2000

A Policy Makers Guide to International Achievement Studies

Margaret Forster

ACER, Forster@acer.edu.au

Follow this and additional works at: https://research.acer.edu.au/policy_makers_guides

Part of the Educational Assessment, Evaluation, and Research Commons, and the International and Comparative Education Commons

Recommended Citation


This Book is brought to you by the ACER Archives at ACEReSearch. It has been accepted for inclusion in Policy Maker's Guides by an authorized administrator of ACEReSearch. For more information, please contact repository@acer.edu.au.
A POLICY MAKER’S GUIDE TO

International Achievement Studies

Margaret Forster
The ACER Core-Funded Research Program

The Australian Council for Educational Research conducts a core program of research funded by an annual grant from the States and Territories and the Commonwealth.

This annual grant allows research to be undertaken into issues of general importance in Australian education and complements research projects commissioned from time to time by individual States, Territories and the Commonwealth.

Priorities for the ACER core research program are reviewed every three years. The three-year program under which this work was completed focused on an overarching question: What can be done to improve learning outcomes? and addressed five priority areas:

- assessment and reporting to improve learning
- improving literacy and numeracy learning
- improving outcomes for Indigenous students
- teaching practices to improve learning
- vocational outcomes and lifelong learning
A POLICY MAKER'S GUIDE TO
International Achievement Studies

Margaret Forster

ACER PRESS
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>3</td>
</tr>
<tr>
<td>What are the purposes of international achievement studies?</td>
<td>3</td>
</tr>
<tr>
<td>Why the interest in international achievement studies?</td>
<td>4</td>
</tr>
<tr>
<td>What are some examples of international achievement studies?</td>
<td>5</td>
</tr>
<tr>
<td>Who reports international achievement data and how are these data reported?</td>
<td>6</td>
</tr>
<tr>
<td>In what ways are international achievement data used and with what impact?</td>
<td>20</td>
</tr>
<tr>
<td>What concerns have been raised?</td>
<td>27</td>
</tr>
<tr>
<td>International achievement data and informed decision making</td>
<td>30</td>
</tr>
<tr>
<td>Useful websites</td>
<td>32</td>
</tr>
<tr>
<td>Endnotes</td>
<td>33</td>
</tr>
</tbody>
</table>
This guide provides policy makers with research-based information about international achievement studies.

Good decision making at all levels of an education system is facilitated by easily accessible, relevant, and reliable information.

Many indicators provide useful input to educational decision making; but the most important indicators are those which address the central concern of education: the promotion of student learning.

Education systems monitor student learning—with the fundamental intention of promoting learning—by collecting, analysing and reporting student achievement data. Given that state, national and international achievement studies are both time consuming and expensive, it seems prudent to reflect on this effort:

What are the purposes of these studies?
How are data reported and used?
What concerns have been raised about these studies?
How can we ensure that data will provide evidence for informed decision making?
INTRODUCTION

International achievement studies focus on the collection and analysis of cross-national information. Two kinds of information usually are collected:

- data on student achievement in particular subject areas at particular ages and stages of schooling; and
- background information including characteristics of students, teachers and schools.

In some studies curriculum information is central also.

Achievement data are collected through tests administered to samples of students selected to be representative of national populations. Background information is collected by means of questionnaires completed by students, their teachers and principals.

Two main agencies direct international achievement studies:

- IEA, the International Association for the Evaluation of Educational Achievement; and
- OECD, the Organisation for Economic Co-operation and Development.

The US Educational Testing Service (ETS) also has conducted several studies (International Assessment of Education Progress, IAEP studies) as has SACMEQ — the South African Consortium for Monitoring Educational Quality.

WHAT ARE THE PURPOSES OF INTERNATIONAL ACHIEVEMENT STUDIES?

The studies aim to

- provide policy makers and educational practitioners with information about their education system in relation to other systems; and
- assist policy makers and educational practitioners to understand the reasons for observed differences in the achievements of students from different educational systems.

To achieve these aims the studies examine the impact and effect on educational systems of policies that are applied consistently (in general) within nations but which may vary across nations:

The understandings we obtain from cross-national comparisons of such policies as age of school entry, hours and methods of instruction, and teacher training, can provide us with new insights into the performance of our own educational system in general, and of the relationship between student performance and its antecedents and consequences in particular:

In international studies, the world is viewed as a global educational laboratory where different national policies and practices yield different educational outcomes. The underlying assumption is that differences in student performance between countries can be linked to characteristics of particular education systems. It is recognised that these characteristics need to be understood in their broader cultural and economic contexts.
WHY THE INTEREST IN INTERNATIONAL ACHIEVEMENT STUDIES?

Governments are convinced that future economic competitiveness depends upon high levels of knowledge and skills in the working population—better performance is seen as essential to a nation’s future standing in the global economy. The link between future economic performance and the current achievement of school populations, particularly in literacy, numeracy, and science, is assumed.

Researchers are interested in using the information provided to explore associations between policies, instructional approaches and achievement that might assist policy makers and teachers to improve student learning. Researchers also recognise that international studies provide a context for the development of research methodologies and analytical approaches that can be of benefit in other educational contexts.

Is this interest well-founded?

Evidence for a causal link between the achievement of a country’s school population and economic performance is inconclusive.

According to OECD research, there is consistent evidence from multiple sources of a strong positive relationship between educational attainment levels and productivity growth at a national level. The relationship is strongest when comparing less developed with more developed countries. However, there is a debate about the direction of causality. There is also evidence at an individual level of a relationship between educational attainment and employment and earnings.

IEA publications indicate that there is ‘very little firm evidence to support the widely-held view that there is a strong and direct causal connection between mean student test scores for nations and their economic competitiveness’. Some recent research studies conclude that there is no evidence to demonstrate this link.

Setting aside considerations of economic performance, it is clear that international achievement studies do provide information with the potential to improve student learning by informing decision making. The ways in which international achievement data are reported and used to assist policy makers is the focus of a large part of this guide.

It also is evident that international studies do provide a context for the development of analytical techniques that can be applied in other educational contexts. For example, pioneering work in the use of regression and multivariate analyses was undertaken in IEA studies.
WHAT ARE SOME EXAMPLES OF INTERNATIONAL ACHIEVEMENT STUDIES?

**IEA reading, mathematics and science studies**

IEA, an international non-government co-operative organisation comprising research centres and ministries of education, has conducted several assessments of reading, mathematics and science as well as other subjects, since 1959.

The recent Third International Mathematics and Science Study (TIMSS) is the largest international comparative study of educational achievement ever undertaken. The 1994–5 testing included 45 countries, more than 15 000 schools, and more than half a million students. Testing at five grade levels (3rd, 4th, 7th, 8th and final year of secondary school) was conducted in more than 30 different languages. Data from student, teacher, and principal questionnaires were collected also.

**OECD/PISA—the Programme for International Student Assessment**

PISA assesses how far students approaching the end of compulsory schooling (defined as 15 year olds) have acquired some of the knowledge and skills essential for full participation in society. Assessments take place every three years (beginning in 2000), in three domains: reading literacy, mathematical literacy and scientific literacy. Thirty-three countries participated in the first PISA cycle.

**International Assessment of Educational Progress (IAEP) mathematics and science studies**

Educational Testing Service (ETS) conducted international assessments of mathematics and science achievements of 13 year olds in 1988 and in 1990–1. Seven countries participated in the first study, 20 in the second.

