Teaching Mathematics in Victoria: What does TIMSS show?

Kaye Stacey
University of Melbourne

What is TIMSS?

TIMSS, the Third International Mathematics and Science Study, is the largest study of educational achievement ever undertaken. It has been conducted by the International Association for the Evaluation of Educational Achievement (IEA), with the co-operation of governments of 48 countries throughout the world. The first main results of TIMSS were released in November, 1996. Many more results and publications of the TIMSS data will appear in the next few years. In Australia, the TIMSS study was conducted by the Australian Council for Educational Research (ACER) and was overseen by an advisory committee consisting principally of representatives of the state and Catholic education systems, the teacher unions, business and academics. The present author was a member of this group. The purpose of this article is to briefly describe some of the first international results of TIMSS and to discuss how they can be interpreted. The position of Victoria will be highlighted. For people in Victoria, information about TIMSS can most readily be obtained from the Australian reports (Lokan, Ford and Greenwood, 1996 and 1997), which also list other TIMSS publications and from the TIMSS website http://wwwcsteep.bc.edu/TIMSS1/TIMSShome.html. Most of the information in this article is from the ACER reports.

History and description of the TIMSS study

TIMSS is the third study in a series of international comparisons. The first mathematics study (FIMS) was conducted in 1964, with junior and upper secondary students. Five states of Australia took part and the results are reported in Keeves (1966). Australia did not take part in the second IEA study (SIMS) for mathematics, although in 1978 ACER re-administered the 1964 test. Comparison of the 1964 and 1978 results showed that there had been a slight decline in performance, especially in geometry (Rosier, 1980). TIMSS testing began late in 1994 in Australia. Some of the test items are "link items", used in all three studies. These items should eventually enable achievement from 1964 to 1994 to be monitored.

TIMSS has a number of separate components. Three main populations of children were tested: nine year olds, thirteen year olds and students in the final year of school, the latter category being broken up into students specialising in mathematics and the general student population. A substantial random sample of children in each of these populations answered written test items. A smaller sample was also tested on performance items, where children had to make, measure or do something, in the presence of an examiner. Children and their teachers both completed a substantial questionnaire about work practices, home background and interests, etc. Before the tests were constructed a survey of curriculum across the world was undertaken.

Strengths of TIMSS

TIMSS derives its strengths from its scale, the level of participation and from the possibilities it offers for comparisons within and between countries and over time. Large numbers of students and teachers have been involved. For example, across Australia 11248 students and their teachers were involved in the testing of 9 year olds. These students were selected through a carefully designed random process, which makes generalisation of results possible. Bias in selection may, of course, be introduced because in Australia, schools and parents can choose whether they wish to be involved. Different types of schools may differentially choose to participate in the study. However, TIMSS has used other information where available (such as the results of statewide achievement monitoring) to decide if any bias has resulted and its likely direction.

A familiar idea to all Victorian teachers is that a sophisticated assessment of the mathematical achievement of a child cannot be undertaken with a short pencil and paper test. TIMSS has addressed this in several ways. Firstly, because it is a measure of achievement of the school system that is to be undertaken, not that of any one child, different tests have been given to different children (with some calibrating items) so that the curriculum coverage of the tests can be substantially increased. Secondly, a range of item response formats have been used, such as multiple choice, short response and longer response items. Using the perfomance items described above supplemented the written items, enabling a wider range of curriculum goals to be included. Any measure of achievement used so widely has to be fairly crude and so it is important to recognise the limitations of the measures.

Item development was preceded by a substantial analysis of the curriculum in many countries around the world. The results of this part of TIMSS, documenting curriculum and curriculum changes, are of substantial interest in that they can offer ideas about what might be done in a subject or in a school system beyond what is traditionally done in a particular part of the world. Why, for example, has proof nearly disappeared from our schools when it features strongly in the French curriculum? The results of this analysis will also be useful in interpreting the findings of the achievement surveys. Does the teaching of a wide curriculum (e.g. including chance and data in primary schools, for example) detract from the teaching of basic number skills?

