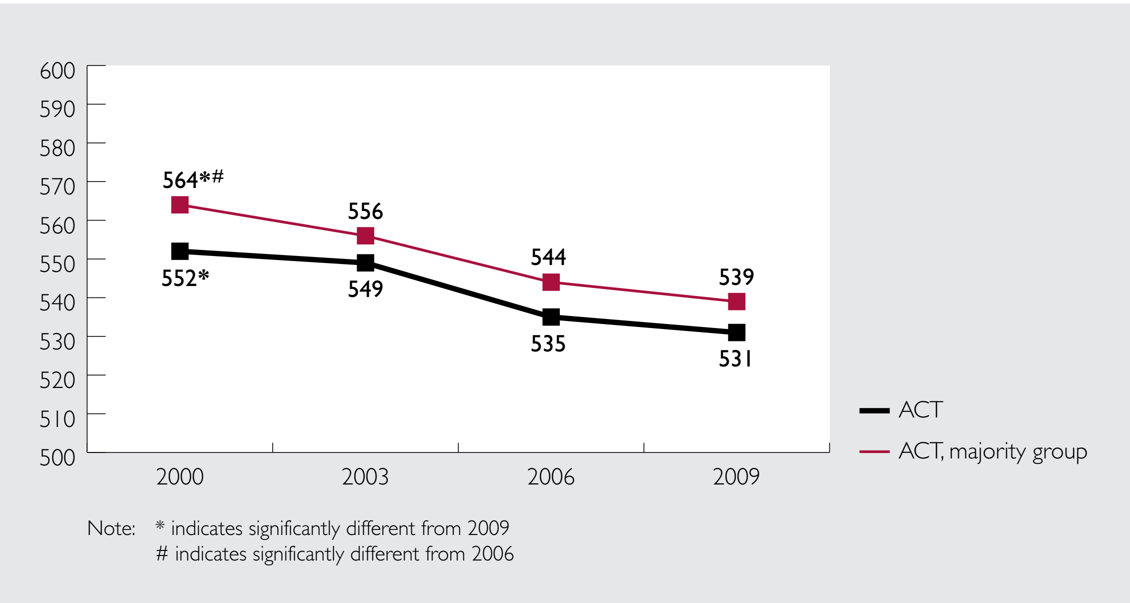


Figure 2.7 Average reading performance since 2000 in the ACT

Western Australia: There was no significant overall change between 2000 and 2003, but significant decline after 2003. The average performance of students in provincial areas was stable between 2000 and 2006, but dropped significantly in 2009 (see Figure 2.8).

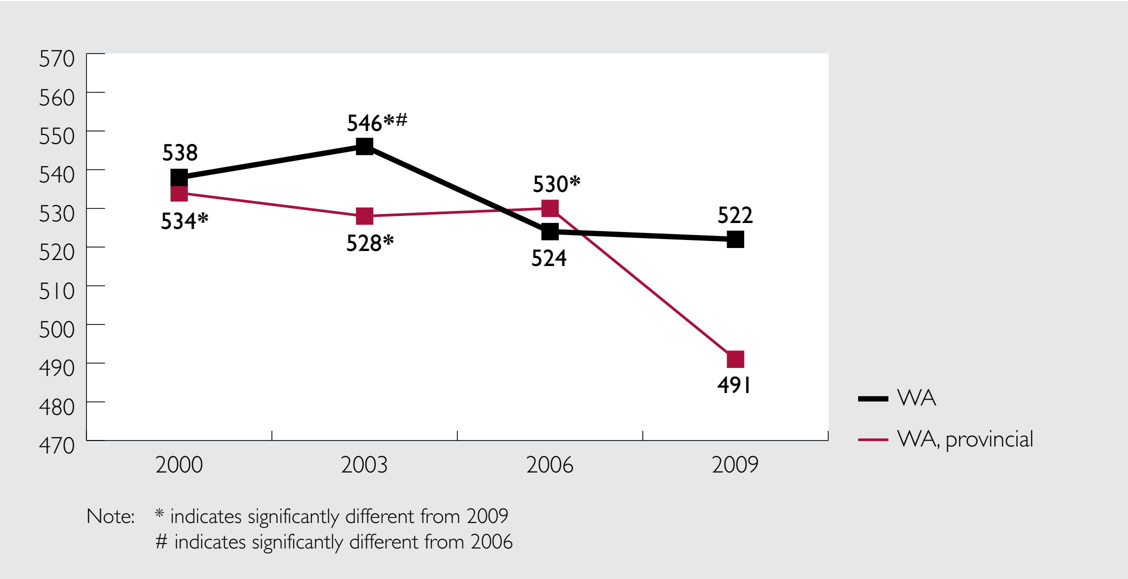


Figure 2.8 Average reading performance since 2000 in Western Australia

For Victoria, Queensland and the Northern Territory there were no significant changes in PISA reading achievement between 2000 and 2009.

CHANGES IN DIFFERENCES BETWEEN SUBGROUPS AT VARIOUS PERCENTILES ON PISA READING

It was of interest to know whether the differences between subgroups were the same across the distribution of scores. It was also of interest to determine whether those differences changed over time. Quantile regression analyses were conducted for each PISA cycle to investigate these issues.

Various pairs of subgroups were compared: (1) male versus female students, (2) Indigenous versus non-Indigenous students, (3) students with a language background other than English versus students with an English language background, (4) students born overseas versus students born in Australia, (5) students attending schools in metropolitan areas versus students attending schools in provincial or remote areas, and (6) students with high socioeconomic background (above the national median on the ESCS measure) versus students with a low socioeconomic background (below the national median on the ESCS measure). Quantile regression was used to estimate the difference in performance between each pair of groups at a predefined set of percentiles of the distribution. The 10th, 25th, 50th, 75th and 90th percentiles were chosen for these analyses.

The graphs presented in this section were used to explore patterns; hence testing for the significance of each of the reported differences and changes in differences was not regarded as necessary.[15] The vertical axis of each graph is 100 points, which is approximately one standard deviation in performance of the total Australian population. The results are shown in Figures 2.9 to 2.14.

15 Standard errors were not adjusted for a complex two-stage sampling design or for the measurement error and are therefore not reported.

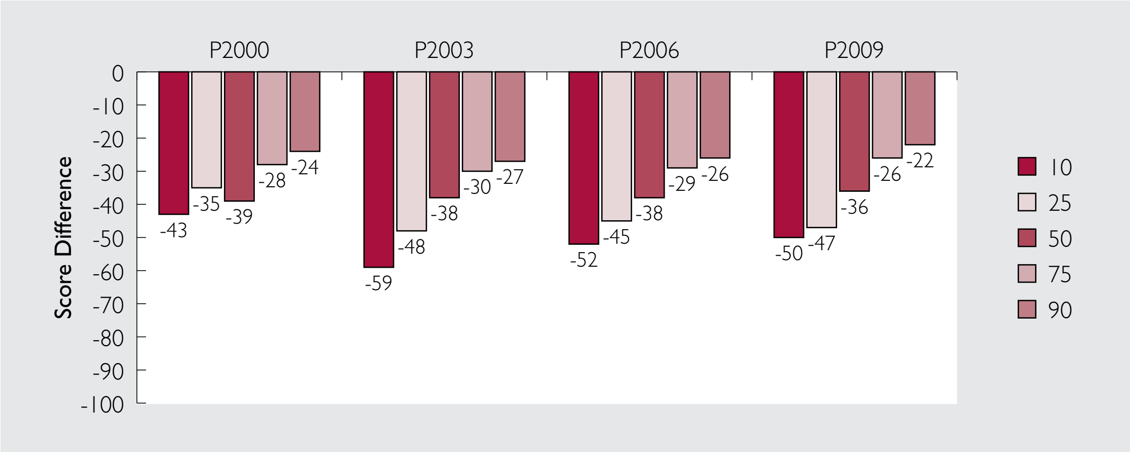


Figure 2.9 Differences in reading achievement between male and female students at different percentiles across time

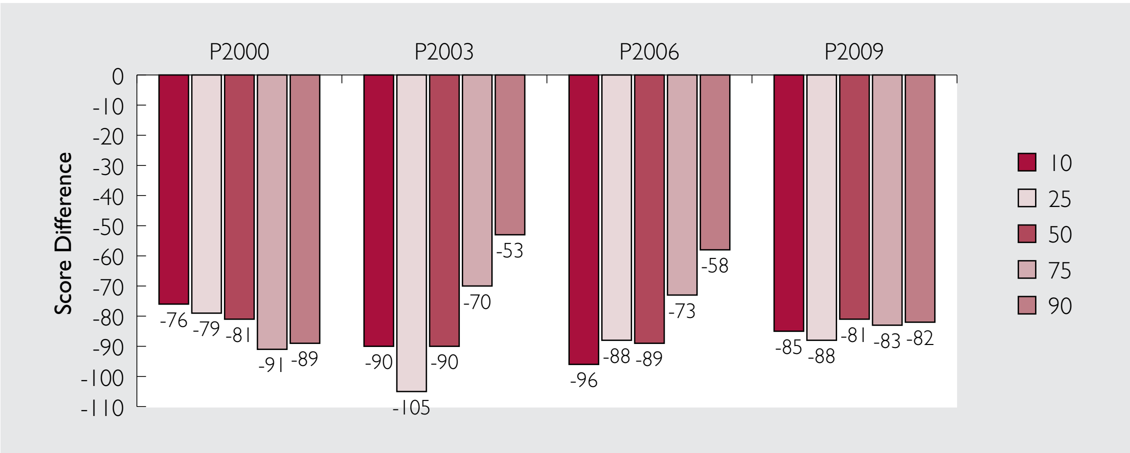


Figure 2.10 Differences in reading achievement between Indigenous and non-Indigenous students at different percentiles across time

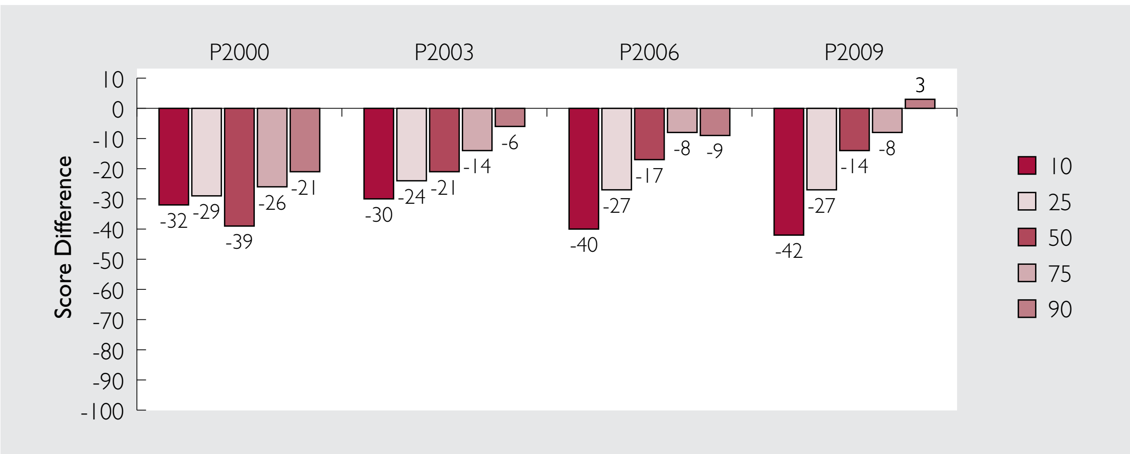


Figure 2.11 Differences in reading achievement between LBOTE students and non-LBOTE students at different percentiles across time

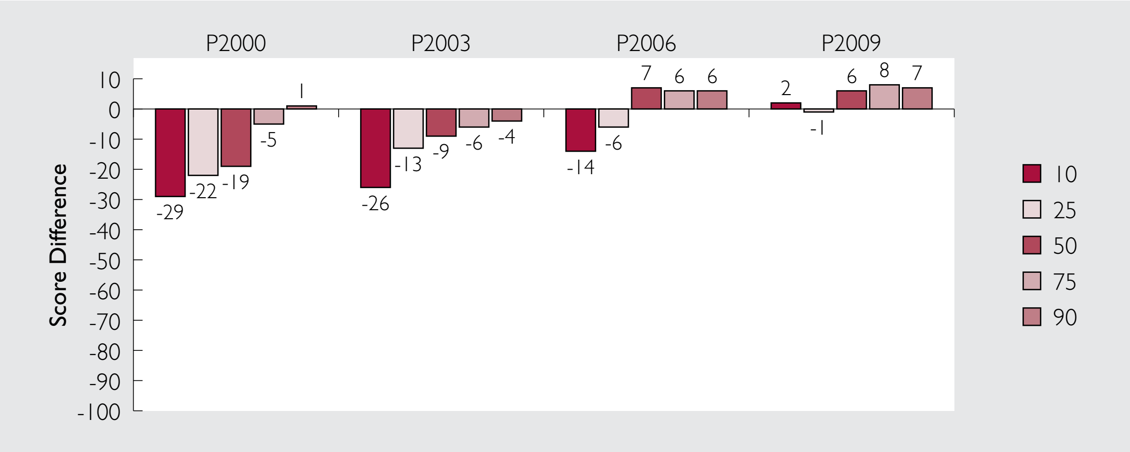


Figure 2.12 Differences in reading achievement of overseas-born and Australian-born students at different percentiles across time

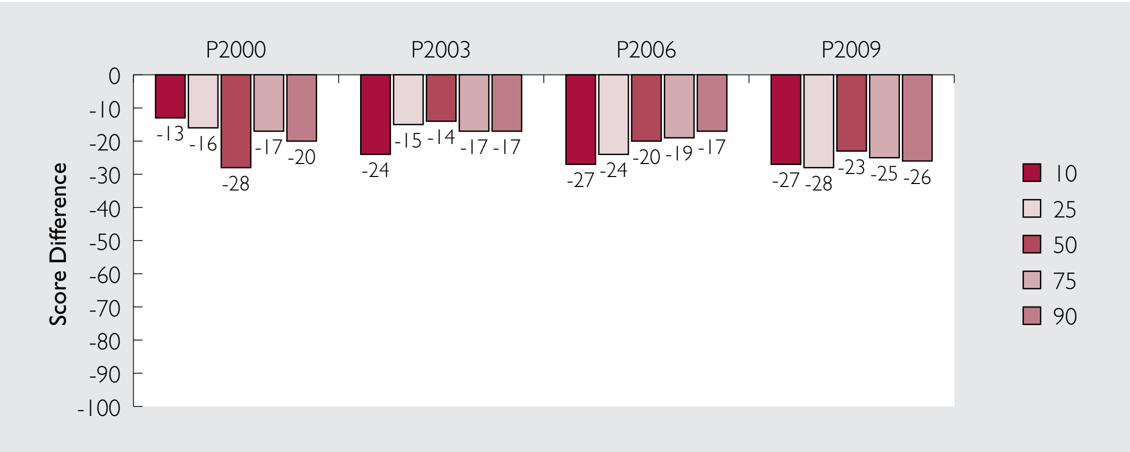


Figure 2.13 Differences in reading achievement of students in provincial or remote areas and metropolitan areas at different percentiles across time

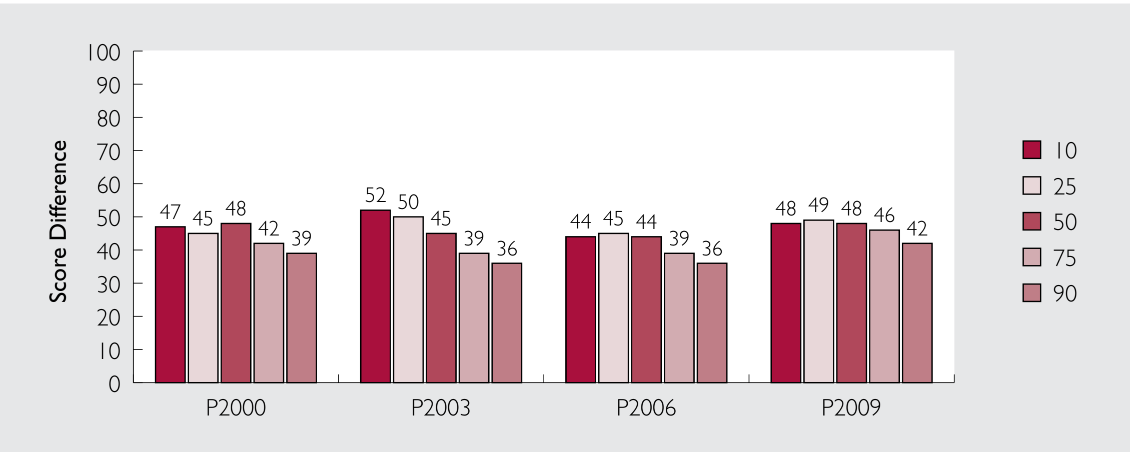


Figure 2.14 Differences in reading achievement of students with high ESCS and students with low ESCS at different percentiles across time

Figure 2.9 shows that males performed less well than females at each percentile and that the difference was smaller among the more able students. At the 10th percentile, males scored about half a standard deviation lower than girls (about 50 score points), while the difference was about a quarter of a standard deviation at the 90th percentile. The differences between male and female students remained stable over time at each percentile.

While the difference between Indigenous students and non-Indigenous students (Figure 2.10) ranged from one standard deviation at the 10th percentile (about 100 points) to a little more than half a standard deviation at the 90th percentile (just over 50 points) in 2003 and 2006, the differences were more equal across the scale in 2000 and 2009. These differences between the cycles may be associated with differences in sampling methods.

In every PISA cycle, there were no differences in reading performance evident between high performing students with a language background other than English and students with an English language background (Figure 2.11) since 2003. However, at the lower end of the scale there were differences of about one third of a standard deviation.

Although there was no difference between students born overseas and students born in Australia at any point of the distribution in 2006 and 2009, there was some difference (about one quarter of a standard deviation) at the low end of the scale in 2000 and 2003 (Figure 2.12).

Students attending schools in remote and provincial areas performed less well than students attending schools in metropolitan areas by about one quarter of a standard deviation (Figure 2.13). This difference was similar at each end of the distribution but may have increased slightly with time across the full distribution.

While students with high socioeconomic background (ESCS above the national median) generally performed better than students with low socioeconomic background, the difference was slightly smaller among higher achieving students than lower achieving students (Figure 2.14). This pattern was quite stable across time, but the difference may have increased slightly at the high end of the reading scale.

Overall, differences in reading achievement between subgroups were usually smaller among higher achievers than among lower achievers. Regarding changes over time, the difference in achievement between students attending schools in provincial or remote areas and those attending schools in metropolitan areas appears to have increased. In addition, the difference in reading achievement at the bottom end of the distribution between students born overseas and those born in Australia appears to have disappeared.

NATIONAL PERSPECTIVES ON READING FROM FIVE YEARS OF NAPLAN

Australia's National Assessment Program – Literacy and Numeracy (NAPLAN) began in 2008 and was a successor to the range of jurisdictional assessment programs that had previously operated. NAPLAN assessments 'broadly reflect aspects of literacy and numeracy within the curriculum in all states and territories' (ACARA, 2011: iv). This means that the assessment frameworks for the domains (e.g. in reading) are pragmatically based in jurisdictional curricula rather than in published assessment frameworks.

NAPLAN reports achievements for reading, numeracy, writing and language conventions (spelling, grammar and punctuation) but this discussion is focused on reading. Results from NAPLAN are reported on measurement scales that constitute a continuum of increasing achievement with each student being placed at a location on the continuum that represents his or her achievement. The difficulties of test items are statistically calibrated on scales that enable performances on different tests, across year levels and over time, of the same construct to be reported and compared on the same numerical scale. Because the Years 3, 5, 7 and 9 reading tests are calibrated on the same measurement scale (using common items), then a good performance on the Year 3 test may lead to exactly the same scale score as an average performance on the Year 5 test. Thus, NAPLAN scales are not limited to a particular year of school, but extend from Year 3 to Year 9 enabling student progress to be monitored. Similarly, a NAPLAN reading score of 345 will represent the same level of reading proficiency in 2014 as it represented in 2011.

NAPLAN reporting metrics and summary statistics

Student reading achievement can be represented as a score on the reading scale, in terms of the location of that score in a range called a proficiency band or whether the score was at or above a level called the national minimum standard.

Ten proficiency bands along each scale are defined for the reading scale using an approach similar to that developed for PISA. The proficiency bands are based on the difficulties of the items and encompass a range of difficulties so that all bands are of equal width. This ensures that the notion of being at a level can be interpreted consistently and in line with the fact that the achievement scale is a continuum. The ten proficiency bands encompass the full range of the scale. For each year level, six of these bands encompass most of the range of student achievement. And for each year level, a national minimum standard is defined and located on the NAPLAN scale. For Year 3, Band 2 is the national minimum standard; for Year 5, Band 4; Year 7, Band 5; and Year 9, Band 6.

The student reports for NAPLAN contain summaries of the skills assessed by the items in each proficiency band in relation to the year level concerned. These summaries are based on skills included in the test for that year level. Thus, the results are curriculum referenced even though the underlying scale is common to all year levels.

National trends in NAPLAN reading

National means for NAPLAN reading over the period from 2008 to 2012 are recorded in Table 2.9. These data show that, nationally, there has been a steady improvement in reading achievement among Year 3 students. The increase over the five years has been 19 points (or 0.22 of standard deviation[16]). There was also a smaller increase of nine points (equivalent to 0.12 of a standard deviation[17]) in Year 5 reading. However, rather than being a result of steady growth this mean has fluctuated over the five years. There were no significant changes in the national means for Year 7 and Year 9 reading.

16 The standard deviations for NAPLAN Reading at Year 3 have been 85, 86, 83, 88 and 88 over the five years from 2008 to 2012.

17 The standard deviations for NAPLAN Reading at Year 5 have been 77, 78, 76, 76, and 78 over the five years from 2008 to 2012.

Table 2.9 National mean scores for NAPLAN reading from 2008 to 2012

	2008	2009	2010	2011	2012	Difference
	Mean	Mean	Mean	Mean	Mean	2012–2008
Year 3	401 (0.6)	411 (0.6)	414 (0.6)	416 (0.6)	420 (0.6)	**19**
Year 5	484 (0.5)	494 (0.5)	487 (0.5)	488 (0.6)	494 (0.6)	**9**
Year 7	537 (0.7)	541 (0.7)	546 (0.7)	540 (0.7)	541 (0.7)	5
Year 9	578 (0.8)	581 (0.7)	574 (0.8)	580 (0.8)	575 (0.8)	-3

Notes:
1 Standard errors are shown in parentheses.
2 Significant differences between 2008 and 2012 are shown in bold.
Sources: ACARA (2012a) and national reports for previous years.

The steady improvements in Year 3 reading give some cause for optimism in terms of the efforts that have been applied to the early years of school and to the years before school. Those efforts have been considerable, have been in place over some time and have encompassed changes outside of, as well as in, education. The lack of any improvement in Years 7 and 9, and the fact that there has only been a very small change in Year 5, suggests that reform on a similar scale to the early years is required at those levels.

Distributions of NAPLAN Year 3 reading scores

Ainley and Khoo (under review) have investigated the change in the distribution of NAPLAN Year 3 reading scores by examining shifts in distributions over proficiency levels between 2008 and 2012. That analysis indicates that the shifts at the top of the distribution were greater than the shifts at other points of the distribution. This can be illustrated in the relative changes in the percentages of students in various proficiency bands shown in Table 2.10. The data in Table 2.10 show that the percentage of students in Bands 1 and 2 combined dropped by four percentage points (from 18 to 14 per cent)[18] whereas the percentage of students in Band 6 increased by eight percentage points (from 18 to 26 per cent).[19]

Table 2.10 Distributions of percentages of Year 3 students across NAPLAN reading proficiency bands: 2008 and 2012

	Bands 1 & 2	Band 3	Band 4	Band 5	Band 6
2008	18.1	17.7	23.0	21.8	17.7
2012	13.8	15.7	21.5	21.5	25.5
Change	-4.3	-2.0	-1.5	-0.3	+7.8

Sources: ACARA (2012a); ACARA (2008)

Hence, it appears that the increase in the percentage in Band 6 is rather greater (almost double) than the decrease in the percentage in Bands 1 and 2 combined. There are several possible interpretations of this shift in the distribution of Year 3 reading scores. One is that some parents make greater use of the educational opportunities in the years before school and that the effects of this become manifest in the early school years of school. Another possibility is that students who have developed greater expertise in reading are better able to benefit from teaching in the early years and grow more rapidly. However, the results do raise doubts about whether the early years' initiatives have been successful in providing a more even start to schooling. The OECD Education Policy Outlook for Australia notes

18 It is necessary to consider Bands 1 and 2 together because there are so few students in Band 1.

19 The standard errors for the percentages in bands are not published for 2012 but based on data for 2008 and other factors. It is estimated that the standard errors for the percentages shown would be approximately 0.4 units. Consequently these shifts would be likely to be statistically significant.

that 'a national early childhood development strategy was developed to give all children the best start in life' (OECD, 2013: 7). These results suggest that, in terms of reading development, there could be greater attention to ensuring an even start.

