

**THIRD INTERNATIONAL
MATHEMATICS AND SCIENCE STUDY
Second National Report
Part 1**



Sue Harris, Wendy Keys and Cres Fernandes

nfer

**THIRD INTERNATIONAL
MATHEMATICS AND SCIENCE STUDY**
Second National Report
Part 1

**Achievement in Mathematics and Science
at Age 9 in England**

Sue Harris, Wendy Keys and Cres Fernandes

nfer

Published in June 1997
by the National Foundation for Educational Research,
The Mere, Upton Park, Slough, Berkshire SL1 2DQ

© National Foundation for Educational Research 1997
Registered Charity No. 313392
ISBN 0 7005 1464 3

CONTENTS

ACKNOWLEDGEMENTS	i
SUMMARY	ii
CHAPTER 1 Introduction to the study	1
1.1 Preface	2
1.2 Design and administration of TIMSS	3
1.2.1 Age groups	3
1.2.2 The schools and pupils	3
1.2.3 Administration	4
1.2.4 Test-curriculum Matching Analysis	4
1.3 The tests and questionnaires	5
1.3.1 The tests of mathematics and science	5
1.3.2 The questionnaires	6
1.4 Interpreting the results	7
CHAPTER 2 The pupils' overall performance on the mathematics tests	8
2.1 Preface	9
2.2 Overall mean scores on the mathematics tests	10
2.2.1 Overall mean mathematics scores: comparisons between England and continental Western European countries	11
2.2.2 Overall mean mathematics scores: comparisons between England and other English-speaking countries	12
2.2.3 Overall mean mathematics scores: comparisons between England and Eastern European countries	13
2.2.4 Overall mean mathematics scores: comparisons between England and Asian Pacific Rim countries	13
2.3 Comparisons of nine-year-olds	14
2.4 Test-curriculum Matching Analysis	16
2.5 Comparisons between nine countries: differences in overall mean mathematics scores between international third and fourth grades (Years 4 and 5 in England)	19
2.6 International marker levels	21
2.7 Comparisons between nine countries: gender differences in overall mean mathematics scores	24
CHAPTER 3 The pupils' performance on the mathematics topics	29
3.1 Preface	30
3.2 Whole numbers	32
3.2.1 Comparisons of total topic scores: <i>Whole numbers</i>	32
3.2.2 Differences in mean scores between international third grade and international fourth grade (Years 4 and 5 in England): <i>Whole numbers</i>	33
3.2.3 Examples of items for <i>Whole numbers</i>	34

3.3	Fractions and proportionality	37
3.3.1	Comparisons of total topic scores: <i>Fractions and proportionality</i>	37
3.3.2	Differences in mean scores between international third grade and international fourth grade (Years 4 and 5 in England): <i>Fractions and proportionality</i>	38
3.3.3	Examples of items for <i>Fractions and proportionality</i>	38
3.4	Measurement, estimation and number sense	42
3.4.1	Comparisons of total topic scores: <i>Measurement, estimation and number sense</i>	42
3.4.2	Differences in mean scores between international third grade and international fourth grade (Years 4 and 5 in England): <i>Measurement, estimation and number sense</i>	43
3.4.3	Examples of items for <i>Measurement, estimation and number sense</i>	43
3.5	Data representation, analysis and probability	47
3.5.1	Comparisons of total topic scores: <i>Data representation, analysis and probability</i>	47
3.5.2	Differences in mean scores between international third grade and international fourth grade (Years 4 and 5 in England): <i>Data representation, analysis and probability</i>	48
3.5.3	Examples of items for <i>Data representation, analysis and probability</i>	48
3.6	Geometry	53
3.6.1	Comparisons of total topic scores: <i>Geometry</i>	53
3.6.2	Differences in mean scores between international third grade and international fourth grade (Years 4 and 5 in England): <i>Geometry</i>	54
3.6.3	Examples of items for <i>Geometry</i>	54
3.7	Patterns, relations and functions	57
3.7.1	Comparisons of total topic scores: <i>Patterns, relations and functions</i>	57
3.7.2	Differences in mean scores between international third grade and international fourth grade (Years 4 and 5 in England): <i>Patterns, relations and functions</i>	58
3.7.3	Examples of items for <i>Patterns, relations and functions</i>	58
3.8	Overview of results for each mathematics reporting category	62
3.9	Gender differences	63
CHAPTER 4	The pupils' overall performance on the science tests	65
4.1	Preface	66
4.2	Overall mean scores on the science tests	67
4.2.1	Overall mean science scores: comparisons between England and continental Western European countries	68
4.2.2	Overall mean science scores: comparisons between England and other English-speaking countries	69
4.2.3	Overall mean science scores: comparisons between England and Eastern European countries	70
4.2.4	Overall mean science scores: comparisons between England and Asian Pacific Rim countries	70
4.3	Comparisons of nine-year-olds	71

4.4	Test-curriculum Matching Analysis	72
4.5	Comparisons between nine countries: differences in overall mean science scores between international third and fourth grades (Years 4 and 5 in England)	76
4.6	International marker levels	77
4.7	Comparisons between nine countries: gender differences in overall mean science scores	79
CHAPTER 5 The pupils' performance on the science topics		85
5.1	Preface	86
5.2	Earth science	88
5.2.1	Comparisons of total topic scores: <i>Earth science</i>	88
5.2.2	Differences in mean scores between international third grade and international fourth grade (Years 4 and 5 in England): <i>Earth science</i>	89
5.2.3	Examples of items for <i>Earth science</i>	89
5.3	Life science	93
5.3.1	Comparisons of total topic scores: <i>Life science</i>	93
5.3.2	Differences in mean scores between international third grade and international fourth grade (Years 4 and 5 in England): <i>Life science</i>	94
5.3.3	Examples of items for <i>Life science</i>	94
5.4	Physical science	99
5.4.1	Comparisons of total topic scores: <i>Physical science</i>	99
5.4.2	Differences in mean scores between international third grade and international fourth grade (Years 4 and 5 in England): <i>Physical science</i>	100
5.4.3	Examples of items for <i>Physical science</i>	100
5.5	Science and environment	104
5.5.1	Comparisons of total topic scores: <i>Science and environment</i>	104
5.5.2	Differences in mean scores between international third grade and international fourth grade (Years 4 and 5 in England): <i>Science and environment</i>	104
5.5.3	Examples of items for <i>Science and environment</i>	105
5.6	Overview of results for each science reporting category	109
5.7	Gender differences	110
CHAPTER 6 Preview of further analyses		112
REFERENCES		113
APPENDIX I National steering committee		115
APPENDIX II Countries taking part in different components of TIMSS		116
ANNEX A Test-curriculum Matching Analysis results: mathematics		118
ANNEX B Test-curriculum Matching Analysis results: science		120

ACKNOWLEDGEMENTS

This study could not have been completed without the help of a large number of people and we would like to express our thanks to:

- ◆ the pupils, teachers and headteachers in the schools which took part in the study
- ◆ members of the National Steering Group for their help and advice at all stages of the study
- ◆ Seamus Hegarty, Director of NFER, for his support throughout the project
- ◆ Professor Al Beaton, the Study Director, and his colleagues at the TIMSS International Study Center, Boston College, Massachusetts, USA; and colleagues in other countries who contributed to the international effort
- ◆ Derek Foxman, the Project Consultant, and Professor Geoffrey Howson for their constructive comments and suggestions
- ◆ the mathematics and science experts, who took part in the item review and Test-curriculum Matching Analysis panels
- ◆ Keith Mason and Steve Sizmur for reading and commenting on the draft report
- ◆ Joan Howell, who organised the printing of the tests and questionnaires and the report
- ◆ Anne Milne and Janet Horgan in NFER's Field Research Services Department for organising the administration of the study
- ◆ Claudia Davis and David Harris, who helped to coordinate the test marking exercise; and the team of mathematics and science practitioners, who marked the pupils' responses to the tests
- ◆ Pauline Pearce and Alison Bannerman for providing valuable secretarial support throughout the project
- ◆ Mary Hargreaves, who prepared the layout, Tim Wright, who designed the cover, and Enver Carim, who coordinated the publication of the report.

SUMMARY

1. Introduction

England is one of 26 countries that have surveyed nine-year-old pupils and their teachers as part of the Third International Mathematics and Science Study (TIMSS). The study is organised by the International Association for the Evaluation of Educational Achievement (IEA), and has included surveys at both primary- and secondary-school level. TIMSS in England is funded by the Department for Education and Employment (DfEE) and carried out by the National Foundation for Educational Research (NFER).

The Second National Report, Part 1, compares the performance of nine-year-old pupils in England on the TIMSS mathematics and science tests with that of pupils of a similar age in other countries.¹ The First National Report, Part 1 (Keys *et al.*, 1996a) made similar comparisons between the performance of 13-year-old students in England and their counterparts in other countries; a second part to each national report (forthcoming) will examine the relationships between home, student and school background factors and performance on the TIMSS mathematics and science tests.

2. Main findings

Pupils in England achieved relatively high mean scores in science but relatively low mean scores in mathematics. This pattern of achievement was also found in the results for 13-year-olds in England which were reported in Keys *et al.* (1996a).

The mean overall mathematics score for pupils in Year 5² in England was:

- ◆ significantly **lower** than those of pupils in about half of the countries that took part in the TIMSS survey of nine-year-olds
- ◆ not significantly different from those in about a quarter of the countries
- ◆ significantly **higher** than those of pupils in about a quarter of the countries.

The mean overall science score of pupils in Year 5 in England was:

- ◆ significantly **lower** than those of pupils in only three other countries in the TIMSS survey of nine-year-olds
- ◆ not significantly different from those of pupils in nine other countries

¹ *With the exception of the sections concerning overall comparisons, the Second National Report, Part 1, focuses mainly on the mathematics and science performance of pupils in England and eight other countries: Canada, Hungary, Japan, the Netherlands, Norway, Singapore, Scotland and the United States.*

² *The results were similar for Year 4.*

- ◆ significantly **higher** than those of pupils in half of the countries that took part.

3. Other important findings

Mathematics

- ◆ Pupils in England scored above the international means for Years 4 and 5 in: *Geometry and Data representation, analysis and probability*.
- ◆ Pupils in England scored below the international means for Years 4 and 5 in: *Whole numbers; Fractions and proportionality; Measurement, estimation and number sense and Patterns, relations and functions*.
- ◆ These results are broadly in line with those of 13-year-old students in England, whose mean scores were above the international means for Years 8 and 9 in one mathematics content area (*Data representation, analysis and probability*), but below the international means for the other five content areas for 13-year-olds.
- ◆ There were no significant gender differences in overall mean mathematics scores in either Year 4 or Year 5, although the boys' mean scores were significantly higher than the girls' in both years for *Whole numbers and Measurement, estimation and number sense*; and in Year 4 only for *Patterns, relations and functions*.