In addition to the achievement data, TIMSS has collected a large amount of data about the conditions under which schooling takes place. The ability to cross-link these items to achievement data may help point to the factors that make a schooling system successful. TIMSS items are also cross-linked to previous international studies, which enables a limited monitoring over time. As time goes on, international studies are becoming more rigorous in their design and standardisation of procedures and wider in scope. The results provide a valuable stimulus for thinking about curriculum, teaching and learning in Australian schools.

Difficulties of international comparisons

Because of the many ways in which schools and school systems are organised across the world and the many different approaches to curriculum and assessment, the world is a natural laboratory for studying the effects of a vast range of factors on educational achievement. However, it is not a controlled laboratory so the study cannot provide simple answers to straightforward educational questions. There are many areas of difficulty, all of which impact on the way in which results can be interpreted. Almost every area is contentious.

Which children, for example, should be compared? Should they be children of the same age or students who have been at school for the same number of years? Given variations in promotion procedures around the world, should sampling be from all children of a given age (as was done for FIMS) or from children in given grades? Even among the states of Australia, 9 (13) year old children have been at school for different lengths of time. The TIMSS study chose to test students in the two adjacent grades with the greatest number of 9 (13) year old children. In Victoria, the testing was done on Years 7 and 8 and Grades 3 and 4 but in other states the grade names are different.

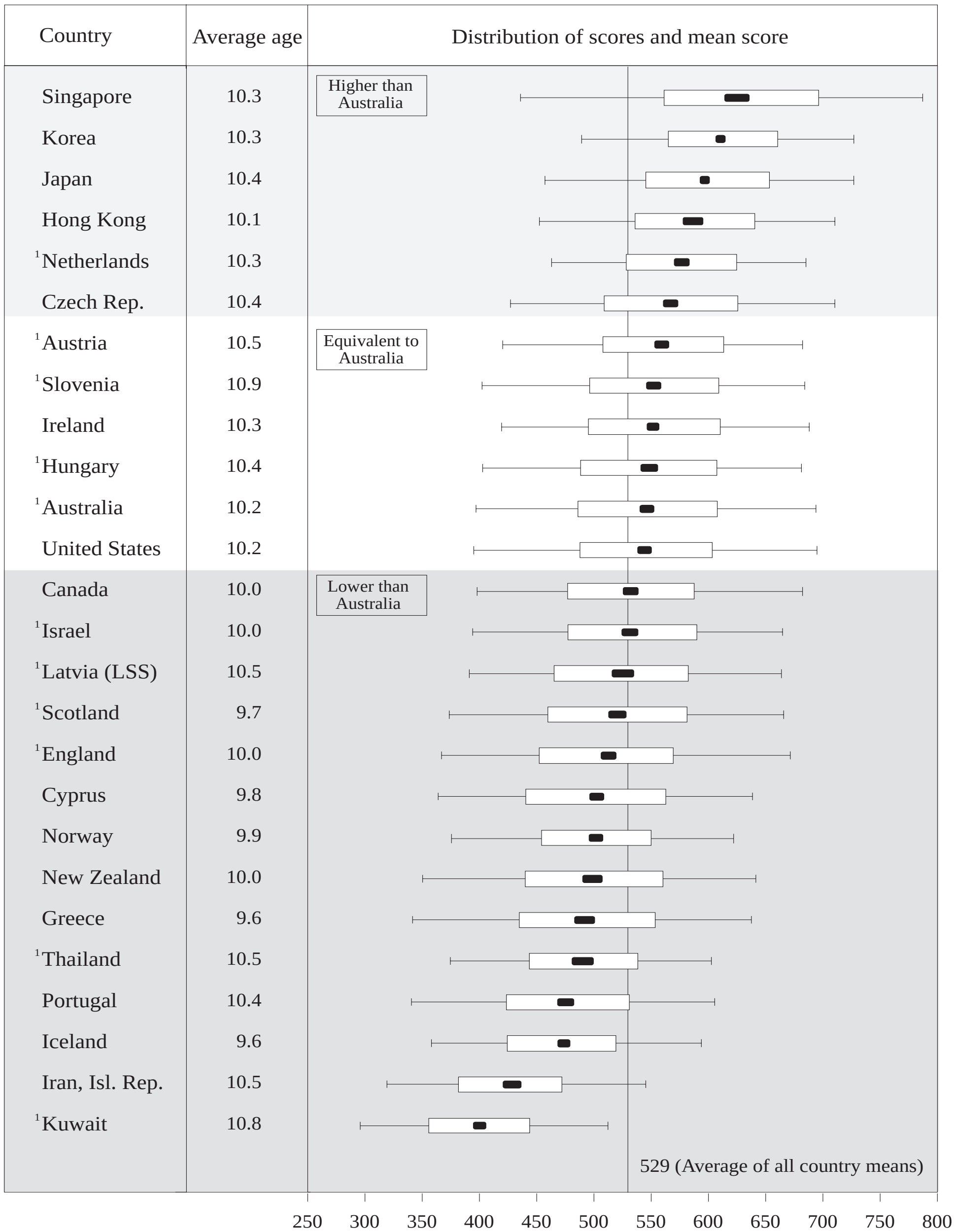
A second area making comparison difficult is the definition of achievement, which TIMSS sets out to measure. Around the world, mathematics curricula have much in common, but they are certainly not the same. Topics are introduced at different ages, some topics are treated in some countries but not in others and different topics receive different emphasis in different countries. Of necessity therefore, the topics to be tested in the TIMSS items are those that are taught in almost all countries. The teams conducting TIMSS in each country were able to comment on the suitability of all items for their students before the final item selection. Each team was also able to nominate a subset of the test items felt to be well represented in the curriculum of their country. Alternative sets of scores were calculated on these item subsets. In fact the use of the alternative scores seems to make little difference to overall country ranking (Lokan *et al*, 1996). There are interesting examples of items which have favoured and not favoured Australian students. For example, Australian students did very well on a science item about which radiation causes sunburn – almost certainly reflecting a high level of public awareness of skin cancer, its causes and prevention. On the other hand, Australian students did very poorly on an item involving multiplication of fractions, presumably because of the very low emphasis that this topic is receiving in current Australian curriculum advice. Whether the tests as a whole have been "fair" to Australia (or to any other country) is a very complex question.