Cross-jurisdictional comparisons of changes in NAPLAN reading

The jurisdictional comparisons in Table 2.11 indicate that there had been significant improvements between 2008 and 2012 in Year 3 reading in New South Wales, Victoria, Queensland, Western Australia, Tasmania and the ACT. All of these had been steady improvements over the five years with the largest being the 37 point increase in Queensland (which is equivalent to 0.42 standard deviations). There was a 23 point increase in the ACT, a 21 point increase in Western Australia, and an 18 point increase in Tasmania. New South Wales and Victoria recorded increases of 14 and 12 points respectively.[20] There was no significant change for South Australia or the Northern Territory.

It is of interest that, while there have been substantial initiatives in early school and pre-school education in all of these jurisdictions, in Queensland there were structural changes to the introduction of Year K (or preparatory year) in schools at this time. Ainley and Khoo (under review) have reported that the increase in Year 3 reading scores in Queensland was more at the 80th and 90th percentiles than at other points of the distribution.

Table 2.11 Jurisdictional mean scores for NAPLAN reading in 2008 and 2012

	Year 3		Year 5		Year 7		Year 9	
	2008	2012	2008	2012	2008	2012	2008	2012
New South Wales	412 (0.9)	**426** (1.0)	495 (1.0)	500 (1.0)	543 (1.5)	546 (1.5)	583 (1.4)	578 (1.4)
Victoria	420 (0.8)	**432** (1.0)	497 (0.8)	504 (0.9)	543 (1.3)	548 (1.3)	585 (1.5)	582 (1.5)
Queensland	371 (1.3)	**408** (1.2)	466 (1.2)	**480** (1.2)	528 (1.1)	533 (1.0)	568 (1.6)	567 (1.6)
Western Australia	387 (1.5)	**408** (1.7)	474 (1.4)	483 (1.5)	527 (1.4)	**538** (1.5)	570 (2.3)	572 (2.4)
South Australia	401 (1.6)	409 (1.8)	478 (1.5)	484 (1.6)	533 (1.4)	537 (1.5)	575 (2.5)	570 (2.5)
Tasmania	401 (2.4)	**419** (3.6)	476 (2.4)	**492** (2.7)	534 (3.6)	541 (3.8)	579 (3.7)	571 (3.8)
ACT	421 (2.9)	**444** (3.0)	503 (2.8)	**519** (3.6)	558 (5.0)	559 (4.2)	602 (5.0)	597 (4.5)
Northern Territory	307 (9.9)	332 (10.0)	405 (9.0)	405 (11.9)	468 (11.0)	474 (11.3)	524 (10.8)	516 (10.3)

Notes:
1 Where differences between 2008 and 2012 are statistically significant the 2012 mean has been highlighted as bold.
2 Standard errors shown in parentheses estimated from published confidence intervals.
Sources: ACARA (2012a); ACARA (2008)

There were other improvements over the five years from 2008 to 2012. There was an overall improvement for Year 5 in Queensland (14 points), Tasmania (16 points) and the ACT (16 points). In addition there was an improvement in the reading score for Year 7 in Western Australia (11 points). The improvement in reading achievement for Year 7 students in Western Australia may have been associated with the changes to school entry (principally a school year prior to Year 1) introduced some years earlier.

The data in Table 2.11 also indicate that the means for New South Wales, Victoria and the ACT are consistently higher than the national mean (see Table 2.9) and that the means for the Northern Territory and Queensland, as well as South Australia and Western Australia for Year 3 and Year 5, are

20 The apparent 25 point increase in the Northern Territory is not statistically significant because of the uncertainty in the estimates of the means for each of 2008 and 2012. This is a consequence of a small population and a large spread of scores.

consistently lower than the national mean. The means for Tasmania in 2012, and in 2008, were not significantly different from the national mean.

Dispersion of NAPLAN reading scores

Table 2.12 records the standard deviations for the distributions of NAPLAN reading scores for each Year level and each cycle of NAPLAN. The standard deviation is a measure of variability in NAPLAN reading scores for the specified group.[21] A larger standard deviation indicates a wider spread of scores. These data are based on the national NAPLAN reports for each assessment cycle (ACARA, 2008; 2012a). The data recorded in Table 2.12 indicate that there has been a very small increase in the dispersion of Year 3 reading scores nationally with small increases in Tasmania, Victoria, New South Wales and the ACT. However, it is not possible from the published data to determine whether these small changes are statistically significant.

Table 2.12 Standard deviations for NAPLAN reading by jurisdiction: 2008 and 2012

	Year 3	Year 5	Year 7	Year 9
NAPLAN Reading 2012				
New South Wales	86	77	70	68
Victoria	82	71	64	65
Queensland	87	76	65	64
Western Australia	91	79	67	67
South Australia	85	74	65	65
Tasmania	94	81	70	69
ACT	88	76	68	69
Northern Territory	127	132	107	101
Australia	88	78	68	67
NAPLAN Reading 2008				
New South Wales	80	75	69	67
Victoria	75	69	63	63
Queensland	85	78	67	68
Western Australia	88	77	67	66
South Australia	81	71	65	64
Tasmania	84	76	69	68
ACT	82	72	70	68
Northern Territory	134	123	108	102
Australia	85	77	68	67

Sources: ACARA (2012a); ACARA (2008)

There did not appear to be any changes in the dispersion of scores at any other year level. Those data also show that there is a wider dispersion of scores in Year 3 than in Year 5 and in turn than Year 7 and Year 9 (which are not significantly different from each other). The dispersion of reading scores is wider in the Northern Territory than in other jurisdictions at each year level and for each cycle.

Differences among groups of students in NAPLAN reading scores

Table 2.13 records the mean scores for groups of students for each year level for NAPLAN reading in 2012. There are substantial missing data on parental education and occupation and especially in some jurisdictions and for the earlier NAPLAN cycles.[22] For that reason this discussion focuses on the national pattern and on data for 2012. It is not possible to report the statistical significance of the differences between groups or across cycles because NAPLAN reports from 2009 onwards do not include confidence intervals for these groups in published reports.

21 In a normal distribution approximately 68 per cent of students' reading scores would be between minus one and plus one standard deviation around the mean.

22 In 2008 data on parental education and occupation were missing in approximately 45 per cent of cases.

On the basis of the confidence intervals shown in the 2008 national report it would be expected that all differences between males and females, Indigenous and non-Indigenous, adjacent location categories, adjacent parental occupation categories and adjacent parental education categories are significant at every year level. The difference between students of a language background other than English and other students may not be consistently significant.

In terms of the magnitude of the differences between groups it appears that the average differences in 2012 between the top and bottom categories of parental education (bachelor degree or higher compared with Year 11 or below) and between Indigenous and non-Indigenous students were each a little greater than one standard deviation. The average difference between top and bottom categories of parental occupation (senior managers and professionals compared with unskilled manual and service) was about four-fifths of a standard deviation and the average difference between metropolitan and remote students was approximately three-fifths of a standard deviation. There were small differences between males and females of about one-fifth of a standard deviation and between LBOTE and non-LBOTE students of about one-tenth of a standard deviation.

Table 2.13 National means on NAPLAN reading by student characteristics: 2012

	Year 3	Year 5	Year 7	Year 9
Sex				
Males	413	486	535	568
Females	427	501	548	582
Difference	-14	-15	-13	-14
Indigenous status				
Indigenous	333	409	475	510
Non-Indigenous	424	498	545	578
Difference	-91	-81	-70	-68
Language background				
Language background other than English	417	486	535	569
Language background English	421	496	543	578
Difference	-4	-10	-8	-9
Location				
Metropolitan	427	500	547	580
Provincial	405	483	532	566
Remote	373	452	508	543
Very Remote	295	355	438	474
Difference (metro–provincial)	22	17	15	14
Difference (metro–remote)	54	48	39	37
Parental occupation				
Senior management and qualified professionals	461	532	577	611
Other business managers and associate professionals	434	507	554	587
Tradespeople, clerks, skilled office, sales and service staff	410	487	534	567
Machine operators, hospitality staff, assistants, labourers	390	468	517	551
Not in paid work in the previous 12 months (very small)	377	455	502	537
Not stated or missing (13%)	390	466	522	556
Difference (senior – unskilled)	71	64	60	60
Parental education				
Bachelor degree or higher	462	533	578	613
Advanced diploma/diploma	423	499	547	581
Certificate 1 to 4	402	480	529	563
Year 12 or equivalent	401	482	531	566
Year 11 or equivalent or below	367	448	502	538
Not stated or missing (9%)	397	472	528	560
Difference (bachelor – Year 11)	95	85	76	75
Dispersion				
Standard deviation	88	78	68	67

Source: ACARA (2012a)

READING ACHIEVEMENT IN PRIMARY SCHOOL: PERSPECTIVES FROM PIRLS IN YEAR 4

The IEA Progress in International Reading Literacy Study (PIRLS) has operated a five-year cycle of assessments of reading literacy in Year 4 (or its equivalent) since 2001. The 2011 cycle (conducted in 2010 in the southern hemisphere) was the first of the three cycles in which Australia has participated. What is assessed in PIRLS 2011 is described in the *PIRLS 2011 Assessment Framework* (Mullis, Martin, Kennedy, Trong & Sainsbury, 2009). It states that reading literacy is:

> *… the ability to understand and use those written language forms required by society and/or valued by the individual. Young readers can construct meaning from a variety of texts. They read to learn, to participate in communities of readers in school and everyday life, and for enjoyment.*
>
> (Mullis, Martin, Kennedy, Trong & Sainsbury, 2009, p. 11)

The framework argues that Year 4 is important in reading development as it represents a stage when most students make the transition from learning to read to reading to learn. The framework is structured around two organising dimensions: purposes and processes. The purpose dimension refers to reading for literacy experience and reading to gain information. There are four elements in the process dimension: focusing on and retrieving explicitly stated information; making straightforward inferences; interpreting and integrating ideas and information; and examining and evaluating content, language and textual elements. In 2011 the assessment was based on 10 texts: five for the literary purpose and five for the informational purpose.

PIRLS is also concerned with investigating reading behaviours and attitudes about which data are based on responses to student and home (completed by parents/caregivers) questionnaires. National policies on literacy education are gathered and reported in an encyclopaedia and the surveys include questionnaires for teachers and school principals.

Comparisons among countries

In Australia PIRLS was conducted in 2010 with a sample of 6126 Year 4 students from 280 schools. Mean scores for the OECD countries and sub-national entities such as England, Northern Ireland and Belgium (French) within OECD countries that participated in PIRLS 2011 are recorded in Table 2.14.

According to the data in Table 2.14 the mean score for Australia is not significantly different from that of New Zealand but is significantly lower than the means of the United States, England and Canada. The differences between the means of these education systems and that of Australia were 20 or more points. The comparison with Canada is of particular interest because of the demographic similarities. It can also be seen in Table 2.14 that Australia has a relatively large dispersion of scores.

Table 2.14 National means on Year 4 PIRLS reading in 2011/2010

	Mean score	Dispersion 10th – 90th percentile
Finland	568 (1.9)	162
Northern Ireland	558 (2.4)	192
United States	556 (1.5)	190
Denmark	554 (1.7)	164
Ireland	552 (2.3)	191
England	552 (2.6)	212
Canada	548 (1.6)	176
Netherlands	546 (1.9)	139
Czech Republic	545 (2.2)	156
Sweden	542 (2.1)	165
Italy	541 (2.2)	167
Germany	541 (2.2)	168
Israel	541 (2.7)	218
Portugal	541 (2.6)	169
Hungary	539 (2.9)	198
Slovak Republic	535 (2.8)	174
New Zealand	531 (1.9)	229
Slovenia	530 (2.0)	180
Austria	529 (2.0)	163
Australia	527 (2.2)	207
Poland	526 (2.1)	187
France	520 (2.6)	176
Spain	513 (2.3)	175
Norway	507 (1.9)	156
Belgium (French)	506 (2.9)	166

Notes:
1 England and Northern Ireland participated as separate entities and are recorded separately.
2 Countries with mean scores not significantly different from Australia are shaded in green.
3 Standard errors are shown in parentheses.
Source: Mullis, Martin, Foy, and Drucker (2012)

Comparisons among Australian jurisdictions

There were differences among Australian jurisdictions in the mean PIRLS reading scores of Year 4 students. Those means are recorded in Table 2.15.

The mean for the ACT is significantly higher than those of all other jurisdictions. The means for Victoria and New South Wales are not significantly different from each other but they are significantly greater than those of South Australia, Western Australia, Queensland and Northern Territory. Greater detail on jurisdictional comparisons is contained in the report by Thomson, Hillman, Wernert, Schmid, Buckley and Munene (2012).

Table 2.15 Jurisdictional means on Year 4 PIRLS reading in 2010

	Mean Score: PIRLS Reading
ACT	558 (5.3)
Victoria	539 (4.0)
New South Wales	535 (4.9)
Tasmania	525 (7.5)
South Australia	518 (4.0)
Western Australia	516 (4.5)
Queensland	511 (5.0)
Northern Territory	509 (10.3)

Note: Standard errors are shown in parentheses.
Source: Thomson, Hillman, Wernert, Schmid, Buckley & Munene (2012)

Differences among groups of students in PIRLS reading scores

Table 2.16 records the mean scores for groups of students on PIRLS reading. Those data show that the difference between Indigenous and non-Indigenous students is approximately three-fifths of a standard deviation and the difference between females and males is approximately two-fifths of a standard deviation. The difference between those of a language background other than English and other students was approximately one-fifth of a standard deviation. There were also differences related to geographic location. There was a small difference between students in metropolitan and provincial locations and a large difference between students in metropolitan and remote locations.

Table 2.16 Year 4 PIRLS reading scores in 2010 by student characteristics

	Mean		Percentage
Sex			
Male	519	(2.7)	51
Female	536	(2.7)	49
Difference	-17		
Indigenous status			
Indigenous	475	(5.5)	7
Non-Indigenous	532	(2.2)	93
Difference	-57		
Language background			
Language of the test	531	(2.0)	79
Language other than the test	513	(5.0)	21
Difference	18		
Location			
Metropolitan	532	(2.6)	72
Provincial	518	(4.5)	27
Remote	462	(17.4)	1
Difference (metro–provincial)	14		
Difference (metro–remote)	70		
Books in the home			
Many	553	(3.9)	19
Average	534	(2.3)	59
Few	489	(2.9)	22
Difference (many – few)	64		

Note:
1 Standard errors are shown in parentheses.
2 Categories for 'books in the home':
'Many' books in the home is more than 200.
'Average' number of books in the home is from 26 to 200.
'Few' books in the home is 25 or fewer books.
Source: Thomson, Hillman, Wernert, Schmid, Buckley & Munene (2012)

PIRLS did not gather data about parental education or occupation but there was an association with the reported number of books in the home. PIRLS classified 25 or fewer books as 'few', from 26 to 200 books as 'average' and more than 200 books as 'many'. The difference between the top and bottom category was a little more than three-fifths of a standard deviation.

COMPARING MEASURES OF READING ACHIEVEMENT

Comparing scale scores

One of the ways of looking at the similarities in measures of achievement is to examine the extent to which scale scores on two or more of the measures are correlated. This would provide an index of concurrent validity of the measures but would require data on two or more measures from a group of students. In principle it could be possible to examine the NAPLAN scores for the PIRLS and PISA samples if the data could be matched. However, in the case of PIRLS there would be a gap of one year in the times of data collection between it and NAPLAN. In the case of PISA it would be possible to have data reasonably close in time for the 70 per cent of the sample who were in Year 10 but one year different for the students who were in either Year 9 or Year 11 when they completed the PISA assessment.

It is possible to compare jurisdictional estimates for the relevant cohorts of students on different measures of reading achievement. Figure 2.15 contains plots of the jurisdictional means for PIRLS reading scores against the means for NAPLAN Year 3 reading in 2009 and NAPLAN Year 5 reading in 2011 (i.e. they reference the same cohort). With the exception of the Northern Territory there is a close alignment of the means.

The discrepancy in the Northern Territory is probably due to issues of sample design and participation rates as well as the large standard errors associated with those estimates. Based on the other seven jurisdictions the correlations of the PIRLS-based means and the NAPLAN-based means are 0.95 (Year 3 NAPLAN) and 0.99 (Year 5 NAPLAN).

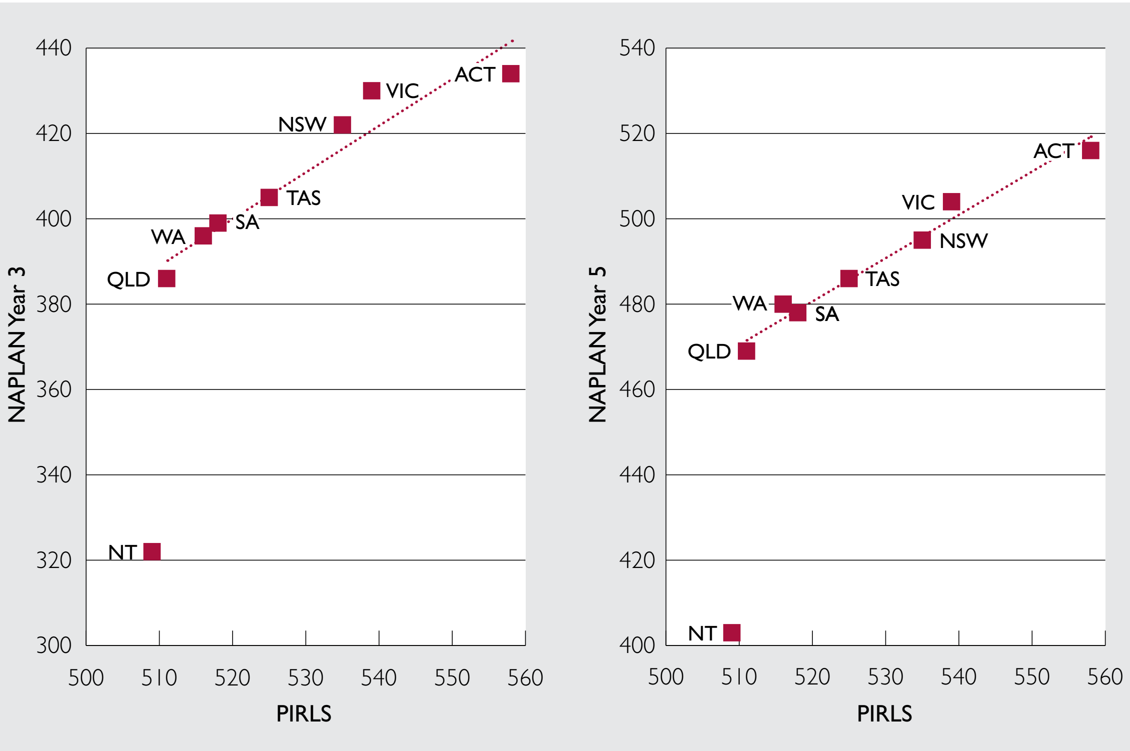


Figure 2.15 Jurisdictional means for PIRLS and NAPLAN reading in Year 3 and Year 5 for the 2010 Year 4 cohort

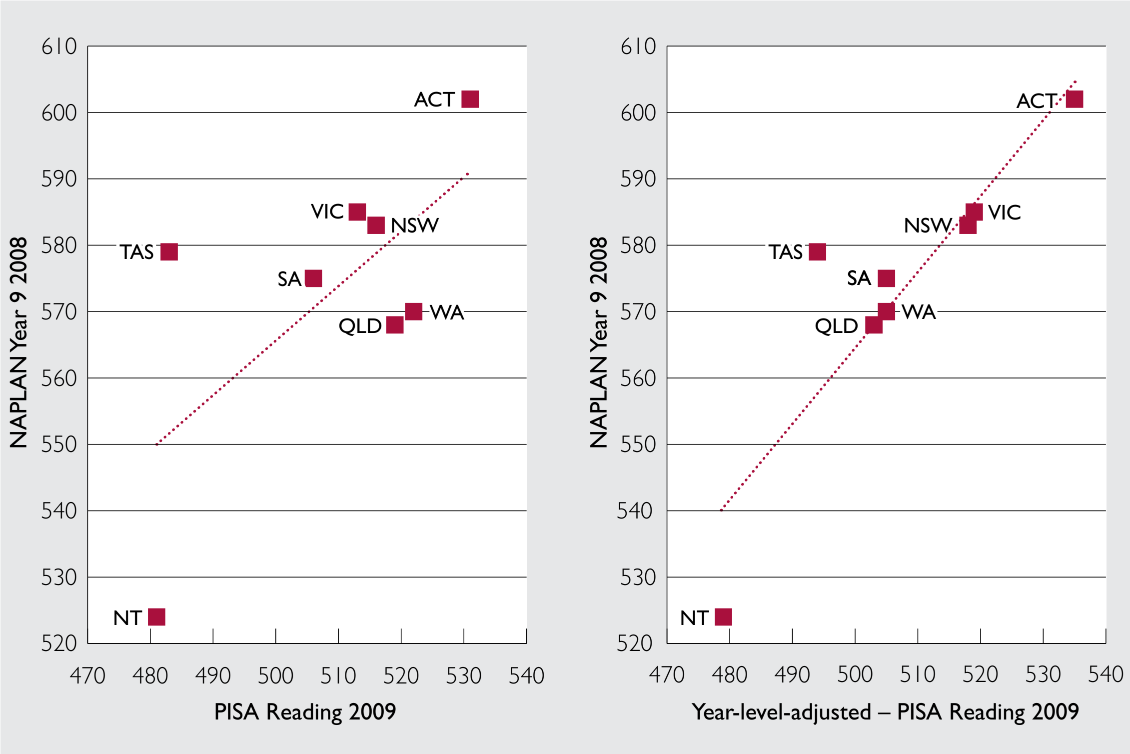


Figure 2.16 Comparisons of NAPLAN Year 9 and PISA reading means for jurisdictions

Part of the lack of match of jurisdictional comparisons based on NAPLAN compared with those based on PISA is a result of the different populations or samples. PISA is based on a sample of 15-year-olds in education where NAPLAN is based on specified year levels. As a result of differences in age by year level distributions among jurisdictions, discrepancies become evident. It is possible to adjust jurisdictional means for PISA so that adjusted scores refer to what would have been the mean had all students been in Year 10. This is based on the estimate that, in Australia, one year of school corresponds to 33 score points on the PISA reading scale (OECD, 2010a: 169). The correlation coefficient for the association between the unadjusted PISA means for reading and NAPLAN Year 9 2008 reading means was 0.66. However, the correlation coefficient for the association between year-level-adjusted PISA reading means and NAPLAN Year 9 2008 reading means was 0.89. Figure 2.16 shows that six of the jurisdictions were very close to the regression line for NAPLAN on adjusted PISA scores. The Tasmanian mean for NAPLAN reading was a little higher than would have been expected from the PISA mean and the NAPLAN reading mean for the Northern Territory was lower than would have been expected on the basis of PISA.

Reporting percentages of students in specified score ranges or above specified scores

PISA 2009 defined level 2 (of seven proficiency levels[23]) as 'a baseline level of proficiency at which students begin to demonstrate the reading literacy competencies that will enable them to participate effectively and productively in life' (OECD, 2009: 52). Across OECD countries 81 per cent of students

23 In 2000 there were six levels but in 2009 the bottom level was subdivided into two levels. The seven levels were labelled as 1b (scores from 262.0 to 334.8 points), 1a (334.9 to 407.5 points), 2 (407.6 to 480.2 points), 3 (480.3 to 552.9 points), 4 (553.0 to 625.6 points), 5 (625.7 to 698.3 points) and 6 (698.4 and above). In addition the percentage of students who scored below level 1b is reported.

could successfully perform tasks at least at Level 2. For Australia 86 per cent of students performed at Level 2 or above.

For NAPLAN reading in 2012 the percentages of students 'at or above the national minimum standard' were 94, 92, 94 and 91 per cent for Years 3, 5, 7 and 9 respectively. The percentages 'above the national minimum standard' were 84, 80, 81 and 74 per cent respectively. The figures most widely referred to by education authorities are the percentages of students 'at or above the national minimum standard' but the percentages of students 'above the national minimum standard' would provide figures more similar to those generated by PISA and PIRLS as representing minimum competence. Those figures are also more similar to those based on two other large-scale assessments. The Australian Studies of School Performance in 1975 reported that 72 per cent of 14-year-old students, and 53 per cent of 10-year-old students, had attained mastery of reading appropriate to their age (Bourke & Keeves, 1977: 54). Similarly a report based on a standards setting exercise using data from the 1996 *National School English Literacy Survey* indicated that 73 per cent of Year 3 students and 71 per cent of Year 5 students met the performance standard appropriate to their year level (Masters & Forster, 1997b).

The student reports for NAPLAN contain summaries of the skills assessed by the items in each proficiency band in relation to the year level concerned. These summaries are based on skills included in the test for that year level. This is so that the results are curriculum referenced even though the underlying scale is common to all year levels.