Science

- ◆ Thirteen per cent of the pupils in England were in the international top ten per cent in both year groups.
- ◆ Pupils in England scored above the international means for Years 4 and 5 in all four science content areas: *Earth science; Life science; Physical science and Science and environment*.
- ◆ These results are consistent with the results for 13-year-olds in England, whose mean scores were above the international means for Years 8 and 9 for all five science content areas.
- ◆ There were no significant gender differences in overall mean science scores in either Year 4 or Year 5, although the boys' mean scores were significantly higher than the girls' in both years for *Earth science*.

CHAPTER 1

Introduction to the study

The structure of this report

- ◆ **This chapter provides a brief description of the design and administration of the study.**
- ◆ **Chapter 2 describes the pupils' overall performance on the mathematics test.**
- ◆ **Chapter 3 focuses on their performance in different mathematics content areas.**
- ◆ **Chapters 4 and 5 provide similar information about pupils' performance on the science tests.**
- ◆ **The final chapter previews the topics to be covered in Part 2 of this report and in subsequent national reports on TIMSS.**
- ◆ **Appendix I lists the members of the TIMSS National Steering Group for England.**
- ◆ **Appendix II provides information on the countries participating in different components of TIMSS.**

A companion volume (Keys *et al.*, 1996b) contains additional appendices giving more detailed information on the design and administration of the study.

Part 2 of this report (forthcoming) will compare the responses to the TIMSS questionnaires of pupils, teachers and headteachers in England with those of their counterparts in other countries.

1.1 Preface

England is one of 26 countries¹ taking part in the survey of nine-year-olds which forms part of the Third International Mathematics and Science Study (TIMSS), a large-scale international comparative study of educational performance. TIMSS is organised by the International Association for the Evaluation of Educational Achievement (IEA). The study in England is funded by the Department for Education and Employment (DfEE) and carried out by the National Foundation for Educational Research (NFER).

This is the second national report on TIMSS. Its main purpose is to compare the performance on the TIMSS mathematics and science tests of nine-year-old pupils in England with the performance of pupils of the same age in other countries. Most of the information contained in this report has been drawn from the international reports on TIMSS (Mullis *et al.*, 1997 and Martin *et al.*, 1997), which provide a broad brush picture of comparisons between the mathematics and science performance of nine-year-olds (TIMSS Population 1) in 26 educational systems. Where appropriate, comparisons have been made with the results of previous large-scale comparisons of achievement, the most important of which are listed in Chapter 2 (mathematics) and Chapter 4 (science).

In order to provide clear and coherent comparisons it was decided to focus many of the comparisons in this report on the following nine countries:

- | | |
|---------------|-------------|
| • Canada | • Norway |
| • England | • Scotland |
| • Hungary | • Singapore |
| • Japan | • USA |
| • Netherlands | |

The countries selected represent the four main groups of countries taking part in TIMSS: Western European countries; other English-speaking countries; Eastern European countries; and countries from the Asian Pacific Rim. These countries' mean scores on the TIMSS tests illustrate the wide range of scores achieved by the industrialised nations taking part in TIMSS.

¹ The countries taking part in the different components of TIMSS are listed in Appendix II.

1.2 Design and administration of TIMSS

The international research focused on three different stages of education: upper primary (mainly nine-year-olds), lower secondary (mainly 13-year-olds) and pupils in their final year of schooling (mainly 17-year-olds). England participated in the survey of two age groups: nine-year-olds (Years 4 and 5 in England) and 13-year-olds (Years 8 and 9 in England) and in the Performance Assessment (practical activities) for 13-year-olds. A summary of the design and administration of the TIMSS survey of nine-year-olds (TIMSS Population 1), which forms the subject of this report, is given below. Appendix III, which is included in the companion volume to this study (Keys *et al.*, 1996b) gives full details of the design and administration of TIMSS in England.

1.2.1 Age groups

The TIMSS Population 1 sample was drawn from pupils in all schools who were in international third and fourth grades (equivalent to Years 4 and 5 in England).² At the time of testing (early March 1995), the age of the Population 1 pupils taking part in the study in England ranged from eight years seven months to ten years six months.

1.2.2 The schools and pupils

The samples of schools and pupils taking part in TIMSS in each country were required to conform to rigorous procedures. Each country's sampling plans and all details of samples had to be documented and approved by an independent sampling referee. The sample design in England for Population 1 involved the selection of an initial sample of 150 primary or middle schools and two matching sets of 150 replacement schools which were selected at the same time. All three samples were representative of schools in England containing nine-year-olds in terms of type of school, size and background characteristics. Each school in the initial sample was matched with two schools (one in each of the replacement samples) which were as similar as possible to the initial school in terms of type of school, size and background characteristics. If a school in the initial sample declined to participate, it was replaced by a matched school from one of the replacement samples. This ensured that the final sample of schools taking part in the TIMSS Population 1 study remained representative of schools in England irrespective of whether any individual school was from the initial or a replacement sample.³ Full details are given in Appendices III and V (Keys *et al.*, 1996b).

² In the tables in this report, year groups are identified as 'international third grade (Year 4 in England)' and 'international fourth grade (Year 5 in England)' as appropriate. In the text, year groups are normally referred to in this way at the beginning of each main section, and as Year 4 or Year 5, or as lower or upper grades, as appropriate, thereafter.

³ The sample of schools from England met the TIMSS sampling requirements, as did those in six of the other eight countries on which this report focuses. The participation rate in the Netherlands did not satisfy the TIMSS Guidelines for Sample Participation Rates (Martin *et al.*, 1997; Mullis *et al.*, 1997). Hungary used unapproved sampling procedures at the classroom level.

Once a school agreed to participate in the Population 1 study, one intact class was randomly selected by NFER from each year group (Year 4 and Year 5). (Information about classes selected from schools with mixed-age classes, and from schools with only one of the target year groups is given in Appendix V (Keys *et al.*, 1996b)). A total of 134 schools took part in the study (96 first-choice schools and 38 replacement schools). The response rate for Population 1 was 88 per cent including replacement schools (64 per cent from first-choice schools). Tests and questionnaires were completed by 6,142 pupils.⁴

1.2.3 Administration

The tests and questionnaires were administered in schools by teachers in early March 1995. The tests were subject to strict time limits which were common to all participating countries. These time limits were set so as to allow the majority of pupils to finish the test. The pupil questionnaires had no time limit, but teachers were advised to allow sufficient time for all pupils to complete the whole questionnaire (roughly 20 minutes). Teachers and headteachers were also asked to fill in questionnaires.

To ensure that the survey was administered under the same conditions in all countries, the administration manual (which included a script for the teacher administering the tests and questionnaires and details of time limits) followed the same general pattern in all countries. Each country's manuals had to be agreed with the international co-ordinator. In England, the tests and questionnaires were returned to NFER for marking, coding and data processing. In addition, the International Study Center set up a quality control programme, involving visits by independent scrutineers to a sub-set of schools, to monitor the conduct of the study in each country.

1.2.4 Test-curriculum Matching Analysis

All countries taking part in TIMSS used the same tests of mathematics and science. Although this ensured that comparisons could be made between different countries in terms of attainment on the same tests, it does not recognise the fact that the curricula for mathematics and science vary from one country to another. The Test-curriculum Matching Analysis was set up to collect information about the relevance of the TIMSS test items to the mathematics and science curricula in countries taking part in the study. In order to identify the relevant sub-sets of items, experts in each country had to consider whether or not each item was covered in the 'intended curriculum' for each of the grades (year groups) separately. In England this was interpreted as:

- ◆ whether or not the topic of the item formed part of the National Curriculum for the majority of pupils (at least 50 per cent) in each year group

⁴ Details of sample sizes and participation rates in other countries are given in the international reports (Martin *et al.*, 1997; Mullis *et al.*, 1997).

- ◆ whether or not the content of the item was likely to have been covered by pupils at the time of testing (in England, early March 1995).

Analyses of pupils' performance on the sub-sets of items which matched the National Curriculum in England are provided in Chapters 2 and 4.

1.3 The tests and questionnaires

1.3.1 The tests of mathematics and science

The TIMSS curriculum frameworks for mathematics and science were developed from analyses of the science and mathematics curricula in participating countries (Robitaille, 1993). These frameworks provided a structure which ensured that the tests were as relevant as possible to the curricula of the countries taking part in the study. Categories in the TIMSS mathematics curriculum framework were grouped together to form the mathematics reporting categories used in Chapter 3.

Mathematics	• Whole numbers	(25 items)
	• Fractions and proportionality	(21 items)
	• Measurement, estimation and number sense	(20 items)
	• Data representation, analysis and probability	(12 items)
	• Geometry	(14 items)
	• Patterns, relations and functions	(10 items)

Similarly categories from the science curriculum framework were grouped to form the science reporting categories shown below.

Science	• Earth science	(17 items)
	• Life science	(41 items)
	• Physical science	(30 items)
	• Science and environment ⁵	(9 items)

⁵ Topics covered in this category include: environmental and resource issues; the nature of scientific knowledge; and the influence of science and technology on society. Throughout this report the category is referred to as 'Science and environment' as was done in the national report on 13-year-olds. However, in the international reports for nine-year-olds (Martin et al., 1997) and 13-year-olds (Beaton et al., 1996a), the category is referred to as 'Environmental issues and the nature of science'.

A relatively large number of items was required in order to achieve broad content coverage of curricula. There were 199 test items for Population 1 (102 focusing on mathematics and 97 on science). About three-quarters of both the mathematics items and the science items were multiple-choice; the remainder were either short answer (requiring the pupil to write in a single word or number) or extended response items. In order to avoid placing excessive burdens on individual pupils, eight different versions of the test were prepared for each population, each containing a different combination of items. Some items were included in every version of the test; the remaining items appeared in one, two, three or four version(s) of the test. Each individual pupil completed one version of the test only consisting of about 65 items. Only three or four pupils in each class attempted the same version of the test. Full details of these procedures are given in Appendix IV (Keys *et al.*, 1996b).

Each version of the test was divided into two booklets, each of which contained a mixture of multiple-choice, short answer and extended response items. Each test booklet included both mathematics and science items arranged in clusters. However, a separate mathematics and science score was calculated for each pupil. There were two testing sessions, giving a total testing time of 54 minutes.

1.3.2 The questionnaires

The questionnaire for pupils was designed to obtain information from pupils about:

- ◆ themselves
- ◆ their home background
- ◆ their attitudes towards mathematics and science
- ◆ their perceptions of lessons in these subjects.

The questionnaire for mathematics and science teachers sought information on:

- ◆ teachers' education, training and experience
- ◆ how they divided their time between teaching and teaching-related activities
- ◆ teaching approaches
- ◆ the resources available to them
- ◆ their attitudes towards mathematics or science as subjects
- ◆ their views on teaching and learning mathematics or science
- ◆ their views on teaching as a career.

The questionnaire for schools sought:

- ◆ general background information on the schools
- ◆ details of organisational features relating to the teaching of mathematics and science
- ◆ information on teaching resources and teaching-time devoted to these subjects.

