Australian Government
Department of Education, Science and Training

2004/2

SCHOOLING ISSUES DIGEST

Performance of Australian School Students in International Studies in Science

ABOUT THE DIGESTS

The Australian Government Department of Education, Science and Training (DEST) is publishing a series of brief reports titled "Schooling Issues Digests" which summarise existing research material on selected topics relevant to schooling in Australia. The purpose of these digests is to provide status reports on the results of recent international and national research on selected topics, in a non-technical, easy to read format, which brings together and demystifies complicated research and statistical data. Contact Roger Wright on (02) 6240 7897 or email address roger.wright@dest.gov.au for more information on this series.

The authors of this digest are Professors Richard Gunstone and Peter Fensham. Professor Gunstone is Professor of Science and Technology Education at Monash University. Emeritus Professor Fensham, formerly Professor of Science Education at Monash University, is now Adjunct Professor in the School of Mathematics, Science and Technology at Queensland University of Technology. Both have been involved in science education for many years and have each written many academic papers and chapters in scholarly publications, together with many conference and journal papers and books.

This Digest was produced by Professors Gunstone and Fensham with funding from the Australian Government Department of Education, Science and Training. The views expressed here do not necessarily represent the views of the Department.

OVERVIEW OF THIS DIGEST

In the last decade, Australian primary and secondary students have participated in major projects making international comparisons of various forms of science learning outcomes and other aspects of schooling in science. These projects are TIMSS (Third International Mathematics and Science Study) and PISA (Programme for International Student Assessment). Each project assessed different things, and hence each leads to somewhat different findings.

TIMSS has so far tested science learning outcomes among three groups – middle primary, lower secondary, final year secondary – in the mid 1990s, and then re-tested lower secondary students in 1999. The focus in these tests was essentially on 'what students know'. TIMSS also collected information about the intended science curricula in participating countries, and a range of data about student and teacher perceptions and approaches, and school characteristics.

PISA has taken a different approach. In its first round of data collection in 2000 its focus was strongly on 'what students can do with their knowledge'. PISA is concerned with reading, mathematics and science – all in terms of 'literacy'. This project has involved 15 year olds, mostly in the industrialised countries of the OECD. The literacy focus is about finding out how well students at (or near) the end of compulsory school are prepared for the range of challenges that face them. Hence, 'scientific literacy', the focus of the science data collection, is seen by PISA as "the capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help make decisions".

Although the two projects took different approaches to assessing science learning outcomes, there are interesting common findings.

- In general Australian students performed very well, in both TIMSS and PISA. Only a small number of countries had better average performance than Australia (see Table 1).

The digest is available in HTML and PDF versions on the DEST website at http://dest.gov.au/schools/publications/2004/index.htm

- The spread of Australian scores in both TIMSS and PISA was greater than the case for some other high performing countries. That is, while Australian students have done well overall, some other high-performing countries have fewer individual students who perform less well.
- In TIMSS, Australian students have consistently done very well on questions relating to environmental issues. This is consistent with the nature of intended science curricula around the country.
- In both TIMSS and PISA, girls and boys performed equally well. This contrasts with earlier international science testing in the 1970s and 1980s when Australian boys performed better than Australian girls.
- The form of science testing in PISA made substantial reading demands. PISA reading literacy results showed strong gender differences (in favour of girls), yet there were no such differences in the reading intensive science test results. This suggests that the story form of reading material in the science testing engaged boys to a greater extent than did the reading literacy tests. This is then consistent with the idea of "curriculum as story" as a very important approach to use in the teaching of science.

Table 1: Australian performance in science compared with other countries - TIMSS and PISA