The description of reading at the national minimum standard for Year 3 (Band 2) reads as:

> *Makes some meaning from short texts, such as simple reports and stories, which have some visual support. Makes connections between pieces of clearly stated information.*

Reading performance in Band 1 for Year 3 is described as:

> *Makes some meaning from simple texts with familiar content. Texts have short sentences, common words and pictures to support the reader. Finds clearly stated information.*
>
> (ACARA, 2012c)

Although there is no minimum standard of proficiency specified for PIRLS, Thomson and colleagues adopt the minimum standard set for TIMSS in mathematics and science, which is the Intermediate benchmark, as a reference point. This means that students achieving at or above the low benchmark: 'when reading literary texts, can locate and retrieve an explicitly stated detail' and 'when reading informational texts can locate and reproduce explicitly stated information that is at the beginning of the text' (Thomson, Hillman, Wernert, Schmid, Buckley & Munene, 2012: 13). Seventy-six per cent of Australian Year 4 students had attained this standard compared to 86 per cent in Canada and the United States and 83 per cent in England.

Even though the processes for standards setting have been well documented they appear to be implemented in different ways in different studies and there is nothing common in the meaning that can be associated with apparently similar terms such as 'minimum standard' or 'benchmark'. Consequently there is no basis for comparing the percentages of students above designated levels without a detailed examination of the items that constitute particular difficulty levels.

SUMMARY

Over the period from 2000 to 2009 there was a small decline in average reading achievement for students in the middle of the period of secondary school (those aged 15 years) which was a little more pronounced at the higher levels of achievement than at the lower levels of achievement. Over the same period it appeared that there was an increase in the differences among schools in reading achievement which was associated with differences among schools in socioeconomic background. Although the average reading achievement of students in the middle secondary years remained high compared with other OECD countries this was not the case for students in the middle primary school years (Year 4). The small overall decline in middle secondary school was not observed in all states and territories. Victoria, Queensland and the Northern Territory showed no changes in average performance in the middle secondary years between 2000 and 2009.

Over the period from 2008 to 2012 there was evidence of an improvement in reading achievement in early primary school (Year 3) that was more pronounced at the upper end of the achievement distribution and more pronounced in those jurisdictions where there had been changes in school entry and the first year of school. This general improvement may have been the result of the various initiatives that focused on the early years (in school and prior to school).

MATHEMATICS AND NUMERACY

3

Mathematics occupies a substantial part of the total curriculum in schools. The results of the TIMSS surveys of mathematics teachers indicate that, in Australia, mathematics teaching takes up 23 per cent of teaching time in Year 4 and 14 per cent of curriculum time in Year 8 (Mullis, Martin, Foy & Arora, 2012: 341–345).[24] It is possible that the time devoted to numeracy activities might be greater than this if account is taken of numerical activities in other learning areas.

This chapter incorporates consideration of achievement measures designated as mathematics (in TIMSS), mathematical literacy (in PISA) and numeracy (in NAPLAN). There are subtle differences in the meaning of these terms with numeracy emphasising 'the key role of applications and utility in learning the discipline of mathematics' (ACARA, 2009: 5), and mathematical literacy connoting the 'capacity to identify and understand the role that mathematics play in the world, to make well-founded judgements and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen' (OECD, 2009, p.14). However, mathematics, mathematical literacy and numeracy are rather close in meaning (AAMT, 1997: 11–12; COAG, 2008: 6). Differences may be more evident in curriculum and teaching than in assessments.

MATHEMATICS IN LOWER AND MIDDLE SECONDARY SCHOOL

Achievement in mathematical literacy in PISA 2003 and 2009

On the basis of the data in Table 3.1 it can be inferred that Australian 15-year-olds performed well in mathematical literacy in 2003 and moderately well in 2009. Australian 15-year-olds performed similarly to their peers from New Zealand, Belgium and Germany but significantly less well than 15-year-olds from 12 participating countries including six OECD countries: Korea, Finland, Switzerland, Japan, Canada, and the Netherlands (Thomson et al., 2010: 52). The average score for Australian students in mathematical literacy was 514 scale points in 2009 compared to the average of 496 points for the 28 OECD countries with data for both 2003 and 2009. The scale had a standard deviation a little less than 100 points.

24 Based on teachers' logs of activities Angus, Olney and Ainley (2007) estimated that mathematics took up 18 per cent of available teaching time in Australian primary schools.

The spread of student scores in mathematical literacy for Australia, as indicated by the standard deviations, was not significantly different from the average spread in the other 27 OECD countries in 2003 and 2009. The ratio of variances in 2009 to 2003 indicates that the spread of scores in Australia did not change. This similarity of the spread of student scores in mathematics to the OECD average is different from the pattern observed for reading where the Australian spread of scores was wider. In mathematics there is no evidence of a wider spread.

Table 3.1 OECD country-level PISA mathematics statistics for 2003 and 2009

	PISA 2003		PISA 2009		Change	
	Mean score	Standard deviation	Mean score	Standard deviation	Mean difference	Ratio of variances[2]
Korea	542 (3.2)	92	546 (4.0)	89	4	0.93
Finland	544 (1.9)	84	541 (2.2)	82	-4	0.97
Switzerland	527 (3.4)	98	534 (3.3)	99	7	1.02
Japan	534 (4.0)	101	529 (3.3)	94	-5	0.88
Canada	532 (1.8)	87	527 (1.6)	88	-6	1.01
Netherlands	538 (3.1)	93	526 (4.7)	89	**-12**	0.93
New Zealand	523 (2.3)	98	519 (2.3)	96	-4	0.96
Belgium	529 (2.3)	110	515 (2.3)	104	**-14**	**0.90**
Australia	524 (2.1)	95	514 (2.5)	94	**-10**	0.97
Germany	503 (3.3)	103	513 (2.9)	98	**10**	0.92
Iceland	515 (1.4)	90	507 (1.4)	91	**-8**	1.01
Denmark	514 (2.7)	91	503 (2.6)	87	**-11**	**0.91**
Norway	495 (2.4)	92	498 (2.4)	85	3	**0.86**
France	511 (2.5)	92	497 (3.1)	101	**-14**	**1.21**
Slovak Republic	498 (3.3)	93	497 (3.1)	96	-2	1.06
Poland	490 (2.5)	90	495 (2.8)	88	5	0.96
Sweden	509 (2.6)	95	494 (2.9)	94	**-15**	0.98
Czech Republic	516 (3.5)	96	493 (2.8)	93	**-24**	0.94
Luxembourg	493 (1.0)	92	489 (1.2)	98	-4	**1.13**
Hungary	490 (2.8)	94	490 (3.5)	92	0	0.97
Ireland	503 (2.4)	85	487 (2.5)	86	**-16**	1.01
Portugal	466 (3.4)	88	487 (2.9)	91	**21**	1.09
United States	483 (2.9)	95	487 (3.6)	91	5	**0.91**
Italy	466 (3.1)	96	483 (1.9)	93	**17**	0.95
Spain	485 (2.4)	88	483 (2.3)	91	-2	1.05
Greece	445 (3.9)	94	466 (3.9)	89	**21**	0.91
Turkey	423 (6.7)	105	445 (4.4)	93	**22**	**0.79**
Mexico	385 (3.6)	85	419 (1.8)	79	**33**	**0.85**
Average-28	500 (0.5)	94	499 (0.5)	92	0	0.98

Note:
1 Standard errors are shown in parentheses.
2 This is the ratio of the variances is the squared 2009 standard deviation divided by the squared 2003 standard deviation. Ratios less than one mean a decrease in spread and ratios more than one reflect an increase. The ratios have been tested for significant difference from one. Computations were performed for this report.
3 Differences and variance ratios that are statistically significant have been shown in bold.
4 Countries listed in order of mean scores for 2009. Countries shown with shading are not significantly different from Australia in 2009.
5 The average shown is for 28 OECD countries with data for 2003 and 2009.
6 Differences in means that are recorded may be affected by rounding errors.

Data Source: OECD (2010b) *PISA 2009 Results: Learning Trends*

Change over time

The data in Table 3.1 also indicate that, between 2003 and 2009, the average achievement in mathematical literacy for Australia declined from 524 to 514; a difference that is small but statistically significant. There was no change in the spread of mathematics scores (with the standard deviation being 95 in 2003 and 94 in 2009). Other OECD countries to record a significant decline in mathematics scores from 2003 to 2009 were the Czech Republic (24 points), Ireland (16 points), Sweden (15 points), Belgium (14 points), France (14 points), the Netherlands (12 points), Denmark (11 points) and Iceland (8 points). OECD countries to record a significant increase over the same period were Mexico (33 points), Turkey (22 points), Portugal (21 points), Greece (21 points), Italy (17 points) and Germany (10 points).

For 2003 where mathematical literacy was the major domain it was possible to consider the subscales of mathematical literacy. In that cycle Australian students were strongest on the uncertainty subscale and weakest on the quantity subscale. Scores on the space and shape and change and relationships subscales were very close to the overall mathematical literacy scores (Thomson, Cresswell & De Bortoli, 2004).

Distribution of scores

Data regarding the distributions in percentiles of student scores in mathematical literacy in 2003 and 2009 are recorded in Table 3.2. An examination of differences in percentiles suggests that, superimposed on the overall decline in scores, the negative difference is observed slightly more above the median than below the median. As was observed for reading literacy, but less clearly for mathematics, there did appear to be a decline in the percentage of students in mathematics proficiency level 5 and above (20% in 2003 compared to 16% in 2009) but no significant change in the percentage of students below level 2 (14% in 2003 compared to 16% in 2009). The larger drop in the percentage in the upper proficiency levels compared to the lack of change in the bottom proficiency levels indicates that there has been a small change in the shape of the distribution. However, as shown in Table 3.2, the change in the 95th percentile was not statistically significant.

Table 3.2 Distribution of achievement in PISA mathematical literacy for Australia in 2000 and 2009

	PISA 2003	PISA 2009		
	Mean	Mean	Difference	Significance of difference
Distribution of scores				
95th percentile	676 (3.5)	665 (5.0)	11	No
90th percentile	645 (3.0)	634 (3.9)	11	**Yes**
75th percentile	592 (2.5)	580 (3.1)	12	**Yes**
50th percentile	527 (2.4)	516 (2.5)	11	**Yes**
25th percentile	460 (2.7)	451 (2.5)	9	**Yes**
10th percentile	399 (3.4)	392 (2.8)	6	No
5th percentile	364 (4.4)	357 (3.3)	8	No
Percentage of students in proficiency levels				
Level 5 & above	20 (0.8)	16 (0.7)	4	**Yes**
Below level 2	14 (0.6)	16 (0.7)	2	No

Note: Standard errors are shown in parentheses
Sources: OECD (2010b); Thomson et al. (2011)

Differences among jurisdictions

Table 3.3 indicates that there were differences among jurisdictions in the change in mean mathematical literacy scores between 2003 and 2009. In South Australia (27 points), the ACT (20 points), Western Australia (19 points) and New South Wales (14 points) there were significant declines. There were no significant changes in Tasmania (although the apparent decline was 20 points but not statistically significant), the Northern Territory, Queensland or Victoria (Thomson et al., 2011). There is a high correlation between the jurisdictional declines in mathematics between 2003 and 2009 and the jurisdictional declines in reading between 2000 and 2009 (the between-jurisdiction correlation coefficient was 0.91). This suggests that the decline is not associated with specific curriculum provision but more likely general changes in structures or resources (including the availability of qualified teachers).

Table 3.3 Jurisdictional mean mathematical literacy achievement in Australia: PISA 2003 and PISA 2009

Jurisdiction	PISA 2003 Mean		PISA 2009 Mean		Difference (PISA 2009 – PISA 2003)
Western Australia	548	(4.1)	529	(7.2)	**-19**
ACT	548	(3.5)	528	(6.4)	**-20**
Queensland	520	(6.9)	518	(7.5)	-2
New South Wales	526	(4.3)	512	(5.2)	**-14**
Victoria	511	(5.1)	512	(4.9)	1
South Australia	536	(4.9)	509	(5.3)	**-27**
Tasmania	507	(9.4)	487	(5.1)	-20
Northern Territory	496	(4.9)	487	(4.9)	-9

Notes:
1 Standard errors are shown in parentheses.
2 Differences that are statistically significant are highlighted in bold.
Source: Thomson et al. (2011: 198)

Changes in differences associated with social and demographic characteristics of students

Table 3.4 records the mean scores in mathematical literacy scores for various groups of students in 2003 and 2009 as well as the changes in the means for those groups over the period. Over the period from 2003 to 2009 there was no change in the relative performance of females and males, Indigenous and non-Indigenous, students of different socioeconomic background (as measured by the index of economic, social and cultural status) or students in different geographic locations.

There was a change in the difference in mathematics scores between students whose home language was English and those whose home language was a language other than English. This arose as a result of a decline in the achievement of the former group while there was no significant change in the achievement of the latter group (Thomson et al., 2011).

Table 3.4 PISA mathematical literacy for groups of Australian students in 2003 and 2009

	PISA 2003		PISA 2009		Significance of difference
	Mean		Mean		
Sex					
Females	515	(2.9)	509	(2.8)	
Males	526	(3.2)	519	(3.0)	
Difference	-11	(4.3)	-10	(4.1)	
Indigenous status					
Non-Indigenous	526	(2.1)	517	(2.5)	*
Indigenous	440	(5.4)	441	(5.3)	
Difference	86	(5.8)	76	(5.9)	
Language background					
English language at home	529	(2.0)	516	(2.2)	*
LBOTE	505	(6.1)	517	(8.9)	
Difference	24	(6.4)	-1	(9.2)	*
Immigrant status					
Australian born (AB)	527	(2.1)	511	(2.5)	*
First generation (FG)	522	(4.7)	526	(3.3)	
Overseas born (OB)	525	(4.9)	518	(6.4)	
Difference (AB–FG)	5	(5.1)	-15	(4.1)	*
Difference (AB–OB)	2	(6.8)	-7	(7.2)	
Location					
Metropolitan	528	(2.5)	520	(3.1)	*
Provincial	515	(4.4)	499	(3.7)	*
Remote	493	(9.6)	465	(15.8)	
Difference (metro–provincial)	13	(5.1)	21	(4.8)	
Difference (metro–remote)	35	(10.6)	55	(16.2)	
Economic, social and cultural status (ESCS)					
Top quarter	572	(2.9)	561	(3.1)	
Upper quarter	537	(3.1)	530	(3.0)	*
Lower quarter	513	(2.3)	503	(2.5)	
Bottom quarter	479	(4.1)	471	(2.6)	*
Difference (Top–Bottom)	93	(5.0)	90	(4.0)	

Notes:
3 Standard errors are shown in parentheses.
4 Differences between groups that are significant are shown in bold.
5 Differences across cycles that are significant are designated with an asterisk *.
Sources: Thomson et al. (2011); Thomson, Cresswell & De Bortoli (2004)

CHANGES IN DIFFERENCES BETWEEN SUBGROUPS AT VARIOUS PERCENTILES ON PISA MATHEMATICS

As for PISA reading achievement, quantile regression analyses were conducted for PISA mathematics in each cycle to investigate whether differences between subgroups were the same across the distribution of scores and whether differences at different points on the distribution changed over time. The same pairs of subgroups were compared for mathematics as had been compared for reading. Quantile regression was used to estimate the difference in performance between each pair of groups at the 10th, 25th, 50th, 75th, and 90th percentiles of the distributions.[25] The results are shown in Figures 3.1 to 3.6.

No differences in mathematics achievement were observed between male and female students among the lower performers in three PISA cycles (Figure 3.1). Male students performed a little better than female students at the top end of the distribution (by about one-sixth of a standard deviation) in each PISA cycle. Difference in mathematics achievement between Indigenous and non-Indigenous students were equal at different points of the distributions and stable across time (Figure 3.2). The difference was around four-fifths of a standard deviation.

Differences between students with a non-English, and English, language background were small and seem to be in the advantage of English speakers at the bottom end of the scale and in the advantage of the non-English speakers at the top end of the distribution (Figure 3.3). The differences were small and possibly not statistically significant. As with language background, the difference between students born overseas and students born in Australia was small and possibly non-significant (Figure 3.4). The difference was consistently a little larger at the top end in the advantage of students that were born overseas.

Students attending schools in provincial or remote areas performed a little less well in mathematics than students attending schools in metropolitan areas (Figure 3.5). In 2003, the difference was very small for the low performers and small for the high performers. The gap seems to have increased over time, especially among the low performers. The differences in mathematics scores in favour of students from high compared to low socioeconomic backgrounds were similar for low and high performers (Figure 3.6), especially since 2006. The differences have possibly increased slightly at the higher end of the scale.

In general there were few changes in the differences across the achievement distribution. Overall, differences in mathematics achievement between subgroups were usually larger among high performers than among low performers. This is opposite to the pattern that was observed for differences in reading achievement. The gap in achievement between students attending schools in provincial or remote areas and those in metropolitan areas seemed to increase (as was the case for reading achievement). In mathematics, the increase was especially at the low end of the distribution. The difference in mathematics performance between students with high and low socioeconomic background seemed to increase slightly among the high achievers.

25 As was the case for reading, the graphs presented were used to explore patterns and testing the significance of differences and changes in differences was not conducted. The vertical axis of each graph is 100 points, which is approximately one standard deviation in performance of the total Australian population.

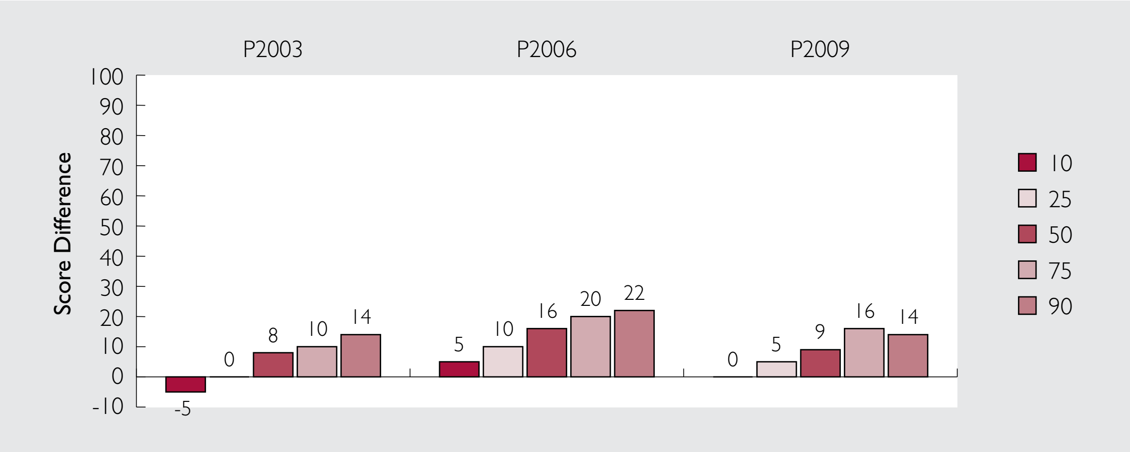


Figure 3.1 Differences in mathematics achievement between male and female students at different percentiles across time

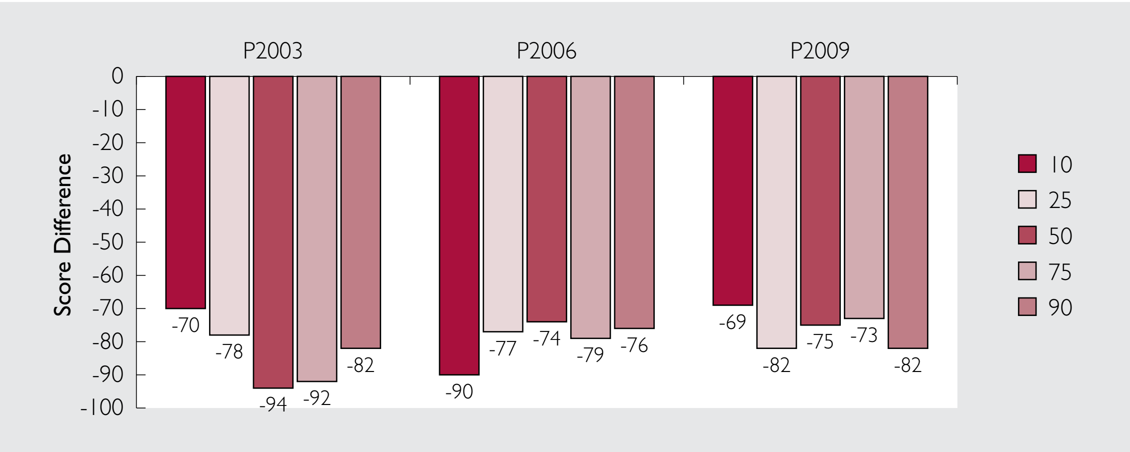


Figure 3.2 Differences in mathematics achievement between Indigenous and non-Indigenous students at different percentiles across time

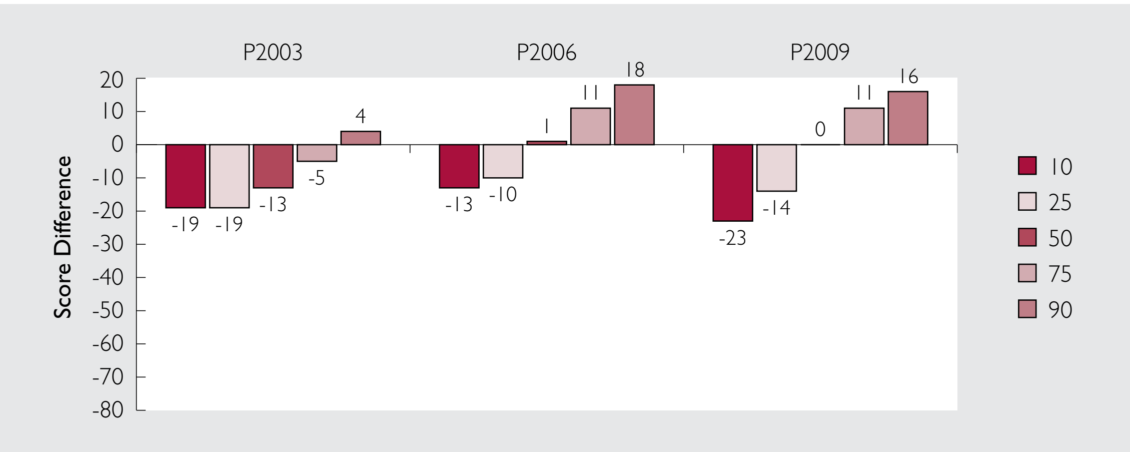


Figure 3.3 Differences in mathematics achievement between LBOTE students and non-LBOTE students at different percentiles across time

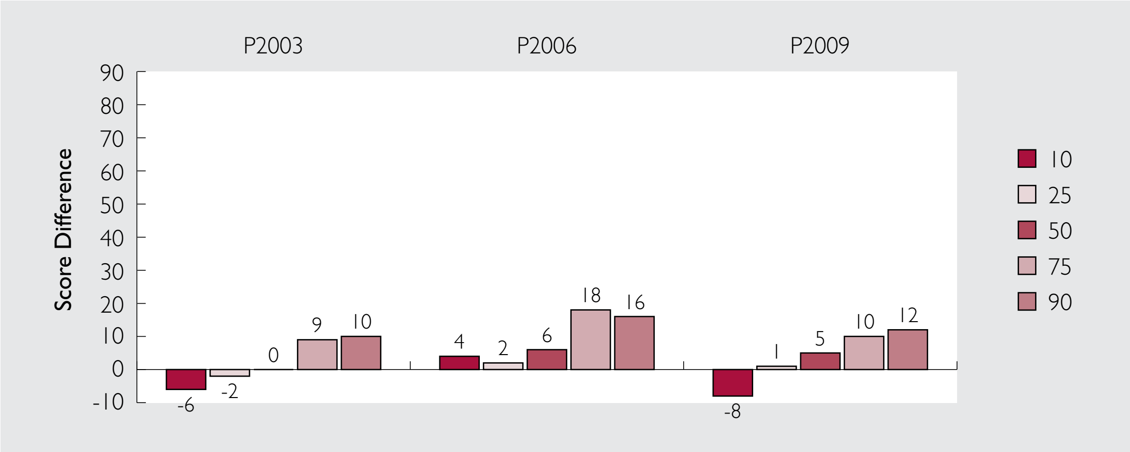


Figure 3.4 Differences in mathematics achievement of overseas-born and Australian-born students at different percentiles across time

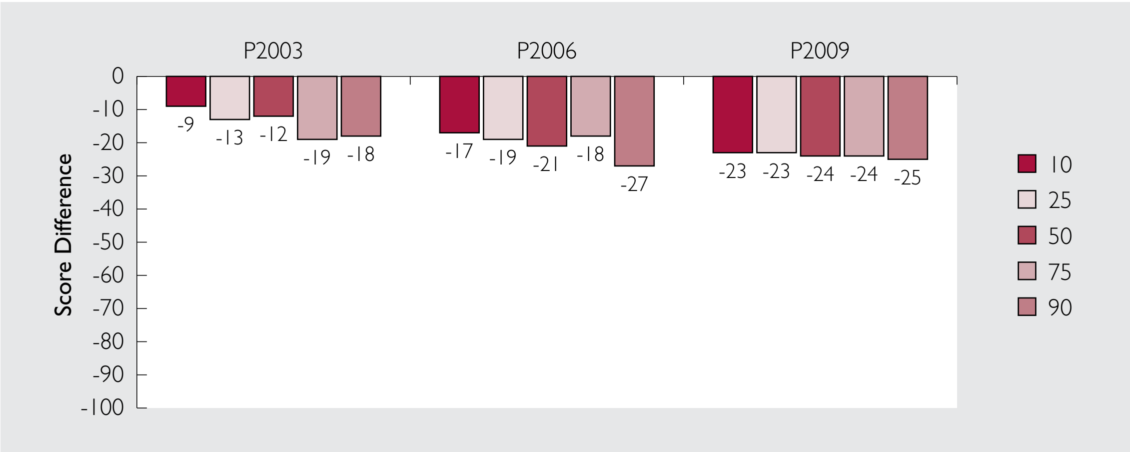


Figure 3.5 Differences in mathematics achievement of students in provincial or remote areas and metropolitan areas at different percentiles across time

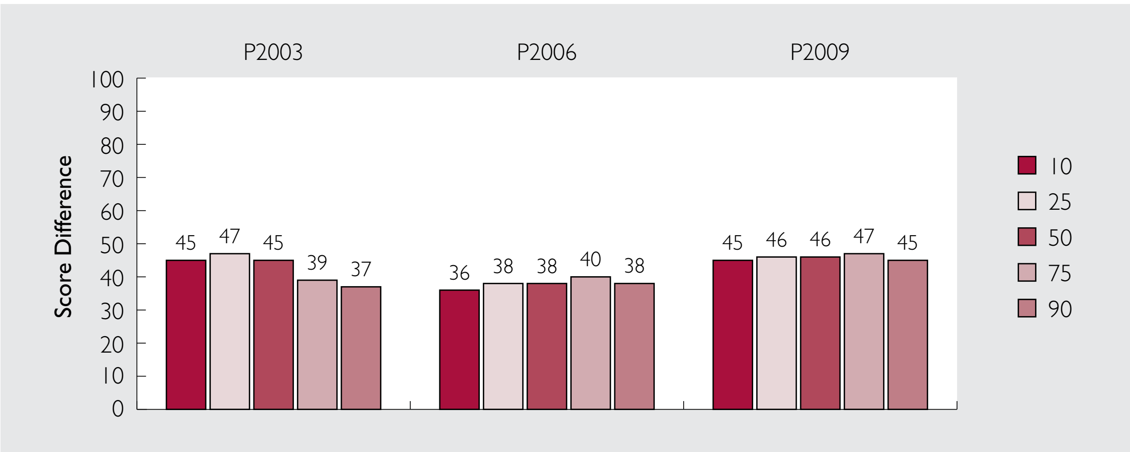


Figure 3.6 Differences in mathematics achievement of students with high ESCS and students with low ESCS at different percentiles across time

Achievement in TIMSS Mathematics in TIMSS among Year 8 students

Overall mathematics achievement in Year 8

Table 3.5 records country means for mathematics in Year 8 for selected TIMSS countries over the period from 1995 to 2011 arranged in descending order of country means for the 2011 cycle of TIMSS. The countries are those that had participated in TIMSS Year 8 in either 1995 or 1999 and also in 2011. Six countries (Korea, Singapore, Chinese Taipei, Hong Kong, Japan and the Russian Federation) achieved significantly higher mean mathematics scores than Australia. Australia had a mean score that was not significantly different from that of eight other countries in Table 3.5 (Finland, United States, England, Hungary, Slovenia, Lithuania, Sweden and Italy), as well as Israel (which has participated in TIMSS only since 2007).