A third area making the conduct of an international study difficult is the fact that items have to be presented in many languages. When an item is translated, its difficulty may be changed in unpredicted ways. For example, if a language has no special word for "herbivore", but it is translated literally as "grass-eater", students would be expected to find easy an item asking what such animals eat. Cultural differences and familiarity with problem contexts provide other sources of incomparability. As far as possible, these sources of advantage and disadvantage were eliminated by international consultation. There are many other ways in which differences can affect national results. Children and teachers in some countries may try harder on a test for outsiders than those in other countries, perceiving participation as an honour rather than a chore. A different group of special education students may be outside the normal education system; students may specialise in maths and science more in some countries than in others.

The factors above (and many others) mean that the results of a study like TIMSS cannot be simply interpreted as "the truth" about school achievement. However, carefully interpreted, the results of the TIMSS study are a valuable source of information to help us think about our curriculum and our school system, about what we want to achieve and about what we value and about what it is possible to achieve.

Summary of international results on achievement

The most publicised results of the TIMSS study are the international rankings of achievement. At the time of writing, the first international results for the 13 year olds and the 9 year olds have been released. Figures 1 and 2 show the results of the written tests of the upper grades tested. These figures give a box and whiskers diagram for the scores on the written tests for each country. Ninety percent of students scored between the whiskers and the middle 50% scored within the box. The mean score is shaded dark within a confidence interval.



[1] These countries did not meet all the sampling criteria.

Figure 1. Upper Grade Achievement in Mathematics for 9 year old students.

[1] These countries did not meet all the sampling criteria.

Figure 2. Upper Grade Achievement in Mathematics for 13 year old students.

Complicated statistical questions arise immediately in the interpretation of the data. Because the data is obtained from samples of students, it is a statistical question whether or not the differences in the average scores of countries are likely to be real or caused by the sampling. The shading in the figures is given from the Australian perspective – countries are grouped according to their statistical difference from Australia. For 9 year olds, Singapore achieved the highest result, statistically higher than every other country except Korea. Japan and Hong Kong performed similarly to each other and statistically better than all other countries below them (except for the Netherlands). Netherlands and Czech Republic formed the next group. The next group, which includes Australia, consists of six countries, a mixture of European and some other English speaking countries. In a sense then, Australia tied in seventh place with five other countries. The results for the lower grade are broadly similar: Australia tied for fifth place along with eight other countries. (Lokan et al, 1997, p13,14).