	TIMSS POPULATION 1 (MIDDLE PRIMARY; 1994)	TIMSS POPULATION 2 (LOWER SECONDARY; 1994)	TIMSS-REPEAT (LOWER SECONDARY; 1998)	PISA* (MID SECONDARY; 2000)
COUNTRIES PERFORMING SIGNIFICANTLY[8] BETTER THAN AUSTRALIA	Republic of Korea Japan	Singapore Czech Republic Japan Republic of Korea	Chinese Taipei	Republic of Korea Japan
NUMBER OF COUNTRIES DOING AS WELL AS AUSTRALIA	6 (USA, Austria, Netherlands, Czech Republic, England, Singapore)	14 (Bulgaria, Netherlands, Slovenia, Austria, Hungary, England, Belgium (Flemish), Slovak Rep., Russian Fed., Ireland, Sweden, USA, Germany, Canada)	15 (Singapore, Hungary, Japan, Republic of Korea, Netherlands, Czech Rep., England, Finland, Slovak Rep., Belgium (Flemish), Slovenia, Canada, Hong Kong, Russian Fed., Bulgaria)	6 (Finland, UK, Canada, NZ, Austria, Ireland)
NUMBER OF COUNTRIES DOING SIGNIFICANTLY MORE POORLY THAN AUSTRALIA	17 (Canada, Slovenia, Ireland, Scotland, Hong Kong, Hungary, NZ, Norway, Latvia, Israel, Iceland, Greece, Portugal, Cyprus, Thailand, Iran, Kuwait)	22 (Norway, NZ, Thailand, Israel, Hong Kong, Switzerland, Scotland, Spain, France, Greece, Iceland, Romania, Latvia, Portugal, Denmark, Lithuania, Belgium (French), Iran, Cyprus, Kuwait, Colombia, Sth Africa)	21 (USA, NZ, Latvia, Italy, Malaysia, Lithuania, Thailand, Romania, Israel, Cyprus, Moldova, Macedonia, Jordan, Iran, Indonesia, Turkey, Tunisia, Chile, Philippines, Morocco, Sth Africa)	22 (Sweden, Czech Rep., France, Norway, USA, Hungary, Iceland, Belgium, Switzerland, Spain, Germany, Poland, Denmark, Italy, Liechtenstein, Greece, Russian Fed., Latvia, Portugal, Luxembourg, Mexico, Brazil)

* Results for countries that participated in PISA in 2000 (ie excluding countries that did the PISA 2000 test in 2002 – known as PISA –Plus)

INTERNATIONAL COMPARISONS OF SCHOOL SCIENCE LEARNING

The origins of international comparisons

Comparisons of school students' learning (including in science) across countries have been undertaken a number of times in the last 50 years. Until recently, these comparisons have been conducted by the International Association for the Evaluation of Educational Achievement (IEA), "an independent, international cooperative of national research institutions and governmental research agencies".[1]

The motivation for creating IEA in 1958 was described by one of the architects of its early work, Neville Postlethwaite, as informing decision makers by providing comparative data about educational performances. "At all levels in an educational system, from the teacher in the classroom through the administrator to the policymaker, decisions have continually to be made most of the time on the basis of very little factual information."[2]

The intent of this digest is to provide information from international comparisons of school science learning for policy makers and other decision makers within education systems.

Recent international studies of science learning

The most recent IEA study of science and mathematics learning, the Third International Mathematics and Science Study or TIMSS, was conducted in 1994-95 and repeated in 1998-99. In Australia data was collected from:

- upper primary and lower secondary in 1994, and
- lower secondary again in 1998.

The inclusion of mathematics and science together in TIMSS enabled some economies in data collection across many countries and opened the possibility, later not pursued, of looking for correlations in student achievements in these subject areas. An unintended consequence of the joint testing was that the science items did not include much quantitative reasoning.

The major focus of the three IEA studies of science, put simply, has been comparison of what students know from the school science curriculum content that is common across the participating countries.

As part of its growing interest in more detailed aspects of the effectiveness of schools, the Organisation for Economic Cooperation and Development (OECD) in the later 1990s launched a different program of international comparisons of learning - the Programme for International Student Assessment (PISA). PISA involves testing in reading, mathematics and science. PISA describes its focus as follows:

"Are students well prepared to meet the challenges of the future? Are they able to analyse, reason and communicate their ideas effectively? Do they have the capacity to continue learning throughout life? These are questions that parents, students, the public and those who run education systems continually ask. PISA, a new three-yearly survey of the knowledge and skills of 15-year-olds in the principal industrialised countries, provides some answers. It assesses how far students near the end of compulsory education have acquired some of the knowledge and skills that are essential for full participation in society …., reveals factors that influence the development of these skills at home and at school, and examines what the implications are for policy development."[3]

In keeping with this emphasis on preparation for life, PISA describes its three areas of interest as 'reading literacy', 'mathematical literacy', and 'scientific literacy.' The preparation-for-life emphasis led PISA to test students around the final year of compulsory schooling, when the vast majority of young people in OECD countries are still at school.

The focus in this digest is on TIMSS and PISA, and some relevant Australian studies that help interpret the TIMSS and PISA findings. There is more space given in the digest to the TIMSS data than the PISA data. Given that TIMSS have, to this point, collected more data and published more analyses of these data, this is not surprising. It is important to remember that the differences in the amount written about TIMSS and PISA below do not reflect anything about the relative merits of the two programs – each is testing different things, as described above.

WHY TIMSS *AND* PISA?

TIMSS and PISA attempted to assess different features of student learning and hence one can reasonably expect different findings from them.

In general terms, TIMSS sought to find *'what students know'* and PISA sought to find *'what students can do with their knowledge'*. These two perspectives are neither 'better' nor 'worse'. Rather they are different, and each has importance as a learning outcome of the study of science at school.

The data gathered in the TIMSS project related to the *intended curriculum* (the curriculum specified by the system or other body), *the implemented curriculum* (the curriculum as taught by teachers, the nature of actual classrooms), and the *attained curriculum* (what students have learned).