Change over time

Table 3.5 also records whether the mean score in Year 8 mathematics represents a significant change from the score for previous cycles of TIMSS. The mean score for Australia in 2011 was not significantly different from that for previous cycles even though there had been a dip in the score in the 2007 cycle which was a significant drop from the 1995 mean.

Countries which had shown an improvement since 1995 were Korea (by 32 points), Hong Kong (by 17 points), Russia (by 15 points), the United States (by 17 points), Slovenia (by 11 points) as well as Chile (by 24 points since 1999) and Italy (by 10 points since 1999). England had improved between 1995 and 2007 but did not maintain that improvement into 2011.

Countries which had shown a decline since either 1995 or 1999 were Sweden (by 46 points), Norway (by 23 points), Hungary (by 22 points), Romania (by 16 points), Japan (by 11 points) and Macedonia (by 21 points) as well as countries where there may have been changes in participation such as Malaysia (by 70 points), Thailand (by 40 points), Jordan (by 22 points) and Tunisia (by 23 points).

For Singapore, Chinese Taipei, Finland, New Zealand and Iran, as well as England (see above), there were no significant changes between 1995 or 1999 and 2011.

The substantial decline in mathematics achievement in Sweden is congruent with the declines in PISA mathematics and reading as well as in national assessments (Skolverket, 2009). An analysis by the Swedish National Agency for Education (Skolverket) noted an increase in the variation in average final grades between schools between 1998 and 2004 with an increase in school-level effects of socioeconomic background. It attributes this in part to increased residential differentiation and the effects of decentralisation from government controlled schools to municipality controlled schools as well as the introduction of greater choice in school enrolment (Skolverket, 2009). Gustafsson (2009) also offers an interpretation that the decline has been associated with 'the increased use of independent learning and decreased teacher-led instruction'.

Table 3.5 Country-level TIMSS Year 8 mathematics statistics from 1995 to 2009

Country	Mean score (with standard error) 1995	1999	2003	2007	2011	Significant change for 2011 since: 1995	1999	2003	2007
Korea	581 (2.0)	587 (2.0)	589 (2.2)	597 (2.7)	613 (2.9)	+	+	+	+
Singapore	609 (4.0)	604 (6.3)	605 (3.6)	593 (3.8)	611 (3.8)				+
Chinese Taipei		585 (4.0)	585 (4.6)	598 (4.5)	609 (3.2)				
Hong Kong SAR	569 (6.1)	582 (4.3)	586 (3.3)	572 (5.8)	586 (3.8)	+			
Japan	581 (1.6)	579 (1.7)	570 (2.1)	570 (2.4)	570 (2.6)	-	-		
Russia	524 (5.3)	526 (5.9)	508 (3.7)	512 (4.1)	539 (3.6)	+		+	+
Finland		520 (2.7)			514 (2.5)				
United States	492 (4.7)	502 (4.0)	504 (3.3)	508 (2.8)	509 (2.6)	+			
England	498 (3.0)	496 (4.1)	498 (4.7)	513 (4.8)	507 (5.5)				
Australia	509 (3.7)		505 (4.6)	496 (3.9)	505 (5.1)				
Hungary	527 (3.2)	532 (3.7)	529 (3.2)	517 (3.5)	505 (5.1)	-	-	-	-
Slovenia	494 (2.9)		493 (2.2)	501 (2.1)	505 (2.2)	+		+	
Lithuania	472 (4.1)	482 (4.3)	502 (2.5)	506 (2.3)	502 (2.2)				
Sweden	540 (4.3)		499 (2.6)	491 (2.3)	494 (1.9)	-		-	-
Italy		479 (3.8)	484 (3.2)	480 (3.0)	489 (2.4)		+	+	+
New Zealand	501 (4.7)	491 (5.2)	494 (5.3)		488 (5.5)				
Norway	498 (2.2)		461 (2.5)	469 (2.0)	475 (2.4)	-		+	
Romania	474 (4.6)	472 (5.8)	475 (4.8)	461 (4.1)	458 (4.0)	-	-	-	
Malaysia		510 (4.4)	508 (4.1)	474 (5.0)	440 (5.4)		-	-	-
Thailand		467 (5.1)		441 (5.0)	427 (4.3)		-		-
Macedonia		447 (4.2)	435 (3.5)		426 (5.2)		-		
Tunisia		448 (2.4)	410 (2.2)	420 (2.4)	425 (2.8)		-	+	
Chile		392 (4.4)	387 (3.3)		416 (2.6)		+	+	
Iran	418 (3.9)	422 (3.4)	411 (2.4)	403 (4.1)	415 (4.3)				
Jordan		428 (3.6)	424 (4.1)	427 (4.1)	406 (3.7)		-	-	-

Notes:
1 Countries not significantly different from Australia in TIMSS 2011 are shaded.
2 Significant improvements shown with **+** and significant declines shown with a **-** sign.
3 Standard errors are shown in parentheses.
Source: Mullis, Martin, Foy & Arora (2012)

Differences between PISA and TIMSS assessments

An inspection of the data in Table 3.5 compared to data in Table 3.1 suggests that some countries perform relatively better on PISA and other countries perform relatively better on TIMSS. There are two factors that contribute to this: differences in age–grade distributions and differences in the balance of what is assessed. PISA is based on 15-year-olds whereas TIMSS is based on a year level (Year 8). As a consequence, countries (and jurisdictions within countries) will have differing balances of Year levels represented in the sample of 15-year-olds in PISA depending on their age–grade distribution. Conversely, countries (and jurisdictions) will have different ages represented in their grade-based samples in TIMSS. Using data from PISA 2003 and TIMSS 2003,[26] Wu (2008) has shown that this has effects on the differences in between-country comparisons based on these studies.[27]

As a consequence of different mathematics assessment frameworks PISA and TIMSS have different balances of item content. Based on an analysis of the items in TIMSS 2003 and PISA 2003, Wu has

26 In this report, the international study is referred to as TIMSS 2003 (because that is how it is referred to in the literature) but the Australian data is referred to as 2002 because that is the year of data collection in Australia (and correspondingly for other cycles). The same protocol has been followed for other cycles of TIMSS. See 'Timing of TIMSS and PIRLS' later in Chapter 1 for further explanation.

27 This also has consequences in comparisons among Australian jurisdictions. Western Australia, which has a younger average age for any Year and a higher percentage of 15-year-olds in Year 11, has relatively high achievement in PISA mathematics or reading and relatively low achievement in TIMSS mathematics or science.

shown that there is a stronger representation of 'data' items in PISA mathematics than in TIMSS Grade 8 mathematics. Wu reports a balance of item content in PISA 2003 that is similar to that reported by Neidforf, Binkley, Gattis and Nohara (2006) as well as Gronmo and Olsen (2008). Data items made up between 31 and 39 per cent of the items in PISA mathematics in 2003 but only 15 per cent of the items in TIMSS 2003 (Wu, 2008). Conversely, algebra items made up 24 per cent of the content of TIMSS mathematics in 2003 but only eight per cent of the content of PISA mathematics in 2003 (Wu, 2008).[28]

Wu also observes that countries, such as Australia, in which students perform well on data record relatively higher scores on PISA than on TIMSS. For Australia, mean student performance on the sub-domains number, algebra, measurement and geometry is close to the OECD average whereas for data mean student performance is more than 30 points higher than the OECD average. Conversely, countries such as the Russian Federation, record relatively higher scores on TIMSS than on PISA. There is no clear answer concerning the correct balance of items across domains but it does mean that comparisons need to be informed by knowledge of assessment frameworks.

Differences among jurisdictions

Table 3.6 records the jurisdictional means for TIMSS Year 8 mathematics in 1994/5 and 2010/11. In 2010/11 the jurisdictional mean scores ranged from 462 to 532. The mean for the ACT was significantly greater than the national mean and the means for South Australia, Tasmania and the Northern Territory were significantly lower than the national mean. A more detailed analysis of differences among jurisdictions is reported by Thomson, Hillman and Wernert (2012). It is also evident from Table 3.6 that the only significant changes over the 16-year period were the declines in the mean scores for Western Australia and South Australia.

Table 3.6 Jurisdictional means for TIMSS Year 8 mathematics in 1994/5 and 2010/11

Jurisdiction	TIMSS 1994/5		TIMSS 2010/11		Difference
ACT	528	(11.4)	532	(9.9)	4
New South Wales	512	(8.6)	518	(11.1)	6
Australia	509	(3.7)	505	(5.1)	-4
Victoria	500	(6.4)	504	(8.0)	4
Queensland	506	(8.5)	497	(8.0)	-9
Western Australia	527	(6.7)	493	(10.6)	**-34**
South Australia	513	(5.6)	489	(5.8)	**-24**
Tasmania	496	(11.5)	475	(6.9)	-21
Northern Territory	470	(19.9)	462	(14.4)	-8

Notes:
1 Significant differences are shown in bold.
2 Standard errors are shown in parentheses.
Source: Thomson, Hillman & Wernert (2012)

28 The other three content categories were number, measurement and geometry which constituted 29, 16 and 16 per cent of TIMSS mathematics. The corresponding percentages for PISA mathematics were 38, 9 and 14 per cent.

Differences associated with social and demographic characteristics of students

Table 3.7 records the mean mathematics scores for various groups of students in 2010/11. In 2010/11 there was a large difference (more than one standard deviation) between those students whose parents had a university degree and those whose parents had not completed secondary school. There was also a substantial difference between Indigenous and non-Indigenous students of similar magnitude to that observed in the PISA mathematical literacy data. There were smaller but statistically significant differences between students attending schools in metropolitan compared to provincial locations but a substantial difference between students in metropolitan and remote locations. The differences between males and females were not significant. Because there was no significant overall change between 1994/5 and 2010/11 our consideration has focused on the patterns for 2010/11.

Table 3.7 TIMSS mathematics scores for groups of Australian Year 8 students: 2010

	Mean score TIMSS 2010/11	
Sex		
Male	509	(7.6)
Female	500	(4.7)
Difference	9	
Indigenous status		
Non-Indigenous	509	(5.3)
Indigenous	438	(4.8)
Difference	**71**	
Language background		
English	504	(5.0)
LBOTE	521	(10.3)
Difference	17	
Location of school		
Metropolitan	512	(5.8)
Provincial	487	(9.1)
Remote	448	(27.4)
Difference (metro–provincial)	**25**	
Difference (metro–remote)	**64**	
Parental education		
University degree	569	(9.9)
Post-secondary but not university	499	(4.9)
Completed upper secondary	480	(7.0)
Not completed upper secondary	437	(9.6)
Difference (Degree–lower secondary)	**132**	

Notes:
1 Standard errors are shown in parentheses.
2 Differences between groups that are significant are shown in bold.
Source: Thomson, Hillman & Wernert (2012)

MATHEMATICS IN PRIMARY SCHOOL: A PERSPECTIVE FROM TIMSS IN YEAR 4

Table 3.8 records country means for mathematics in Year 4 for selected TIMSS countries over the period from 1995 to 2011 arranged in descending order of country means for the 2011 cycle of TIMSS. The countries included are those that had participated in TIMSS Year 4 in any three cycles of TIMSS including 2011 or in both 1995 and 2011.

Table 3.8 Country-level TIMSS Year 4 mathematics statistics from 1995 to 2011

Country	Mean score (with standard error)								Significant change since:		
	1995		2003		2007		2011		1995	2003	2007
Singapore	590	(4.5)	594	(5.6)	599	(3.7)	606	(3.2)	+		
Korea	581	(1.8)					605	(1.9)	+		
Hong Kong SAR	557	(4.0)	575	(3.2)	607	(3.6)	602	(3.4)	+	+	
Chinese Taipei			564	(1.8)	576	(1.7)	591	(2.0)		+	+
Japan	567	(1.9)	565	(1.6)	568	(2.1)	585	(1.7)	+	+	+
England	484	(3.3)	531	(3.7)	541	(2.9)	542	(3.5)	+	+	
Russia			532	(4.7)	544	(4.9)	542	(3.7)			
United States	518	(2.9)	518	(2.4)	529	(2.4)	541	(1.8)	+	+	+
Netherlands	549	(3.0)	540	(2.1)	535	(2.1)	540	(1.7)	-		
Lithuania			534	(2.8)	530	(2.4)	534	(2.4)			
Portugal	442	(3.9)					532	(3.4)	+		
Ireland	523	(3.5)					527	(2.6)			
Australia	496	(3.4)	499	(3.9)	516	(3.5)	516	(2.9)	+	+	
Hungary	521	(3.6)	529	(3.1)	510	(3.5)	515	(3.4)		-	
Slovenia	462	(3.1)	479	(2.6)	502	(1.8)	513	(2.2)	+	+	+
Czech Republic	541	(3.1)			486	(2.8)	511	(2.4)	-		-
Austria	531	(2.9)			505	(2.0)	508	(2.6)	-		
Italy			503	(3.7)	507	(3.1)	508	(2.6)			
Norway	476	(3.0)	451	(2.3)	473	(2.5)	495	(2.8)	+	+	+
New Zealand	469	(4.4)	493	(2.2)	492	(2.3)	486	(2.6)	+	-	
Iran	387	(5.0)	389	(4.2)	402	(4.1)	431	(3.5)	+	+	+
Tunisia			339	(4.7)	327	(4.5)	359	(3.9)		+	+

Notes:
1 Standard errors are shown in parentheses.
2 Significant improvements shown with **+** and significant declines shown with a **-** sign.
3 Countries that were not significantly different from Australia in 2011 are shaded.
Sources: Mullis, Martin, Foy and Arora (2012); Thomson, Hillman, Wernert, Schmid, Buckley and Munene (2012)

Twelve countries in Table 3.8 (Singapore, Korea, Hong Kong, Chinese Taipei, Japan, England, Russian Federation, United States, Netherlands, Lithuania, Portugal and Ireland) had mean achievement scores significantly higher than Australia in 2011. In addition Northern Ireland, Belgium, Finland, Denmark, and Germany (which do not appear in Table 3.8 because they only have data from 2007 onwards) also have mean achievement scores significantly higher than Australia. Four countries in Table 3.8 had Year 4 mathematics achievement in 2011 not significantly different from Australia (Hungary, Slovenia, Czech Republic and Austria) as well as Serbia that only has data for 2011.

Change over time

Table 3.8 also records whether the mean score in Year 4 mathematics represents a significant change from the score for previous cycles of TIMSS. The achievement of Year 4 students in Australia in the 2011 cycle of TIMSS was higher than in 1995 (by 20 points) and 2003 but not different from the 2007 cycle. An inspection of the time series suggests that the improvement took place between the 2003 and 2007 cycles and was then maintained into the 2011 cycle.

Other countries to record improvements between 1995 and 2011 were Portugal (90 points), England (58 points), Slovenia (51 points), Hong Kong (45 points), Iran (44 points), Korea (24 points), the United States (23 points), Japan (18 points), Norway (19 points), New Zealand (17 points) and Singapore (16 points). However, the trajectories of these countries differ. The United States followed a similar trend to Australia except that it continued to improve between 2007 and 2011. England recorded an improvement from 1995 to 2003 as well as from 2003 to 2007 and then flattened off. Chinese Taipei did not participate in 1995 but recorded an improvement of 27 points between 2003 and 2011. Declines in Year 4 mathematics achievement from 1995 to 2011 were evident in the Czech Republic (30 points), Austria (23 points) and the Netherlands (9 points).

Differences among jurisdictions

Table 3.9 records the jurisdictional means for TIMSS Year 4 mathematics in 1994/5 and 2010/11 in descending order of the 2010/11 means. The means for 2010/11 ranged from 489 to 545 points. The means for the ACT and Victoria were significantly greater than the national mean and those for Queensland, Western Australia, South Australia and the Northern Territory were lower than the national mean. A more detailed analysis of differences among jurisdictions is reported by Thomson, Hillman, Wernert, Schmid, Buckley and Munene (2012). Table 3.9 also shows the changes in jurisdictional means from 1994/5 to 2010/11. It is evident that there was a significant improvement for Australia as a whole, and significant improvements for all jurisdictions except Queensland, Western Australia and the Northern Territory.

Table 3.9 Jurisdictional means for TIMSS Year 4 mathematics in 1994/5 and 2010/11

Jurisdiction	TIMSS 1994/5		TIMSS 2010/11		Difference 2010–1994
ACT	527	(5.8)	545	(5 9)	**18**
Victoria	507	(7.8)	531	(5.6)	**24**
New South Wales	496	(6.7)	525	(6.0)	**29**
Tasmania	486	(8.5)	517	(7.7)	**31**
Australia	496	(3.4)	516	(2.9)	**20**
South Australia	485	(7.0)	502	(5.2)	**17**
Queensland	484	(7.7)	499	(5.5)	15
Western Australia	483	(7.6)	499	(6.4)	16
Northern Territory	491	(8.4)	489	(12 8)	-2

Note:
1 Significant differences are shown in bold.
2 Standard errors are shown in parentheses.
Source: Thomson, Hillman, Wernert, Schmid, Buckley & Munene (2012)

Differences associated with social and demographic characteristics of students

Table 3.10 records the mean mathematics scores for various groups of students in 2002 and 2010. The base point of 2002 was chosen because the increase in Year 4 mathematics achievement took place between 2002 and 2006. In 2010 there was a substantial difference (of over 60 scale points) between students with many books in their homes and students with few books in their homes (71 points), Indigenous and non-Indigenous students (66 points) and students in metropolitan and remote locations (64 points). The differences between males and females were not significant. The differences between students with a language background other than English and other students were also not statistically significant. Over the eight-year period there was an increase in the mean score for students with many books in their homes and a decline in the mean score for students with few books in their homes. As this measure is taken as a measure of home resources this may suggest a possible increase in differences by social background but it could also be interpreted in terms of a change in the meaning of this measure over time, for example due to the emergence of literary resources through digital media.

Table 3.10 Mean TIMSS mathematics scores for groups of Australian Year 4 students in 2002 and 2010

	TIMSS 2002/3		TIMSS 2010/11		Significance of difference across cycles
Sex					
Male	500	(4.3)	519	(3.6)	*
Female	497	(4.5)	513	(3.3)	*
Difference	3		6		na
Indigenous status					
Non-Indigenous	504	(3.3)	522	(2.7)	*
Indigenous	444	(11.2)	458	(7.8)	
Difference	**60**		**64**		na
Language spoken at home					
English	500	(4.3)	520	(2.6)	*
LBOTE	501	(9.4)	507	(6.2)	
Difference	-1		13		na
Location of school					
Metropolitan	Data not available		521	(3.2)	
Provincial			505	(5.6)	
Remote			457	(7.8)	
Difference (metro–provincial)			**16**		
Difference (metro–remote)			**64**		
Books in the home					
Many (19%)	517	(4.7)	544	(4.4)	*
Average (59%)	519	(4.2)	525	(3.0)	
Few (22%)	496	(3.8)	473	(4.3)	*
Difference (many–few)	**21**		**71**		na

Notes:
1 Differences between groups that are significant are shown in bold.
2 Standard errors are shown in parentheses.
3 Differences across cycles that are significant are designated with an asterisk *.
Sources: Thomson, Hillman, Wernert, Schmid, Buckley & Munene (2012); Mullis, Martin, Gonzalez & Chrostowski (2004)

NATIONAL PERSPECTIVES ON NUMERACY FROM FIVE YEARS OF NAPLAN

It was noted in Chapter 2 that Australia's national assessment program in literacy and numeracy (NAPLAN) began in 2008 and reports achievements annually for numeracy as well as aspects of literacy. Results from NAPLAN are reported on common measurement scales that extend from Year 3 to Year 9.

National trends in NAPLAN numeracy

National means for NAPLAN numeracy over the period from 2008 to 2012 are recorded in Table 3.11. These data show that, nationally, there has been no change in numeracy achievement among Year 3 students or Year 9 students. However, there was a small improvement in Year 5 numeracy and a very small decline in Year 7 numeracy. The increase in Year 5 numeracy over the five years has been 13 points (a little less than one-fifth of a standard deviation[29]). However, rather than being a

29 The standard deviations for NAPLAN Numeracy at Year 5 have been 69, 68, 70, 68, and 71 over the five years from 2008 to 2012.

result of steady growth this mean increased between 2008 and 2009 and has not changed since then. There was also a smaller decrease of seven points (less than one-tenth of a standard deviation[30]) in Year 7 numeracy. In this case the mean had been relatively constant until 2011 but there was a significant decline from 2011 to 2012.

Table 3.11 National mean scores for NAPLAN numeracy from 2008 to 2012

	2008	2009	2010	2011	2012	Difference 2012-2008
	Mean	Mean	Mean	Mean	Mean	
Year 3	397 (0.5)	394 (0.5)	395 (0.5)	398 (0.5)	395 (0.5)	2
Year 5	476 (0.5)	487 (0.5)	489 (0.5)	488 (0.5)	489 (0.5)	**13**
Year 7	545 (0.8)	544 (0.8)	548 (0.8)	545 (0.8)	538 (0.8)	**-7**
Year 9	582 (0.9)	589 (0.9)	585 (0.9)	583 (1.0)	584 (1.0)	2

Notes:
1 Differences between 2008 and 2012 that are statistically significant are shown in bold.
2 Standard errors (based on published confidence intervals) are shown in parentheses.
Sources: ACARA (2012a) and national reports for previous years.