The results for 13 year olds again place Australia well amongst the western countries, but with Singapore, Eastern and European countries achieving the higher scores (see Figure 2). In a sense, Australia tied for ninth place with thirteen other countries (upper grade).

The full reports of the Australian results for 13 year olds and 9 year olds (Lokan et al, 1996, 1997) contain many fascinating results. Here there is space to list just a few:

- The top 75% of Singapore children perform approximately as well as the top half of Australian children on the written tests. Singapore students performed very highly on all areas on the curriculum, including data representation.
- Children seem to make more progress in mathematics from the lower to the upper primary grade (Grade 3 to 4) than from the lower to the upper secondary grade (Year 7 to 8).
- Australian results in each content area are in line with its overall position, except for Geometry which is lower (13 year olds).
- For Australia, no item showed a statistically significant difference between boys' and girls' performance. Around the world, differences favouring boys are common.
- Australian students did relatively well on meaurement and data representation, and on conceptualisation of fractions, but not fraction operations. (13 year olds).
- Better mathematics achievement is associated with occasional work in small groups, occasional projects, moderate or low television watching, frequent calculator use, frequent homework and orderly classes. (These results need more careful analysis).

Summary of comparisons within Australia

Many of the most interesting results of TIMSS from an Australian perspective will come from the comparison of students' performance in different parts of Australia. Comparison within Australia reduces the effects of the most gross causes of difference that are discussed above. Despite the fact that Australia is diverse and multi-cultural, it is nevertheless a reasonably homogeneous society, across its geographically dispersed parts and the differences between areas are reasonably well understood by us. Schooling takes place within broadly similar social parameters, teacher qualifications are comparable, school resources are broadly comparable across systems, English is almost everywhere the language of instruction and the national statements on curriculum are the basis of curriculum in most states, indicating a similarity of goals. The sample of schools across

Australia was designed to be large enough so that comparisons between the answers on the written tests and questionnaires of students in the states and territories could be made. Unfortunately, collection of data was too expensive to allow an adequate sample size for interstate comparisons on performance items.

Inter-state comparison of mathematics for the 9 year olds

Table 1 presents the comparison of performance on the written test of students in the Australian states and territories for 9 year olds (upper and lower grade levels combined). The international average score was 500. The table shows that Victoria had the second lowest mean score. This score was statistically below the scores of Queensland, Western Australia, South Australia, ACT and the Northern Territory. It was statistically not different to either NSW or Tasmania. In international terms, Western Australia had a mean score similar to that of Hong Kong in Figure 1, whilst Victoria's score is near Canada's and very close to the international average. Previous international and national surveys of student achievement have found differences in mathematics achievement between the states and territories of Australia. Victoria has generally had low achievement in Mathematics (Keeves, 1966; Rosier, 1980).

The pattern of differences for 13 year olds is similar although the Northern Territory is in the lower group (see Lokan, 1996, p 33) and the state results are spread even more widely amongst other countries. Western Australia's mean score is approximately that of Flemish Belgium (see Figure 2), whilst Victoria's is between Germany and the USA, significantly below the international average.

Insert TABLE 1 about here

Table 1. Mathematics Achievement score for 9 year old sample for states and territories.

	QLD	WA	SA	ACT	NT	NSW	VIC	TAS
Score	546	545	540	530	528	498	498	492

An immediate explanation for Victoria's relatively low performance presents itself: although they have been to school for a similar length of time, children in the lower performing states are about six months younger than those elsewhere (except for ACT – see below). A selection of this data is presented in Table 2 (derived from Table 3.5 of Lokan et al (1997, p 43)) which shows that children in WA and QLD with the same number of years of schooling do better (approximately 45 points) than children in NSW and Victoria. From the scores of NSW and Victorian children, an estimate of a score of 499 for children aged 9.5 can be obtained, which is again below the WA/QLD scores of 513 and 514. It therefore seems that there is a difference between achievement in these pairs of states that is not accounted for by age or length of schooling. Whether children's age is in fact an adequate explanation is one of the many fascinating questions that await further analysis of the TIMSS data. Internationally there is no correlation between average age and country perfomance for 13 year olds, although there is a small correlation (0.25) for 9 year olds. Age seems to matter more in science than in mathematics, probably reflecting the greater number of science issues that could be learned outside of school.