The PISA project is not directly focused on any of these aspects of curricula. Rather PISA is concerned with how well 15 year old students can make use of science knowledge acquired from school and from other sources, in situations in everyday life that involve science and technology.

The intentional differences in the projects are evident in how each project structured and developed its tests.

The approach of TIMSS

TIMSS gathered data from samples of the student population at three levels (middle primary, lower secondary, final year secondary), from the teachers of these students, and from their schools and systems.

Development of the tests of student learning outcomes for each student population began with an analysis of science curriculum guides and textbooks from many countries to 'identify priority topics' for the tests. An international panel of science curriculum specialists then produced a framework to guide test development. The framework had:

- a *content* dimension which indicated the proportions of test questions required for each of the areas of science (life science, earth science, physical science, etc.), and
- a *performance expectations* dimension for what was likely to be involved in answering the items (understanding simple information, solving problems, using science processes, etc.).

These two dimensions are described in Table 2.

There was also a *perspectives* dimension that included science-related details about the individual student and his / her classroom and school contexts (attitudes, interests, habits of mind, and so on). Tests for different levels had different proportions of items on the content and performance expectations dimensions, and included multiple choice, short answer and free response items. The free response items usually required students to answer a question and then explain their answer.

On a much smaller scale, some of the students from the first two populations undertook a 'Practical Performance Test' that required them to carry out a range of tasks and experiments involving simple scientific equipment. No comments are made about this performance testing in this digest as no international results have been published.

A School Questionnaire was used to gather information about the intended and implemented curriculum, and a range of school characteristics (location, size, resources, curriculum offerings, etc.). A Teacher Questionnaire asked about qualifications, levels taught, approaches to planning and carrying out teaching, use of textbooks and other resources, views on current curricular issues, etc. A Student Questionnaire sought information about demographic details, how students spent time, attitudes to science, expectations, etc. Finally, data for curriculum analyses were collected at the system-level, together with details about structural aspects that differed across the countries.

Table 2: The structure of TIMSS tests and the distribution of items

	TIMSS Population 1 (1994)	TIMSS Population 2 (1994)	TIMSS-Repeat (1998)
Content dimension:			
Life science	42%	30%	27%
Earth science	18%	16%	15%
Chemistry	} combined as "Physical sci" 31%	14%	14%
Physics		30%	27%
Environmental	9%	} 10%	"& Resource" 9%
Nature of science	–		"& sci inquiry" 8%
Performance expectations dimension:			
Understanding simple information	45%	40%	39%
Understanding complex information	31%	29%	31%
Theorising, analysing, solving problems	15%	21%	19%
Using tools, routine & science processes	6%	6%	7%
Investigating the natural world	3%	4%	4%

The approach of PISA

The first round of PISA data collection was in 2000. Because of the emphasis on scientific literacy and 'preparation for life', the beginning point for the science testing was quite different to TIMSS; it was to determine what aspect of scientific literacy would be the focus for the short testing time available for science in 2000.[(4)] After consideration by the science expert group, the following definition of scientific literacy was adopted by PISA for its testing.

"Scientific literacy is the capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity."[(5)]

This definition comprises three aspects (see Table 3):

- *scientific* processes (recognising questions, identifying evidence, drawing conclusions, etc. - that is knowledge about how scientific investigations are done),
- scientific concepts (with a very strong emphasis on being able to use concepts, as described above), and
- *situations* or *contexts* in the world for which science is a key component.

The next step in developing the test was to determine the range of processes and concepts to be assessed.

The selection of scientific concepts involved four criteria: relevance to everyday situations; 'enduring relevance' to life in the next decade at least; relevance to situations that involve scientific literacy; the concepts should be combined with scientific processes in the assessment. This set of criteria differs markedly from TIMSS, and shows clearly the different approach of PISA. It was possible that a test based on these criteria would relate to only some of the topics that make up the intended curriculum of a given country or system.

A large number of short media reports involving science was collected and sorted against the first three of the criteria. A number of items requiring use of the chosen scientific processes and application of the related science concepts, was developed for each of these media reports, and trialled with students in several countries. These procedures led to the actual test, with items of varying difficulty and a good coverage of the three aspects of scientific literacy. Multiple choice, 'closed' short answer and 'open' free response items were used. Table 4 provides information on where to find examples of test items for both PISA and TIMSS.

Other data collected by PISA involved a School Questionnaire (resources, structure, parental involvement, etc.) and a Student Questionnaire (demographic data, perceptions of school and teaching, expectations, language spoken at home, home resources, and time spent on homework).