Table 3.12 Jurisdictional mean scores for NAPLAN numeracy in 2008 and 2012

	Year 3		Year 5		Year 7		Year 9	
	2008	2012	2008	2012	2008	2012	2008	2012
New South Wales	409	405	**488**	**498**	551	543	591	591
Victoria	417	409	490	498	**552**	**544**	591	591
Queensland	**368**	**381**	**458**	**476**	**539**	**532**	571	575
Western Australia	382	384	**461**	**478**	534	535	**571**	**582**
South Australia	**389**	**377**	**460**	**472**	536	529	571	573
Tasmania	400	392	**465**	**480**	534	526	568	568
ACT	412	410	**484**	**504**	556	546	595	597
Northern Territory	338	323	416	418	488	475	533	532

Note: Where differences between 2008 and 2012 are statistically significant the means have been highlighted as bold.
Sources: ACARA (2012a); ACARA (2008)

Cross-jurisdictional comparisons of changes in NAPLAN numeracy

The jurisdictional comparisons in Table 3.12 indicate that there had been significant improvements between 2008 and 2012 in Year 5 numeracy in New South Wales, Queensland, Western Australia, South Australia, Tasmania and the ACT. The largest gain was the 20 point increase in the ACT followed by the 18 point increase in Queensland and the 17 point increase in Western Australia. In Year 3 numeracy there was a 13 point increase in Queensland and a 12 point decline in South Australia but no other significant change. In Year 7 there was a significant decline in Victoria and Queensland and in Year 9 there was a significant increase in Western Australia.

Differences among groups of students in NAPLAN numeracy scores

Table 3.13 records the mean scores for groups of students for each year level for NAPLAN numeracy in 2012. As noted in the corresponding discussion of reading achievement, there are substantial missing data on parental education and occupation and especially in some jurisdictions and for the earlier NAPLAN cycles. For that reason this discussion focuses on the national pattern and on data for 2012. It is not possible to report the statistical significance of the differences between groups or

30 The standard deviations for NAPLAN Numeracy at Year 7 have been 73, 71, 72, 74 and 74 over the five years from 2008 to 2012.

across cycles because NAPLAN published reports from 2009 onwards do not include confidence intervals for these groups.

On the basis of the confidence intervals shown in the 2008 reports it would be expected that all differences between males and females, Indigenous and non-Indigenous, adjacent location categories, adjacent parental occupation categories and adjacent parental education categories are significant at every year level. The differences between students of a language background other than English and students with an English language background are probably significant in Years 7 and 9 but may not be significant in Years 3 and 5.

In terms of the magnitude of the differences between groups it appears that the average differences in 2012 between the top and bottom categories of parental education, and Indigenous and non-Indigenous students, were large (more than one standard deviation within each year level[31]). The average difference between top and bottom categories of parental occupation (senior managers and professionals compared with unskilled manual and service) were also substantial (nine-tenths of a standard deviation) and the average difference between metropolitan and remote students was large (about seven-tenths of a standard deviation).

There were small differences between males and females and between students with a language background other than English and other students of around one-tenth of a standard deviation.

31 The standard deviation in each NAPLAN domain was initially set to 100 for all students (Years 3, 5, 7 and 9). The standard deviation within each year level was approximately 70 for numeracy.

Table 3.13 National means on NAPLAN numeracy by student characteristics: 2012

	Year 3	Year 5	Year 7	Year 9
Sex				
Males	400	492	544	590
Females	391	485	532	578
Difference	9	7	12	12
Indigenous status				
Indigenous	320	414	469	518
Non-Indigenous	400	493	542	588
Difference	-80	-79	-73	-70
Language background				
Language background other than English	400	494	549	600
Language background English	395	488	536	581
Difference	5	6	13	19
Location				
Metropolitan	403	496	545	591
Provincial	381	476	524	570
Remote	353	448	503	550
Very Remote	292	380	445	501
Difference (metro–provincial)	22	20	21	21
Difference (metro–remote)	50	48	42	41
Parental occupation				
Senior management and qualified professionals	430	524	574	621
Other business managers and associate professionals	408	501	550	596
Tradespeople, clerks, skilled office, sales and service staff	388	482	529	575
Machine operators, hospitality staff, assistants, labourers	372	466	515	561
Not in paid work in the previous 12 months (very small)	359	454	498	547
Not stated or missing (13%)	369	464	520	566
Difference (senior – unskilled)	58	58	59	60
Parental education				
Bachelor degree or higher	431	526	578	626
Advanced diploma/diploma	399	492	541	589
Certificate 1 to 4	381	474	522	569
Year 12 or equivalent	381	478	529	576
Year 11 or equivalent or below	351	448	497	546
Not stated or missing (9%)	373	469	525	570
Difference (bachelor – Year 11)	80	78	81	80

Source: ACARA (2012a)

SUMMARY

There are data covering a longer span of time with respect to mathematics than reading. Data on Australian student achievement in mathematics for Year 4 and Year 8 date back to 1994 from TIMSS (which has antecedent studies going back to 1962[32]). PISA provides perspectives on mathematics achievement from 2003 to 2009 among 15-year-olds and NAPLAN numeracy extends from 2008 to 2012.

There was no overall change in Australian Year 8 mathematics achievement in TIMSS from 1994/5 to 2010/11 although there had been a dip in 2006/7. The same 16-year period had seen improvements in a number of countries including Korea, the United States, Hong Kong and the Russian Federation. Correspondingly there had been declines in a number of countries including Sweden, Norway, Hungary and Japan. Data from PISA indicate a small decline in the mathematics achievement of one-tenth of a standard deviation among 15-year-olds from 2003 to 2009. There was a small change in the shape of the distribution of mathematics scores in Australia with a relatively larger decline at the top of the distribution than at the bottom of the distribution, although the change was not as clearly evident as for reading. There was no change in NAPLAN numeracy achievement for Year 9 over the period from 2008 to 2012.

The achievement of Year 4 students in Australia in 2010 in TIMSS was higher than in 1994 but not different from 2006. The improvement took place between 2002 and 2006 and was then maintained. Other countries to record improvements were Portugal, England, Slovenia, Hong Kong, Iran, Korea, the United States, Japan, Norway, New Zealand and Singapore. Although Year 4 mathematics achievement improved for Australia over the period, it remained below a group of countries including Singapore, Korea, England, the United States, the Netherlands, Portugal and Ireland. NAPLAN numeracy scores for Year 5 students improved between 2008 and 2012 by a little less than one-fifth of a standard deviation. However, this increase took place between 2008 and 2009 and has not changed since then. This suggests that it may have been a method effect.

32 Afrassa and Keeves (1999) showed that there was a substantial decline in mathematics achievement for 13-year-old students in Australia between 1964 and 1994. They linked scales from the First and Second International Mathematics Studies conducted by the IEA with TIMSS. The decline over 30 years was approximately 0.30 logits or 0.29 standard deviations. It was also evident that the change varied among jurisdictions.

OTHER FIELDS

4

International and national large-scale assessments cover fields of learning other than reading and mathematics. This chapter reviews evidence about the achievements of Australian school students from large-scale assessments in science as well as the cross-curricular fields of digital literacy and civics and citizenship.

SCIENCE

There has been some national policy interest in the health of Australian science including the quality of education in science (Office of the Chief Scientist, 2012). A review of Australian science devoted one of its chapters to science in secondary schools and noted the enduring challenges including the tension between developing broad science literacy and preparing students for 'university science' (Office of the Chief Scientist, 2012: 42). This interest in science education parallels that in other countries and is often connected to a belief that too few young people are preparing for careers in science, technology, engineering and mathematics (STEM) (Global Science Forum, 2006; US Department of Labour, 2007; Bybee, 2010). A concern with the uptake of science studies has led to a focus on levels of achievement in science in secondary schools, and interest in science that is generated during the secondary school years (Ainley & Ainley, 2011). It has also been argued that the foundations for ongoing participation in science studies are formed through experiences in primary school and before school (Ainley & Ainley, in press).

This section of the report focuses on science achievement in secondary school from the perspective of data from two international large-scale achievement studies: TIMSS and PISA. TIMSS provides information about science achievement in lower secondary school (Year 8) in Australia and a number of other countries from 1994/5 through to 2010/11. PISA provides information about science achievement in middle secondary school (15-year-olds) especially from 2006 (when its science framework became defined as a major domain for the first time) to 2009. It also draws on information from large-scale studies in primary school. Again, TIMSS provides information about science achievement in middle primary school (Year 4) for Australia and a number of other countries from 1994/5 through to 2010/11. It also utilises data from the National Assessment Program – Science Literacy (NAP-SL) at the end of primary school (Year 6) in 2006 and 2009.

Science achievement in lower secondary school: TIMSS science from 1994 to 2010

Overall science achievement in Year 8

Table 4.1 records country means for science in Year 8 for selected TIMSS countries over the period from 1995 to 2011 arranged in descending order of country means for the 2011 cycle of TIMSS. The countries are those that had participated in TIMSS Year 8 in either 1995 or 1999 and also in 2011.

Table 4.1 TIMSS Year 8 science statistics from 1994/5 to 2010/11

Country	Mean score										Significant change for 2011 since:			
	1994/5		1998/9		2002/3		2006/7		2010/11		1995	1999	2003	2007
Singapore	580	(5.5)	568	(8.0)	578	(4.3)	567	(4.4)	590	(4.3)		+	+	+
Chinese Taipei			569	(4.4)	571	(3.5)	561	(3.7)	564	(2.3)				
Korea	546	(2.0)	549	(2.6)	558	(1.6)	553	(2.0)	560	(2.0)	+	+		+
Japan	554	(1.8)	550	(2.2)	552	(1.7)	554	(1.9)	558	(2.4)		+		
Finland[a]			535	(3.5)					552	(2.5)				
Slovenia	514	(2.7)			520	(2.7)	538	(2.2)	543	(2.7)	+		+	
Russia	523	(4.5)	529	(6.4)	514	(3.7)	530	(3.9)	542	(3.2)	+		+	+
Hong Kong SAR	510	(5.8)	530	(3.7)	556	(3.0)	530	(4.9)	535	(3.4)	+		-	
England	533	(3.6)	538	(4.8)	544	(4.1)	542	(4.5)	533	(4.9)				
United States	513	(5.6)	515	(4.6)	527	(3.1)	520	(2.9)	525	(2.6)	+			
Hungary	537	(3.1)	552	(3.7)	543	(2.8)	539	(2.9)	522	(3.1)	-	-	-	-
Ontario	496	3.7	518	(3.1)	533	(2.7)	526	(3.6)	521	(2.5)	+		-	
Australia	514	(3.9)			527	(3.8)	515	(3.6)	519	(4.8)				
Lithuania	464	(4.0)	488	(4.1)	519	(2.1)	519	(2.5)	514	(2.6)	+	+		
New Zealand	511	(4.9)	510	(4.9)	520	(5.0)			512	(4.6)				
Sweden	553	(4.4)			524	(2.7)	511	(2.6)	509	(2.5)	-		-	
Italy			493	(3.9)	491	(3.1)	495	(2.8)	501	(2.5)			+	
Norway	514	(2.4)			492	(2.2)	487	(2.2)	494	(2.6)	-			+
Iran	463	(3.6)	448	(3.8)	453	(2.3)	459	(3.6)	474	(4.0)	+	+	+	+
Romania	471	(5.1)	472	(5.8)	470	(4.9)	462	(3.9)	465	(3.2)				
Chile			420	(3.7)	413	(2.9)			461	(2.5)		+	+	
Thailand			482	(4.0)			471	(4.3)	451	(3.9)		-		-
Jordan			450	(3.8)	475	(3.8)	482	(4.0)	449	(4.0)		-	-	
Tunisia			430	((3.4)	404	(2.1)	445	(2.1)	439	(2.5)		+	+	
Malaysia			492	(4.4)	510	(3.7)	471	(6.0)	426	(6.3)		-	-	-
Macedonia			458	(5.2)	449	(3.6)			407	(5.4)		-	-	

Notes:

[a] Finland assessed Year 7 in 1999 and both Year 7 and Year 8 in 2011. The mean for Year 7 in 2011 was 529 (3.2) so the change from 1999 to 2011 was not statistically significant.

1 Standard errors are shown in parentheses.

2 Countries shown as shaded are not significantly different from Australia in 2011.

3 Countries that showed a significant improvement for a designated time period are shown with a **+**.

4 Countries that showed a significant decline for a designated time period are shown with a **-**.

Source: Martin, Mullis, Foy & Stanco (2012)

Nine countries (Singapore, Chinese Taipei, Korea, Japan, Finland, Slovenia, the Russian Federation, Hong Kong and England) achieved significantly higher mean science scores than Australia. Australia had a mean score that was not significantly different from that of five other countries (United States, Hungary, Lithuania, New Zealand and Sweden) as well as the Canadian province of Ontario.[33]

In the discussion of TIMSS Year 8 and PISA mathematics the effects of differences in the content balance in the assessments were discussed as well as the effects of differences in age–grade

33 Several Canadian provinces participated in TIMSS 2011 even though the whole of Canada did not participate as a country. We have focused on Ontario as the largest province and a province which has demographic similarities with Australia.

distributions. In the case of science, allowing for differences in age–grade distributions appears to remove most discrepancies between country means on TIMSS science and PISA science.

Change over time

Table 4.1 also records whether the mean score in Year 8 science represents a significant change from the score for previous cycles of TIMSS. The mean score for Australia in 2010/11 was not significantly different from that for previous cycles even though there had been a 'blip' (the opposite of a temporary dip) in the score in 2002/3.

Countries which had shown an improvement since 1994/5 were Lithuania (50 points), Slovenia (29 points), Hong Kong (25 points), as well as the province of Ontario (25 points), Russian Federation (19 points), Korea (14 points) and Iran (11 points). The United States improved by 14 points between 1995 and 2003 but did not maintain that improvement and slipped back just a little from 2003 to 2011. Countries which had shown a decline since either 1995 or 1999 were Sweden (by 44 points), Norway (by 20 points) and Hungary (by 15 points).

Among the countries that entered the TIMSS program in 1998/9 (rather than 1994/5) Chile had shown substantial improvement through to 2011 (by 41 points). Malaysia, Macedonia and Thailand all declined by substantial amounts.

Differences among jurisdictions

Table 4.2 records the jurisdictional means for TIMSS Year 8 science in 1994/5 and 2010/11. In 2010 the jurisdictional mean scores ranged from 481 to 551. The mean for the ACT was significantly greater than the national mean and the means for South Australia, Tasmania and the Northern Territory were significantly lower than the national mean. A more detailed analysis of differences among jurisdictions is reported by Thomson, Hillman and Wernert (2012). There were no significant changes in science achievement means over the 16-year period for jurisdictions or for Australia as a whole.

Table 4.2 Jurisdictional means for TIMSS Year 8 science in 1994/5 and 2010/11

Jurisdiction	TIMSS 1994/5		TIMSS 2010/11		Difference
ACT	529	(12.7)	551	(9.2)	22
New South Wales	517	(8.2)	532	(10.1)	15
Australia	514	(3.9)	519	(4.8)	5
Victoria	497	(6.2)	513	(7.5)	16
Queensland	510	(8.4)	516	(7.5)	6
Western Australia	531	(6.7)	514	(9.2)	-17
South Australia	510	(5.9)	506	(5.0)	-4
Tasmania	496	(10.7)	496	(6.4)	0
Northern Territory	466	(16.8)	481	(14.4)	15

Notes:
1 There were no significant differences between 2010/11 and 1994/5.
2 Standard errors are shown in parentheses.
Source: Thomson, Hillman & Wernert (2012)

Differences associated with social and demographic characteristics of students

Table 4.3 records the mean science scores for various groups of students in 2010/11. There was a large difference (more than one standard deviation) between those students whose parents had a university degree and those whose parents had not completed secondary school. There was also a substantial difference (three-fifths of a standard deviation) between Indigenous and non-Indigenous students.

Table 4.3 TIMSS science scores for groups of Australian Year 8 students: 2010

	Mean Score TIMSS 2010/11	
Sex		
Male	527	(6.5)
Female	511	(4.5)
Difference	**16**	
Indigenous status		
Non-Indigenous	524	(5.0)
Indigenous	459	(4.5)
Difference	**65**	
Language background		
English	521	(4.8)
Language other than English	500	(9.2)
Difference	21	
Location of school		
Metropolitan	523	(5.3)
Provincial	511	(8.6)
Remote	466	(32.5)
Difference (metro–provincial)	12	
Difference (metro–remote)	**57**	
Parental education		
University degree	580	(8.3)
Post-secondary but not university	521	(4.9)
Completed upper secondary	495	(6.2)
Not completed upper secondary	446	(10.8)
Difference (Degree–lower secondary)	**134**	

Notes:
1 Standard errors are shown in parentheses.
2 Differences between groups that are significant are shown in bold.
Source: Thomson, Hillman & Wernert (2012)

There was a substantial difference between students attending schools in metropolitan locations compared to those attending schools in remote locations (three-fifths of a standard deviation). However, there was no significant difference between students in metropolitan locations and those in provincial locations. The differences between males and females were significant and amounted to about one-sixth of a standard deviation. Because there was no significant overall change between 1994 and 2010 our consideration has focused on the patterns for 2010.

Science achievement in middle secondary school: PISA science in 2006 and 2009

Overall achievement

On the basis of the data in Table 4.4 it can be inferred that Australian 15-year-olds performed well in scientific literacy in 2009. Australian 15-year-olds performed less well than five OECD countries in Table 4.4 (Finland, Japan, Korea, New Zealand and Canada).34 It performed similarly to (i.e. not significantly different from) five OECD countries including Estonia, the Netherlands, Germany, Switzerland and the United Kingdom. It performed better than a number of countries such as the United States. Overall, the Australian performance on scientific literacy is better relative to other countries than was the case for mathematical literacy. However, it is also of interest that, as was noted for reading literacy, Australia has a relatively wide dispersion of scores. The Australian variance in

34 Australia also performed less well in 2009 than the non-OECD countries Singapore and Hong Kong (as well as the city of Shanghai). Singapore and the city of Shanghai were not participants in PISA 2006.

science achievement was significantly larger than the average of 32 other OECD countries. There was no change in that variance between 2006 and 2009.

Table 4.4 OECD country-level PISA science statistics for 2006 and 2009

	PISA 2006		PISA 2009		Change	
	Mean score	Standard deviation	Mean score	Standard deviation	Mean difference	Ratio of variances[2]
Finland	563 (2.0)	86	554 (2.3)	89	**-9**	1.09
Japan	531 (3.4)	100	539 (3.4)	100	8	0.99
Korea	522 (3.4)	90	538 (3.4)	82	**16**	**0.83**
New Zealand	530 (2.7)	107	532 (2.6)	107	2	1.00
Canada	534 (2.0)	94	529 (1.6)	90	-6	**0.91**
Estonia	531 (2.5)	84	528 (2.7)	84	-4	1.01
Australia	527 (2.3)	100	527 (2.5)	101	0	1.03
Netherlands	525 (2.7)	96	522 (5.4)	96	-3	1.01
Germany	516 (3.8)	100	520 (2.8)	101	5	1.01
Switzerland	512 (3.2)	99	517 (2.8)	96	5	0.93
United Kingdom	515 (2.3)	107	514 (2.5)	99	-1	**0.86**
Slovenia	519 (1.1)	98	512 (1.1)	94	**-7**	**0.92**
Ireland	508 (3.2)	94	508 (3.3)	97	0	1.06
Poland	498 (2.3)	90	508 (2.4)	87	**10**	0.93
Belgium	510 (2.5)	100	507 (2.5)	105	-4	1.11
Hungary	504 (2.7)	88	503 (3.1)	87	-1	0.96
United States	489 (4.2)	106	502 (3.6)	98	**13**	**0.85**
Czech Republic	513 (3.5)	98	500 (3.0)	97	**-12**	0.98
Norway	487 (3.1)	96	500 (2.6)	90	**13**	**0.87**
Denmark	496 (3.1)	93	499 (2.5)	92	3	0.97
France	495 (3.4)	102	498 (3.6)	103	3	1.02
Iceland	491 (1.6)	97	496 (2.7)	95	5	0.97
Sweden	503 (2.4)	94	495 (2.7)	100	-8	**1.12**
Portugal	474 (3.0)	89	493 (2.9)	83	**19**	**0.89**
Slovak Republic	488 (2.6)	93	490 (3.0)	95	2	1.05
Italy	475 (2.0)	96	489 (1.8)	97	**13**	1.02
Spain	488 (2.6)	91	488 (2.1)	87	0	**0.93**
Luxembourg	486 (1.1)	97	484 (1.2)	104	-2	**1.17**
Greece	473 (3.2)	92	470 (4.0)	92	-3	0.99
Israel	454 (3.7)	111	455 (3.1)	107	1	0.92
Turkey	424 (3.8)	83	454 (3.6)	81	**30**	0.94
Chile	438 (4.3)	92	447 (2.9)	81	9	**0.79**
Mexico	410 (2.7)	81	416 (1.8)	77	6	**0.92**
Average	498 (0.5)	96	501 (0.5)	94	3	0.97

Notes:
1 Standard errors for means are shown in parentheses.
2 This is the ratio of the variances is the squared 2009 standard deviation divided by the squared 2006 standard deviation. Ratios less than one mean a decrease in spread and ratios more than one reflect an increase. The ratios have been tested for significant difference from one. Computations were performed for this report.
3 Differences and variance ratios that are statistically significant have been shown in bold.
4 Countries listed in order of mean scores for 2009. Countries shown with shading are not significantly different from Australia in 2009.
5 'Average' refers to 33 OECD countries with data for 2006 and 2009.
Data Source: OECD (2010b) *PISA 2009 Results: Learning Trends*

Change over time

The data in Table 4.4 also indicate that, between 2006 and 2009, there was no change in the average scientific literacy scores of Australian 15-year-olds. In addition there was no significant change in the measure of dispersion of scores: the standard deviation. A number of countries recorded a significant improvement in scientific literacy scores: Portugal, Korea, United States, Italy, Norway, Poland and Turkey. The Czech Republic, Finland and Slovenia recorded declines in science achievement.

Differences among jurisdictions

Table 4.5 records, in descending order of 2009 means, the mean scientific literacy scores in 2006 and 2009 for each jurisdiction. In 2006, when science was the major domain, the ACT, Western Australia and New South Wales had mean scores significantly higher than the national mean and Victoria, Tasmania and the Northern Territory had scores below the national mean. In 2009, the differences from the national mean for Western Australia, New South Wales and Victoria were no longer statistically significant. However, there was no jurisdictional change between 2006 and 2009 that was statistically significant.

Table 4.5 Jurisdictional mean scientific literacy achievement in Australia: PISA 2006 and PISA 2009

	PISA 2006		PISA 2009		Difference in means
Jurisdiction	Mean		Mean		(PISA 2009 – PISA 2006)
ACT	549	(4.9)	546	(6.0)	-3
Western Australia	543	(6.8)	539	(7.3)	-4
New South Wales	535	(4.6)	531	(5.7)	-4
Queensland	522	(4.2)	530	(7.5)	8
Australia	527	(2.3)	527	(2.5)	0
Victoria	513	(4.9)	521	(4.9)	8
South Australia	532	(4.9)	519	(5.0)	-13
Tasmania	507	(4.6)	497	(5.3)	-10
Northern Territory	490	(6.6)	492	(7.7)	2

Note: Standard errors are shown in parentheses.
Source: Thomson et al. (2011)

Differences associated with social and demographic characteristics of students

Table 4.6 records the mean scientific literacy scores for various groups of students in 2006 and 2009. In each cycle there were large differences between the top and bottom quarters of socioeconomic background (close to one standard deviation) and between Indigenous and non-Indigenous students (about four-fifths of a standard deviation). There were also differences associated with location (mainly between metropolitan and remote locations) and with language background (those for whom English was the main language spoken at home scored approximately one-fifth of a standard deviation higher than other students). There was no significant difference between the two cycles of PISA for any of the subgroups.

Table 4.6 PISA scientific literacy for groups of Australian students in 2006 and 2009

	PISA 2006		PISA 2009	
	Mean		Mean	
Sex				
Females	527	(2.7)	528	(2.8)
Males	527	(3.2)	527	(3.1)
Difference	0		1	
Indigenous status				
Indigenous	441	(7.8)	449	(6.2)
Non-Indigenous	529	(2.3)	530	(2.4)
Difference	**88**		**81**	
Language background				
English language at home	530	(2.0)	532	(2.1)
LBOTE	507	(7.6)	512	(9.9)
Difference	**23**		**20**	
Immigrant status				
Australian born (AB)	528	(2.1)	526	(2.4)
First generation (FG)	531	(3.5)	538	(3.3)
Overseas born (OB)	526	(5.7)	524	(6.9)
Difference (AB-FG)	-3		12	
Difference (AB-OB)	2		2	
Location				
Metropolitan	531	(2.8)	532	(3.2)
Provincial	521	(3.5)	515	(4.0)
Remote	474	(15.6)	479	(13.0)
Difference (metro–provincial)	**10**		**17**	
Difference (metro–remote)	**57**		**53**	
Economic, social and cultural status (ESCS)				
Top quarter	572	(2.8)	577	(3.1)
Upper quarter	540	(2.8)	545	(3.1)
Lower quarter	516	(2.4)	515	(2.6)
Bottom quarter	485	(2.2)	481	(2.8)
Difference (Top–Bottom)	**86**		**96**	

Notes:
1 Differences between groups that are significant are shown in bold.
2 Standard errors are shown in parentheses.
3 No differences across cycles were significant.
Sources: Thomson et al. (2011); Thomson and De Bortoli (2008)

Science achievement in middle primary school: TIMSS science in Year 4

Overall science achievement in Year 4

Table 4.7 records country means for science in Year 4 for selected TIMSS countries over the period from 1995 to 2011 arranged in descending order of country means for the 2011 cycle of TIMSS. The countries are those that had participated in two cycles of TIMSS Year 4.