It will be seen then that the TIMSS research encompassed much more than the mere testing of pupil attainment. Indeed, to comprehend the data on pupil attainment, it is essential to take into account fully such other factors as, for example, timetabled hours allocated to the study of mathematics and science. These issues will be examined in Part 2 of this report, which will be published later this year.

1.4 Interpreting the results

Two important points should be taken into account when interpreting the results of this study.

Firstly, comparisons between countries have been made in terms of the mathematics and science performance of pupils in the two adjacent grades (year groups) which contain the majority of nine-year-olds in each country (equivalent to Years 4 and 5 in England). Children start school at different ages in different countries. Thus, the amount of formal schooling received by pupils in grades equivalent to Year 4 in England prior to their taking the TIMSS tests ranged from two years in Norway to four years in England and Scotland (and from three to five years for those in grades equivalent to Year 5). It is particularly important to take these differences into account when making comparisons between the performance of pupils in grades equivalent to Year 4.

Secondly, in most of the countries taking part in the study, mathematics and science are taught to nine-year-olds by their own class teachers, rather than by specialist subject teachers as in the secondary school. This means that the majority of nine-year-olds taking the TIMSS test were taught both subjects by the same teacher. This fact should be borne in mind when making comparisons between a country's relative position in terms of the mathematics performance of pupils at this level and its relative position in science.

CHAPTER 2

The pupils' overall performance on the mathematics tests

- ◆ The mean overall mathematics scores of pupils in England were:
 - significantly *lower* than those of pupils in about half of the 26 countries taking part in TIMSS
 - not significantly different from those of pupils in about a quarter of the countries
 - significantly *higher* than the mean scores of pupils in about a quarter of the countries.
- ◆ The relative position of England in terms of the mathematics scores of nine-year-olds was similar to its position based on the mathematics scores of 13-year-olds.
- ◆ The relative position of England in mathematics appears to have deteriorated slightly compared with some countries (for example, United States, Canada and Ireland) since previous comparative studies were carried out.
- ◆ There were no significant gender differences in overall mean mathematics scores in either Year 4 or Year 5 in England, in common with the majority of other countries.

2.1 Preface

This chapter compares the mean scores on the TIMSS mathematics tests of pupils in England with pupils of the same age in other countries. Separate mean scores are provided for the two year groups tested (Years 4 and 5 in England, which are equivalent to international third and fourth grades) and for nine-year-olds (i.e. pupils aged 9.0–9.11 years at the time of testing). Comparisons with other countries are made in terms of:

- ◆ pupils' overall mean¹ scores on the TIMSS mathematics test
- ◆ pupils' mean scores on the sub-set of items identified by mathematics experts in England as being covered within the mathematics curriculum for each year group (the Test-curriculum Matching Analysis)
- ◆ the ranges of scores in different countries (percentages of pupils reaching the following international marker levels: top ten per cent, top 25 per cent, top half and bottom 25 per cent of the international distribution)
- ◆ differences in mean mathematics scores of pupils in the international third and fourth grades (Years 4 and 5 in England)
- ◆ differences in mean mathematics scores of boys and girls.

Where appropriate, pupils' results in TIMSS are compared with the results of the only previous large-scale international study of mathematics achievement at age nine, as detailed below; within this chapter references to this earlier study will be abbreviated to IAEP2.

Date of testing	Study	References to reports
1991	The second study carried out by the International Association for the Evaluation of Educational Progress (IAEP2)	Lapointe <i>et al.</i> , (1992a) Foxman (1992)

When drawing comparisons between the results of TIMSS and IAEP2, the following points should be borne in mind:

- ◆ 26 countries participated in the TIMSS survey of nine-year-olds as compared with 14 countries in IAEP2
- ◆ only ten countries (including England) took part in both studies
- ◆ some of the countries that participated in TIMSS have not taken part in any previous international comparative studies.

¹ Throughout this report the term 'mean' has been used when referring to national and international results, whereas the international report uses the term 'average'.

2.2 Overall mean scores on the mathematics tests

In this chapter, the overall mean mathematics scores of pupils in England are compared with those of pupils of similar ages in four distinct groups of other countries, as follows:

continental Western European countries	Table 2.2.1
English-speaking countries	Table 2.2.2
Eastern European countries	Table 2.2.3
Asian Pacific Rim countries	Table 2.2.4

The information presented in the above tables is derived from Tables 1.1 (Mathematics) and 1.2 (Mathematics) and Figures 1.1 (Mathematics) and 1.2 (Mathematics) from the international report on TIMSS (Population 1) (Mullis *et al.*, 1997); these tables and figures, which give detailed information on pupils' mean mathematics scores in all the countries taking part in TIMSS together with the average age and numbers of years of formal schooling for each sample, have been reproduced at the end of this chapter. The international tables and figures include all the countries that participated in TIMSS. In the international tables, countries which did not meet the international sampling requirements are shown 'below the line', i.e. the mean scores of their pupils are not included in the rank orders of results. In the figures, however, these countries are included in the rank ordering of results, but are listed in italics to differentiate them. The reasons why each country did not meet the sampling requirements are explained in the footnotes to the international tables.

Overall mean mathematics scores (and the overall mean science scores given in Chapter 4) are expressed in terms of a scale with a mean of 500 and a standard deviation² of about 100, based on the combined performance of pupils in both grades (year groups). The international means (i.e. based on the performance of pupils in all participating countries) were 470 for the lower grade and 529 for the upper grade.

The relative position of England in terms of the mathematics scores of nine-year-olds was similar to its position based on the scores of 13-year-olds (Keys *et al.*, 1996a).

² This means that about two-thirds of pupils in both age groups in all participating countries combined achieved scores between 400-600 and about 95 per cent achieved scores between 300-700.

2.2.1 Overall mean mathematics scores: comparisons between England and continental Western European Countries

The overall mean mathematics scores of pupils in England were similar to or significantly higher than the mean scores of pupils in four of the six Western European countries that took part in the TIMSS survey of nine-year-olds (Greece, Iceland, Norway and Portugal), as shown in Table 2.2.1. At the same time, however, the overall mean mathematics scores of pupils in Austria and the Netherlands were significantly higher than those of pupils in England in both the international third grade and the international fourth grade (equivalent to Years 4 and 5 in England). The younger pupils in Austria and the Netherlands outperformed pupils in Year 4 in England by 31 scale points and 37 scale points respectively; for the older pupils, the performance of pupils in Austria and the Netherlands was higher than that of pupils in Year 5 in England by 46 scale points and 64 scale points respectively.

Table 2.2.1 Overall mean mathematics scores: comparisons between England and continental Western European countries

Compared with England	International third grade (Year 4 in England)	International fourth grade (Year 5 in England)
Significantly higher	Austria; Netherlands	Austria; Netherlands
Not significantly different	None	Norway
Significantly lower	Greece; Iceland; Norway; Portugal	Greece; Iceland; Portugal

Source: Figures 1.1 and 1.2 (Mullis et al., 1997)

Comparisons with previous studies

Only one of the continental Western European countries taking part in TIMSS at the Population 1 level took part in IAEP2. Compared with Portugal, the relative position of England appears to have improved.

- ◆ In IAEP2, pupils in Portugal scored at approximately the same level as those in England, whereas in TIMSS pupils in England had mean mathematics scores which were significantly higher than those of pupils in Portugal.

2.2.2 Overall mean mathematics scores: comparisons between England and other English-speaking countries

The performance of pupils in four of the other English-speaking countries (Australia, Canada, Ireland and the United States) was significantly higher than that of pupils in England in both the lower and the upper grades, as shown in Table 2.2.2. The overall mean mathematics scores achieved by pupils in these countries exceeded those of pupils in Year 4 in England by 13–27 scale points and those of pupils in Year 5 by 19–37 scale points. There were no significant differences between the performance of pupils in England and pupils in Scotland (both year groups) and New Zealand (upper grade only).

Table 2.2.2 Overall mean mathematics scores: comparisons between England and other English-speaking countries

Compared with England	International third grade (Year 4 in England)	International fourth grade (Year 5 in England)
Significantly higher	Australia; Canada; Ireland; United States	Australia; Canada; Ireland; United States
Not significantly different	Scotland	New Zealand; Scotland
Significantly lower	New Zealand	None

Source: Figures 1.1 and 1.2 (Mullis et al., 1997)

Comparisons with previous studies

Comparisons suggest that, compared with Scotland, the relative performance of pupils in England may have improved since IAEP2 was carried out.

- ◆ in IAEP2, pupils in Scotland achieved a mean score six percentage points above that of pupils in England, whereas in TIMSS, pupils in Scotland achieved mean mathematics scores which were similar to those of pupils in England.

However, the relative mathematics performance of pupils in England appears to have deteriorated compared with that of pupils in several other English-speaking countries.

- ◆ In IAEP2, pupils in Canada, Ireland and the United States performed at about the same level as pupils in England, whereas in TIMSS pupils in these countries outperformed those in England.

2.2.3 Overall mean mathematics scores: comparisons between England and Eastern European countries

In three of the four Eastern European countries pupils achieved significantly higher overall mean mathematics scores than pupils in Years 4 and 5 in England (see Table 2.2.3). At the lower grade, pupils' performance in the Czech Republic, Hungary and Slovenia was 20–41 scale points higher than the performance of pupils in Year 4 in England; at the upper grade, pupils in these same three countries outperformed pupils in Year 5 in England by 35–54 scale points. Only pupils in Latvia showed no significant difference in achievement as compared with pupils in England (both years tested).

Table 2.2.3 Overall mean mathematics scores: *comparisons between England and Eastern European countries*

Compared with England	International third grade (Year 4 in England)	International fourth grade (Year 5 in England)
Significantly higher	Czech Republic; Hungary; Slovenia	Czech Republic; Hungary; Slovenia
Not significantly different	Latvia	Latvia
Significantly lower	<i>None</i>	<i>None</i>

Source: Figures 1.1 and 1.2 (Mullis et al., 1997)

Comparisons with previous studies

Two of the Eastern European countries taking part in TIMSS at the Population 1 level also took part in IAEP2.

- ◆ As in IAEP2, pupils in Hungary achieved higher mean mathematics scores than pupils in England in TIMSS.
- ◆ In IAEP2, the mean mathematics score of pupils in Slovenia was four percentage points below that of pupils in England, whereas in TIMSS, their mean mathematics scores were significantly higher than those of pupils in England.

2.2.4 Overall mean mathematics scores: comparisons between England and Asian Pacific Rim countries

Pupils in four of the Asian Pacific Rim countries (Hong Kong; Japan; Korea and Singapore) achieved significantly higher overall mean mathematics scores than pupils in both Years 4 and 5 in England, as shown in Table 2.2.4. The differences in performance between pupils in any one of these countries and pupils in England (both year groups) were greater than those for any of the other countries discussed in the preceding sections. At the lower grade, the higher scores achieved by pupils in these

four countries exceeded those of pupils in Year 4 in England by 68–105 scale points, and at the upper grade, the differences ranged from 74–112 scale points.