**SACMEQ studies of mathematics and reading**

SACMEQ, a network of ministries of education in the South African sub-region, has conducted two assessments (1995 and 1999) of the mathematics and reading achievement of Grade 6 students. Fourteen countries participated in the 1999 study.
WHO REPORTS INTERNATIONAL ACHIEVEMENT DATA AND HOW ARE THESE DATA REPORTED?

Data are reported by

• agencies directing the studies (official study reports);
• national agencies co-ordinating the studies within participating countries (official national reports);
• independent educational researchers; and
• the press.

Data usually are reported showing

• rank ordering of countries (sometimes called ‘league tables’) by whole test and sub-test including item level;
• the spread of student achievement (distributions);
• subgroup differences in achievement; and
• relationships between achievement and background variables.

A description of the scales against which student achievement is reported sometimes accompanies official study reports.

Examples of the ways in which data are reported are shown on pages 7–19.

Rank ordering of countries

1 Countries are shown in rank order according to the average (mean) achievement of their students on the complete set of test items.

Where league tables appear in newspaper articles they are presented in their most basic form as tables or graphs without qualifiers (see example 1 opposite). Tables and graphs of this kind are simple and visually dramatic but are open to misinterpretation (see page 27).

Where league tables appear in official reports they usually are presented with qualifiers including years of formal schooling, average age of participants, mean and confidence intervals, sampling irregularities, and statistical significance of country differences (see examples 2 and 3). Tables and graphs of this kind are complex, but provide information that assists readers to interpret the relative positions of countries.
Example 1 Reporting student achievement: The press—league table
(The Age, Melbourne, Australia, 12 June 1997)

Graphs of this kind are simple and visually dramatic but are open to misinterpretation (see page 29).

Primary students among world’s best in science

By ALEX MESSINA, education reporter

Australian primary school science students were among the best in a world study involving 26 countries.

The 11,500 Australian nine-year-old students from 179 schools outperformed two-thirds of countries. Australia was ranked equal-second with three other countries, behind Korea.

However, Australian students were only just above the average in mathematics, according to the Third International Maths and Science Study, which tested 180,000 nine-year-olds worldwide in 1994.

Within Australia, Victoria came in the bottom group of three states in both disciplines. In science, Western Australia outclassed all states and tied with Korea.

Asian nations — notably Singapore, Korea and Japan — tended to excel, compared with Western nations. A survey of 14,000 secondary students aged 13, released in November, had similar results.

A trend was the discontent among Australian teachers. Almost half (45 per cent) of primary teachers said they would give up teaching if they had the chance. Of 26 countries, only New Zealand and Israeli teachers were more discontented. Last year’s results showed that up to 60 per cent of lower secondary teachers in Australia felt the same.

The TIMSS test sorts nations in eight groups. In science, Australia was in the top one-third of nations and in the second-strongest group with Japan, the United States and Austria.

In maths, Australia was ranked in group three with seven other countries.

Of the states, the next best group in science behind WA included Queensland, the Australian Capital Territory and the Northern Territory, followed by the lowest group with Victoria, New South Wales and Tasmania.

In mathematics, the states finished in the same positions. Victoria and Tasmania also finished in the bottom group of 13-year-olds.
Example 2  Reporting student achievement: O fficial study reports—league table with qualifiers

TIMSS Science achievement

Tables and graphs of this kind are complex, but provide information that assists readers to interpret the relative positions of countries. For example, without a qualifier related to sample participation rates, Bulgaria and the Netherlands would have appeared towards the top of the country rankings. Slovenia would have appeared in fifth place without a caveat related to age grade specifications—in Slovenia a high percentage of older students participated.

<table>
<thead>
<tr>
<th>Country</th>
<th>Mean</th>
<th>Years of Formal Schooling</th>
<th>Average Age</th>
<th>Science Achievement Scale Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singapore</td>
<td>607</td>
<td>8</td>
<td>14.5</td>
<td></td>
</tr>
<tr>
<td>Czech Republic</td>
<td>574</td>
<td>8</td>
<td>14.4</td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>571</td>
<td>8</td>
<td>14.4</td>
<td></td>
</tr>
<tr>
<td>Korea</td>
<td>565</td>
<td>8</td>
<td>14.2</td>
<td></td>
</tr>
<tr>
<td>Hungary</td>
<td>554</td>
<td>8</td>
<td>14.3</td>
<td></td>
</tr>
<tr>
<td>-2England</td>
<td>552</td>
<td>9</td>
<td>14.0</td>
<td></td>
</tr>
<tr>
<td>- Belgium (FI)</td>
<td>550</td>
<td>8</td>
<td>14.1</td>
<td></td>
</tr>
<tr>
<td>Slovak Republic</td>
<td>544</td>
<td>8</td>
<td>14.3</td>
<td></td>
</tr>
<tr>
<td>Russian Federation</td>
<td>538</td>
<td>7 or 8</td>
<td>14.0</td>
<td></td>
</tr>
<tr>
<td>Ireland</td>
<td>538</td>
<td>8</td>
<td>14.4</td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>536</td>
<td>7</td>
<td>13.9</td>
<td></td>
</tr>
<tr>
<td>- United States</td>
<td>534</td>
<td>7</td>
<td>14.2</td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>531</td>
<td>8</td>
<td>14.1</td>
<td></td>
</tr>
<tr>
<td>Norway</td>
<td>527</td>
<td>7</td>
<td>13.9</td>
<td></td>
</tr>
<tr>
<td>New Zealand</td>
<td>525</td>
<td>8.5-9.5</td>
<td>14.0</td>
<td></td>
</tr>
<tr>
<td>Hong Kong</td>
<td>522</td>
<td>8</td>
<td>14.2</td>
<td></td>
</tr>
<tr>
<td>1 Switzerland</td>
<td>522</td>
<td>7 or 8</td>
<td>14.2</td>
<td></td>
</tr>
<tr>
<td>Spain</td>
<td>517</td>
<td>8</td>
<td>14.3</td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>498</td>
<td>8</td>
<td>14.3</td>
<td></td>
</tr>
<tr>
<td>Iceland</td>
<td>494</td>
<td>8</td>
<td>13.6</td>
<td></td>
</tr>
<tr>
<td>1 Latvia (LSS)</td>
<td>485</td>
<td>8</td>
<td>14.3</td>
<td></td>
</tr>
<tr>
<td>Portugal</td>
<td>480</td>
<td>8</td>
<td>14.5</td>
<td></td>
</tr>
<tr>
<td>Lithuania</td>
<td>476</td>
<td>8</td>
<td>14.3</td>
<td></td>
</tr>
<tr>
<td>Iran, Islamic Rep.</td>
<td>478</td>
<td>8</td>
<td>14.6</td>
<td></td>
</tr>
<tr>
<td>Cyprus</td>
<td>463</td>
<td>8</td>
<td>13.7</td>
<td></td>
</tr>
</tbody>
</table>

Countries Not Satisfying Guidelines for Sample Participation Rates (See Appendix A for Details):