Fifteen countries (Korea, Singapore, Japan, Chinese Taipei, Russian Federation, United States, the Czech Republic, Hong Kong, Hungary, Sweden, Austria, Slovak Republic, Netherlands, England, Germany) and the Canadian province of Ontario[35] achieved significantly higher mean science scores than Australia in 2011. Australia had a mean score that was not significantly different from that of five other countries (Portugal, Slovenia, Ireland, Italy and Lithuania).

35 Several Canadian provinces participated in TIMSS 2011 even though the whole of Canada did not participate as a country.

When looking at the rankings, the mean for Australia appears to be relatively lower in Year 4 science than was the case in Year 8 science (see Table 4.1). However, when comparing the mean scores for Australia with those of the highest performing countries the difference is approximately the same for Year 4 and Year 8. In Year 8 the Australia mean science achievement had been similar to that of the United States, Hungary, Lithuania, New Zealand and Sweden as well as the province of Ontario. In Year 4 the Australian mean science achievement was lower than that of the United States, Hungary, and Sweden, as well as the province of Ontario.

Table 4.7 Country-level TIMSS Year 4 science statistics from 1994/5 to 2010/11

Country	Mean score (with standard error)				Significant change for 2011 since:		
	1994/5	2002/3	2006/7	2010/11	1995	2003	2007
Korea	576 (2.1)			587 (2.0)	+		
Singapore	523 (4.8)	565 (5.5)	587 (3.4)	583 (3.4)	+	+	
Japan	553 (1.8)	543 (1.5)	548 (2.1)	559 (1.9)	+	+	+
Chinese Taipei		551 (1.7)	557 (2.0)	552 (2.2)			
Russia		526 (5.2)	546 (4.8)	552 (3.5)		+	
United States	542 (3.3)	536 (2.5)	539 (2.7)	544 (2.1)		+	
Czech Republic	532 (3.0)		515 (3.1)	536 (2.5)			+
Hong Kong SAR	508 (3.3)	542 (3.1)	554 (3.5)	535 (3.8)	+		-
Hungary	508 (3.4)	530 (3.0)	536 (3.3)	534 (3.7)	+		
Sweden			525 (2.9)	533 (2.7)			+
Austria	538 (3.6)		526 (2.5)	532 (2.8)			
Slovak Republic			526 (4.8)	532 (3.8)			
Netherlands	530 (3.2)	525 (2.0)	523 (2.6)	531 (2.2)		+	+
England	528 (3.1)	540 (3.6)	542 (2.9)	529 (2.9)		-	-
Germany			528 (2.4)	528 (2.9)			
Ontario	516 (3.7)	540 (3.7)	536 (3.7)	528 (3.0)	+	-	
Portugal	452 (4.1)			522 (3.9)	+		
Slovenia	464 (3.1)	490 (2.5)	518 (1.9)	520 (2.7)	+	+	
Ireland	515 (3.5)			516 (2.8)			
Australia	521 (3.8)	521 (4.2)	527 (3.3)	516 (2.8)			-
Italy	515 (3.5)			516 (3.4)			
Lithuania		512 (2.6)	514 (2.4)	515 (2.4)			
Belgium (Flemish)		518 (1.8)		509 (2.0)		-	
New Zealand	505 (5.3)	520 (2.5)	504 (2.6)	497 (2.3)		-	-
Norway	504 (3.7)	466 (2.6)	477 (3.5)	494 (2.3)	-	+	+
Georgia			418 (4.6)	455 (3.8)			+
Iran	380 (4.6)	414 (4.1)	436 (4.3)	453 (3.7)	+	+	+
Armenia		437 (4.3)		416 (3.8)		-	
Tunisia		314 (5.7)	318 (5.9)	346 (5.3)		+	+

Notes:
1 Countries shown as shaded are not significantly different from Australia in 2011.
2 Standard errors are shown in parentheses.
3 A '+' sign indicates that the mean for 2011 is significantly greater than the mean for the designated year.
4 A '-' sign indicates that the mean for 2011 is significantly less than the mean for the designated year.
Source: Martin, Mullis, Foy & Stanco (2012)

Change over time

Table 4.7 also records whether the mean score in Year 4 science represents a significant change from the score for previous cycles of TIMSS. The mean score for Australia in 2010 (2011) was not significantly different from 1994 (1995) or 2002 (2003) but there had been a fall from the mean in 2006 (2007).

Countries which had shown an improvement since 1995 were Iran (73 points), Portugal (70 points), Singapore (60 points), Slovenia (56 points), Hong Kong (27 points), Hungary (26 points), Korea (11 points) and Japan (6 points). The Canadian province of Ontario also recorded an improvement since 1995 (12 points), In addition the Russian Federation that began participation in TIMSS science from 2003 also showed improvement over that shorter period (by 26 points). Norway declined since 1995 (by 10 points) and England declined since 2003 (by 11 points).

Differences among jurisdictions

Table 4.8 records the jurisdictional means for TIMSS Year 4 science in 1994/5 and 2010/11. In 2010/11 the jurisdictional mean scores ranged from 491 to 547. The means for the ACT and Victoria were significantly greater than the national mean and the means for Queensland and Western Australia were significantly lower than the national mean. A more detailed analysis of differences among jurisdictions is reported by Thomson, Hillman, Wernert, Schmid, Buckley and Munene (2012). The only significant (but negative) change over the 16-year period in jurisdictional science achievement means was for Western Australia. However, there had been a rise between 2006/7 and 2010/11 for the ACT following a drop between 1994/5 and 2006/7 and a decline between 2006/7 and 2010/11 for New South Wales back to levels similar to those in 2002/3 and 1994/5.

Table 4.8 Jurisdictional means for TIMSS Year 4 science in 1994/5 and 2010/11

Jurisdiction	TIMSS 1994/5		TIMSS 2010/11		Difference
ACT	557	(6.0)	547	(5.0)	-10
Victoria	529	(10.7)	529	(4.9)	0
New South Wales	522	(6.1)	522	(5.5)	0
Tasmania	523	(8.7)	518	(7.3)	-5
Australia	521	(3.8)	516	(2.8)	-5
South Australia	519	(7.1)	506	(5.1)	-13
Western Australia	527	(6.2)	502	(6.1)	**-25**
Queensland	503	(7.6)	501	(5.9)	-2
Northern Territory	512	(11.2)	491	(12.7)	-21

Notes:
1 Significant differences are shown in bold.
2 Standard errors are shown in parentheses.
Source: Thomson, Hillman, Wernert, Schmid, Buckley & Munene (2012)

Differences associated with social and demographic characteristics of students

Table 4.9 records the mean science scores for various groups of Year 4 students in 2010/11.

Table 4.9 TIMSS science scores for groups of Australian Year 4 students: 2010/11

	TIMSS 2010/11 Mean score	
Sex		
Male	516	(3.7)
Female	516	(3.1)
Difference	0	
Indigenous status		
Non-Indigenous	522	(2.6)
Indigenous	458	(7.7)
Difference	**64**	
Language background		
English	522	(2.6)
LBOTE	498	(5.6)
Difference	**24**	
Location of school		
Metropolitan	520	(3.1)
Provincial	507	(5.9)
Remote	459	(8.7)
Difference (metro–provincial)	13	
Difference (metro–remote)	**61**	
Books in the home		
Many books (more than 200)	545	(5.0)
Average number of books (26 to 200)	523	(2.9)
Few books (25 or fewer)	478	(3.3)
Difference (many–few)	**67**	

Notes:
1 Differences between groups that are significant are shown in bold.
2 Standard errors are shown in parentheses.
Source: Thomson, Hillman, Wernert, Schmid, Buckley & Munene (2012)

In 2010/11 the differences between those students with many books in their homes compared with those with few books in their home were large and between Indigenous and non-Indigenous students were large (about two-thirds of a standard deviation). There was a substantial difference between students attending schools in metropolitan locations compared to those attending schools in remote locations (three-fifths of a standard deviation) but there was no significant difference between those in metropolitan and provincial locations. There was also a significant difference between students who mainly spoke English at home and those who spoke a language other than English (a quarter of a standard deviation).

Science literacy in Year 6

National sample surveys of Science Literacy in Year 6 were conducted in 2003, 2006 and 2009 as part of the NAP sample surveys (ACARA, 2010). This discussion focuses on the 2006 and 2009 surveys which are equated on the same scale. Results are reported on that scale and can be compared across the two cycles. The results are also reported in terms of five proficiency levels and the percentage of students who attain (i.e. are at or above the level defined as the standard) a defined 'proficient standard'. In 2009, 52 per cent of Year 6 students had attained the proficient standard in science which was not significantly different from the percentage attaining this standard in 2006. The construct 'science literacy' has been adapted from the corresponding PISA construct and refers to the capacity to think scientifically. It covers strands concerned with formulating or identifying investigable questions and hypotheses and planning investigations of these, interpreting evidence and drawing conclusions from data, and using scientific understanding to describe and explain natural phenomena (ACARA, 2010: 3). These strands draw on four content areas: earth and space, energy and force, living things and matter. In addition to paper-based assessments organised in a rotated block design were two practical assessment tasks.

Differences among jurisdictions

Table 4.10 records the jurisdictional means for science literacy in 2006 and 2009. In 2009 the mean for the ACT was significantly greater than the national mean and the means for South Australia and the Northern Territory were significantly lower than the national mean. A more detailed analysis of differences among jurisdictions is reported by ACARA (2010). The only significant change over the three-year period in jurisdictional science achievement means was for Tasmania where there was a decline of one-fifth of a standard deviation.

Table 4.10 Jurisdictional means Year 6 NAP – Science literacy in 2006 and 2009

Jurisdiction	Science literacy 2006		Science literacy 2009		Difference
ACT	418	(7.2)	415	(5.4)	-3
Victoria	408	(5.2)	398	(4.6)	-10
New South Wales	411	(6.3)	396	(6.1)	-15
Western Australia	381	(5.1)	393	(4.8)	12
Australia	400	(6.1)	392	(2.6)	-8
Tasmania	406	(4.3)	386	(6.8)	**-20**
Queensland	387	(5.1)	385	(4.5)	-2
South Australia	392	(17.0)	380	(5.3)	-12
Northern Territory	325	(2.7)	326	(14.4)	1

Notes:
1 Significant differences are shown in bold.
2 Standard errors are shown in parentheses.
Source: ACARA (2010).

Differences associated with social and demographic characteristics of students

Table 4.11 records the mean science literacy scores for various groups of Year 6 students in 2009. Because there was no overall change between 2006 and 2009 our consideration has focused on the patterns for 2009.

In 2010 there was a large difference (about one standard deviation) between Indigenous and non-Indigenous students. There was a difference between students attending schools in metropolitan locations compared to those attending schools in remote locations (three-fifths of a standard deviation) but there was no significant difference between those in metropolitan and provincial locations. There was no significant difference between students who mainly spoke English at home and those who spoke a language other than English and there was no significant difference between males and females. There are no data reported for parental education, parental occupation or home literacy resources.

Table 4.11 Mean NAP science literacy scores for groups of Australian Year 6 students: 2009

	Science literacy	
Sex		
Male	393	(3.0)
Female	391	(2.6)
Difference	2	
Indigenous status		
Non-Indigenous	397	(2.5)
Indigenous	297	(8.1)
Difference	**100**	
Language background		
English	396	(2.4)
LBOTE	384	(6.6)
Difference	12	
Location of school		
Metropolitan	395	(3.1)
Provincial	389	(4.0)
Remote	336	(11.9)
Difference (metro–provincial)	6	
Difference (metro–remote)	**59**	

Notes:
1 Differences shown in bold are statistically significant.
2 Standard errors are shown in parentheses.
Source: ACARA (2010)

CIVICS AND CITIZENSHIP

National Assessment Program: Civics and citizenship

National assessments of civics and citizenship, based on large representative samples of students in Year 6 and Year 10, have been conducted in 2004, 2007 and 2010. An assessment domain was initially developed to frame 2004 assessment around two key performance measures: civic knowledge and understanding and participatory skills and civic values (Print & Hughes, 2001). The assessment domain was revised in 2008 taking into account two key curriculum developments: the Statements of Learning for Civics and Citizenship (2006) and the Melbourne Declaration on Educational Goals for Young Australians (MCEETYA, 2008). The assessment incorporates a test of civics and citizenship (covering knowledge and understanding) and a student questionnaire (covering attitudes and participation). As with other large-scale assessments conducted as part of the National Assessment Program (NAP), results are reported in terms of scale scores, distributions across proficiency levels and the percentages attaining a proficient standard that was set as part of the 2004 cycle. In 2010, 52 per cent of Year 6 students and 49 per cent of Year 10 students attained the proficient standard for the relevant Year. National means for the three cycles are recorded in Table 4.12. The data in Table 4.12 indicate that from 2004 to 2010 there was a significant improvement in civics and citizenship scores for Year 10 students.

Table 4.12 Mean NAP civics and citizenship scores: 2004, 2007 and 2010

	Year 6		Year 10		Difference	
2004	400	(3.4)	496	(3.6)	**96**	(4.9)
2007	405	(2.8)	502	(4.4)	**97**	(5.6)
2010	408	(3.4)	519	(5.8)	**111**	(6.7)
Difference (2010–2007)	3	(6.9)	**17**	(8.4)		

Notes:
1 Differences shown in bold are statistically significant.
2 Standard errors are shown in parentheses.

Table 4.13 records the jurisdictional means for the 2007 and 2010 cycles of the NAP civics and citizenship assessment. The data are organised in descending order of the 2010 means. For Year 6, the three top performing states – ACT, New South Wales and Victoria – did not differ from each other significantly. The two lowest performing states, Queensland and the Northern Territory, performed significantly less than each of the other states and territories and the Northern Territory performed less well than Queensland. For Year 10, New South Wales outperformed all other states and territories. While the ACT, Victoria and Western Australia did not differ from each other significantly, the ACT performed better than Tasmania, South Australia and Queensland, Victoria performed better than South Australia only and Western Australia was in the middle of the range of jurisdictional means. These data also indicate that only in Western Australia and the Northern Territory, each at Year 6, was there a significant difference (an improvement) between 2007 and 2010. None of the apparent changes at Year 10 were significant.

Table 4.13 Jurisdictional mean NAP civics and citizenship scores: 2007 and 2010

	2007		2010		Difference	
Year 6						
ACT	425	(8.4)	442	(10.5)	16	(14.4)
New South Wales	432	(6.6)	426	(5.6}	-6	(10.2)
Victoria	418	(7.2)	422	(5.2)	4	(10.4)
Tasmania	401	(7.4)	411	(9.0)	10	(12.8)
Western Australia	369	(7.6)	402	(5.6)	**33**	(10.8)
South Australia	385	(6.5)	396	(7.7)	11	(11.4)
Queensland	376	(8.6)	374	(6.9)	-2	(12.2)
Northern Territory	266	(16.0)	316	(16.7)	**50**	(23.7)
Year 10						
New South Wales	529	(8.7)	558	(12.1)	29	(15.5)
ACT	523	(10.0)	523	(12.3)	0	(16.4)
Victoria	494	(8.7)	514	(9.8)	20	(13.8)
Western Australia	478	(11.5)	509	(10.8)	32	(16.3)
Tasmania	485	(8.2)	492	(7.8)	7	(12.0)
South Australia	505	(11.9)	487	(9.3)	-18	(15.8)
Northern Territory	464	(19.4)	483	(16.4)	20	(25.9)
Queensland	481	(7.1)	482	(14.4)	2	(16.7)

Notes:
1 Differences shown in bold are statistically significant.
2 Standard errors are shown in parentheses.
Sources: MCEETYA (2009); ACARA (2011)

Differences associated with social and demographic characteristics of students

Table 4.14 records the mean civics and citizenship scores for various groups of Year 6 and Year 10 students in 2010. In 2010 large differences were evident between Indigenous and non-Indigenous students (more than one standard deviation) and between the two parental occupation categories of 'senior managers and professionals' and 'unskilled labourers, office, sales, service' staff (about one standard deviation). There were also differences between school location categories (especially between metropolitan and remote locations).

Females scored higher than males by one-fifth of a standard deviation at Year 6 and three-tenths of a standard deviation at Year 10. There were no differences between students who mainly spoke English at home and other students.

Table 4.14 Mean NAP civics and citizenship scores for groups of Australian students: 2010

	Year 6		Year 10	
Sex				
Male	398	(3.7)	504	(7.3)
Female	418	(3.2)	534	(6.9)
Difference	**-20**		**-30**	na
Indigenous status				
Non-Indigenous	414	(3.4)	523	(5.8)
Indigenous	276	16.2)	405	(13.6)
Difference	**138**		**117**	na
Country of birth				
Australia	410	(3.5)	523	(5.8)
Overseas	404	(13.9)	488	(21.8)
Difference	5		35	na
Language background				
English	411	(4.2)	522	(6.4)
LBOTE	401	(7.6)	518	(12.8)
Difference	10		4	na
Location of school				
Metropolitan	418	(1.7)	531	(2.8)
Provincial	391	(2.8)	488	(4.7)
Remote	318	(4.4)	462	(14.1)
Difference (metro–provincial)	**27**		**43**	na
Difference (metro–remote)	**100**		**69**	na
Parental occupation				
Senior managers and professionals	467	(5.3)	583	(7.6)
Other managers and associate prof.	437	(5.0)	536	(7.2)
Trades and skilled office, sales, service	388	(5.0)	502	(8.2)
Unskilled labourers office, sales, service	366	(6.7)	480	(8.9)
Not in paid work in last year	345	(10.7)	446	(13.9)
Not stated or unknown	387	(5.5)	481	(8.4)
Difference (Senior manager–unskilled)	**101**		**103**	na

Notes:
1 Standard errors are shown in parentheses.
2 Significant differences shown in bold.
Source: ACARA (2011)

IEA Civic Education Study

Australia participated in the IEA Civic Education Study in 1999 (Mellor, Kennedy & Greenwood, 2001; Torney-Purta, Lehmann, Oswald & Schulz, 2001). The assessment was structured around three broad domains: democracy and citizenship; national identity and international relations; social cohesion and diversity. The items were concerned with content knowledge as well as interpretative skills and conceptual understanding. Questionnaire items asked about attitudes and expected participation. The sample was nationally representative and consisted of 3331 Year 9 students[36] from 142 schools. In each school one class (usually from history or another civic-related area) was selected at random.

Table 4.15 Mean national civic knowledge scores by domain: 1999

Country	Civic knowledge		Content knowledge		Interpretive skills		Year
Poland	111	(1.7)	112	(1.3)	106	(1.7)	8
Finland	109	(0.7)	108	(0.7)	110	(0.7)	8
Cyprus	108	(0.5)	108	(0.5)	108	(0.5)	9
Greece	108	(0.8)	109	(0.7)	105	(0.8)	9
Hong Kong SAR	107	(1.1)	108	(1.0)	104	(1.1)	9
United States	106	(1.2)	102	(1.1)	114	(1.2)	9
Italy	105	(0.8)	105	(0.8)	105	(0.8)	9
Slovak Republic	105	(0.7)	107	(0.7)	103	(0.7)	8
Norway	103	(0.5)	103	(0.5)	103	(0.5)	8
Czech Republic	103	(0.8)	103	(0.8)	102	(0.8)	8
Australia	102	(0.8)	99	(0.7)	107	(0.8)	9
Hungary	102	(0.6)	102	(0.6)	101	(0.6)	8
Slovenia	101	(0.5)	102	(0.5)	99	(0.5)	8
Denmark	100	(0.5)	100	(0.5)	100	(0.5)	8
Germany	100	(0.5)	99	(0.5)	101	(0.5)	8
Russian Federation	100	(1.3)	102	(1.3)	96	(1.3)	9
England	99	(0.6)	96	(0.6)	105	(0.6)	9
Sweden	99	(0.8)	97	(0.8)	102	(0.8)	8
Switzerland	98	(0.8)	96	(0.8)	102	(0.8)	8 & 9
Bulgaria	98	(1.3)	99	(1.1)	95	(1.3)	8
Portugal	96	(0.7)	97	(0.7)	95	(0.7)	8
Belgium (French)	95	(0.9)	94	(0.9)	96	(0.9)	8
Estonia	94	(0.5)	94	(0.5)	95	(0.5)	8
Lithuania	94	(0.7)	94	(0.7)	93	(0.7)	8
Romania	92	(0.9)	93	(1.0)	90	(0.9)	8
Latvia	91	(0.9)	92	(0.9)	92	(0.9)	8
Chile	88	(0.7)	89	(0.6)	88	(0.7)	8
Colombia	86	(0.9)	89	(0.8)	84	(0.9)	8

Notes:
1 Countries shown with shading have means on the overall civic knowledge scale not significantly different from Australia.
2 Standard errors are shown in parentheses.
Source: Torney-Purta, Lehmann, Oswald & Schulz (2001)

Results are recorded in Table 4.15. Those data show overall mean scores for civic knowledge as well as scores for the subscales concerned with content knowledge and interpretative skills. The year level that was tested is also recorded.

The data in Table 4.15 show that Australian students recorded means on the civic knowledge not significantly different from a group of 11 countries (Norway, Czech Republic, Hungary, Slovenia, Denmark, Germany, Russian Federation, England, Sweden, Switzerland and Bulgaria). The Australian mean was lower than a group of countries that included Poland, Finland and the United States.

Australian students performed relatively better on the interpretative skills subscale than on the content knowledge subscale. On interpretative skills Australian students were outperformed by only the United States and Finland.[37]

36 The population was students in the Year that contained the most 14-year-olds. This meant that some countries assessed Year 8 students.

37 Although Cyprus appears to have a higher mean score on interpretive skills than Australia the difference is not statistically significant.

In Australia and in most countries the differences between female and male students were not statistically significant. In the large majority of countries, and in Australia, the more books students reported in the home the better they performed on the civic knowledge test.

Australian students' scores were significantly below the international mean on three of the four attitude scales which made up the civic engagement dimension and were also below the international mean on the scale concerned with expected participation in political activities.

DIGITAL OR ICT LITERACY

Information and communication technologies (ICT) have changed, and continue to change, the education, work and social lives of people. Education authorities have come to see competence in ICT as a key preparation for young people's futures and several have established assessment programs in this field. There are two large-scale assessment surveys that provide perspectives on digital or ICT literacy among school students in Australia. One is the PISA study of digital reading conducted as part of the 2009 cycle of PISA among 16 countries (OECD, 2011). The other is the NAP ICT literacy surveys which have been conducted with Australian Year 6 and Year 10 students over three cycles in 2005, 2008 and 2011 (ACARA, 2012b).

PISA digital reading

An assessment of digital reading was included as an international option in the PISA 2009 cycle. The report of this assessment covers 16 OECD countries including Australia and three partner countries (OECD, 2011). The assessment was concerned with reading in a digital medium rather than being a computer-delivered assessment of print-based reading in the print medium. Digital texts make use of dynamic windows and frames, non-linear sequences that use hyperlinks and networks within texts, multimedia and augmented images, and provide for online discussion and social networking. The processes that are most affected in reading digital texts are concerned with accessing through navigation tools and devices (including manipulating windows containing information and scrolling), locating information through searching for phrases and other units, integrating and comparing information from different pieces of text, and evaluating web-based documents.

The PISA digital reading framework also recognised that many of the skills in reading digital texts were similar to those involved in reading printed texts. Its framework saw digital reading as an extension of its reading framework. Country-level patterns for performance on digital reading were similar to those for print reading. However, there were some countries (including Australia, New Zealand, Korea, Ireland, Sweden and Iceland) where students performed relatively better in digital than print reading.

Table 4.16 Country-level statistics for PISA digital reading: 2009

Country	All students					Gender differences				
				Males		Females		Difference (Female–Male)
	Mean		Standard deviation	Mean		Mean		
Korea	568	(3.0)	68	559	(4.3)	577	(3.5)	**-18**
New Zealand	537	(2.3)	99	518	(3.5)	558	(2.7)	**-40**
Australia	537	(2.8)	97	522	(3.6)	550	(2.9)	**-28**
Japan	519	(2.4)	76	508	(3.2)	531	(2.9)	**-23**
Hong Kong-China*	515	(2.6)	82	511	(3.2)	519	(3.2)	**-8**
Iceland	512	(1.4)	91	497	(2.1)	527	(1.8)	**-30**
Sweden	510	(3.3)	89	497	(3.5)	524	(3.5)	**-26**
Ireland	509	(2.8)	87	494	(3.7)	525	(2.9)	**-31**
Belgium	507	(2.1)	94	496	(3.0)	520	(2.4)	**-24**
Norway	500	(2.8)	83	483	(3.2)	518	(3.0)	**-35**
France	494	(5.2)	96	484	(5.2)	504	(5.7)	**-20**
Macau-China*	492	(0.7)	66	486	(1.0)	498	(1.1)	**-12**
Denmark	489	(2.6)	84	486	(3.1)	492	(2.9)	**-6**
Spain	475	(3.8)	95	466	(4.3)	485	(3.8)	**-19**
Hungary	468	(4.2)	103	458	(5.0)	479	(4.8)	**-21**
Poland	464	(3.1)	91	449	(3.4)	478	(3.3)	**-29**
Austria	459	(3.9)	103	447	(4.6)	469	(5.1)	**-22**
Chile	435	(3.6)	89	425	(4.3)	444	(3.8)	**-19**
Colombia*	368	(3.4)	83	367	(4.5)	370	(3.8)	-3
OECD average	499	(0.8)	90	487	(1.0)	511	(0.9)	**-24**

Notes:
1 Shading indicates result for New Zealand is not significantly different from Australia.
2 Standard errors are shown in parentheses.
3 OECD average based on 13 participating OECD countries
4 * Hong Kong-China, Macau-China and Colombia were non-OECD participants in the digital reading assessment.
Data Source: OECD (2011) *PISA 2009 Results: Students On Line: Digital Technologies and Performance*

The results in Table 4.16 indicate that Australian 15-year-old students performed very well on digital reading. Korea was the top performing country on digital reading but was followed by Australia and New Zealand which were in turn higher than the countries below Australia in the table (including Japan, Hong Kong, Iceland and Sweden). Korea was also the top performing country in PISA 2009 for print reading. The difference between Korea and Australia was 31 points for electronic reading and 24 points for print reading. From this perspective, relative to Korea, Australia performs approximately as well on digital as print reading.