Table 2.2.4 Overall mean mathematics scores: comparisons between England and the Asian Pacific Rim countries

Compared with England	International third grade (Year 4 in England)	International fourth grade (Year 5 in England)
Significantly higher	Hong Kong; Japan; Korea; Singapore	Hong Kong; Japan; Korea; Singapore
Not significantly different	Thailand	None
Significantly lower	None	Thailand

Source: Figures 1.1 and 1.2 (Mullis et al., 1997)

Comparisons with previous studies

Only one of the Asian Pacific Rim countries taking part in TIMSS, Korea, also took part in IAEP2.

- ◆ In IAEP2, as in TIMSS, the mean mathematics score of pupils in Korea was substantially higher than that of pupils in England.

2.3 Comparisons of nine-year-olds

The preceding sections have compared the performance of pupils in England with that of pupils in the other countries that participated in TIMSS. In the rest of this chapter, comparisons will be limited to the nine countries identified in Chapter 1.

One of the aims of TIMSS was to compare the performance of nine-year-old pupils in the countries taking part, irrespective of their grade (year group) within their respective countries. In some countries, including England, the two grades (year groups) tested included almost all nine-year-olds, whereas in other countries substantial percentages of nine-year-olds were in other grades, most frequently in the grade below (equivalent to Year 3 in England). Countries with fewer than 75 per cent of nine-year-olds in the two grades tested were excluded from this analysis. However, for countries where the grades tested contained at least 75 per cent of nine-

year-olds, TIMSS estimated the median performance of nine-year-olds. The median is the point on the mathematics scale which separates the higher-performing 50 per cent of pupils from the lower-performing 50 per cent; like the mean, it provides a useful summary statistic which can be used to compare performance across countries. It is used in Table 2.3 because it can reliably be estimated even when scores from some pupils within the population are not available³ (i.e. nine-year-olds outside the two grades tested). Table 2.3 shows how the rank order of median scores of nine-year-olds compared with the median score for the international third grade (Year 4 in England) for England and the eight other countries identified in Chapter 1.

Table 2.3 Median mathematics achievement in ten countries: the scores of nine-year-olds compared with the scores of pupils in international third grade (Year 4 in England)

COUNTRY	nine-year-olds		International third grade (Year 4 in England)		Mean age (years)
	Median score	Rank	Median score	Rank	
Singapore	569	1	556	1	9.3
Japan	544	2	538	2	9.4
Netherlands	512	3	492	3	9.3
Canada	504	4	472	6	9.1
United States	503	5	479	4	9.2
Scotland	502	6	456	7	8.7
Hungary	491	7	476	5	9.4
England	476	8	455	8	9.1
Norway	473	9	422	9	8.8

Source: Tables 1.2 and 1.8 and C.2 (Mullis et al., 1997)

It is apparent from Table 2.3 that although there are minor differences in the two rank orders, the countries ranked first, second and third are the same in both cases. In both rankings, the performance of pupils in England was placed within the lowest third amongst those countries selected for comparison.

³ 'Because TIMSS sampled students in the two adjacent grades with the most 9-year-olds within a country, it was possible to estimate the median for the 9-year-old students when the two tested grades included at least an estimated 75 per cent of the 9-year-olds in that country. To compute the median, TIMSS assumed that those 9-year-olds in the grades below the tested grades would score below the median and those in the grades above the tested grades would score above the median. The percentages assumed to be above and below the median were added to the tails of the distribution before calculating the median using the modified distribution.' (Mullis et al., 1997)

2.4 Test-curriculum Matching Analysis

The Test-curriculum Matching Analysis was designed and carried out to investigate the extent of the match between the curriculum covered within the TIMSS mathematics tests for pupils in the two grades tested (Years 4 and 5 in England) and the mathematics curriculum within participating countries for the same two grades, and to show how pupils' performance varied when based only on test items that were judged to be relevant to their own curriculum.⁴ A summary of how the Test-curriculum Matching Analysis was carried out in England is given in Chapter 1; more detailed information is provided in Appendix III (section III.6), which is included in a separate publication presenting all the technical appendices for the TIMSS study in England (Keys *et al.*, 1996b). In the Test-curriculum Matching Analysis, comparisons are made in terms of percentages of correct items rather than points on a scale, since the costs and international delays caused by opting for the latter would have been prohibitive (Mullis *et al.*, 1997).

The scores of pupils in the lower grade (Year 4 in England) on the sub-set of items selected by mathematics curriculum experts in England as being relevant to the National Curriculum for this age are shown in Table 2.4.1 for the nine selected countries, together with their respective scores on the entire mathematics test. Similar information for the upper grade (Year 5 in England) is shown in Table 2.4.2. The complete matrices, showing how pupils in every country performed on their own and other countries' selected sub-sets of items are reproduced from the international report in Annex A at the end of this report.

Pupils in Year 4 in England achieved a higher score (in terms of mean percentage correct) on the sub-set of 54 items selected by mathematics curriculum experts in England as being relevant to the National Curriculum for Year 4 than they achieved on the entire TIMSS mathematics test of 113 items. However, the means achieved by pupils in the other selected countries also improved when based on this sub-set of items. When the rank orders of pupils' mean percentages of correct responses are compared (i.e. based on the full set of mathematics items, and based on the sub-set chosen as appropriate to pupils in England), the two orders are virtually the same.

⁴ 'Because there may also be curriculum areas covered in some countries that are not covered by the TIMSS tests, the Test-curriculum Matching Analysis does not provide complete information about how well the TIMSS tests cover the curricula of the participating countries. The purpose of the Test-curriculum Matching Analysis was to investigate whether or not the content covered by the actual test items was in the intended curricula of participating countries.' (Beaton *et al.*, 1996b)

Table 2.4.1 Mathematics Test-curriculum Matching Analysis results for international third grade (Year 4 in England): the scores of pupils in nine selected countries on the sub-set of items selected by curriculum experts in England as addressing the National Curriculum compared with their scores on the entire mathematics test

COUNTRY	Mean percent correct: for all 113 items	Rank	Mean percent correct: for 54 items selected by England	Rank	Difference in mean percent correct
	%		%		%
Japan	63	1	68	2	+5
Singapore	62	2	69	1	+7
Netherlands	52	3	60	3	+8
United States	49	4	56	4	+7
Hungary	49	4	55	5	+6
Canada	47	6	54	6	+7
Scotland	45	7	53	7	+8
England	45	7	52	8	+7
Norway	36	9	40	9	+4
International mean	47		53		+6

Some of the 102 items in the TIMSS mathematics tests consisted of two or more parts. The total number of score points available for analysis was therefore 113. Differences may be subject to rounding errors.

Source: Table B2 (Mullis et al., 1997)

Table 2.4.2 shows similar information for pupils in the upper grade (Year 5 in England). In this instance, the sub-set of items selected was somewhat larger than that selected for Year 4: 86 items were chosen as being representative of the National Curriculum for mathematics for Year 5. Although the performance of pupils in each of the selected countries was higher on the sub-set of items selected by England, the increase in mean percentages of correct responses was less than it had been for pupils in the lower grade. The rank orders prepared using both the full set of mathematics items and the sub-set of items selected by England were identical.

Table 2.4.2 Mathematics Test-curriculum Matching Analysis results for international fourth grade (Year 5 in England): the scores of pupils in nine selected countries on the sub-set of items selected by curriculum experts in England as addressing the National Curriculum compared with their scores on the entire mathematics test

COUNTRY	Mean percent correct: for all 113 items	Rank	Mean percent correct: for 86 items selected by England	Rank	Difference in mean percent correct
	%		%		%
Singapore	76	1	78	1	+2
Japan	74	2	77	2	+3
Netherlands	69	3	72	3	+3
Hungary	64	4	66	4	+2
United States	63	5	65	5	+2
Canada	60	6	63	6	+3
Scotland	58	7	61	7	+3
England	57	8	59	8	+2
Norway	53	9	55	9	+2
International mean	59		61		+2

Some of the 102 items in the TIMSS mathematics tests consisted of two or more parts. The total number of score points available for analysis was therefore 113.

Differences may be subject to rounding errors.

Source: Table B1 (Mullis et al., 1997)

Minor differences in performance, but essentially unchanged rankings, were also found when comparisons were made using the sub-sets of items selected by other countries as being relevant to their own national curricula for each of the two grades tested. These results suggest that the TIMSS mathematics test provided a reasonable basis for comparing achievement in participating countries: this is not surprising as one of the main considerations during the development of the tests was to make them as fair as possible to pupils in all participating countries.

The number of items selected by each of the nine countries as being relevant to their own curriculum are shown below for both grades tested (equivalent to Years 4 and 5 in England). To date, there is no more detailed information about the specific items selected and rejected by different countries.

	International third grade (Year 4 in England)	International fourth grade (Year 5 in England)
United States	113	113
Japan	84	101
Norway	81	100
Hungary	74	96
Canada	63	99
Singapore	58	84
England	54	86
Scotland	46	92
Netherlands	26	59
International test	113	113

Source: Tables B1 and B2 (Mullis et al., 1997)

It is apparent that, at both grades tested, there was considerable variation in the numbers of items deemed relevant to countries' respective mathematics curricula. It is difficult to determine whether or not there is a positive association between curriculum coverage and achievement: in the two countries which selected the largest numbers of items at both the lower and the upper grades (United States and Japan), the overall mathematics performance of their pupils was significantly higher than that of pupils in England (both grades). On the other hand, experts in Singapore selected similar numbers of items to those selected in England, yet their pupils in both grades also achieved significantly higher overall mathematics scores than pupils in England. In view of these results, it seems likely that factors other than curriculum coverage have an impact on pupils' performance. This issue will be explored in Part 2 of this report (forthcoming).