<table>
<thead>
<tr>
<th>Country</th>
<th>Mean</th>
<th>Years of Formal Schooling</th>
<th>Average Age</th>
<th>Science Achievement Scale Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>543</td>
<td>8</td>
<td>14.2</td>
<td></td>
</tr>
<tr>
<td>Austria</td>
<td>558</td>
<td>8</td>
<td>14.3</td>
<td></td>
</tr>
<tr>
<td>Belgium (Fr)</td>
<td>471</td>
<td>8</td>
<td>14.3</td>
<td></td>
</tr>
<tr>
<td>Bulgaria</td>
<td>566</td>
<td>8</td>
<td>14.0</td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>560</td>
<td>8</td>
<td>14.3</td>
<td></td>
</tr>
<tr>
<td>Scotland</td>
<td>517</td>
<td>9</td>
<td>13.7</td>
<td></td>
</tr>
</tbody>
</table>

Countries Not Meeting Age/Grade Specifications (High Percentage of Older Students; See Appendix A for Details):

<table>
<thead>
<tr>
<th>Country</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colombia</td>
<td>411</td>
</tr>
<tr>
<td>Germany</td>
<td>531</td>
</tr>
<tr>
<td>Romania</td>
<td>466</td>
</tr>
<tr>
<td>Slovakia</td>
<td>560</td>
</tr>
</tbody>
</table>

Countries With Unapproved Sampling Procedures at Classroom Level (See Appendix A for Details):

<table>
<thead>
<tr>
<th>Country</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denmark</td>
<td>478</td>
</tr>
<tr>
<td>Greece</td>
<td>497</td>
</tr>
<tr>
<td>Thailand</td>
<td>525</td>
</tr>
</tbody>
</table>

Unapproved Sampling Procedures at Classroom Level and Not Meeting Other Guidelines (See Appendix A for Details):

<table>
<thead>
<tr>
<th>Country</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iraq</td>
<td>524</td>
</tr>
<tr>
<td>Kuwait</td>
<td>430</td>
</tr>
<tr>
<td>South Africa</td>
<td>326</td>
</tr>
</tbody>
</table>

Percentiles of Performance

Mean and Confidence Interval (L2SE)

1. Eighth grade in most countries; see Table 2 for information about the grades tested in each country.
2. Met guidelines for sample participation rates only after replacement schools were included (see Appendix A for details).
3. Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

Source: IEA Third International Mathematics and Science Study (TIMSS), 1994-95
Example 3 Reporting student achievement: Official study reports—
showing statistical significance of country differences

TIMSS Science achievement

Although they are visually demanding, reports of this kind assist readers to understand
the significance of reported differences in the achievements of countries. For example,
this report shows that although the Czech Republic, Japan and Korea were ranked
second, third and fourth (Example 2) there was no significant difference in their scores.

Multiple Comparisons of Achievement in the Sciences – Upper Grade (Eighth Grade*)

Instructions: Read across the row for a country to compare performance with the countries listed in the heading of the chart. The
symbols indicate whether the mean achievement of the country in the row is significantly lower than that of the comparison country,
significantly higher than that of the comparison country, or if there is no statistically significant difference between the countries.¹

| COUNTRY            | Singapore | Czech Republic | Japan | Korea | Bulgaria | Netherlands | Slovenia | Austria | Hungary | England | Belgium (Fi) | Australia | Slovak Republic | Russian Federation | Ireland | Sweden | United States | Germany | Canada | Norway | New Zealand | Thailand | Israel | Hong Kong | Switzerland | Scotland | Spain | France | Tenerife | Iceland | Romania | Latvia (LSS) | Portugal | Denmark | Lithuania | Belgium (Fr) | Iran, Islamic Rep. | Cyprus | Kuwait | Côte d'Ivoire | South Africa |
|--------------------|-----------|----------------|-------|-------|----------|------------|----------|---------|---------|---------|----------|-----------|-----------|------------------|---------------------|---------|--------|--------------|---------|--------|--------|------------|----------|--------|-----------|-------------|--------------|--------|--------|-------------|-------------|--------|---------|------------|-------------|---------------|--------|---------|-------------|-------------|
|                    | Mean achievement significantly higher than comparison country | Mean achievement significantly lower than comparison country | No statistically significant difference from comparison country |

Countries are ordered by mean achievement across the heading and down the rows.

* Eighth grade in most countries: see Table 2 for information about the grades tested
  in each country.
  ¹ Statistically significant at .05 level, adjusted for multiple comparisons.
  Because coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking School only.
  Countries shown in italics did not satisfy one or more guidelines for sample participation rates,
  age/grade specifications, or classroom sampling procedures (see Appendix A for details).

Source: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.
Countries are shown in rank order according to the average (mean) achievement of their students on subsets of test items. For example, country rankings in a science study may be reported separately on subsets of items addressing earth science, life science, physics, chemistry, environmental issues, and the nature of science. Displays of this kind usually are found in national reports and independent research reports.

Given that the means rarely are reported with qualifiers, and that the items may not have been selected to adequately represent a curriculum area, the information needs to be interpreted with caution. Nevertheless, these breakdowns have the advantage of providing information that is masked by overall means. Students may perform particularly well, or poorly, in certain content areas and this may inform curriculum review.

Examples 4 and 5 illustrate increasingly detailed content-level reporting.

---

**Example 4 Reporting student achievement: Independent research reports—league tables showing mean scores for content areas of the test**

Average mathematics scores for specific content areas for fourth grade students (national per cent correct in each area).  

<table>
<thead>
<tr>
<th>Common Fractions</th>
<th>Geometry: position and shapes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singapore</td>
<td>Australia</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>England</td>
</tr>
<tr>
<td>Japan</td>
<td>Netherlands</td>
</tr>
<tr>
<td>Korea</td>
<td>Hong Kong</td>
</tr>
<tr>
<td>Hungary</td>
<td>Canada</td>
</tr>
<tr>
<td>Netherlands</td>
<td>USA</td>
</tr>
<tr>
<td>USA</td>
<td>Czech Republic</td>
</tr>
<tr>
<td>Israel</td>
<td>Japan</td>
</tr>
<tr>
<td>Australia</td>
<td>Singapore</td>
</tr>
<tr>
<td>Canada</td>
<td>New Zealand</td>
</tr>
<tr>
<td>International</td>
<td>Korea</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>Hungary</td>
</tr>
<tr>
<td>England</td>
<td>International</td>
</tr>
<tr>
<td>Thailand</td>
<td>Israel</td>
</tr>
<tr>
<td>New Zealand</td>
<td>Norway</td>
</tr>
<tr>
<td>Norway</td>
<td>Thailand</td>
</tr>
</tbody>
</table>

---
Countries are shown in rank order according to the percentage of students scoring in the top ten per cent of students internationally. Reports of this kind usually are found in independent research articles. They provide a different perspective on relative performance. Example 6 shows the percentage of students in each country scoring in the top 10 per cent of students in the world. About 45% of the participating students from Singapore scored in this category.