Table 3: The structure of the PISA 2000 test and the distribution of items

BROAD FRAMEWORK ASPECT	COMPONENTS OF THE ASPECT	PERCENTAGE OF ITEMS ON THE TEST
Science Major Area Or Theme (Content)	Earth and environment Life and health Technology	37% 37% 26%
Science Area Of Application (Content)	Atmospheric, chemical, physical, geological change Biodiversity and ecosystems Earth and universe Energy transfer Form and function Genetic control Human biology, physiological change Structure of matter - ~ 17%	20% 11% 14% 11% 9% 6% 11% 17%
Science Process	Communicating conclusions, evidence or data to others Drawing or evaluating conclusions Identifying evidence or data Recognising questions Demonstrating conceptual understanding	9% 20% 14% 14% 43%
Situation (Context)	Global Historical Personal Public	46% 11% 23% 20%

Table 4: Where to find examples of items used in tests

TIMSS - population 1 **- population 2** **- population 3**	http://timss.bc.edu/timss1995i/TIMSSPDF/ASitems.pdf http://timss.bc.edu/timss1995i/TIMSSPDF/BSItems.pdf science literacy: http://timss.bc.edu/timss1995i/TIMSSPDF/CitemMSL.pdf physics: http://timss.bc.edu/timss1995i/TIMSSPDF/CitemPhy.pdf
TIMSS - Repeat	http://timss.bc.edu/timss1999i/pdf/t99science_items.pdf
PISA	http://www.pisa.oecd.org/science/struct.htm

RESULTS FROM TIMSS

The three student levels involved in TIMSS were termed Population 1 (middle primary students; tested in 1994), Population 2 (lower secondary students; tested in 1994, then repeated in 1998), and Population 3 (final year secondary students). Because the Population 3 findings are much harder to consider in terms of comparisons (as explained below), we concentrate here on Populations 1 and 2. We first give separate brief descriptions of the overall findings for each Population, and then look at broad trends across the two levels.

Population 1 (Middle Primary School Students)'[6]

The Population 1 sample was drawn from "the two adjacent grade levels containing the largest proportion of 9 year old students at the time of testing."[7] Australia was one of 26 countries who tested Population 1 students.

Overall results

Australian students did very well. Only two countries, Republic of Korea and Japan, significantly outperformed Australia, as shown in the table. (Note that 'significantly' is used in its statistical sense.)[8]

Australia also differed from Korea and Japan in its spread of student achievement – the range of performance from bottom 5% to top 5% in Australia was about 25% greater than for Korea and Japan. USA, England and Singapore, countries with student performances no different from Australia, had somewhat wider spreads than Australia.

Comparing states

TIMSS in Australia increased the sample size for the smaller states and territories to obtain state comparisons, but it should be remembered the states have some important structural differences such as the total time students had actually spent in school by the time of testing.

All states performed above the international average, but there was substantial variation between states. WA in fact achieved at a marginally higher level than Korea (the top ranked country), and SA, Queensland and ACT all performed above or as well as Japan (the second ranked country). ACT, Tasmania and Victoria had greater spreads of scores, while WA had the least.

The remarkably high Population 1 achievement in a number of states was surprising to some. State-wide science testing in primary schools around the same time as TIMSS (1994) had been interpreted as showing problems with the levels of student science achievement. For example, in 1993, WA students at Years 3, 7, and 10 were tested in science as part of a standards monitoring program, with results that were seen as very disappointing in terms of the students' levels of understanding of concepts. An analysis of the question forms and the learning emphases in these various tests gives one plausible explanation for this great difference between TIMSS performance and performance on the state-developed tests[9] - the emphasis in the WA testing was on understanding of concepts, not a strong aspect in TIMSS. However the plausibility of this explanation is diminished by the data from PISA. In PISA, the test focus was on ability to use knowledge and thus there was a stronger focus on understanding of concepts than in the TIMSS tests. State comparisons of PISA performance show WA performance was significantly higher than four other states.

Population 2 (Lower Secondary School Students)'[10]

The Population 2 sample for 1994 testing was drawn from "the two adjacent grade levels containing the largest proportion of 13 year old students at the time of testing."[11] Australia was one of 40 countries who tested Population 2 students at this time. A second testing of this age group took place in 1998. This testing was called TIMSS-Repeat (or TIMSS-R).[12] The results from the two testings were very similar, and so these are discussed together below. There is however one difference in the reporting. In 1994 the performances of students in each individual state and territory were reported, while in 1998 reporting was confined to the national level.

Overall results

Australian students did very well. However the spread of results (the range from the bottom 5% of students to the top 5% of students) was even greater than for Population 1. It appears that the longer total time the Population 2 students had spent in schooling was accentuating differences between high achieving and low achieving students, rather than alleviating the educational 'differences' students have when they enter school.