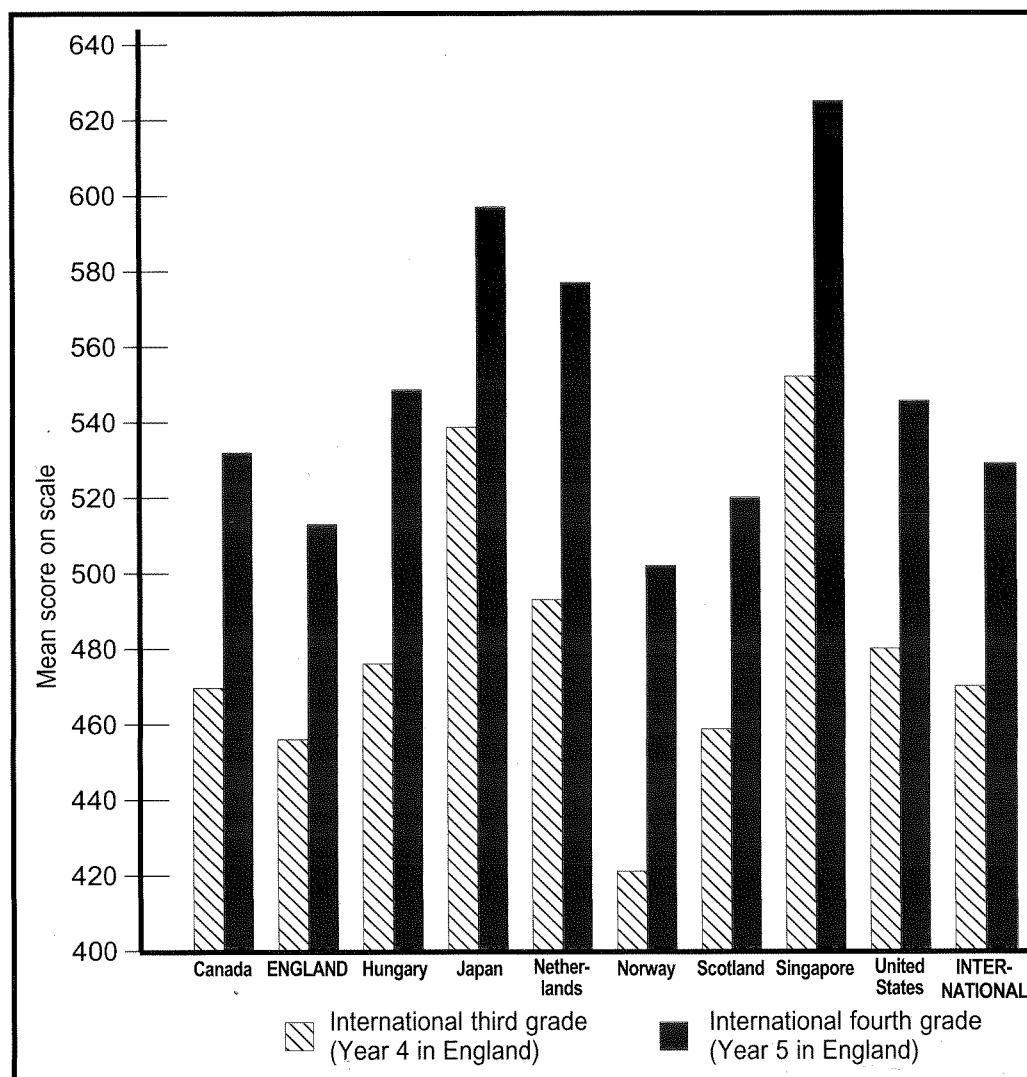
2.5 Comparisons between nine countries: differences in overall mean mathematics scores between international third and fourth grades (Years 4 and 5 in England)

One of the questions which TIMSS set out to answer was: how do pupils in different countries vary in the amount of progress they make over time? The most reliable way to answer this question would have been to test the same group of pupils in each country on two separate occasions, several years apart. This approach places additional demands on schools, and incurs considerable extra costs. The TIMSS study design provided an alternative, however: each country participating in the survey of Population 1 tested pupils in two adjacent grades (equivalent to Years 4 and 5 in England) in each of the schools taking part in the study. For each

country, the difference between the mean scores of pupils in Year 4 and Year 5 can be considered to be a proxy measure for the progress achieved between Year 4 and Year 5 (one school year) by pupils in that country: the progress score.

The differences in the mean scores achieved by pupils in the lower and upper grades tested in each of the selected countries are shown in Figure 2.5. In each country the mean score for the upper grade was higher than the mean score for the lower grade: the differences ranged from 56 to 84 scale points, with a median of 65 scale points. The difference between the mean scores of pupils in Year 4 and Year 5 in England not only represented the smallest increase in performance amongst the selected countries, but was also the only one that was less than the difference between the international means for the upper and lower grades (59 scale points).

Figure 2.5 Comparisons between nine countries: difference in overall mean mathematics score of international third and fourth grades (Years 4 and 5 in England)



Differences may be subject to rounding errors.

Source: Tables 1.1, 1.2 and 1.3 (Mullis et al., 1997)

There does not seem to be a clear relationship between the level of pupils' performance, relative to that of pupils in other countries, and their progress score. The two countries with the highest progress scores, the Netherlands and Norway, represented the upper and lowest levels of achievement respectively, when measured in terms of pupils' mean score for both the lower and the upper grades. The high progress score in Norway may be a function of the fact that at the time of testing children in Norway did not start school until they were seven; the lower grade pupils would therefore have had only seven terms of schooling at the time of the testing. The two selected countries with the highest mean scores for pupils in both the lower and upper grades (Singapore and Japan) had considerably different progress scores: Singapore had one of the highest (73 scale points) whereas Japan's progress score was the same as the difference between the international third and fourth grades (59 scale points). These results suggest that different sets of factors may have affected the size of the Year 4/5 increases in different countries. Possible factors include:

- ◆ the pupils' prior knowledge and understanding (e.g. curricula covered in previous years)
- ◆ curricula experienced in the international fourth grade (Year 5 in England) (e.g. the range and depth of topics covered)
- ◆ the amount of time spent on mathematics at school and at home
- ◆ the nature and quality of mathematics teaching for nine-year-olds.

The relationship between pupils' scores on the mathematics tests and some of these factors will be explored in Part 2 of this report (forthcoming).

2.6 International marker levels

One of the questions which TIMSS was designed to answer was: how do the proportions of very high-scoring and very low-scoring pupils vary from country to country? In order to answer this question, international marker levels, corresponding to the percentiles computed from the combined data of all participating countries, were calculated as shown below.

	International third grade (Year 4 in England)	International fourth grade (Year 5 in England)
• Top 10 per cent benchmark (90th percentile)	592	658
• Top 25 per cent benchmark (75th percentile)	538	601
• Top half benchmark (50th percentile or median)	474	535
• Bottom 25 per cent benchmark (25th percentile)	409	464

Source: Tables 1.4 and 1.5 (Mullis et al., 1997) and additional analyses provided by the International Study Centre.

Table 2.6 presents information from the international report showing the percentages of international fourth grade pupils (Year 5 in England) in each of the selected countries in the top half and the top ten per cent of the international distribution. The pattern was generally similar for the international third grade (Year 4 in England).

Table 2.6 Comparisons between nine countries: percentages of pupils reaching international benchmarks (top half and top ten per cent) of mathematics achievement in international fourth grade (Year 5 in England)

COUNTRY	Top half		Top ten per cent	
	%	Rank	%	Rank
Singapore	82	1	39	1
Japan	79	2	23	2
Netherlands	72	3	13	3
Hungary	56	4	11	4
United States	56	4	9	5
Canada	49	6	7	6
Scotland	43	7	6	8
England	39	8	7	6
Norway	33	9	2	9

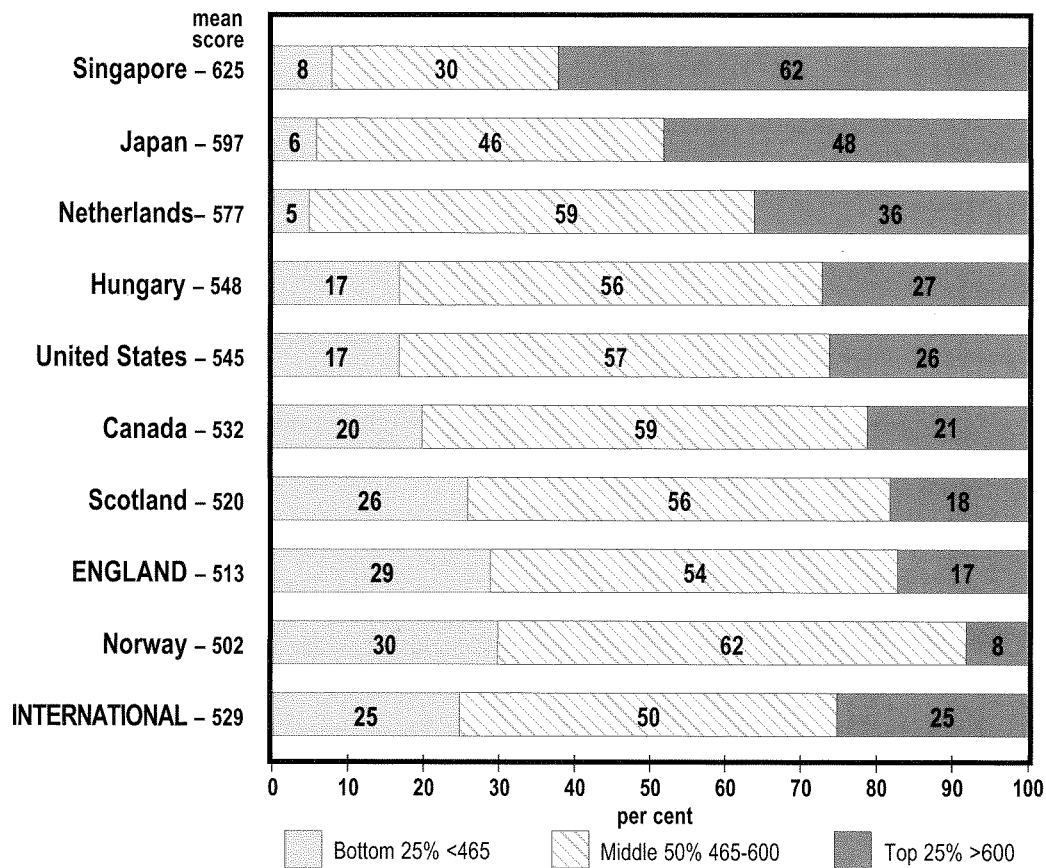
Differences may be subject to rounding errors

Source: Table 1.4 (Mullis et al., 1997)

It is apparent from Table 2.6 that although Canada and Scotland both had more pupils in the top half of the international distribution (49 per cent and 43 per cent respectively as compared with 39 per cent in England), the proportions of pupils in the top ten per cent were very close to those in England. Within the European countries selected for comparison, the Netherlands achieved the highest proportion of pupils in both the top half and the top ten per cent: approximately twice as many of their pupils reached the top ten per cent marker level compared with pupils in England (13 per cent and seven per cent respectively). The proportions of pupils that reached both the top half and top ten per cent marker levels in Singapore and Japan were considerably higher than in England.

Figure 2.6 shows the percentages of pupils in the upper grade (Year 5 in England) in each of the selected nine countries that were in the top 25 per cent, middle 50 per cent and bottom 25 per cent of the international distribution. The percentages for the lower grade (Year 4 in England) showed the same general pattern.

Figure 2.6 Percentages in the bottom 25%, middle 50% and top 25% of the international distribution for mathematics (international fourth grade : Year 5 in England): *comparisons between nine countries*



Source: Table 1.4 (Mullis et al., 1997) and additional analyses provided by the International Study Center

It can be seen that there was a very close correlation between the proportion of pupils in the international top 25 per cent and a country's mean mathematics score. Thus, Figure 2.6 illustrates differences in mean scores very clearly: Singapore, the highest scoring country had the highest proportion (62 per cent) of pupils in the top 25 per cent and Norway, the lowest scoring of the nine countries, had the lowest (eight per cent).