The highest scoring ten per cent contained about 13% of Australian students. ‘Perhaps we [in Australia] are not doing as much for our best students as we could, to challenge them to achieve their potential,’ concludes the researcher who presented these data.13

<table>
<thead>
<tr>
<th>Country</th>
<th>Overall mean for task</th>
<th>Q1 Complete Table</th>
<th>Q2 Describe Pattern</th>
<th>Q3 Apply algorithm</th>
<th>Q4 Count frequencies</th>
<th>Q5A Identify most frequent number</th>
<th>Q5B Explain findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singapore</td>
<td>84</td>
<td>97</td>
<td>90</td>
<td>95</td>
<td>84</td>
<td>95</td>
<td>44</td>
</tr>
<tr>
<td>England</td>
<td>79</td>
<td>97</td>
<td>83</td>
<td>93</td>
<td>73</td>
<td>90</td>
<td>38</td>
</tr>
<tr>
<td>Switzerland</td>
<td>79</td>
<td>91</td>
<td>86</td>
<td>94</td>
<td>69</td>
<td>86</td>
<td>45</td>
</tr>
<tr>
<td>Canada</td>
<td>77</td>
<td>92</td>
<td>84</td>
<td>90</td>
<td>75</td>
<td>88</td>
<td>31</td>
</tr>
<tr>
<td>Netherlands</td>
<td>76</td>
<td>97</td>
<td>82</td>
<td>96</td>
<td>72</td>
<td>87</td>
<td>21</td>
</tr>
<tr>
<td>Scotland</td>
<td>76</td>
<td>93</td>
<td>73</td>
<td>93</td>
<td>70</td>
<td>87</td>
<td>41</td>
</tr>
<tr>
<td>Sweden</td>
<td>74</td>
<td>94</td>
<td>65</td>
<td>92</td>
<td>71</td>
<td>81</td>
<td>44</td>
</tr>
<tr>
<td>Czech Rep.</td>
<td>73</td>
<td>93</td>
<td>75</td>
<td>83</td>
<td>73</td>
<td>78</td>
<td>39</td>
</tr>
<tr>
<td>United States</td>
<td>71</td>
<td>89</td>
<td>76</td>
<td>88</td>
<td>69</td>
<td>77</td>
<td>29</td>
</tr>
<tr>
<td>International Mean</td>
<td>73</td>
<td>90</td>
<td>71</td>
<td>90</td>
<td>71</td>
<td>83</td>
<td>33</td>
</tr>
</tbody>
</table>
Box and whisker displays sometimes are used to illustrate the distributions of student achievement for each participating country. Displays of this kind, which are found in official study reports, illustrate the spread of results between the lowest and highest five per cent of students in any country (see example 7, opposite).
Example 7 Reporting student achievement: Official study reports—illustrating the distributions of student achievement

Lower grade mathematics achievement in TIMSS

<table>
<thead>
<tr>
<th>Country</th>
<th>Average age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Korea</td>
<td>9.3</td>
</tr>
<tr>
<td>Singapore</td>
<td>9.3</td>
</tr>
<tr>
<td>Japan</td>
<td>9.4</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>9.1</td>
</tr>
<tr>
<td>Czech Rep.</td>
<td>9.4</td>
</tr>
<tr>
<td>*Netherlands</td>
<td>9.3</td>
</tr>
<tr>
<td>*Slovenia</td>
<td>9.9</td>
</tr>
<tr>
<td>*Austria</td>
<td>9.5</td>
</tr>
<tr>
<td>*Australia</td>
<td>9.2</td>
</tr>
<tr>
<td>United States</td>
<td>9.2</td>
</tr>
<tr>
<td>*Hungary</td>
<td>9.4</td>
</tr>
<tr>
<td>Ireland</td>
<td>9.3</td>
</tr>
<tr>
<td>Canada</td>
<td>9.1</td>
</tr>
<tr>
<td>*Latvia</td>
<td>9.7</td>
</tr>
<tr>
<td>*Scotland</td>
<td>8.7</td>
</tr>
<tr>
<td>*England</td>
<td>9.1</td>
</tr>
<tr>
<td>*Thailand</td>
<td>9.7</td>
</tr>
<tr>
<td>New Zealand</td>
<td>9.0</td>
</tr>
<tr>
<td>Cyprus</td>
<td>8.8</td>
</tr>
<tr>
<td>Greece</td>
<td>8.6</td>
</tr>
<tr>
<td>Portugal</td>
<td>9.1</td>
</tr>
<tr>
<td>Norway</td>
<td>8.8</td>
</tr>
<tr>
<td>Iceland</td>
<td>8.6</td>
</tr>
<tr>
<td>Iran, Isl. Rep.</td>
<td>9.4</td>
</tr>
</tbody>
</table>

* These countries did not meet all the sampling criteria.

470 (Average of all country means)
Showing subgroup differences
Subgroup differences are shown in tabular form or illustrated graphically. Displays of this kind usually are found in official reports. They reveal between- and within-country differences that can be masked by country means. Example 8 below shows the achievements of male and female students in the 1988 IAEP mathematics study.

Describing achievement scales
An elaboration of the scales against which student achievement is reported sometimes accompanies results in official reports. Different positions on the scale are described in words (see example 9) and/or illustrated with tasks (see example 10). These displays help readers to understand the underlying dimension against which achievement is being monitored.

Example 8  Reporting student achievement: Official study reports—illustrating sub-group achievement
Displays of this kind can illustrate within-country differences that can be masked by country means.

Average mathematics proficiency by gender, age 13 IAEP 1988
Jackknifed standard errors are presented in parentheses.

Background data are missing from 31% of the New Brunswick (English) students. The resultant effect on the differences between groups from this particular province is estimated to be less than one standard error.

Reprinted by permission of Educational Testing Service, the copyright owner.

<table>
<thead>
<tr>
<th>Country</th>
<th>MALE</th>
<th>FEMALE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Score</td>
<td>Score</td>
</tr>
<tr>
<td></td>
<td>(SE)</td>
<td>(SE)</td>
</tr>
<tr>
<td>Korea</td>
<td>576.7</td>
<td>558.0</td>
</tr>
<tr>
<td>(French)</td>
<td>(3.4)</td>
<td>(3.9)</td>
</tr>
<tr>
<td>Quebec (Fr.)</td>
<td>546.3</td>
<td>539.3</td>
</tr>
<tr>
<td>(Eng.)</td>
<td>(4.2)</td>
<td>(4.3)</td>
</tr>
<tr>
<td>British</td>
<td>539.6</td>
<td>541.3</td>
</tr>
<tr>
<td>Columbia</td>
<td>(2.8)</td>
<td>(2.6)</td>
</tr>
<tr>
<td>Quebec (Eng.)</td>
<td>534.1</td>
<td>537.3</td>
</tr>
<tr>
<td>(Eng.)</td>
<td>(3.0)</td>
<td>(2.3)</td>
</tr>
<tr>
<td>New Brunswick (Eng.)</td>
<td>526.6</td>
<td>529.0</td>
</tr>
<tr>
<td>(Eng.)</td>
<td>(4.3)</td>
<td>(3.4)</td>
</tr>
<tr>
<td>Ontario (Eng.)</td>
<td>517.8</td>
<td>514.6</td>
</tr>
<tr>
<td>(Eng.)</td>
<td>(4.4)</td>
<td>(3.3)</td>
</tr>
<tr>
<td>New Brunswick (Fr.)</td>
<td>516.8</td>
<td>513.4</td>
</tr>
<tr>
<td>(Fr.)</td>
<td>(4.2)</td>
<td>(3.6)</td>
</tr>
<tr>
<td>Spain</td>
<td>523.2</td>
<td>499.9</td>
</tr>
<tr>
<td>(Fr.)</td>
<td>(5.3)</td>
<td>(5.3)</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>507.0</td>
<td>512.5</td>
</tr>
<tr>
<td>(Fr.)</td>
<td>(5.0)</td>
<td>(3.4)</td>
</tr>
<tr>
<td>Ireland</td>
<td>508.2</td>
<td>499.6</td>
</tr>
<tr>
<td>(Eng.)</td>
<td>(5.7)</td>
<td>(3.9)</td>
</tr>
<tr>
<td>Ontario (Fr.)</td>
<td>480.6</td>
<td>482.7</td>
</tr>
<tr>
<td>(Eng.)</td>
<td>(3.4)</td>
<td>(3.0)</td>
</tr>
<tr>
<td>United States</td>
<td>474.6</td>
<td>473.2</td>
</tr>
<tr>
<td>(Fr.)</td>
<td>(6.4)</td>
<td>(5.1)</td>
</tr>
</tbody>
</table>
Example 9  Reporting achievement: Official study reports—describing positions on achievement scales

The first IAEP Science Scale.16

Descriptions of this kind help readers to understand the underlying dimension against which achievement is being monitored.