Comparing states

All states performed above the international average, but their variation was much wider than for Population 1. WA was again the top ranked state, and ranked higher than all countries except Singapore. Victoria was just above the international average, and the other states were between WA and Victoria. The range of scores in Tasmania, NT and ACT was wide, while Victoria and SA had smaller ranges.

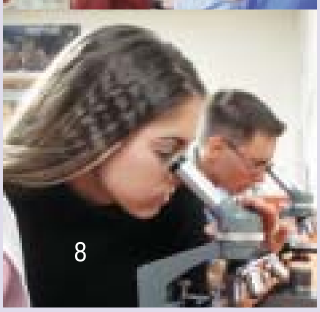
Looking across both Population 1 and Population 2

Performance on the categories within the tests

At both levels, Australian students performed better than the international average on all the content categories, and all the performance expectations. The differences at both levels for the content category 'Environmental issues' were substantial. This high achievement is consistent with data about the intended curriculum in Australia where an emphasis on environmental matters has been a feature for the last 15 years. For Population 2, Australian performance on the 'Earth Science' category was relatively poorer than for the other content categories (Australian students' 'Earth Science' performance was only just above the international average) – again this is consistent with the intended curriculum across the states.

Considering other variables

Gender: In the two previous international comparisons of science achievement (early 1970s and mid 1980s) Australian boys had significantly outperformed girls. No such gender differences existed in the overall TIMSS Population 1 and 2 science findings, a feature of performance found in only a handful of countries. Furthermore, there were no significant gender differences in any category of item at either level, also a rare finding internationally.



Student perceptions: There was a significant positive correlation between achievement and enjoyment of science, but no significant relations between achievement and students' attribution of success to effort or ability, or achievement and belief in the importance of science. There was a significant negative correlation between achievement and the attribution of success to luck.

Family/home variables: The findings for both Populations were remarkably similar. Science achievement correlated strongly with parent occupation (categorised into 6 groups, from unemployed/low-skilled to professional/semi-professional). Students reporting English as their home language achieved higher than students whose home language was not English. Of students whose home language was English, those born in a non-English speaking country outperformed those born in an English speaking country.

Indigenous students: Indigenous students in both TIMSS Populations had significantly lower average scores than non-Indigenous students. In Population 1, Indigenous students omitted 25% of the open-ended types of items, compared with 14% for other Australian students. Where these questions were answered, the former had only 37% correct, compared with 56% for the latter. In Population 2 this trend was even more extreme (55% omitted extended response items compared with 22% overall, 8% correct compared with 31% overall). This pattern of response was not found for the multiple-choice questions that made up a majority of the test. This difference in forms of performance may possibly relate to lesser proficiency with the written language demands of the extended response format, or a relatively lesser familiarity with this format of item.

Curriculum and contextual data

TIMSS has reported findings about the nature of the intended curriculum, characteristics of the schools, teachers and classes sampled, structure of science classes, teachers' perceptions of problems, and students' perceptions of the nature of the tasks they undertake in science. More recent findings about Australian teacher and student perceptions of primary and secondary science classrooms can be found in 'The Status and Quality of Teaching and Learning of Science in Australian Schools'.[13]

We note here some particularly interesting international comparisons.

- There is a tendency for countries doing well on TIMSS to have curricula that are more 'focused' (e.g. Japan), that is, have fewer topics than some other countries, and for countries whose science curricula are more 'diverse' (i.e. cover a wider range of topics) to do less well. (Australia is somewhat in the middle of this dimension of the intended curriculum.)
- Textbooks play a stronger role in the science classrooms of many TIMSS countries than they do in Australia. In many countries students are expected to use their textbooks for independent learning, something rare in Australia.

- Science textbooks in Australia have changed considerably over the last 10 to 15 years, from factually based approaches towards a much stronger activity based focus and to topics related to issues and social contexts. This general trend was not found in many countries.
- Assessment of science learning in Australia has been more concerned with formative assessment than in many other countries.
- Some countries reported data from teachers of the sampled classes about career preferences, and beliefs concerning appreciation of their work. In both Populations, Australian teachers were around average in teaching being a first choice career, but they reported considering a career change more commonly than in most countries. Australian teachers were around or just below average in terms of belief that their students appreciated their work, but were much more likely than most countries to believe that society did not appreciate their work. As the data for these findings were collected at the same time as severe budget cuts were being implemented in some states, it may be that these rather disturbing findings reflect a view of that change in conditions.

Population 3 (Final year of secondary schooling)

TIMSS chose 'final year of secondary schooling' as Population 3 to enable some indication of what students know about science as they complete schooling. However the findings for Population 3 are harder to consider in terms of international comparisons. Only 22 countries participated, and both the age of students in 'final year' and numbers still at school varied considerably. In Australia there is evidence that some of the Year 12 sample did not seriously engage with the test[14], presumably because of their views of the personal importance of the TIMSS test in relation to their Year 12 studies and assessments.