In all countries, including Australia, females performed better than males. In Australia the gap was similar to the average gap for the 16 OECD countries (about one quarter of a standard deviation). For the print reading assessment the corresponding gaps were just a little less than two-fifths of a standard deviation. In other words the gap in favour of females was less for digital reading than for print reading in most OECD countries. Both Australian males and females performed better in electronic reading than in print reading.

Digital reading was associated with socioeconomic background in a similar way to the association of print reading with socioeconomic background. Relevant data are recorded in Table 4.17. Differences between the top and bottom quarters of the socioeconomic distribution are above four-fifths of a standard deviation for digital and print reading and for both Australia and the 16 OECD countries.

Table 4.17 PISA digital and print reading by socioeconomic background: 2009

	Australia				OECD – 16 countries			
	Digital		Print		Digital		Print	
Top quarter	581	(3.5)	562	(3.1)	542	(1.1)	545	(0.9)
Upper quarter	552	(3.2)	532	(3.0)	513	(0.9)	512	(0.9)
Lower quarter	527	(3.0)	504	(2.4)	490	(1.0)	488	(0.9)
Bottom quarter	497	(3.0)	471	(2.7)	457	(1.1)	456	(1.0)
Difference (Top–Bottom)	**84**		**91**		**85**		**89**	

Notes:
1 Socioeconomic background is measured using the index of economic, social and cultural status (ESCS) (OECD, 2010c).
2 Standard errors are shown in parentheses.
Data Source: OECD (2011) *PISA 2009 Results: Students On Line: Digital Technologies and Performance*

NAP ICT literacy

Assessments of ICT literacy in Australia were conducted in Years 6 and 10 with large nationally representative samples of approximately 10,000 students in 2005, 2008 and 2011 (ACARA, 2012b). These assessments were computer-based and incorporated the performance of discrete tasks as well as the creation of digital products using multiple tasks. In this assessment ICT literacy is seen as a set of generalisable and transferable knowledge, skills and understandings concerned with the use of computer technology to investigate, create and communicate information in a variety of contexts. The definition of ICT literacy adopted for the NAP was: 'the ability of individuals to use ICT appropriately to access, manage, integrate and evaluate information, develop new understandings, and communicate with others in order to participate effectively in society' (MCEETYA, 2005).

The assessments combined the performance of specific software functions with the creation of digital products. Each assessment involved tasks embedded in seven thematic modules. In 2011 six of the modules were built around a narrative and required the production of a digital product: sports picnic, friend's PC, saving electricity, wiki builder, language preservation and art show. One module focused on the performance of discrete tasks concerned with general computing skills. The six thematic modules involved using blogs and other websites, search engines, graphics software, mapping software, and photo management software; producing a video; editing, formatting and updating a wiki; and using collaboration software to work on a project with others. Each assessment cycle incorporated three modules from previous cycles and four new modules. In this way it was possible to incorporate new technologies and applications and to measure changes over time.

Table 4.18 ICT literacy mean scale scores for Years 6 and 10 from 2005 to 2011

	Year 6		Year 10	
2011	435	(2.9)	559	(2.9)
2008	419	(3.5)	560	(3.6)
2005	400	(3.2)	551	(2.9)
Difference (2011–2008)	**16**	(7.3)	-1	(7.3)
Difference (2011–2005)	**35**	(8.4)	9	(8.3)

Notes:
1 Differences that are statistically significant are shown in bold.
2 Standard errors are shown in parentheses.
Source: ACARA (2012b)

The ICT literacy assessment was highly reliable and the resulting scale was characterised by descriptions of proficiency levels based on descriptions of the items. It appeared to measure one underlying construct. Common tasks were used to compare the relative performance of the Year 6 and Year 10 students and other common tasks were used to link the 2011 results to those from 2008 and 2005. The comparisons of achievement over time in a rapidly developing field were made possible through instruments that reflect relevant technological changes and maintained integrity to the core processes of the ICT literacy construct.

The national means in Table 4.18 show an improvement in the ICT literacy among Year 6 students but no change for Year 10 students. Among Year 6 students the change was also evident in the rise in the percentage of students attaining the proficient standard from 49 per cent in 2005 to 62 per cent in 2011. In 2011, 65 per cent of Year 10 students had attained the proficient standard but this had not changed significantly from 2005.

Table 4.19 Jurisdictional ICT literacy mean scale scores for Years 6 and 10 in 2011

	Year 6		Year 10	
New South Wales	445	(6.3)	565	(6.4)
Victoria	448	(4.7)	568	(6.3)
Queensland	415	(7.0)	553	(4.8)
Western Australia	424	(6.8)	548	(5.4)
South Australia	436	(5.2)	552	(7.4)
Tasmania	405	(6.2)	534	(7.8)
ACT	466	(11.4)	582	(8.1)
Northern Territory	367	(18.8)	490	(24.7)
Australia	435	(2.9)	559	(2.9)

Note: Standard errors are shown in parentheses.
Source: ACARA (2012b)

Jurisdictional mean scores are shown in Table 4.19. At Year 6, there were differences among jurisdictions in ICT literacy. Mean scores in the ACT, Victoria and New South Wales were higher than those for Western Australia, Queensland, Tasmania and the Northern Territory. For all jurisdictions except Tasmania and the Northern Territory there was an increase in mean scores between 2005 and 2011. In Year 10 the range in mean scores for ICT literacy was smaller than in Year 6. On average, ICT literacy scores for Year 10 students in the ACT, Victoria, New South Wales were higher than in Tasmania and the Northern Territory.

Table 4.20 records the differences among social and demographic groups of Australian students for 2005 and 2011. Even though the methods of collecting student characteristics in 2005 (student reports) were different from the methods for collecting similar data in 2011 (school records based on information from parents) the patterns were very similar in 2005 and 2011.

There was a large effect associated with parental occupation. In Year 6, the difference in the mean score of students with parents in 'unskilled labourers, office, sales, and service' occupational groups was four-fifths of a standard deviation lower than that for students with parents from the 'senior managers and professionals' occupational group. In Year 10 the corresponding difference was two-thirds of a standard deviation.

There was also a substantial gap in ICT literacy between Indigenous and non-Indigenous students in both year levels and at both time points. Although the means are not strictly comparable over time, because of differences in data collection methods, the results are similar.

There was also evidence of differences in ICT literacy among geographic locations. At both Year 6 and Year 10 higher ICT literacy scores were recorded for metropolitan students than for students in provincial areas, as well as those in remote areas, and this gap seemed to increase over time. Females recorded increasingly higher levels of ICT literacy than males.

Table 4.20 ICT literacy scores for Australian Year 6 and Year 10 students in 2005 and 2011

	Year 6					Year 10				
	2005		2011		Diff.	2005		2011		Diff.
Sex										
Females	407	(3.3)	446	(3.4)	*	555	(3.7)	566	(3.8)	
Males	393	(4.6)	425	(3.6)	*	546	(3.8)	553	(3.7)	
Difference	**15**		**22**		na	9		**14**		na
Indigenous status										
Non-Indigenous	405	(3.2)	441	(2.8)	*	553	(2.8)	563	(2.8)	
Indigenous	339	(11.8)	343	(11.2)		482	(11.8)	469	(17.7)	
Difference	**66**		**98**		na	**71**		**92**		na
Language background										
English language at home	400	(3.0)	434	(3.4)	*	553	(3.0)	560	(3.2)	
LBOTE	400	(6.3)	448	(6.3)	*	545	(5.6)	558	(7.2)	
Difference	**0**		14		na	8		2		na
Location										
Metropolitan	408	(4.1)	448	(3.4)	*	555	(3.7)	569	(3.2)	
Provincial	386	(4.9)	404	(4.3)		545	(6.0)	536	(6.3)	
Remote	345	(24.0)	381	(22.4)		504	(11.6)	483	(31.8)	
Difference (metro–provincial)	**22**		**44**		na	10		**33**		na
Difference (metro–remote)	**63**		**67**		na	**51**		**86**		na
Parental occupation										
Senior managers and qualified professionals	450	(6.0)	485	(4.6)	*	586	(4.8)	599	(4.2)	
Other managers and associate professionals	424	(3.1)	454	(4.4)	*	560	(3.6)	571	(4.2)	
Tradespeople, skilled office, sales & service	392	(4.0)	428	(4.4)	*	542	(3.4)	554	(5.0)	
Unskilled labourers, office, sales & service	363	(4.3)	402	(5.8)	*	521	(5.5)	535	(8.8)	
Not in paid work in last year	333	na	381	(8.7)	na	476	na	507	(10.2)	
Not stated or unknown	na		406	(6.6)	na	na		541	(5.9)	
Difference (Senior manager – unskilled)	**87**		**83**			65		**66**		na

Notes:
1 Student background characteristics were gathered by different methods in 2005 (student questionnaire) and 2011 (school records based on parent-supplied data).
2 Differences between groups that are significant are shown in bold. Differences across cycles that are significant are designated with an asterisk *
3 In 2005 there was 3.0 per cent of Year 6 students and 1.9 per cent of Year 10 students who reported their parents had not been in paid work for 12 months.
4 In 2005 there was 3.3 per cent of Year 6 students and 3.9 per cent of Year 10 students for whom these data were missing.
5 Standard errors are shown in parentheses.
Sources: ACARA (2012b); MCEETYA (2007)

There were no differences between students speaking a language other than English at home and those with an English-speaking background.

SUMMARY

There are data that can inform judgements about student outcomes in science, digital literacy and civics and citizenship. Generally, among secondary school students, achievement in science had not changed for Year 8 students since 1994 and had not changed for 15-year-olds between 2006 and 2009. With regard to primary school students, science achievement for Year 4 students had been stable between 1994 and 2006 but declined slightly between 2006 and 2010 to revert to levels similar to those of 1994. No change was evident in the national assessment of science literacy among Year 6 students between 2006 and 2009. Based on results for those countries that participated in both Year levels, it appeared that Year 8 Australian students performed better in science than Year 4 Australian students. However, the spread in science performance is relatively high in Australia compared to other countries.

In 1999, Australian Year 9 students demonstrated civic knowledge similar to the international average (and similar to England, New Zealand, Norway and Sweden) but lower than countries such as Finland and the United States. Australian students performed relatively better on the interpretative skills subscale than on the content knowledge subscale. Results from the NAP in civics and citizenship indicate that, from 2004 to 2010, there was a small but significant improvement in civics and citizenship scores for Year 10 students.

Australian 15-year-old students performed very well in digital reading, being similar to New Zealand, and only behind Korea, among the 15 OECD countries that participated. In addition, the NAP in ICT literacy also showed steady improvement in ICT literacy from 2005 to 2011 among Year 6 students. This field appears to be one in which Australian students perform well and it could be argued that it is a field that is important for the future.

5 SUMMARY AND CONCLUSION

Large-scale assessments play an important role in education policy and education planning in Australia and in many other countries. They have become increasingly used as tools for monitoring the effectiveness of educational systems. These large-scale assessments include international assessment surveys, national assessment programs and assessment programs implemented within particular education systems. Although they differ in purpose, approach and methods these assessment programs share the common features of utilising a common assessment tool administered to large numbers of students (either samples or populations) under uniform conditions. In practice these assessments typically utilise similar features of test design and methods of analysis and reporting but they differ in much of the detailed aspects of design and method. Most recent large-scale assessments embody methods for measuring change on an absolute scale (rather than just in a relative sense) but they differ in the extent to which they measure higher-order expertise in the field. Masters and Forster (2000: 1) argue that large-scale assessments will be most useful if they 'incorporate assessments of higher-order skills and thinking' and 'if results are reported in ways that recognise and encourage high achievement'.

The purpose of this report has been to provide a synthesis or appraisal of results from large-scale national and international assessments in Australia over the past 20 years. The international assessments included were the Programme for International Student Assessment (PISA) since 2000, the Trends in International Mathematics and Science Studies (TIMSS) since 1994/5, the IEA Progress in International Reading Literacy Study (PIRLS) since 2010 in Australia (but since 2001 internationally) and the IEA Civic Education Study (CIVED). National assessments that have been included were the National Assessment Program – Literacy and Numeracy (NAPLAN) since 2008, and the NAP sample studies in science literacy since 2003, civics and citizenship since 2004 and ICT literacy since 2005. Those studies covered a range of learning areas, and cross-curricular aspects of learning, in primary and secondary school. The paper has been mainly based on published results from those assessment programs but in a few places it has made use of secondary analyses of publicly available data files.

This final chapter of the report summarises the findings and offers some interpretations of those findings. As for the report as a whole, the emphasis in this chapter is on examining change over time where that is possible but it also takes account of comparative data regarding other countries, differences among jurisdictions and differences among groups of students. Rather than providing a full summary the chapter focuses on what appears to us to be the major highlights in the data.

READING ACHIEVEMENT

Perspectives on reading achievement in secondary schools can be inferred from the PISA surveys conducted every three years since 2000 as well as from NAPLAN surveys conducted annually since 2008.[38] Consideration of reading achievement in primary schools is informed by data from NAPLAN surveys conducted annually since 2008 as well as the internationally comparative data provided by PIRLS in 2010/11.

Lower and middle secondary schooling

Although reading achievement by Australian 15-year-olds (the modal Year for 15-year-olds is Year 10) was considered high compared to most other OECD countries in 2000, there has been a small but significant decline from 2000 to 2009. This decline has been a little more pronounced in the upper part of the distributions of achievement scores than at the lower part of those distributions. Reading achievement among Year 9 students in NAPLAN has not changed over the period from 2008 to 2012. The spread of reading scores among Australian students is relatively large compared to other countries and this did not change between 2000 and 2009.

Clearly the PISA and NAPLAN results refer to different time periods and so cannot be directly compared. However, the test designs also differ with PISA assessments containing a higher proportion of items concerned with higher-order processes and NAPLAN containing a higher proportion of items around the national minimum standard. Consequently, it would be possible for the mean for PISA reading to decline (along with the performance of students at the upper performance levels) but for there to be no appreciable change at the lower performance levels. In those circumstances different trends in means for PISA and NAPLAN reading could be expected.

Two other interesting features of the change in PISA reading scores were that there was:

- an increase in percentage of the variation in student scores that was associated with differences among schools and the association of the between-school differences with the average socioeconomic background of the students at each school; and
- a decline in reading achievement that was not evident in all jurisdictions which suggests that there could be some organisational and curricular aspects of school systems associated with the decline in reading achievement.

These two conclusions have implications for the organisation and operation of Australian school systems.

Finally, the relative differences in reading achievement among groups of students defined by personal, social and demographic characteristics other than jurisdiction did not appear to change over the time from 2000 to 2009. An explanation for change does not appear to reside in changes for particular groups of students.

Primary schooling

Data that can inform judgements about reading achievement among primary school students cover a more limited period of time than data regarding students in secondary schools. NAPLAN reading covers the period from 2008 onwards and data from PIRLS references 2010.

38 NAPLAN assessments are conducted in Year 7 and Year 9 which form part of secondary schools in most jurisdictions. However, it is not clear whether NAPLAN for Year 7 should be considered as providing perspectives on the effectiveness of secondary schools since, even where Year 7 is part of secondary schooling, the assessment is administered in May of the first Year of secondary school.

NAPLAN reading data show that there has been a steady improvement in reading achievement among Year 3 students from 2008 to 2012 amounting to one-fifth of a standard deviation. There has also been a smaller less steady increase of one-eighth of a standard deviation in Year 5 reading achievement over the same period. These improvements give some cause for optimism in terms of the efforts that have been applied to the early years of schooling and to the years before school. These efforts have included expanding provision for pre-school education (Maguire & Hayes, 2011), the emergence of an *Early Years Learning Framework* (DEEWR, 2009), and a *National Quality Framework for Early Childhood Education and Care* (ACECQA, 2013). The *Australian Early Development Inventory* was implemented in 2009 to gather data about school readiness and inform planning decisions (Goldfeld, Sayers, Brinkman, Silburn & Oberklaid (2009). Since the late 1990s education authorities provided smaller classes in the early years of school[39] and emphasised the teaching of literacy in those years.

A closer inspection of the improvements in Year 3 reading shows that the shifts at the top of the distribution were greater than the shifts at other points of the distribution. Between 2008 and 2012 the percentage of students in NAPLAN Bands 1 and 2 combined dropped by four percentage points (from 18% to 14%) whereas the percentage of students in Band 6 increased by eight percentage points (from 18% to 26%). It is possible that some parents make greater use of the educational opportunities in the years before school or that students who have developed greater expertise in reading at an early age are better able to benefit from teaching in the early years and grow more rapidly. However, the results do raise doubts about whether the early years' initiatives have been successful in providing a more even start to schooling.

The extent to which Year 3 reading scores improved over time differed among jurisdictions with the largest being the increase in Queensland of two-fifths of a standard deviation, which is a considerable improvement. In Queensland there had been an introduction of a Year K (or preparatory year) before Year 1 prior to this period of time with a focus on literacy development.

ACHIEVEMENT IN MATHEMATICS AND NUMERACY

There are data covering a longer span of time with respect to mathematics than reading. Data on achievement in mathematics for Year 4 and Year 8 date back to 1994 from TIMSS (which has antecedent IEA mathematics studies going back to 1961). PISA provides perspectives on mathematics achievement from 2003 to 2009 among 15-year-olds and NAPLAN numeracy extends from 2008 to 2012.

Secondary schooling

There was no overall change in Australian Year 8 mathematics achievement on TIMSS from 1994/5 to 2010/11 although there had been a dip in 2006. The same 16-year period had seen improvements in a number of countries including Korea, the United States, Hong Kong and the Russian Federation. Correspondingly there had been declines in a number of countries including Sweden, Norway, Hungary and Japan. Within Australia there were declines in the mathematics achievement of Year 8 students in Western Australia and South Australia but there were no changes in the relative performance of groups of students based on personal, social and demographic characteristics.

Data from PISA indicate a small decline in the mathematics achievement of one-tenth of a standard deviation among 15-year-olds from 2003 to 2009, a change that is small but statistically significant. The spread of mathematics scores was not different from that of other OECD countries (which is

39 For example, in New South Wales government schools the average class sizes in 1997 for Years K, 1 and 2 were 24.1, 25.5 and 26.2 and for Years 3 through 6 the average was 26.8. In 2011 the average class sizes in Years K, 1 and 2 were 19.2, 21.2 and 22.6 respectively compared with an average of 26.1 across Years 3 through 6 (DEC, 2011).

different than that for reading). Other OECD countries to record a significant decline in mathematics over the same period included the Czech Republic, Ireland, Belgium, Sweden, France, Denmark and the Netherlands. OECD countries to record a significant increase over the same period included Portugal, Greece, Italy and Germany. There was a small change in the shape of the distribution of mathematics scores in Australia with a relatively larger decline at the top of the distribution than at the bottom of the distribution, although the change was not as clearly evident as for reading.

One would not expect to observe the same patterns among countries, or jurisdictions, in PISA mathematics as in TIMSS mathematics. This is both because PISA samples an age group whereas TIMSS samples a Year level (countries differ in their age by Year level distributions), and because PISA mathematics places relatively greater emphasis on 'data' items and TIMSS places relatively greater emphasis on 'algebra'.

There were differences among jurisdictions in the change in PISA mathematical literacy scores between 2003 and 2009. In South Australia, the ACT, Western Australia and New South Wales there were significant declines. The jurisdictional declines in mathematics between 2003 and 2009 matched those for reading between 2000 and 2009. Over the period from 2003 to 2009 there was no change in the relative performance of designated groups of students except that the mathematics achievement of students whose home language was other than English did not decline whereas that for other students did decline.

There was no change in NAPLAN numeracy achievement for Year 9 over the period from 2008 to 2012. There was a very small decline of approximately one-tenth of a standard deviation in Year 7 numeracy. This mainly took place between 2011 and 2012 but it is not clear to what extent, if at all, this can be attributed to secondary schooling because in some jurisdictions Year 7 is part of primary school and even for the other jurisdictions the assessment takes place in May of the first year of secondary school.

Primary schooling

The achievement of Year 4 students in Australia in 2010/11 in TIMSS was higher than in 1994/5 (by one-fifth of a standard deviation) but not different from 2006/7. The improvement took place between 2002/3 and 2006/7 and was then maintained. Other countries to record improvements were Portugal, England, Slovenia, Hong Kong, Iran, Korea, the United States, Japan, Norway, New Zealand and Singapore. Although Year 4 mathematics achievement improved for Australia over the period, it remained below a group of countries including Singapore, Korea, England, the United States, the Netherlands, Portugal and Ireland. Within Australia there were significant increases over the 16-year period in Tasmania, New South Wales, Victoria, the ACT and South Australia. There was some evidence that the increase had been greater for students with more abundant literacy resources in their homes than for students with few literacy resources.

NAPLAN numeracy scores for Year 5 students improved between 2008 and 2012 by a little less than one-fifth of a standard deviation. However, this increase took place between 2008 and 2009 and has not changed since then. There was also a smaller decrease of one-tenth of a standard deviation in Year 7 numeracy. In this case the mean had been relatively constant until 2012 but there was a significant decline from 2011 to 2012. The improvements in Year 5 numeracy were observed in New South Wales, Queensland, Western Australia, South Australia, Tasmania and the ACT. It is notable that there was no increase in NAPLAN numeracy achievement among Year 3 students to correspond with the improvement in reading achievement in Year 3.

SCIENCE ACHIEVEMENT

There is a broad interest in science achievement in Australia based largely on concerns with skill formation as well as broader scientific literacy. There are data for Year 4 and Year 8 from 1994 onwards that can inform perspectives on science achievement as well as data from PISA in 2006 and 2009 and data from the NAP sample study of science literacy in Year 6 in 2006 and 2009.

Secondary schooling

According to TIMSS the science achievement of Australian Year 8 students did not change from 1994/5 to 2010/11 despite a ‘blip’ in 2002/3. Countries which improved over the period included Lithuania, Slovenia, Hong Kong, the Russian Federation, Korea and Iran as well as the Canadian province of Ontario. Countries which declined over the same period included Sweden, Norway and Hungary.

Data from PISA suggest that Australian 15-year-olds perform well in scientific literacy. In 2009 Australian 15-year-olds performed less well than Finland, Hong Kong, Japan, Korea and Singapore and similarly to seven countries including New Zealand, Canada, the Netherlands and Germany. It performed better than a number of countries such as the United Kingdom and the United States. However, as was noted for reading literacy, Australia has a relatively wide dispersion of scores. There was no change in the average scientific literacy scores of Australian 15-year-olds between 2006 and 2009 although a number of countries recorded a significant improvement including Portugal, Korea, United States, Italy, Norway, Poland and Turkey. The differences among groups of students based on personal, social and demographic characteristics were similar to the differences observed for PISA reading and mathematics.

Primary schooling

The mean for science achievement in Australia appears to be relatively lower in Year 4 than was the case in Year 8. In Year 8 the Australian mean science achievement had been similar to that of the United States, Hungary, Lithuania, New Zealand and Sweden as well as the province of Ontario. In Year 4 the Australian mean score for science achievement was lower than that of the United States, Hungary, Ontario and Sweden. Fifteen countries (Korea, Singapore, Japan, Chinese Taipei, Russian Federation, United States, the Czech Republic, Hong Kong, Hungary, Sweden, Austria, Slovak Republic, Netherlands, England and Germany) and the Canadian province of Ontario achieved significantly higher mean science scores for Year 4 than Australia in 2010/11. The mean score science achievement among Year 4 students in Australia did not change between 1994 and 2010 despite a dip in 2006. In 2010, Year 4 science achievement in the ACT and Victoria was significantly greater than the national mean and in Queensland and Western Australia it was significantly lower than the national mean.

Data from the national sample surveys of science literacy in Year 6 show no change in average achievement between 2006 and 2009 (ACARA, 2010). Only in Tasmania was there a change and that was a decline of one-fifth of a standard deviation. In 2009 the mean for the ACT was significantly greater than the national mean and the means for South Australia and the Northern Territory were significantly lower than the national mean.