**Level 700  Integrate scientific information and experimental evidence**

Students at this level can interpret experimental data that involve several variables. They also can interrelate information represented in a variety of forms—text, graphs, figures, and diagrams. Students can make predictions based on data and observations and are aware of limitations of extrapolations. Students demonstrate a growing understanding of more advanced scientific knowledge and concepts, such as the definition of a calorie or the concept of chemical change.

**Level 600  Understand and apply intermediate scientific knowledge and principles**

Students at this level demonstrate an understanding of intermediate scientific facts and principles and can apply this understanding in designing experiments and interpreting data. They also can interpret figures and diagrams used to convey scientific information. Students at this level can infer relationships and draw conclusions by applying facts and principles, particularly from physical science.

**Level 500  Use scientific procedures and analyse scientific data**

Students at this level have a grasp of experimental procedures used in science, such as designing experiments, controlling variables, and using equipment. They can identify the best conclusions drawn from data on a graph and the best explanation for observed phenomena. Students also understand some concepts in a variety of science content areas, including the Life Sciences, Physical Sciences, and Earth and Space Sciences.

**Level 400  Understand and apply simple scientific principles**

Students at this level exhibit growing knowledge in the Life Sciences, particularly human biological systems, and can apply some basic principles from the Physical Sciences, including force. They also display a beginning understanding of some of the basic methods of reasoning used in science, including classification and interpretation of statements.

**Level 300  Know everyday science facts**

Students at this level know some general science facts of the type that can be learned from everyday experiences. For example, they exhibit some rudimentary knowledge concerning the environment and animals.

Reprinted by permission of Educational Testing Service, the copyright owner.
TIMSS international science scale

Displays of this kind help readers to understand the underlying dimension against which achievement is being monitored.

International Difficulty Map for Earth Science Example Items Lower and Upper Grades (Seventh and Eighth Grades*)

Example 1A  River on the plain: Good place for farming.
Scale value = 383
International Average Per cent Correct:
Eighth Grade = 79%
Seventh Grade = 76%  K01A

Example 1B  River on the plain: Bad place for farming.
Scale value = 632
International Average Per cent Correct:
Eighth Grade = 42%
Seventh Grade = 38%  W01B

Example 2  Fossil fuels.
Scale value = 526
International Average Per cent Correct:
Eighth Grade = 62%
Seventh Grade = 55%  K15

Example 3  Ozone layer.
Scale value = 583
International Average Per cent Correct:
Eighth Grade = 53%
Seventh Grade = 43%  R04

Example 4  Diagram of Earth’s water cycle.
Scale value = 659
International Average Per cent Correct:
Eighth Grade = 32%
Seventh Grade = 27%  W02

Example 5  Gases in air.
Scale value = 750
International Average Per cent Correct:
Eighth Grade = 27%
Seventh Grade = 22%  012

* Seventh and Eighth grades in most countries; see Table 2 for information about the grades tested in each country.
NOTE: Each item was placed onto the TIMMS international science scale based on students’ performance in both grades. Items are shown at the point on the scale where students with that level of proficiency had a 65 per cent probability of providing a correct response.
Showing relationships with background variables

The relationships between student achievement and background variables are shown either in correlation tables (see example 11 below) or graphically (see examples 12 and 13). Displays of this kind are found in official study reports and analyses by independent researchers.

Example 11 Reporting student achievement: Official study reports—correlations between TIMSS written tests and student background variables for Australian students

This table shows the relationship between student achievement on TIMSS written tests in mathematics and science and a number of background variables. For example, there is a low correlation between students’ attraction to the subject (‘like maths’, ‘like science’) and their achievement.

<table>
<thead>
<tr>
<th></th>
<th>Population 1</th>
<th></th>
<th>Population 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Maths</td>
<td>Science</td>
<td>Maths</td>
<td>Science</td>
</tr>
<tr>
<td>Number of books in home</td>
<td>.18</td>
<td>.21</td>
<td>.27</td>
<td>.28</td>
</tr>
<tr>
<td>Family size</td>
<td>-.14</td>
<td>-.17</td>
<td>-.10</td>
<td>-.11</td>
</tr>
<tr>
<td>Parents’ education status</td>
<td>-</td>
<td>-</td>
<td>.28</td>
<td>.30</td>
</tr>
<tr>
<td>Parents’ occupation status</td>
<td>.27</td>
<td>.27</td>
<td>.30</td>
<td>.29</td>
</tr>
<tr>
<td>Home background composite</td>
<td>.36</td>
<td>.37</td>
<td>.40</td>
<td>.39</td>
</tr>
<tr>
<td>Language background composite</td>
<td>.11</td>
<td>.15</td>
<td>.10</td>
<td>.15</td>
</tr>
<tr>
<td>Word knowledge</td>
<td>.61</td>
<td>.60</td>
<td>.47</td>
<td>.48</td>
</tr>
<tr>
<td>Like mathematics</td>
<td>.16</td>
<td>.15</td>
<td>.22</td>
<td>.10</td>
</tr>
<tr>
<td>Like science</td>
<td>.14</td>
<td>.16</td>
<td>.19</td>
<td>.25</td>
</tr>
<tr>
<td>Attribute success to luck</td>
<td>-.32</td>
<td>-.30</td>
<td>-.29</td>
<td>-.28</td>
</tr>
<tr>
<td>Self-efficacy belief</td>
<td>.16</td>
<td>.15</td>
<td>.32</td>
<td>.28</td>
</tr>
</tbody>
</table>
Example 12 Reporting achievement: Official study reports—the relationship between mathematics and science TIMSS achievement and time spent watching TV for Australian students

Achievement and daily TV watching
Population 1

- Science
- Maths
Example 13 Reporting achievement: Official study reports—correlations between reading achievement and the size of school libraries, IEA reading literacy study

This graph illustrates the regular increase in average score with increases in school library size. This finding was across all countries participating in the study and within most of them. There was also a difference between wealthier and poorer countries as defined by the Composite Development Index (CDI).
International achievement data are used to

• motivate ‘improvement’;
• confirm performance expectations;
• inform policy making; and
• initiate further within-country analyses.

Comparisons and analyses of different kinds are made by the media, in official study reports, in participating country reports and by independent educational researchers. Examples of the impact of reports are provided below.