Two forms of test were used. One involved final year students, regardless of whether or not they were currently studying some science. Not surprisingly this was termed the 'scientific literacy' test. The second test assessed what those still studying physics in the final year knew.

The performance of Australian students on both these tests was not significantly different from the international average.

RESULTS FROM PISA

In 2000 PISA tested 15 year olds in 28 OECD and 4 other countries. The test was structured around a framework including aspects of content, process, and context, as shown in the table.

PISA attracted considerable interest around the world as it developed its approaches. This interest led to a repeating in 2002 of the testing used in 2000 to allow a further 11 non-OECD countries to participate. This 2002 testing is referred to as PISA+.

Overall results

Once again, Australian students did very well, with only Korea and Japan performing significantly better. Australian students did as well as or better than the OECD average on all but one of the items on the test. The top 5% of Australian students did as well as in any country, and the range of scores for the top 75% of Australians students was less than the OECD average range for the top 75%. This is a very positive result – the very best of the Australian students did as well as the very best from other countries and the spread of scores for the top 75% of Australian students was less than the OECD average. However, for the lowest achieving 25% of students the range of scores was similar to the range across the OECD for these lower achieving students.

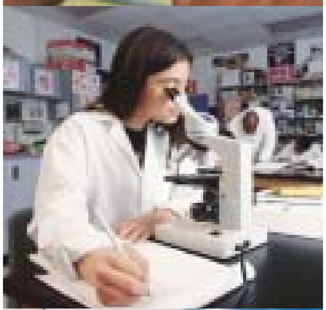
Comparing states

State average performance was above the OECD average for all states except NT, who were at the OECD average. NT also had a larger range of scores than the other states, and NSW the smallest range. ACT had the highest average performance, with their performance differing only from Korea in terms of having a larger range in their scores.

Considering other variables

Gender: There were no gender differences in the science performance of Australian students. However, this brief statement is only part of a very interesting and significant set of gender-related PISA findings. The PISA science test required considerably more reading than is common in school science testing. The PISA reading literacy results showed overwhelming gender differences (in favour of girls) on all three of the sub-scales involved in the reading test. Among the 32 countries in the original round of PISA 2000 there were only 5 cases of no gender difference on a sub-scale. Australian gender differences in reading were above the OECD average on 2 of the 3 sub-scales.

It is therefore quite remarkable that among the great majority of the participating countries (including Australia) there was no significant gender difference in science. Only



7 countries of the original 32 had gender difference in science performance, with 3 of these differences favouring girls.

The only explanation we can see in the case of Australia, where gender differences in science achievement have not shown up in other testings such as TIMSS, is that the PISA tasks have presented such intrinsically interesting reading that the usual reading bias has not applied. The PISA tasks begin with a journalistic short 'story' involving science. This PISA finding is a strong affirmation of the suggestions by the Curriculum Corporation in the mid-1990s that 'curriculum as story' should be a central approach for presenting science to students at all levels of schooling.[15]

Family/home variables: In Australia students whose main language at home was not English did significantly worse in science than those who spoke English at home, and the difference between these two groups was, surprisingly, greater in science than in reading. This fact is consistent with the comments above about the impact of intrinsic interest on reading. Students who lived in cities (including provincial cities) did significantly better than students from remote areas.

Indigenous students: These results were very similar to those from TIMSS. The average performance of students identifying themselves as Indigenous was clearly significantly poorer than the average for all other Australian students.

OTHER FINDINGS SUGGESTED BY THE INTERNATIONAL COMPARISONS FROM PISA

In an analysis of the full range of PISA data (student performance and other data), McGaw[16] (Director, Directorate for Education, OECD) has advanced some conclusions of broad and great educational significance. These conclusions are noted here because they speak powerfully to all school education, science and other.

In summary,[17] McGaw notes that

- not only do countries differ in both average and range of performance, some countries have both a high average performance and a low range of scores (e.g. Finland, Korea, Canada),
- the correlation between student social background and student performance varies considerably between countries – in some countries (e.g. Korea, Finland) this correlation is much lower than in others.

From these observations and other data McGaw concludes that

- quality and equity can be achieved together – 'levelling up' is achievable
- some countries ameliorate the effects of home background better than others – some high performing countries achieve quality and equity together
- organising students into hierarchical streams, by school type, exacerbates the differences among students and produces lower average performances.

SOME IMPORTANT ASPECTS OF SCIENCE LEARNING STILL TO BE COMPREHENSIVELY CONSIDERED

Some obvious aspects of science learning yet to be comprehensively explored in these international studies are the range of things involved in practical activity — skills in using equipment for measuring, designing investigations, collecting and analysing data, etc. As these things are such common goals of school science education it might seem strange that comprehensive testing has not already occurred. However, such testing is particularly complex. Not only is it very much more expensive to assess students actually doing investigations, it is also very much harder to ensure that the testing is done in the same way, and under the same conditions, across a range of contexts and countries.