England had 17 per cent of pupils in the top 25 per cent and 29 per cent in the bottom 25 per cent, reflecting the fact that the mean mathematics score of pupils in England was below the international mean. The pattern in Scotland was similar.

2.7 Comparisons between nine countries: gender differences in overall mean mathematics scores

The gender differences in the overall mean mathematics scores were not statistically significant in most of the countries for either the lower or the upper grades (equivalent to Years 4 and 5 in England). The differences between the boys' and girls' mean scores in England were eight scale points in Year 4 and five scale points in Year 5, both in favour of boys. Neither of these differences was statistically significant, however. Out of the eight countries selected for comparisons, two showed significant differences in favour of boys at the lower grade: Canada (14 scale points) and Norway (18 scale points), and two showed significant differences (again, in favour of boys) at the upper grade: Japan and the Netherlands, eight scale points and 15 scale points respectively.

Table 2.7 shows the gender differences in overall mean mathematics score for upper grade pupils in the selected countries.

Table 2.7 Comparisons between nine countries: gender differences in overall mean mathematics score for international fourth grade (Year 5 in England)

COUNTRY	Boys' Mean	Girls' Mean	Difference in favour of boys
Singapore	620	630	-10
Scotland	520	520	0
United States	545	544	2
Canada	534	531	3
Hungary	552	546	5
England	515	510	5
Norway	504	499	5
Japan	601	593	8*
Netherlands	585	569	15*

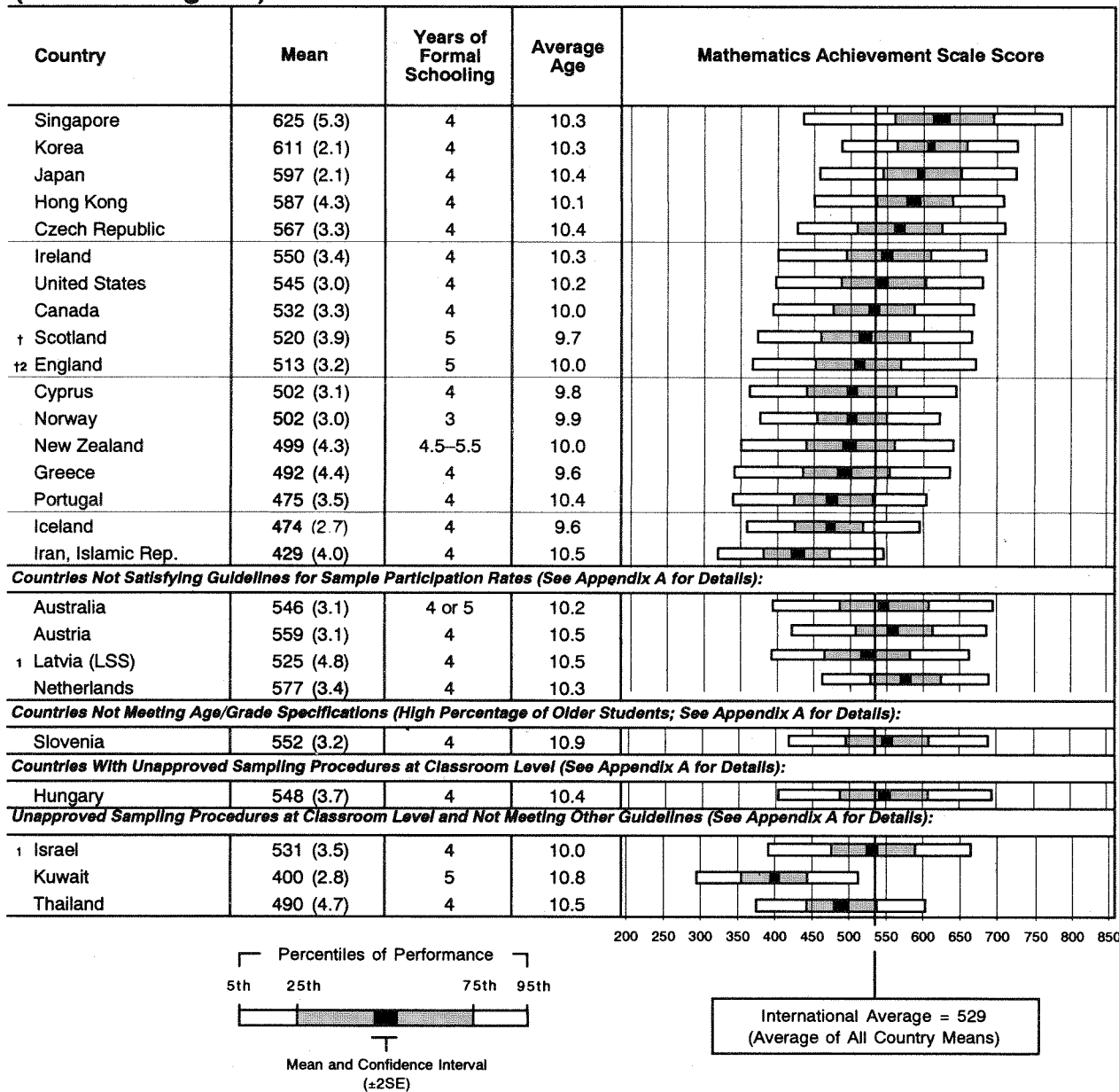
*Gender difference statistically significant at the five per cent level⁵.

Differences may be subject to rounding errors.

Source: Table 1.6 (Mullis et al., 1997)

⁵ There is always a possibility that the given difference between two samples could have arisen by chance. The level of statistical significance gives an indication of how likely this is. 0.05 is generally taken as the cut-off between differences that are thought to be real and those which could have arisen by chance, i.e. a difference of this size is only likely to have arisen by chance five times out of a hundred.

Table 1.1 (Mathematics)
Distributions of mathematics achievement—international fourth grade*
(Year 5 in England)



*Fourth grade in most countries.
 †Met guidelines for sample participation rates only after replacement schools were included (see Appendix A for details).
 ‡National Desired Population does not cover all of International Desired Population (see Table A.2). Because coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.
 §National Defined Population covers less than 90 percent of National Desired Population (see Table A.2).
 () Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.




SOURCE: Mullis et al., 1997

Figure 1.1 (Mathematics)
Multiple comparisons of mathematics achievement—international fourth grade*
(Year 5 in England)

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 two countries.[†]

Country	Singapore	Korea	Japan	Hong Kong	Netherlands	Czech Republic	Austria	Slovenia	Ireland	Hungary	Australia	United States	Canada	Israel	Latvia (LSS)	Scotland	England	Cyprus	Norway	New Zealand	Greece	Thailand	Portugal	Iceland	Iran, Islamic Rep.	Kuwait	
Singapore		●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	
Korea	●		▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Japan	▼	▼		●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Hong Kong	▼	▼	●		●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Netherlands	▼	▼	▼	●		●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Czech Republic	▼	▼	▼	▼	●		●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Austria	▼	▼	▼	▼	▼	●		●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Slovenia	▼	▼	▼	▼	▼	▼	●		●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Ireland	▼	▼	▼	▼	▼	▼	●	●		●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Hungary	▼	▼	▼	▼	▼	▼	●	●	●		●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Australia	▼	▼	▼	▼	▼	▼	●	●	●	●		●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
United States	▼	▼	▼	▼	▼	▼	▼	●	●	●	●		●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Canada	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●		●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Israel	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●		●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Latvia (LSS)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●		●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Scotland	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●		●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
England	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●		●	●	●	●	▲	▲	▲	▲	▲
Cyprus	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●		●	●	●	●	▲	▲	▲	▲
Norway	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●		●	●	●	▲	▲	▲	▲
New Zealand	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●		●	●	▲	▲	▲	▲
Greece	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●		●	●	▲	▲	▲
Thailand	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●		●	●	▲	▲
Portugal	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●		●	▲	▲
Iceland	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●		▲	▲
Iran, Islamic Rep.	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼		▲
Kuwait	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	

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

 Mean achievement significantly higher than comparison country
  No statistically significant difference from comparison country
  Mean achievement significantly lower than comparison country

*Fourth grade in most countries.

[†]Statistically significant at .05 level, adjusted for multiple comparisons.

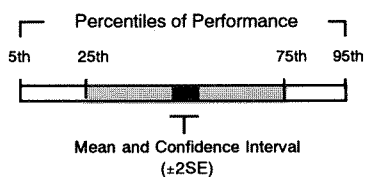
Because coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.

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: Mullis et al., 1997

Table 1.2 (Mathematics)
Distributions of mathematics achievement—international third grade*
(Year 4 in England)

Country	Mean	Years of Formal Schooling	Average Age	Mathematics Achievement Scale Score
Korea	561 (2.3)	3	9.3	
Singapore	552 (4.8)	3	9.3	
Japan	538 (1.5)	3	9.4	
Hong Kong	524 (3.0)	3	9.1	
Czech Republic	497 (3.3)	3	9.4	
United States	480 (3.4)	3	9.2	
Ireland	476 (3.6)	3	9.3	
Canada	469 (2.7)	3	9.1	
¹² England	456 (3.0)	4	9.1	
New Zealand	440 (4.0)	3.5–4.5	9.0	
Cyprus	430 (2.8)	3	8.8	
Greece	428 (4.0)	3	8.6	
Portugal	425 (3.8)	3	9.1	
Norway	421 (3.1)	2	8.8	
Iceland	410 (2.8)	3	8.6	
Iran, Islamic Rep.	378 (3.5)	3	9.4	
Countries Not Satisfying Guidelines for Sample Participation Rates (See Appendix A for Details):				
Australia	483 (4.0)	3 or 4	9.2	
Austria	487 (5.3)	3	9.5	
¹ Latvia (LSS)	463 (4.3)	3	9.7	
Netherlands	493 (2.7)	3	9.3	
Scotland	458 (3.4)	4	8.7	
Countries Not Meeting Age/Grade Specifications (High Percentage of Older Students; See Appendix A for Details):				
Slovenia	488 (2.9)	3	9.9	
Countries With Unapproved Sampling Procedures at Classroom Level (See Appendix A for Details):				
Hungary	476 (4.2)	3	9.4	
Unapproved Sampling Procedures at Classroom Level and Not Meeting Other Guidelines (See Appendix A for Details):				
Thailand	444 (5.1)	3	9.7	



International Average = 470
 (Average of All Country Means)

*Third grade in most countries.
¹Met guidelines for sample participation rates only after replacement schools were included (see Appendix A for details).
¹²National Desired Population does not cover all of International Desired Population (see Table A.2). Because coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.
¹National Defined Population covers less than 90 percent of National Desired Population (see Table A.2).
 () Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.