**Motivating ‘improvement’ by comparing international rankings**

Countries compare their ranking with other countries and decide whether they are happy with their position (sometimes called ‘description’ or ‘mirror’ function of the studies, or the ‘cognitive Olympics’).

Comparisons of this kind are made by researchers

…where would we want to be placed in the next major cross-national study? What would we need to do to ensure that we can achieve that goal? and the press. For example, widespread media publicity was given to the results of 10 of the 23 countries participating in TIMSS. Minor media publicity was given in a further six.

**What is the impact of this kind of reporting?**

Reporting international rankings can have a major impact in generating public engagement, in effecting policy decisions at ministerial level, and in motivating schools to change practice. Examples 14 and 15 indicate the impact of reporting Second International Mathematics Study (SIMS) rankings in Sweden and TIMSS rankings in the United States.

---

**Example 14 Motivating ‘improvement’ by comparing international rankings: Sweden**

The relatively low mean score for 13 year old students in Sweden in the Second International Mathematics Study (SIMS) was publicised widely in the press: ‘Sweden at the Level of Developing Countries’, ‘Sweden close to Swaziland’. A task force was appointed by the minister of school education to investigate mathematics education, and a study of the competence of teachers in the middle grades was undertaken. Changes in resource allocation to in-service mathematics training in response to SIMS has been credited with the improvement in TIMSS results.
Example 15  Motivating ‘improvement’ by comparing international rankings: United States

One of the six goals of education proclaimed at federal level in the United States in 1990 was that American youth would perform at the top of the competency ladder in mathematics and science by the year 2000.

One consortium of 17 school districts located in the north suburbs of Chicago and the Illinois Math and Science Academy aims to become first in the world in mathematics and science achievement.

The consortium has created a forum for dialogue with business and government leaders to clarify standards for being first in the world, and has established a network of learning communities involving math and science staff, research and development personnel, parents and community leaders.

The consortium begins with the question: How do students in consortium schools perform in comparison to students in countries around the world on international tests? Other questions to be answered include: Do school programs in the consortium reflect a world-class curriculum? How do the preparation and instructional practices of consortium teachers compare with those of teachers from countries that successfully prepare their students for the global market place? Does curriculum in consortium schools ‘fit’ with international standards? How do the social and cultural contexts differ between consortium schools and countries around the world? What lessons can be learned from analysing standardised test data that can be used to improve science and mathematics education in consortium schools?

Teachers in participating schools work with specialists from the North Central Regional Educational Laboratory (NCREL) in four learning networks:

1. Curriculum analysis—making explicit, comparing, and aligning grade-level and district-level curriculum with national and international standards.

2. Assessment strategies—using assessment data to establish school improvement plans and to integrate assessment (including alternative assessment techniques) and instruction.

3. Instructional practices—promoting practices that engage learners, including problem based learning, hands-on science and activity centred teaching

4. Technology—using technology to support learning.
Motivating ‘improvement’ by disseminating information widely

Countries disseminate information widely to encourage stakeholders at all levels of the education system to reflect on the findings. Example 16 below describes TIMSS dissemination strategies in Sweden.

Comparing international expectations: ‘Benchmarking’ to confirm performance expectations

The achievements of students in other countries are used as a ‘benchmark’ for the development of a particular country’s performance standards or expectations.

For example, during the development of the Australian Year 3, 5, 7 and 9 numeracy ‘benchmarks’, the Commonwealth (Federal) Government commissioned a study to compare the draft performance standards with international achievement. Similarly, the Victorian (State) Board of Studies commissioned a 1999 study to compare the English, science and mathematics expectations contained in its revised Curriculum Standards Framework with international achievement data. In both instances the intention was to confirm the level of expectation with reference to actual student achievement.

Example 16 Motivating ‘improvement’ by disseminating information widely: Sweden

The TIMSS results were disseminated widely in Sweden to gain public attention and to encourage schools to make the greatest possible use of information. The National Agency for Education issued national reports at the same time as the international results were released. These reports were advertised on the Agency’s home page and sent to all participating schools.

Other reports including all released items with gender separated statistics for both international and Swedish students were sent to all schools in the country to encourage them to use data from TIMSS for comparison in their classrooms.
Examining factors likely to influence achievement to inform policy

In international studies, five broad clusters of factors likely to influence educational achievement usually are explored: home background, school characteristics, teacher characteristics, teaching conditions and practices, and student motivation.

Both bivariate (e.g. amount of homework assigned and achievement in particular subject) and multivariate analyses, which include assessments of the joint effects of background and potential causal factors on achievement, usually are undertaken. More recently, as the statistical techniques have become available, assessments of contextual and potential causal factors at multiple levels (e.g. class performance, school performance, country performance), also have been carried out.

What is the impact of these analyses?

Results of these kinds of analyses have led to extensive policy debate and to changes in educational policy. For example, Hungary’s participation in IEA studies has been credited with curriculum reform in reading—the finding that home factors accounted for more variance in student achievement than school factors credited with undermining the Marxist-Leninist curricular ideologies.

Some key research findings from 35 years of IEA research and their implications for policy are shown on page 25. Example 17, below, illustrates the impact of international achievement results on policy debate and policy decisions in Germany, Japan and Norway. Example 18 illustrates the impact in the United States.

---

Example 17 Influencing policy debate and policy decisions FIMS Germany, SIMS Japan and TIMSS Norway

In the Federal Republic of Germany, vigorous debate followed the publication of the mean scores in Science for the various Länder (states). Different Länder have different school structures and conclusions were drawn about the effectiveness of the different systems.

In Japan, the relatively higher achievement of students in mechanical operations than in higher mental processes required in problem solving became the focus of commissions composed of mathematics teachers and specialists and led to a curriculum revision.

In Norway, the TIMSS results have been used extensively by the Ministry of Education to inform policy on three fronts:

- Curricula for upper secondary education have been revised.
- New curriculum guidelines have been adopted in response to greater gender differences in achievement than expected.
- A revision of curriculum guidelines for teacher education has been undertaken in response to findings that primary teachers’ backgrounds in mathematics and science needed to be strengthened.
Example 18 Influencing policy debate — TIMSS United States

Three characteristics, in particular, of the US education system have been debated by independent researchers as a result of international comparisons drawn from TIMSS: curriculum, student instruction, and ongoing teacher professional development.32

Science curricula and textbooks are unfocused and contain too many topics. They are ‘a mile wide and an inch deep’. As a result, US teachers cover more topics, spend less time on topics, and provide more teaching activities per lesson than high achieving countries.33 Curricula and textbooks also emphasise routine procedures rather than challenging concepts.

Student instruction does not provide sufficient opportunity for student engagement. Teachers instruct students in a concept or skill, solve example problems with the class, and then have students practise on their own while the teacher assists individuals. In high achieving countries, teachers pose a complex thought provoking problem, students struggle with the problem, various students present ideas or solutions to the class, the class discusses the various solutions and methods, the teacher summarises the class’s conclusions, and students practise similar problems.

US teachers lack the long and carefully mentored introduction to teaching that Japanese and German teachers receive. Nor do they have the rich informal opportunities to learn from each other and to share questions about teaching related issues that Japanese teachers enjoy.34

Example 19 Within country analyses — US: state by state comparisons

Although in TIMSS the US as a whole ranked below the international mean in mathematics achievement of 13 year-olds, and although the US did not sample in order to report state differences, secondary analyses claim that one group of districts tied for first ranking.