CIVICS AND CITIZENSHIP

Perspectives on student achievement in civics and citizenship can be derived from the NAP civics and citizenship assessments conducted among students in Year 6 and Year 10 in 2004, 2007 and 2010. In addition the IEA Civic Education Study (CIVED) from 1999 provides international comparisons.

Secondary schooling

In the IEA Civic Education Study of 1999, Australian Year 9 students demonstrated civic knowledge similar to the international mean and a group of 11 countries (Norway, Czech Republic, Hungary, Slovenia, Denmark, Germany, Russian Federation, England, Sweden, Switzerland and Bulgaria). The Australian average was lower than a group of countries that included Poland, Finland and the United States. However, Australian students performed relatively better on the interpretative skills subscale than on the content knowledge subscale. On interpretative skills Australian students were outperformed by only the United States and Finland. In Australia and in most countries there were no differences between female and male students. In the large majority of countries, and in Australia, the more books students reported in the home (which is a measure of social and cultural background) the better they performed on the civic knowledge test. Australia did not participate in the IEA International Civic and Citizenship Education Study in 2009.

Results from the NAP civics and citizenship assessments indicate that, from 2004 to 2010, there was a small but significant improvement of approximately one-sixth of a standard deviation in civics and citizenship scores for Year 10 students. Approximately half of the Year 10 students attained the proficient standard.

Primary schooling

Between 2004 and 2010 there was no change in the average scores of Year 6 students on the national assessment of civics and citizenship. Just over half the Year 6 students attained the proficient standard.

In 2010 there were large differences between Indigenous and non-Indigenous students and between the two parental occupation categories of 'senior managers and professionals' and 'unskilled labourers, office, sales and service staff'. In each case the differences were more than one standard deviation. There were also differences between school location categories (especially between metropolitan and remote locations). Females scored higher than males by one-fifth of a standard deviation at Year 6 and one-third of a standard deviation at Year 10. There were no differences between students who mainly spoke English at home and other students.

DIGITAL LITERACY

There are two large-scale assessment surveys that provide perspectives on digital or ICT literacy among school students in Australia. One is the PISA study of digital reading conducted as part of the 2009 cycle of PISA among 16 countries (OECD, 2011). The other is the NAP ICT literacy surveys which have been conducted with Australian Year 6 and Year 10 students over three cycles in 2005, 2008 and 2011 (ACARA, 2012b).

Secondary schooling

Australian 15-year-old students performed well on digital reading. Korea had the highest mean on digital reading but was followed by Australia and New Zealand which were in turn higher than the other 13 OECD countries that participated. The difference between Korea and Australia was a little larger for electronic reading (31 points) than print reading (24 points). In all countries, including Australia, females performed better than males but by less than for print reading. Digital reading was associated with socioeconomic background, and other characteristics, in a similar way to the association of print reading with socioeconomic background.

The national assessment program in ICT literacy also showed high levels of achievement in ICT literacy among Year 10 students with 65 per cent of those students having attained the proficient standard. This percentage had not changed significantly since 2005. There were fewer differences among jurisdictions for Year 10 students than for Year 6 students but it was evident that achievement was higher in the ACT, Victoria and New South Wales than in Tasmania and the Northern Territory. Patterns of association with student background characteristics were similar to those observed for other outcomes such as reading.

Primary schooling

Between 2005 and 2011 there was an improvement in ICT literacy among Year 6 students. The percentage of students attaining the proficient standard rose from 49 per cent in 2005 to 62 per cent in 2011. Among Year 6 students ICT literacy was higher in the ACT, Victoria and New South Wales than in Western Australia, Queensland, Tasmania and the Northern Territory. For all jurisdictions except Tasmania and the Northern Territory there was an increase in mean score between 2005 and 2011.

Student background characteristics were related to ICT literacy and the patterns are similar in Year 6 and Year 10. Furthermore the relative scores between groups did not appear to change between 2005 and 2011. There was a large effect associated with parental occupation. In Year 6, the difference in the mean score of students with parents in 'unskilled manual, office and sales' occupational groups was four-fifths of a standard deviation lower than that for students with parents from the 'senior managers and professionals' occupational group. There is also a substantial gap in ICT literacy between Indigenous and non-Indigenous students. Females recorded higher levels of ICT literacy than males.

CONTEXT

The medium-term context in which these varied shifts in achievement outcomes have occurred is one which has seen the emergence of a national perspective on educational governance with an increasing role for federal structures, and national companies established by relevant commonwealth and state ministers, in educational policy and governance.

The period considered has also seen a greater emphasis on assessment-based accountability through jurisdictional assessments that began as exercises independent of each other and then became linked through 'benchmarking' exercises. From 2008 a National Assessment Program – Literacy and Numeracy (NAPLAN) was implemented and, in some senses, can be seen as a consolidation of those jurisdictional assessments. The publication of NAPLAN results for individual schools on the public My School website has emphasised the accountability purpose of this program. Data which were previously used for planning and monitoring by authorities are now publicly visible. It is rather too early to look for associations between trends in achievement and these recent developments.

Similarly, it is too early to attribute any of the trends discussed in this paper to the National Partnerships[40] (on Improving Teacher Quality, Literacy and Numeracy, or Youth Attainment and Transitions as well as Early Childhood) that are directed to specific areas of reform and outline agreed policy objectives, outputs and performance benchmarks. The earliest of these were implemented from 2009 over several years.

There have been increases in educational expenditure since 2000. The OECD notes that over the period from 2000 to 2009 expenditure per student for primary, secondary and post-secondary non-tertiary education increased in every country by an average of 36 per cent. The increase for Australia over the period was 44 per cent (OECD, 2012: 224). Data from Year Book Australia for 2012

40 National Partnerships are agreements between the Commonwealth and States that are directed to specific areas of action and contain agreed policy objectives, outputs, performance benchmarks and financial commitments.

show that the increase in the Consumer Price Index over the same period from 2000 to 2009 was 31 per cent (or 35 per cent to 2010) (ABS, 2012b: Table 29.8). Thus there appears to have been a net increase in educational expenditure but an increase that is not as large as sometimes suggested. In 2009 the spending on primary education in Australia was US$8,328 per student which was just a little greater than the OECD average of US$7719. (OECD 2012: 219). For secondary education the corresponding figure was US$10,134 per student which was also a little greater than the OECD average of $US9,312 (OECD, 2012: 219).[41] However, in terms of teaching resources in primary and secondary schools, measured by student–teacher ratios, the change does not appear to have been so large. Over the period from 2001 to 2011 there was an overall decrease in student-to-teacher staff ratios in primary schools from 17.0 in 2001 to 15.6 in 2011 which represents an eight per cent improvement (ABS, 2012a). As noted in Chapter 2, many of those increased resources were directed to the early years of school (Years K through 2). In secondary schools the change in the student-to-teacher ratios was from 12.5 to 12.0 which is a four per cent improvement (ABS, 2012a).

Over the period being considered there has been a shift in the distribution of enrolments between government and non-government schools. In 1996, 29 per cent of school enrolments were in non-government schools but by 2012 the percentage had risen to 35 per cent (ABS, 2013; ABS, 2010). For primary school enrolments the shift was from 26 to 31 per cent and for secondary school enrolments the shift was from 34 to 40 per cent. In addition there appears to have been a growth in selective entry and specialist schools within the government school sector. It is not clear what influence these shifts may have had on school outcomes, or on total expenditure on primary and secondary schooling, but they do represent a significant change in context. Furthermore, the report of PISA 2009 suggests that school systems with low levels of differentiation are more likely to perform above the OECD average and show less pronounced associations of achievement with socioeconomic background (OECD, 2010e: 62–68). The OECD Education Policy Outlook for Australia observes this as an issue and counsels that it is 'important to ensure that there are mechanisms to mitigate this negative effect' (OECD, 2013: 8).

The emerging focus of educational reform is on improvement in school and teaching processes. A national school improvement tool looks at indicators of school practice (Masters, 2012). The tool involves assessments of the quality of practice (low, medium, high and outstanding) on nine aspects of school practice: an explicit improvement agenda, analysis and discussion of data, a culture that promotes learning, targeted use of school resources, an expert teaching team, systematic curriculum delivery, differentiated classroom learning, effective teaching practices and school–community partnerships. The initiative known as *'Better Schools: A National Plan for School Improvement'* that incorporates this tool, has the potential to impact on student outcomes across a range of areas.

USING DATA FROM INTERNATIONAL AND NATIONAL ASSESSMENTS

Data from international and national assessment studies can be used in many ways. In this report we have focused on examining trends and making use of the fact that the assessment instruments are equated over time. Used in that way it is possible to interpret changes in achievement in relation to changes in policy, provision, practice and context. To us this seems to be the most productive way to use these data but it is dependent on the strength of the equating processes. Of course, it is also useful to compare statistics from international and national assessments at a point in time and we have also

41 If the focus is on expenditure per student as a percentage of per capita Gross Domestic Product, Australian expenditure at the primary and secondary levels is just a little below, rather than a little above, the OECD average (OECD, 2012: 222).

made use of those sorts of comparisons. These comparisons may be made in terms of countries, jurisdictions or groups of students based on personal, social and geographic characteristics.

Any data obtained from assessments has associated uncertainty arising from measurement error and sampling error. We have been careful to emphasise the uncertainty in the estimates we have referenced in our report so that we do not claim differences about which we cannot be sure. Comparing rankings of countries or jurisdictions typically does not take such uncertainties into account and can be misleading. In addition, for rankings one cannot be sure whether a country missed a place on the podium by a millimetre or a kilometre. Rankings can also change as a result of new entrants to or withdrawals from the tournament. We have not made much use of rankings in this report. In addition we have tried to keep in mind that differences among countries can be a consequence of factors other than policies in education systems or practices in classrooms.

CONCLUSION

Data from international and national large-scale assessments can provide important broad indications of the progress and status of the outcomes of school systems. We have argued that the most fruitful way to use these data is to examine changes in achievement over time (both improvements and declines) and to relate those changes to developments in policy, practice and context that took place in the immediately preceding years. This seems most likely to generate insights that can inform practice elsewhere. It is surprising that this approach has not been more widely adopted.

APPENDIX A

LIST OF ACRONYMS

ABS	Australian Bureau of Statistics
ACARA	Australian Curriculum, Assessment and Reporting Authority
ACECQA	Australian Children's Education and Care Quality Authority (ACECQA)
ARIA	Accessibility/Remoteness Index of Australia
CIVED	IEA Civic Education Study
COAG	Council of Australian Governments
DEC	Department of Education and Communities (New South Wales)
DEEWR	Department of Education, Employment and Workplace Relations
ESCS	index of economic, social and cultural status
ICCS	IEA International Civic and Citizenship Education Study
ICT	information and communications technology
IEA	International Association for the Evaluation of Educational Achievement
LBOTE	language background other than English
MCEETYA	Ministerial Council on Education, Employment, Training and Youth Affairs
MCEECDYA	Ministerial Council for Education, Early Childhood Development and Youth Affairs
NAEP	National Assessment of Educational Progress
NAP	National Assessment Program
NAP-CC	National Assessment Program: Civics and Citizenship
NAP-ICTL	National Assessment Program: ICT Literacy
NAP-SL	National Assessment Program – Science Literacy
NAPLAN	National Assessment Program – Literacy and Numeracy
OECD	Organisation for Economic Cooperation and Development
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
SCSEEC	Standing Council on School Education and Early Childhood
STEM	science, technology, engineering and mathematics
TIMSS	Trends in International Mathematics and Science Studies

REFERENCES

Afrassa, T., & Keeves, J. (1999). Changes in students' mathematics achievement in Australian lower secondary schools over time. *International Education Journal*, *1* (1): 1–21.

Ainley, J., & Ainley, M. (2011). The influence of interest in science on the uptake of science studies. Paper presented to the European Association for Research in Learning and Instruction, Exeter, United Kingdom.

Ainley, J., & Khoo, S. K. (under review). *Five years of the National Assessment Program in Literacy and Numeracy*. Sydney: ACARA.

Ainley, M., & Ainley, J. (in press). The contribution of opportunity and support to triggering and maintaining interest in science. In K. A. Renninger and M. Nieswandt (Eds.) *The Handbook of Interest, the Self, and K–16 Mathematics and Science Learning*. Washington, DC: American Educational Research Association.

Angus, M., Olney, H., and Ainley, J. (2007). *In the Balance: The Future of Australia's Primary Schools*. Kaleen, ACT: Australian Primary Principals Association

Australian Association of Mathematics Teachers (AAMT) (1997). *Numeracy=Everyone's Business*. Adelaide: Australian Association of Mathematics Teachers.

Australian Bureau of Statistics (ABS) (2001). *Outcomes of ABS Views on Remoteness Consultation*, Australia, 2001. (ABS Information Paper, Cat. No. 1244.0.00.001). Canberra: ABS.

Australian Bureau of Statistics (ABS) (2010). *School Australia 2009 (Catalogue No. 4221.0)*. Canberra: ABS.

Australian Bureau of Statistics (ABS) (2012a). *School Australia 2011 (Catalogue No. 4221.0)*. Canberra: ABS.

Australian Bureau of Statistics (ABS) (2012b). *Year Book Australia, 2012* (Catalogue Number 1301.0) Canberra : ABS.

Australian Bureau of Statistics (ABS) (2013). *School Australia 2012 (Catalogue No. 4221.0)*. Canberra: ABS.

Australian Children's Education and Care Quality Authority (ACECQA) (2013). *Newsletter (May)*. Canberra: Author.

Australian Curriculum, Assessment and Reporting Authority (ACARA) (2008). *NAPLAN Achievement in Reading, Persuasive Writing, Language Conventions and Numeracy: National Report for 2008*. Sydney: ACARA.

Australian Curriculum, Assessment and Reporting Authority (ACARA) (2009). *Shape of the Australian Curriculum: Mathematics*. Sydney: ACARA.

Australian Curriculum, Assessment and Reporting Authority (ACARA) (2010). *National Assessment Program – Science Literacy Year 6 Report: 2009*. Sydney: ACARA.

Australian Curriculum, Assessment and Reporting Authority (ACARA) (2011). *National Assessment Program – Civics and Citizenship Years 6 and 10 Report: 2010*. Sydney: ACARA.

Australian Curriculum, Assessment and Reporting Authority (ACARA) (2012a). *NAPLAN Achievement in Reading, Persuasive Writing, Language Conventions and Numeracy: National Report for 2012*. Sydney: ACARA.

Australian Curriculum, Assessment and Reporting Authority (ACARA) (2012b). *National Assessment Program – ICT Literacy Years 6 and 10 Report: 2011*. Sydney: ACARA.

Australian Curriculum, Assessment and Reporting Authority (ACARA) (2012c). *Student Report 2012: National Assessment Program Literacy and Numeracy*. Sydney: ACARA.

Bourke, S. F., & Keeves, J. P. (1977). *The Mastery of Literacy and Numeracy: Final Report*. Australian Studies in School Performance. Volume III. ERDC Report no. 13. Canberra: AGPS.

Bourke, S. F, Mills, J. M., Stanyon, J., & Holzer, F. (1981). *Performance in Literacy and Numeracy: 1980*. A Report to the Australian Educational Council on the Australian Studies in Student Performance Project. Canberra: AGPS.

Bybee, R. (2010). Advancing STEM education: a 2020 vision. *Technology and Engineering Teacher*, *70* (1): 30–35.

Council of Australian Governments (COAG) (2008). *National Numeracy Review Report*. Canberra: COAG (Human Capital Working Group).

Cox, C. (2004). *Innovation and Reform to Improve the Quality of Primary Education: Chile. 2005 Education for all Global Monitoring Report*. Santiago, Chile: Ministry of Education.

Department of Education and Communities (DEC) New South Wales (2011). *Average Class Sizes: 1997 to 2011*. Sydney: Author.

Department of Education, Employment and Workplace Relations (DEEWR) (2009). *Belonging, Being and Becoming: The Early Years Learning Framework for Australia*. Canberra: Author.

Global Science Forum (OECD) (2006). *Evolution of Student Interest in Science and Technology Studies: Policy Report*. Paris: OECD Global Science Forum.

Goldfeld S., Sayers M., Brinkman S., Silburn S., & Oberklaid F. (2009). The process and policy challenges of adapting and implementing the Canadian Early Development Instrument in Australia. *Early Education and Development*, *20*(6): 978–991.

Gronmo, L., & Olsen, R. (2008). *TIMSS versus PISA: The case of pure and applied mathematics*. Retrieved 26 March 2013 from http://www.timss.no/publications/IRC2006_Gronmo&Olsen.pdf

Gustafsson, J. (2009). *Report explains declining school performance.* Gothenburg, Sweden: University of Gothenburg.

Husen, T. (Ed.). (1967). *International study of achievement in mathematics: A comparison of twelve countries* (Vols. 1–2). Stockholm: Almqvist & Wiksell.

Jakubowski, M., Patrinos, H., Porta, E., & Wisniewski, J. (2010). The impact of the 1999 education reform in Poland. *OECD Education Working Papers*, *No. 49*.

Jones, R. (2004). *Geolocation Questions and Coding Index*. Canberra: Quantitative Evaluation and Design.

Keeves J. P., & Bourke, S. F. (1976). *Literacy and Numeracy in Australian Schools: A First Report*. Australian Studies in School Performance Volume I. ERDC Report no. 8. Canberra: AGPS.

Lohmar, B., & Eckhardt, T. (2011). *The Education System in the Federal Republic of Germany 2010/2011*. Bonn: Secretariat of the Standing Conference of Ministers of Education and Cultural Affairs of the Landers in the Federal Republic of Germany.

Lokan, J., Greenwood, L., & Cresswell, J. (2001). *15-up and Counting, Reading, Writing, Reasoning: How Literate are Australia's Students?* Melbourne: ACER.

Maguire, B., & Hayes, A. (2011). Access to preschool education in the year before full-time school. In Australian Institute of Family Studies, *Longitudinal Study of Australian Children: Annual Statistical Report, 2011*. Melbourne: AIFS.

Martin, M., & Mullis, I. (2013). *Methods and Procedures in TIMSS and PIRLS 2011*. Boston: Lynch School of Education, Boston College.

Martin, M., Mullis, I., Foy, P., & Stanco, G. (2012). *TIMSS 2011 International Results in Science*. Boston: Lynch School of Education, Boston College.

Masters, G. (2012). *Measuring and Rewarding School Improvement*. Canberra: Department of Education, Employment and Workplace Relations.

Masters, G., & Forster, M. (1997a). *Mapping Literacy Achievement: Results of the 1996 National School English Literacy Survey*. Canberra: Department of Employment, Education, Training and Youth Affairs.

Masters, G., & Forster, M. (1997b). *Literacy Standards in Australia*. Canberra: Department of Employment, Education, Training and Youth Affairs.

Masters, G., & Forster, M. (2000). *The Assessments We Need*. Melbourne: ACER.

Masters, G., Lokan, J., Doig, B., Khoo, S. K., Lindsey, J., Robinson, L., & Zammit, S. (1990). *Profiles of Learning: The Basic Skills Testing Program in New South Wales 1989*. Melbourne: ACER.

Matthews, P., Klaver, E., Lannert, J., Ó Conluain, G., & Ventura, A. (2008). *Policy Measures Implemented in the First Cycle of Compulsory Education in Portugal (International Evaluation).* Lisbon: Office for Education Statistics and Planning, Ministry of Education.

McGaw, B., Long, M., Morgan, G., & Rosier, M. (1989). *Literacy and Numeracy in Victorian Schools: 1975–1988*. Melbourne: ACER.

McLarty, K. L., Way, W. D., Porter, A.C., Beimers, J. N., & Miles, J. A. (2013). Evidence-based standard setting: Establishing a validity framework for cut scores. *Educational Researcher, 42* (78) (originally published online 1 March 2013). DOI: 10.3102/0013189X12470855.

Mellor, S., Kennedy, K., & Greenwood, L. (2001) *Citizenship and Democracy: Student's Knowledge and Beliefs: Australian Fourteen Year Olds and the IEA Civic Education Study*. Canberra: Department of Education, Training and Youth Affairs.

Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) (2005). *National Assessment Program – Information and Communication Technology Literacy Years 6 and 10: An Assessment Domain for ICT Literacy*. Carlton: Curriculum Corporation.

Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) (2007). *National Assessment Program – Civics and Citizenship Years 6 & 10 2004 Report*. Carlton: Curriculum Corporation.

Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) (2008). *National Assessment Program – ICT Literacy Years 6 & 10 2005 Report.* Carlton: Curriculum Corporation.

Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) (2009). *National Assessment Program – Civics and Citizenship Years 6 & 10 2007 Report*. Carlton: Curriculum Corporation.

Mourshed, M., Chijioke, C., & Barber, M. (2010). *How the World's Most Improved School Systems Keep Getting Better.* London: McKinsey & Company.

Mullis, I., Martin, M., Foy, P., & Arora, A. (2012). *TIMSS 2011 International Results in Mathematics*. Chestnut Hill, MA: Lynch School of Education, Boston College.

Mullis, I., Martin, M., Foy, P., & Drucker, K. (2012). *PIRLS 2011 International Results in Reading*. Chestnut Hill, MA: Lynch School of Education, Boston College.

Mullis, I., Martin, M., Gonzalez, E., & Chrostowski, S. (2004). *Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: Lynch School of Education, Boston College.

Mullis, I., Martin, M., Kennedy, A., Trong, K., & Sainsbury, M. (2009). *PIRLS 2011 Assessment Framework*. Chestnut Hill, MA: Lynch School of Education, Boston College.

Neidorf, T.S., Binkley, M., Gattis, K., & Nohara, D. (2006). *Comparing Mathematics Content in the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA) 2003 Assessments* (NCES 2006–029). Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Office of the Chief Scientist (2012). *Health of Australian Science*. Australian Government: Canberra.

Organisation for Economic Cooperation and Development (OECD) (2009). *PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science*. Paris: OECD.

Organisation for Economic Cooperation and Development (OECD). (2010a). *PISA 2009 Results: What Students Know and Can Do*. Paris: OECD.

Organisation for Economic Cooperation and Development (OECD). (2010b). *PISA 2009 Results: Learning Trends*. Paris: OECD.

Organisation for Economic Cooperation and Development (OECD). (2010c). *PISA 2009 Results: Overcoming Social Background*. Paris: OECD.

Organisation for Economic Cooperation and Development (OECD). (2010d). *Pathways to Success: How Knowledge and Skills at Age 15 Shape Future Lives in Canada*. Paris: OECD.

Organisation for Economic Cooperation and Development (OECD). (2010e). *PISA 2009 Results: What Makes a School Successful? Resources, Policies and Practices*. Paris: OECD.

Organisation for Economic Cooperation and Development (OECD). (2011). *PISA 2009 Results: Students On Line: Digital Technologies and Performance*. Paris: OECD.

Organisation for Economic Cooperation and Development (OECD). (2012). *Education at a Glance 2012: OECD Indicators*. Paris: OECD.

Organisation for Economic Cooperation and Development (OECD). (2013). *Education Policy Outlook: Australia*. Paris: OECD.

Print, M., & Hughes, J. (2001). *Key Performance Measures in Civics and Citizenship Education: Report to the National Education Performance Monitoring Taskforce*. Sydney: Centre for Research and Teaching in Civics, University of Sydney.

Skolverket (Swedish National Agency for Education) (2009). *What Influences Educational Achievement in Swedish Schools? A Systematic Review and Summary Analysis*. Stockholm: Skolverket.

Thomson, S., Cresswell, J., & De Bortoli, L. (2004). *Facing the Future: A Focus on Mathematical Literacy among Australian 15-year-old Students in PISA 2003*. Melbourne: ACER.

Thomson, S., & De Bortoli, L. (2008). *Exploring Scientific Literacy, How Australia Measures Up: The PISA 2006 Survey of Students' Scientific, Reading and Mathematical Literacy Skills*. Melbourne: ACER

Thomson, S., De Bortoli, L., Nicholas, M., Hillman, K., & Buckley, S. (2011). *Challenges for Australian Education: Results from PISA 2009*. Melbourne: ACER.

Thomson, S., Hillman, K., & Wernert, N. (2012). *Monitoring Australian Year 8 Student Achievement Internationally: TIMSS 2011*. Melbourne: ACER.

Thomson, S., Hillman, K., Wernert, N., Schmid, M., Buckley, S., & Munene, A. (2012). *Monitoring Australian Year 4 Student Achievement Internationally: TIMSS and PIRLS 2011*. Melbourne: ACER.

Torney-Purta, J., Lehmann, R., Oswald, H., and Schulz, W. (2001) *Citizenship and Education in Twenty-eight Countries: Civic Knowledge and Engagement at Age Fourteen*. IEA: Amsterdam.

U.S. Department of Labor. (2007). *The STEM Workforce Challenge: The Role of the Public Workforce System in a National Solution for a Competitive Science, Technology, Engineering, and Mathematics (STEM) Workforce*. US Department of Labor: Washington DC. Retrieved 2012-12-21.

Wu, M. (2008). *A Comparison of PISA and TIMSS 2003 achievement results in Mathematics*. Paper presented to the annual meeting of the American Educational Research Association, New York, April.