SOURCE: Mullis et al., 1997

Figure 1.2 (Mathematics)
Multiple comparisons of mathematics achievement—international third grade*
(Year 4 in England)

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 two countries.¹

Country	Korea	Singapore	Japan	Hong Kong	Czech Republic	Netherlands	Slovenia	Austria	Australia	United States	Hungary	Ireland	Canada	Latvia (LSS)	Scotland	England	Thailand	New Zealand	Cyprus	Greece	Portugal	Norway	Iceland	Iran, Islamic Rep.
Korea		●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Singapore	●		●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Japan	▼	●		▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Hong Kong	▼	▼	▼		▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Czech Republic	▼	▼	▼	▼		●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Netherlands	▼	▼	▼	▼	●		●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Slovenia	▼	▼	▼	▼	●	●		●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Austria	▼	▼	▼	▼	●	●	●		●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Australia	▼	▼	▼	▼	●	●	●	●		●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
United States	▼	▼	▼	▼	▼	●	●	●	●		●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Hungary	▼	▼	▼	▼	▼	●	●	●	●	●		●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Ireland	▼	▼	▼	▼	▼	●	●	●	●	●	●		●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Canada	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●		●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲
Latvia (LSS)	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●		●	●	●	▲	▲	▲	▲	▲	▲	▲
Scotland	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●		●	●	▲	▲	▲	▲	▲	▲	▲
England	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●		●	▲	▲	▲	▲	▲	▲	▲
Thailand	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●		●	●	●	●	▲	▲	▲
New Zealand	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●		●	●	●	▲	▲	▲
Cyprus	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●		●	●	●	▲	▲
Greece	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●		●	●	▲	▲
Portugal	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●		●	▲	▲
Norway	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●		●	▲
Iceland	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●		▲
Iran, Islamic Rep.	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	

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

 Mean achievement significantly higher than comparison country
  No statistically significant difference from comparison country
  Mean achievement significantly lower than comparison country

*Third grade in most countries.

¹Statistically significant at .05 level, adjusted for multiple comparisons.

Because coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools 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: Mullis et al., 1997

CHAPTER 3

The pupils' performance on the mathematics topics

Summary of main points

Compared with the international means for each of the six mathematics reporting categories, pupils in England showed:

- ◆ higher scores in both Year 4 and Year 5 for two content areas: *Geometry* and *Data representation, analysis and probability*
- ◆ lower scores in both Year 4 and Year 5 for four content areas: *Whole numbers; Fractions and proportionality; Measurement, estimation and number sense* and *Patterns, relations and functions*.

These results were broadly similar to those for 13-year-olds in England, in which the mean scores of students in Years 8 and 9 were above the international mean for *Data representation, analysis and probability*, but below the international means for the other five content areas for 13-year-olds.

With regard to gender differences in achievement:

- ◆ the boys' mean scores were significantly higher than the girls' in both Year 4 and Year 5 in two reporting categories: *Whole numbers* and *Measurement, estimation and number sense*; and in Year 4 only for *Patterns, relations and functions*.

3.1 Preface

The mathematics items used in the TIMSS tests were matched to the curriculum frameworks which were drawn up to ensure that the tests covered a wide range of the mathematics curriculum relevant to pupils in the international third and fourth grades (Years 4 and 5 in England). The curriculum frameworks for TIMSS are described in detail in Robitaille (1993). The total set of mathematics items has been organised into six reporting categories, each of which represents a different aspect of the mathematics curriculum. These categories are listed below, together with the number of items within each category and the nearest corresponding attainment targets from the 1995 National Curriculum Order for mathematics¹.

Reporting category	No. of items	Attainment target
Whole numbers	25	Ma 2: Number and Algebra
Fractions and proportionality	21	Ma 2: Number and Algebra
Measurement, estimation and number sense	20	Ma 3: Shape, Space and Measures Ma 2: Number and Algebra
Data representation, analysis and probability	12	Ma 4: Handling Data
Geometry	14	Ma 3: Shape, Space and Measures
Patterns, relations and functions	10	Ma 2: Number and Algebra
Total 102		

It can be seen that none of the mathematics items was closely linked with Ma1: Using and Applying Mathematics. Pupils' activities related to this part of the National Curriculum for mathematics would usually involve more extended and open-ended investigative work than is possible within the format of written tests. An additional component to the TIMSS survey

¹ At the time of testing (February-March 1995), the 1991 Statutory Orders for mathematics were still in effect. The attainment targets at that time were as follows:

Ma1: Using and applying mathematics

Ma2: Number

Ma3: Algebra

Ma4: Shape and space

Ma5: Handling data.

The expert panels convened for the Test-curriculum Matching Analysis (which included professional officers from the School Curriculum and Assessment Authority and inspectors from the Office for Standards in Education as well as practising teachers) considered that at the time of testing schools were more likely to be working to the 1995 Statutory Order than the 1991 Order.

² The reporting categories for Population 1 were slightly different from those for Population 2 (13-year-olds) which were as follows:

Fractions and number sense

Geometry

Algebra

Data representation, analysis and probability

Measurement

Proportionality.

consisted of a number of practical activities known collectively as the Performance Assessment element of the study; tasks within this component were somewhat closer to the approach of Ma1 than the items within the written tests. However, about half of the countries (including England) which participated in the survey of nine-year-olds did not take part in the Performance Assessment of this age group, although England took part in the field trials.

This chapter presents information about the mean scores of pupils in England for each of the six mathematics reporting categories and draws comparisons with the performance of pupils in the eight other countries listed in Chapter 1. Separate comparisons are made for each of the following for each reporting category:

- ◆ comparisons of pupils' mean percentages of correct answers
- ◆ differences in mean scores between Year 4 and Year 5 (a proxy measure for progress in achievement)
- ◆ gender differences in mean scores for both Year 4 and Year 5: these are presented for all six mathematics content areas in the final section of this chapter.

Within each of the following sections relating to mathematics reporting categories, illustrative examples of items are presented after the discussion of mean scores. For each item shown, the percentages of pupils making correct responses are given for England and the countries with the median and highest scores (see Mullis *et al.*, 1997), together with the international percentage correct. Details of the international coding guidelines for free response items are included where appropriate. Additional information, for example, relating to incorrect answers that were frequently given, is provided for each illustrative item. Each of the examples was one of those selected by the national panel of mathematics experts as being representative of the National Curriculum for mathematics for both years tested (Years 4 and 5 in England) unless otherwise stated.

Within this chapter, comparisons are made using the mean percentages of correct responses³ attained by pupils in the nine selected countries and differences are expressed in percentage points. This is in contrast to Chapter 2, where scores on a scale were quoted, and differences were expressed in scale points. Differences of five percentage points in the mean percentages of correct responses have been taken as evidence of real differences in performance.

³ Pupils who omitted to answer specific items were included in the percentage of pupils offering incorrect responses. Full details are provided in the international report (Mullis *et al.*, 1997).

3.2 Whole numbers

The mean percentages of correct answers achieved by pupils in Years 4 and 5 in England and in the eight other selected countries are shown in Table 3.2, together with the international means and the differences in mean scores between the two year groups. Countries are listed in order according to pupils' score in the lower grade, although it can be seen there is little difference in the ranking for the upper grade.

Table 3.2 Mean percentages of correct answers for *Whole numbers* in nine countries (25 items)

COUNTRY	Mean score of international third grade (Year 4 in England) %	Mean score of international fourth grade (Year 5 in England) %	Difference between Years 4 and 5
Singapore	75	83	8
Japan	72	82	10
Hungary	62	76	14
United States	57	71	14
Netherlands	57	75	18
Canada	53	68	15
Scotland	47	61	14
England	46	58	12
Norway	40	61	2
International mean	54	67	13

Differences may be subject to rounding errors

Source: Tables 2.1 and 2.2 (Mullis et al., 1997)

3.2.1 Comparisons of total topic scores: *Whole numbers*

The mean scores of pupils in England were amongst the lowest of the selected countries in both Year 4 and Year 5 for *Whole numbers*. In only two other countries (Scotland and Norway) were the pupils' mean scores also below the international means in both year groups. The mean scores of pupils in other countries exceeded those of pupils in England by up to 29 percentage points for the international third grade and by up to 25 percentage points for the international fourth grade (Years 4 and 5 respectively). Pupils in six of the selected countries achieved mean scores which were more than five percentage points higher than those of their counterparts in England in both the lower and upper grades. *Whole numbers* was the only content area for which the mean scores of pupils in England were more than five percentage points below the international means for both Year 4 and Year 5.

Comparisons with previous international studies

Five of the selected countries, including England, took part in IAEP2.

- ◆ The United States and Canada appear to have improved their relative position compared with England: in IAEP2 pupils in these two countries scored at about the same level as those in England, whereas in TIMSS their mean scores were substantially higher.
- ◆ On the other hand, the performance of pupils in England appears to have improved compared with that of pupils in Scotland. In IAEP2, pupils in Scotland outperformed those in England, whereas in TIMSS there was very little difference between the performance of pupils in England and Scotland.
- ◆ As in IAEP2, pupils in Hungary outperformed those in England.

3.2.2 Differences in mean scores between international third grade and international fourth grade (Years 4 and 5 in England): *Whole numbers*

One of the questions which TIMSS sought to answer was: how do pupils in different countries vary in the amount of progress they make over time? The difference in the mean scores of pupils in the lower and upper grades can be used as a proxy measure for progress in achievement: in the previous chapter, comparisons were drawn between pupils' performance for mathematics overall; within this chapter comparisons will be made for each of the reporting categories.

The mean scores of the pupils in the upper grade were higher than those of pupils in the lower grade in each of the selected countries (see Table 3.2). However, there was a considerable range in the increase in mean scores: from eight percentage points (Singapore, the highest-scoring country out of those selected for comparison in both grades) to 21 percentage points (Norway, the lowest-scoring country in the lower grade).

3.2.3 Examples of items for *Whole numbers*

Example items⁴ S2, U5, V3, and J9 are shown below, together with national and international percentages correct.

Item S2

S2. Here is a number sentence.

$$2000 + \square + 30 + 9 = 2739$$

What number goes in the \square to make this sentence true?

Answer: 700

S2	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	27.9	49.1
International	44.3	62.6
Median country	40.6	60.8
Highest-scoring country	86.2	92.0

Approximately one-quarter of pupils in both Years 4 and 5 in England gave '7' as their answer (28.4 per cent and 26.2 per cent respectively). Ten per cent of the younger pupils and five per cent of the older pupils did not attempt this question.

* * * * *

⁴ Some of the items used in TIMSS may not be quoted because of restrictions imposed by the IEA (which holds the copyright). The examples of items shown throughout this report have all been selected from the ones the IEA has released. All the released items have been published by the International Study Center (TIMSS International Study Center, 1997a, 1997b). NFER will publish an additional document presenting item statistics for the released items (forthcoming).

Item U5

U5.