The US Department of Education is encouraging school districts to use TIMSS at the district level to evaluate how local students are doing compared with their international peers.

Initiating further within-country analyses

Countries undertake their own analyses to expose within-country variations that are obscured by international reports which typically focus on the achievements of the country as a whole. Analyses of this kind are conducted by national agencies responsible for coordinating the international studies and by independent researchers. State by state comparisons are common (see examples 19 and 20) as are population sub-group analyses (example 21) and analyses by curriculum area (example 22).
Some key research findings from 35 years of IEA research

**Effects of curricula time**
Student achievement in Mathematics, French as a foreign language and Science is positively related to the time given to the study of the subject at school, both in comparisons across countries and between students within countries.

Implication: School curriculum design must take into account the level of achievement sought and relative emphasis given to each subject area.

**Homework**
Student achievement is related to the time spent on homework after other factors influencing achievement have been taken into account.

Implication: Careful consideration needs to be given to the amount of time assigned to homework—to support instruction without prejudicing motivation.

**Sex differences**
Sex differences in achievement are found to vary in size and direction across countries, school subjects and over time. Programs can be effective in reducing the gender gap in science achievement.

Implication: Programs to reduce the gender gap in achievement need to be maintained where they exist and introduced where they do not exist.

**Learning conditions**
Although the effects of home background variables are similar across subject areas, the effects of learning conditions in schools differ between subject areas, and in some subject areas are equivalent to, or greater in size than, the effects of the home.

Implication: Learning conditions in schools within a country should be raised and equalised.

**Opportunity to learn**
The average level of student achievement across countries is positively related to the opportunity that students had to learn the content of the items tested.

Implication: The content and skills considered important must be identified in the curriculum and students must be provided with the opportunity to learn that content and skills.
Example 21 Within country analyses— New Zealand: ethnic group analysis

Findings from previous New Zealand research have shown that students identifying themselves as Pakeha/European achieved significantly better, on average, than students from other ethnic groupings. 37 The New Zealand TIMSS report investigated the achievements of students from each Australian State. 36

### Mean mathematics scores (mean per cent) for form 2 students, by ethnic grouping

<table>
<thead>
<tr>
<th>Student Gender</th>
<th>Other ethnic groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pakeha/European</td>
</tr>
<tr>
<td>Girls</td>
<td>50</td>
</tr>
<tr>
<td>Boys</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
</tr>
</tbody>
</table>

Note: Standard errors (by gender) range from 1.0% to 1.3% for Pakeha/European and Maori students and 1.8% to 2.8% for Pacific Islands and Asian students.
A number of concerns have been raised about the quality of data on which international analyses are based, and on the ways in which data are reported. The main concerns are discussed below.

**Comparability of translated assessments**

**The problem:** Translation may produce items that differ in difficulty across languages and the validity of international comparisons of achievement depends on tasks maintaining their relative difficulties across countries. For example, an easy question in English may become a difficult question when translated into French. Differences in student performance on the tests then could be due to differences in the difficulty of the language tests rather than differences in the achievements of the groups.

**A solution:** This concern is addressed at the test development and data analysis stages of the studies. At the test development stage, in IEA studies for example, translations are made from the source language to the target language, then tests are back-translated to the source language. The original and the back-translations are then compared. PISA has introduced even more rigorous procedures. Two source versions, English and French, are used.

---

**Example 22 Within country analyses—South Africa: curriculum analysis**

The TIMSS South African Project Team analysed variations in achievement across the science content areas compared with the international average of achievement. The graph below illustrates the findings.

---

**Average per cent correct by the science content areas for Standard 6**

- **International Average**
- **SA Average**

<table>
<thead>
<tr>
<th>Science Content Area</th>
<th>International Average</th>
<th>SA Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earth Science</td>
<td>55</td>
<td>26</td>
</tr>
<tr>
<td>Life Science</td>
<td>59</td>
<td>27</td>
</tr>
<tr>
<td>Physics</td>
<td>55</td>
<td>27</td>
</tr>
<tr>
<td>Environmental Science Issues</td>
<td>51</td>
<td>26</td>
</tr>
<tr>
<td>Chemistry</td>
<td>53</td>
<td>26</td>
</tr>
</tbody>
</table>
Translations into the third language from both source versions are compared and reconciled. At the analysis stage, statistical techniques are used to detect items which, relative to other items, are unusually difficult or unusually easy in particular countries. These items are not included in the study.

**Match to curriculum (validity of achievement measures)**

**The problem:** Are the tests measures of curriculum knowledge? If they are, do the tests address each participating country’s curriculum? Have students in participating countries had the opportunity to learn what is being assessed?

Where tests aim to measure curriculum knowledge, the validity of comparisons depends on the degree to which the tests used in a particular study reflect the curriculum of each country in the study. Because there are large numbers of participating countries, there is compromise over the content and coverage of the tests, and tasks may match the curriculum of some countries better than others.

**A solution:** In order to address this concern, tests in IEA focus on a central or key body of knowledge and skills agreed by participating countries. In the more recent IEA studies, subject matter experts have met on several occasions to establish the content of the tests. This central body may not represent everything that is taught in a particular country but questions can be asked about how well this body of skills is being taught and learned in different countries.

It is interesting to note that the scoring of students on either items appropriate to their own curricula, or items appropriate to the curricula in different countries, did not substantially change a country's position in international standings in the TIMSS study.

In the PISA study, the aim is to go beyond the curriculum and to focus on more generalised skills in reading, mathematical and scientific literacy. The content domain of the test is defined by international expert opinion, not curriculum.

**Comparability of target populations**

**Alignment of populations**

**The problem:** The validity of comparisons may depend on students having the same exposure to schooling. Countries have different policies with regard to school age entry, grade repetition, promotion and graduation. Both length of time in school (grade level) and age can be expected to influence achievement, particularly at primary level. Countries also have differences in school retention rates and enrolment in particular courses.

**A solution:** Studies address age/grade concerns by clarifying the relationship between age and grade. For example, in TIMSS three groups of students were selected for the study: students midway through elementary (and grades containing most nine year olds), students midway through secondary (grades containing most 13 year olds), and students completing secondary (regardless of age).

**The problem:** If one aim of the study is to examine the effect of curriculum exposure, a difficulty remains. The first group studied, for example, includes both third and fourth graders in some countries and second and third in others, depending on which grades contained the greatest percentage of nine year olds. One indicator of curricular effect
may be the performance differences between students in the lower and upper grades assessed.

**A solution:** In order to address this problem, it has been suggested that studies provide scores that allow the separation of grade status from growth, rather than providing single mean scores of achievement. It has been suggested also that comparative tables provide information indicating retention rates and enrolment in particular courses where relevant, as was done when reporting the TIMSS results for the final year of secondary schooling.

**Exclusions**

**The problem:** Sometimes countries exclude sections of the defined target population, making comparisons less valid. For example, different definitions of disability (physical, emotional and intellectual) may result in the exclusion of different groups of students.

**A solution:** In order to address this problem, it has been suggested that there should be clear standards to regulate the implementation of population exclusions.