TIMSS data includes many measures to be used in later analyses to search for explanations for differences in achievement. For example, from the questionnaire an index of students' home background has been computed, which puts ACT markedly at the top (in accordance with 1991 census data), followed by WA, Victoria and NSW and then Queensland. For example, 77% of ACT students said they had a computer at home, 65% of WA students, 61% of NSW and Victorian students and 55% of QLD students. Queensland therefore scores lower on the socio-economic indicators than Victoria, but higher on the written tests. The questionnaires to the teachers, about themselves and their teaching practices may also assist in finding explanations.

Interestingly in Table 2, the amount of growth between one year and the next (crudely measured by the difference in scores of the two grades) is highest (71) in Victoria and lowest in NSW (57). If we use an average of these scores as a measure of one year's growth in achievement, it is interesting to note from Figure 1 that Singapore students are over a year ahead of Australian children in maths.

INSERT TABLE 2 ABOUT HERE

Table 2. Comparison of four states on achievement, age, years of schooling and index of home background (HB) (9 yr old sample)

State	Maths scores			Mean age			Mean years of schooling			HB index
	Yr 3	Yr 4	Yr 5	Yr 3	Yr 4	Yr 5	Yr 3	Yr 4	Yr 5	
NSW	469	526		9.0	10.0		3.75	4.75		9.1
Vic	465	536		9.1	10.1		3.75	4.75		9.2
QLD		513	574		9.5	10.5		3.75	4.75	8.9
WA		514	577		9.5	10.5		3.75	4.75	9.3

There are other potential explanations for Victoria's relatively weak performance:

(i) The test may not suit Victoria. For example, Victoria's curriculum emphasis on problem solving and real world mathematics could mean that students are not be prepared for written items, which may test an out-of-date skills-based curriculum. On this hypothesis, Victoria would be expected to have done better on the performance items, but the first analysis (Lokan *et al*, 1996, p 114) indicates that this is not so.

(ii) The curriculum sequence and timing may differ. For example, the Victorian curriculum may not cover topics as early as in other states. Possibly the nomenclature of grade levels indicates this: children in WA and QLD begin in "Grade 1" whereas in Victoria children of the same age begin in "prep".

(iii) Differences may be due to characteristics of teachers or aspects of teaching methods or assessment or goals.

(iv) There may be critical differences of school organisation such as less time spent on Mathematics, different policies on student grouping etc. (Note: The international data shows no simple relationship between achievement and in-class instructional time (Beaton *et al*, 1996, p5))

(v) Socio – economic features of Victoria and NSW (both for example being industrial, high migration states) or the way of life may be less conducive to high educational achievement (but see below).
(vi) There may be a bias in the sampling caused by certain types of schools being more likely not to agree to participate in some states. (This appears to be unlikely (Lokan *et al*, 1997, p 46)).

For the benefit of Victorian education, these hypotheses should be fully explored. In particular, the possibility that children are successfully learning topics elsewhere at an earlier age should be explored with a view to potential revision of expectations in Victoria.

Future directions

In this article, it has only been possible to give a short introduction to the vast amount of data that has been produced by the TIMSS study. I have attempted to show that the data has definite limitations and it is generally not possible to get simple answers directly from it. Later analyses will enable the interactions of the many variables to be clarified. However, I have also attempted to show that the international perspective provides us with both some useful measures of how we are going and some ideas for how things might be improved and what might be possible. Australia and Victoria can use this information to set achievable goals that are in harmony with its own values and culture. For Victoria, my tentative conclusion is that a serious examination of curriculum content and its timing could well lead to improved performance.

References

Beaton, A.E., Mullis, I.V.S., Martin, M.O., Gonzales, E.J., Kelly, D.L. & Smith, T.A. (1996) *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS).* Chestnut Hill, Massachusetts: Boston College.

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*. Melbourne: ACER

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.* Melbourne: ACER

Keeves, J. (1966). *Evaluation of Achievement in Mathematics*. Hawthorn, Victoria: ACER.

Rosier, M.J. (1980) *Changes in Secondary school Mathematics in Australia. 1964 – 1980.* Hawthorn, Victoria: ACER.