The high levels of performance of Australian students on the 'Environmental issues' content category in TIMSS, and the surprising lack of any gender differences on the high-reading-demand PISA test of scientific literacy all imply relationships between curriculum content and structure and student engagement and learning. Student perceived relevance seems to be significant in engagement and learning. These are potentially very important relationships, and need further investigation.

Another issue of importance that relates to learning outcomes is suggested by the finding of above average Australian student performance on all the TIMSS performance expectation dimensions (see Table 2). What are the relations between teaching approaches and the forms of learning outcomes described in TIMSS as 'performance expectations'?

Finally, an aspect of science learning warranting further investigation comes from more detailed considerations of students' responses to some items in these tests. An analysis of some Australian students' responses to four free response items from TIMSS[18] revealed new things about student learning and their problems with concepts. The common approach in diagnostic tests of student learning is to use formats such as multiple-choice, so that common misconceptions can be presented in the alternatives for student choice. What the analysis of these responses to free response items shows is that there are clearly important ways in which open-ended questions and the other more complex tasks of PISA can also be used as diagnostic tools for learning.

THE STORY IS NOT YET ENDED – THE FUTURE FOR TIMSS AND PISA

There are more results to come from both TIMSS and PISA. The former has now become Trends in International Mathematics and Science Study (thus still TIMSS) and will involve junior secondary students being regularly tested against the criteria of learning science defined in the original development of TIMSS. These comparisons across time have their use, particularly when major interventions occur such as concerted professional development to improve teachers' pedagogical practice. They become less relevant when new emphases for learning science are introduced in the intended curriculum. The latest round of TIMSS testing was in 2002-03, with Australian students from years 4 and 8 being tested in 2002.

PISA collected data again in 2003 which, for science, is essentially a repeat of the 2000 assessment. In 2006, however, science has the major role in PISA and plans are being developed to include the testing of a wider set of science competencies within the project's meaning of scientific literacy. There will still be a strong emphasis on authentic science situations in the life of modern society. In addition to the measuring of these cognitive science competencies, there is also likely to be an exploration of students' attitudes to the science involved in these situations.

FOOTNOTES

(1) IEA website: http://www.iea.nl/

(2) The first major study conducted by IEA was in Mathematics, in 1964. Science was part of its second, six-subject, study in the early 1970s. This study investigated the learning of science by students at 10 years, 14 years, final year of schooling. This structure of 3 different ages has remained a feature of IEA science studies. The second IEA science study took place in the early-mid 1980s.

(3) PISA website: http://www.pisa.oecd.org/

(4) PISA's approach to testing the three areas of reading, mathematics and science is to focus primarily on one (the "major" domain) and less on the other two ("minor" domains). In 2000 the major domain was reading literacy; mathematics will be the major domain in 2003 and science in 2006.

(5) OECD (1999). *Measuring student knowledge and skills: A new framework for assessment*. Paris: OECD Publications Service.

(6) The full Australian report of Population 1 design, data and discussion, including international comparisons, is J. Lokan, P. Ford & L. Greenwood (1997). *Maths and science on the line: Australian middle primary students' performance in the Third International Mathematics and Science Study.* Melbourne: Australian Council for Educational Research.

(7) This meant that different grade levels were sampled in different states. In the TIMSS reports the achievement in the two grades are reported separately as "upper grade" and "lower grade", but we do not give this level of detail.

(8) Because TIMSS collected data from samples, and not, for obvious reasons, from every student who fitted the criterion for this population, the data obtained can only be estimates of what would have been found if all the population had been involved. "Significant" here means that there is a 95% chance that the difference between Australia and the countries doing significantly better is a real difference, and only a 5% chance that the difference is the result of the particular samples involved in each country.

(9) A comprehensive analysis of this issue is given by P. Fensham (1999). 'International success, but is it science? Identifying strengths and weaknesses in Australian primary school science from TIMSS and other data. '*Australian Science Teachers'* Journal 45(2): 39-44.

(10) The full Australian report of Population 2 design, data and discussion, including international comparisons, is J. Lokan, P. Ford & L. Greenwood (1996). *Maths and science on the line: Australian junior secondary students' performance in the Third International Mathematics and Science Study.* Melbourne: Australian Council for Educational Research.

(11) Therefore, as for Population 1, different grade levels were sampled in each state.

(12) In TIMSS-R the students sampled had been in Population 1 in 1994. Only one grade level was used (the upper year level of the two used before in Population 2 – Year 8 in ACT, NSW, Tas, Vic; Year 9 in NT, Qld, WA, SA).