**Response rates**

**The problem:** The validity of comparisons also will depend on the degree to which the selected samples are representative samples. Although samples are defined to be representative, in practice, response rates vary. Comparisons between countries with widely differing response rates need to be treated cautiously.

**A solution:** Countries which do not meet sample requirements are excluded from reports or their results are reported with caveats as in TIMSS and PISA.

**Comparisons over time**

**The problem:** Concern has been raised about the limitations of international surveys for making causal inferences about patterns of student achievement over time. Concerns raised include: studies are not adequately linked, countries can change relative positions as a result of chance, a country can rise or fall in its relative position as a result of changes another country has made.

**A solution:** Studies need to be carefully designed to ensure that comparisons over time are justified. The extent and quality of links made in the achievement measures used are critical. PISA, for example, has been designed from the outset to give trend data for each country—each country can use its own previous performance as a basis for comparison.

**Misleading reporting - league tables**

**The problem:** There is concern that reporting overall (mean) results, particularly in overall league tables, encourages readers to draw inappropriate conclusions about the differences between countries and the strength of correlations between achievement and background variables.

**A solution:** Where league tables appear, they need to include qualifiers which assist readers to interpret the relative positions of countries.

Data need to be disaggregated in ways that provide information otherwise masked by overall means. For example, within country correlations between achievement and background variables need to be reported. Variables that are positively correlated with achievement across countries may be negatively correlated with achievement in any one country.
INTERNATIONAL ACHIEVEMENT DATA AND INFORMED DECISION MAKING

For a country, knowledge about student performance is of great importance, not in the form of ranking lists, but as structured information that can be used for internal improvement.

I see great challenges nationally in creating a working link from both the international and national level down to the classroom level. How can we—nationally and internationally—manage to make the results from international surveys to be useful for both teachers and schools in such a way that both instruction and practice are improved? 40

Data will be most useful to participating countries if they provide sufficient reliable information to inform debate and decision-making in a meaningful way. The debate about the quality of different school structures in different German Lander (example 17) was conducted in a context where the number of schools drawn for each Land was generally too small to permit inferences—a point overlooked in the debate. 41

The final section of this guide provides a checklist of considerations for ensuring that international achievement data provide participating countries with evidence for informed decision-making.

1 **Clear study aims and research questions**

The aims of the study and the research questions to be addressed need to be clearly stated. Ideally, the research questions should address important policy and theory-oriented issues for all participating countries.

2 **Study design that addresses research questions**

The study design needs to allow the research questions to be answered. For example, if data are to be used to track changes in achievement over time, then, where possible, each new international study needs to be linked to a previous international study.

3 **National research extensions**

Research extensions to the main study to meet national objectives should be considered; for example, the collection of data to facilitate linking the study to earlier studies allowing achievement changes over time to be investigated, or oversampling so that internal comparisons can be made. For example, Germany has a large sample for PISA to ensure that comparisons between Lander can be made.

**Will the data be valid and reliable?**

4 **Valid instruments**

If the instruments are to assess curriculum, then they need to address the common intended curriculum of participating countries. Ideally, the common elements will be important or ‘core’ elements of each country’s curriculum. The tasks will be fair (that is,
they will allow both male and female students from different ethnic, cultural, social, and economic backgrounds to demonstrate what they know and can do), translated appropriately, and meaningful to students. Ideally they also will be sensitive to instruction (that is, effective instruction will produce improvements in performance).

Questions in the background questionnaires that address policy issues outlined in the study will be included.

5 Reliable data and sampling procedures

Procedures need to be in place for collecting reliable (comparable) achievement data, including the trial testing of all instruments, and the development of uniform collection and recording procedures.

Sampling needs to be conducted so that the standard errors are acceptable in relation to the policy decisions that will be based on results. Recent standards require sampling precision at the same as, or better than, a simple random sample of 400 students for educational outcome measures.42

Will like comparisons be facilitated?

6 Target population

When interpreting comparisons, like needs to be compared with like. The extent of school and student-level exclusions needs to be detailed and the impact of these on comparisons of means and distributions assessed.

Will the reporting be comprehensive?

7 Accessible and useful reporting at all levels

When results are reported, policy issues need to be addressed directly and analyses described clearly. Arguments about the interpretation of analyses should be reported also. Reports should attempt to make the results useful for teachers and schools so that instruction and practice can be improved.

8 Comprehensive reporting

Analyses need to be comprehensive to reflect the varied nature and complexity of education systems.

League tables should be used with caution and interpreted carefully.

Is the complexity of the data considered before policy conclusions are drawn?

9 Responses to findings

When promoting policy change on the basis of study findings, it is important to consider the complexity of findings. For example, it could be misleading to argue on the basis of TIMSS results in favour of formal teaching, because it is a characteristic of Japanese education, without arguing for mixed ability groupings, which are also a characteristic of Japanese education.

Are there procedures to monitor the usefulness of the study findings?

10 Collect data on the impact of international studies

As part of the commitment to participating in international studies, research programs at a national level to monitor the usefulness/impact of findings should be considered.
USEFUL WEBSITES

Achieve
— a not for profit organisation created in 1996 by American governors and corporate leaders to provide advice and assistance to states on educational reform. Emphasises strengthening academic expectations by benchmarking students to the highest performing nations, promotes cross-state collaboration on curriculum, accountability and assessment.
www.achieve.org

Eisenhower National Clearinghouse
—a clearinghouse for mathematics and science education located at the Ohio State University and funded by the US Department of Education’s Office of Educational Research and Improvement. The clearinghouse has an extensive TIMSS site.
http://timss.enc.org/

First in the World Consortium
—a consortium of 17 school districts located in the north suburbs of Chicago and the Illinois Math and Science Academy. They aim to become first in the world in math and science achievement.
www.ncrel.org/ftw/homepage.htm

IEA International Association for the Evaluation of Educational Achievement
www.iea.nl/publications.htm

NCES National Center for Educational Statistics
US Dept of Education TIMSS site
Contains a comprehensive list of TIMSS publications.
http://nces.ed.gov/timss/publist.html

PISA 2000
— Programme for International Student Assessment
www.pisa.oecd.org

Regional Alliance Network
Resources on the web related to TIMSS
http://ra.terc.edu/alliance/TEMPLATE/regional_networks/CIA/Assessment/timss.cfm

Thomas B. Fordham Foundation
—a private foundation that supports research, publications, and action projects in elementary/secondary education reform at a national level and in the Dayton area.
www.edexcellence.net

TIMSS The Third International Mathematics and Science Study
This site contains international results, technical reports, international databases, achievement items, TIMSS publications, and links to related TIMSS sites.
http://timss.bc.edu
Endnotes


25 First in the World Consortium www.ncrel.org/fitw/homepage.htm


This guide provides policy makers with research-based information about international achievement studies.

Good decision-making at all levels of an education system is facilitated by easily accessible, relevant, and reliable information. Many indicators provide useful input to educational decision-making; but the most important indicators are those which address the central concern of education: the promotion of student learning.

Education systems monitor student learning—with the fundamental intention of promoting learning—by collecting, analysing and reporting student achievement data. Given that state, national and international achievement studies are both time consuming and expensive, it seems prudent to reflect on this effort:

What are the purposes of these studies?
How are data reported and used?
What concerns have been raised about these studies?
How can we ensure that data will provide evidence for informed decision-making?