(13) D. Goodrum, M. Hackling & L. Rennie (2001). *The status and quality of teaching and learning of science in Australian schools.* Canberra: Commonwealth of Australia (DETYA)

(14) P. Fensham (1998). 'Student response to the TIMSS test.' *Research in Science Education* 28: 481-489.

(15) See for example Malcolm, C. (ed.) (1996). Could we? Should we? Year 10 science. Melbourne: Curriculum Corporation; Egan, K. (1988). Teaching as story telling: An alternative approach to teaching and curriculum. Chicago: Univ Chicago Press; Malcolm, C. (2002). Science for all: Learner-centred science. In R, Cross (ed.) A vision for science education: Responding to the work of Peter Fensham. London: RoutledgeFalmer.

(16) McGaw, B. (2002, October). Raising the bar and reducing failures: A possible dream. Invited paper given at the ACER conference "Providing world-class education: What can Australia learn from international achievement studies?", Sydney.

(17) The data from which McGaw draws these observations and conclusions, and elaborations of the conclusions, are in OECD (2001). Knowledge and skills for life: First results from PISA 2000. Paris: OECD Publications. (Also available electronically at http://www.pisa.oecd.org/knowledge/download.htm); see in particular Chapter 8 and Appendix B1

(18) Fensham, P. & Haslam, F. (1998). Issues of assessment in TIMSS: The scoring of short answer and free response items. Unpublished research paper, Monash University.

FURTHER INFORMATION ON TIMSS AND PISA

TIMSS website: http://timss.bc.edu/

PISA website: http://www.pisa.oecd.org/

Full reports of the various testings:

1. TIMSS (international reports)

- [TIMSS-R] Michael O. Martin, Ina V.S. Mullis, Eugenio J. Gonzalez, Kelvin D. Gregory, Teresa A. Smith, Steven J. Chrostowski, Robert A. Garden & Kathleen M. O'Connor. (2000). TIMSS 1999 International Science Report Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade. Boston: International Study Center, Boston College (also available electronically at http://timss.bc.edu/timss1999i/science_achievement_report.html)
- [TIMSS Population 3] Mullis, I., Martin, M., Beaton, A., Gonzalez, E., Kelly, D. & Smith, T. (1998). Mathematics and science achievement in the final year of secondary school. Boston: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College (also available electronically at http://timss.bc.edu/timss1995i/MathScienceC.html)
- [TIMSS Population 2] Beaton, A., Martin, M. Mullis, I., Gonzalez, E., Smith, T. & Kelly, D. (1996). Science achievement in the middle school years: IEA's third international mathematics and science study (TIMSS). Boston: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College (also available electronically at http://timss.bc.edu/timss1995i/SciencB.html)
- [TIMSS Population 1] Martin, M. Mullis, I., Beaton, A., Gonzalez, E., Smith, T. & Kelly, D. (1997). Science achievement in the primary school years: IEA's third international mathematics and science study (TIMSS). Boston: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College (also available electronically at http://timss.bc.edu/timss1995i/TIMSSPDF/astimss.pdf)

2. TIMSS (Australian reports)

- [TIMSS-R] Zammit, S., Routitsky, A. & Greenwood, L. (2002). Mathematics and science achievement of junior secondary school students in Australia. Melbourne: Australian Council for Educational Research.
- Lokan, J., Ford, P. & Greenwood, L. (1996). Maths and Science on the Line - Australian Junior Secondary Students' Performance in the Third International Mathematics and Science Study, TIMSS Monograph No. 1. Melbourne: Australian Council for Educational Research.
- Lokan, J., Ford, P. & Greenwood, L. (1997). Maths and Science on the Line - Australian Middle Primary Students' Performance in the Third International Mathematics and Science Study, TIMSS Australia Monograph No. 2. Melbourne: Australian Council for Educational Research.
- Lokan, J. & Greenwood, L. (2001). Maths and Science on the Line - Australian Year 12 Students' Performance in the Third International Mathematics and Science Study, TIMSS Australia Monograph No. 3. Melbourne: Australian Council for Educational Research.

3. PISA (International report)

- OECD (2001). Knowledge and skills for life: First results form PISA 2000. Paris: OECD Publications. (Also available electronically at http://www.pisa.oecd.org/knowledge/download.htm)

4. PISA (Australian report)

- Lokan, J., Greenwood, L. & Cresswell, J. (2001). 15-up and counting, reading, writing, reasoning: How literate are Australian students?: The PISA 2000 survey of students' reading, mathematical and scientific literacy skills. Melbourne: ACER (also available electronically at http://www.pisa.oecd.org/NatReports/PISA2000/Australianatrep.pdf)

5. PISA+ (International report)

- Literacy Skills for the World of Tomorrow - Further results from PISA 2000 A copy of the international report on the PISA+ data, released in mid 2003, is available at http://www.pisa.oecd.org/literacy/download.htm

ISSN 1445-1603