Addition Fact
 $4 + 4 + 4 + 4 + 4 = 20$

Write this addition fact as a multiplication fact.

$\underline{5} \times \underline{4} = 20$

U5	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	39.0	52.9
International	63.2	76.6
Median country	62.8	80.2
Highest-scoring country	90.7	94.8

Both $5 \times 4 = 20$ and $4 \times 5 = 20$ were acceptable answers. In England a higher percentage of pupils in both Year 4 and Year 5 gave the former answer (25.7 per cent and 36.4 per cent respectively).

* * * * *

Item V3

V3. What is 5 less than 203 ?

Answer: 198

V3	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	43.6	61.5
International	47.9	61.9
Median country	47.4	63.6
Highest-scoring country	77.7	89.2

There was no incorrect answer which was widespread amongst pupils in either Year 4 or Year 5 in England.

* * * * *

Item J9

J9. Here is part of a wall chart that lists numbers from 1 to 100.

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25					

Below is part of the same wall chart. What number should be in the box with the question mark inside?

43	
53	
	?

- A. 34
 B. 44
 C. 54
 D. 64

J9	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	72.7	82.5
International	64.3	73.3
Median country	70.3	77.0
Highest-scoring country	89.8	92.0

It is possible that the widespread use of number squares showing all numbers up to 100 in rows within classrooms in England accounts for the high percentage of pupils in England answering this item correctly.

* * * * *

3.3 Fractions and proportionality

The mean percentages of correct answers achieved by pupils in the two year groups tested are shown for England and selected other countries in Table 3.3, together with the international means and the differences between the lower and upper grades. Countries are listed in rank order according to the mean score for the lower grade tested.

Table 3.3 Mean percentages of correct answers for *Fractions and proportionality* in nine countries (21 items)

COUNTRY	Mean score of international third grade (Year 4 in England) %	Mean score of international fourth grade (Year 5 in England) %	Difference between Years 4 and 5
Singapore	55	74	19
Japan	52	65	13
Netherlands	39	60	21
United States	36	51	15
Hungary	34	49	15
England	34	45	11
Canada	33	48	15
Scotland	33	46	13
Norway	24	38	14
International mean	36	49	13

Differences may be subject to rounding errors

Source: Tables 2.1 and 2.2 (Mullis et al., 1997)

3.3.1 Comparisons of total topic scores: *Fractions and proportionality*

The mean scores of pupils in England were slightly below the international means for *Fractions and proportionality* in both grades tested, and for Year 5 were the second lowest of the nine selected countries. Within those countries where pupils' mean scores were higher than those of pupils in England, their mean scores were up to 21 percentage points higher than those of Year 4 pupils and up to 29 percentage points higher than those of Year 5 pupils in England. In three of the other selected countries, pupils in the lower grade achieved mean scores which were at least five percentage points higher than those of pupils in England; in the upper grade, pupils in four countries achieved mean scores which were at least five percentage points higher than those of pupils in England.

Comparisons with previous international studies

There was no equivalent category within the IAEP2 study to allow comparisons to be made between the performance of pupils in that study as compared with TIMSS: the small number of items addressing fractions were subsumed within the content area covering number work.

3.3.2 Differences in mean scores between international third grade and international fourth grade (Years 4 and 5 in England): *Fractions and proportionality*

There was an increase in mean scores achieved by pupils in the upper grade as compared with the lower grade tested in each of the nine selected countries. The difference in mean scores ranged from 11 to 21 percentage points (see Table 3.3).

3.3.3 Examples of items for *Fractions and proportionality*

Example items S3, I8 and K9 are shown below, together with national and international percentages correct. In the opinion of the Test-curriculum Matching Analysis panel (see Chapter 1, section 1.2.4 and Chapter 2, section 2.2.4) none of the released items was considered to be representative of the National Curriculum for Year 4 pupils; each of these examples was, however, considered to be representative of the National Curriculum for Year 5 pupils.

Item S3

S3. Julie put a box on a shelf that is 96.4 centimetres long. The box is 33.2 centimetres long. What is the longest box she could put on the rest of the shelf? Show all your working out.

Answer: 63.2

$$\begin{array}{r} 96.4 \\ - 33.2 \\ \hline 63.2 \end{array}$$

S3 fully correct	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	9.5	21.9
International	12.2	25.7
Median country	8.4	25.1
Highest-scoring country	45.8	61.3

S3 fully or partially correct*	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	18.5	34.4
International	23.6	40.2
Median country	20.2	37.9
Highest-scoring country	56.5	70.1

The international marking guidelines stipulated that only responses which showed the correct answer '63.2' (with or without centimetres) **and** showed the calculation $96.4 - 33.2$ were fully correct. Responses which showed the correct answer with no explanation or working out were deemed partially correct, together with responses showing the correct calculation set out, but where the pupil had reached an incorrect answer.

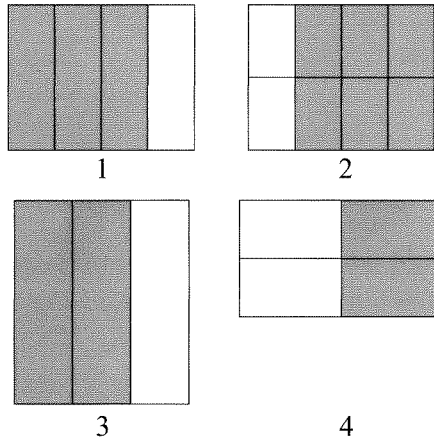
Eighteen per cent of Year 4 and 16 per cent of Year 5 pupils in England did not attempt this question.

* * * * *

* For this question, pupils' responses were marked as either fully correct, partially correct or incorrect. The international data tables showed the percentages of pupils who gave fully correct answers and the percentages who gave partially correct OR fully correct responses, together with the percentages providing incorrect answers.

Item 18

18. A fraction of each of these figures is shaded.



In which two figures is the same fraction shaded?

- A. 1 and 2
- B. 1 and 4
- C. 2 and 3
- D. 3 and 4

18	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	50.5	56.9
International	46.2	53.8
Median country	45.1	52.9
Highest-scoring country	61.5	71.5

Option (D) was a popular wrong answer, with approximately one-quarter of pupils in England selecting this response (26.9 per cent of Year 4 and 24.7 per cent of Year 5).

* * * * *

Item K9

K9. There are 54 marbles. They are put into 6 bags, so that there is the same number of marbles in each bag. How many marbles would there be in 2 bags?

- A. 108 marbles
- B. 18 marbles
- C. 15 marbles
- D. 12 marbles
- E. 9 marbles

K9	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	36.2	35.5
International	27.3	37.5
Median country	27.4	35.7
Highest-scoring country	56.7	75.75

The most popular wrong answer was option (A), which was chosen by approximately one-fifth of pupils in England: 21.6 per cent of Year 4 and 20.3 per cent of Year 5.

* * * * *

3.4 Measurement, estimation and number sense

The mean percentages of correct responses attained by the two year groups tested are shown for England and selected other countries in Table 3.4, together with the international means and the differences between the two year groups.

Table 3.4 Mean percentages of correct answers for *Measurement, estimation and number sense* in nine countries (20 items)

COUNTRY	Mean score of international third grade (Year 4 in England) %	Mean score of international fourth grade (Year 5 in England) %	Difference between Years 4 and 5
Japan	60	72	12
Netherlands	54	70	16
Singapore	52	67	15
Hungary	47	64	17
Canada	42	54	12
England	42	52	10
United States	41	53	12
Scotland	41	53	12
Norway	38	56	18
International mean	45	56	11

Differences may be subject to rounding errors

Source: Tables 2.1 and 2.2 (Mullis et al., 1997)

3.4.1 Comparisons of total topic scores: *Measurement, estimation and number sense*

The mean scores of pupils in England were slightly below the international means for this topic in both year groups.

In both the lower and upper grades tested, pupils in four of the selected countries achieved mean scores which were five percentage points or more higher than those of pupils in England. The highest-scoring pupils in the lower grade achieved a mean score 18 percentage points higher than their counterparts in England. At the upper grade, the difference between the mean scores of the highest-scoring pupils and pupils in England was 20 percentage points.

Comparisons with previous international studies

Comparisons between England and the four selected countries which took part in IAEP2 reveal few changes in relative performance; although the relative performance of Scotland appears to have deteriorated slightly compared with England.

3.4.2 Differences in mean scores between international third grade and international fourth grade (Years 4 and 5 in England): *Measurement, estimation and number sense*

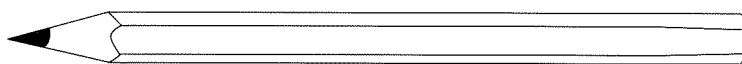
All of the selected countries showed higher mean scores for *Measurement, estimation and number sense* for pupils in the upper grade. Differences in mean scores ranged from ten to 18 percentage points (see Table 3.4).

3.4.3 Examples of items for *Measurement, estimation and number sense*

Example items K5, S5 and L8 are shown below, together with national and international percentages correct.

Item K5

K5. About how long is this picture of a pencil?



- A. 5 cm
- B. 10 cm
- C. 20 cm
- D. 30 cm


K5	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	65.9	67.8
International	68.6	76.7
Median country	66.1	77.6
Highest-scoring country	89.1	91.1

Option (C) was the most popular wrong answer, chosen by 20.0 per cent of Year 4 pupils and 18.4 per cent of Year 5 pupils in England.

* * * * *


Item S5

S5. Here is a paper clip.



← Length →

About how many lengths of the paper clip would fit along this line?



Answer: _____ *Four.* _____

S5	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	44.2	57.7
International	33.8	47.9
Median country	33.6	52.3
Highest-scoring country	58.8	67.6

The international marking guidelines indicated that both '4' and '5' were acceptable answers. Slightly more pupils in both year groups in England gave the former answer (21.4 per cent of the younger pupils and 28.6 per cent of the older pupils offered '4' whereas 20.0 per cent of the younger pupils and 25.2 per cent of the older pupils offered '5'). Other answers giving a number between 4 and 5.5 were also accepted as correct.

* * * * *

Item L8

L8. Four children measured the width of a room by counting how many paces it took them to cross it. The chart shows their measurements.

Name	Number of Paces
Stephen	10
Edward	8
Anna	9
Charles	7

Who had the longest pace?

- A. Stephen
- B. Edward
- C. Anna
- D. Charles

L8	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	19.5	26.8
International	21.2	32.0
Median country	18.0	27.3
Highest-scoring country	45.4	65.3

The most popular wrong answer was option (A), chosen by a sizeable majority of pupils in both years in England (72.5 per cent of Year 4 and 70.2 per cent of Year 5).

This item was also included in the tests for 13-year-olds (Population 2). Information on the percentages of correct responses achieved by students nationally and internationally in the international seventh and eighth grades (equivalent to Years 8 and 9 respectively in England) is presented for comparative purposes below, although it should be remembered that more countries took part in the survey of 13-year-olds than surveyed nine-year-olds (see Appendix II).

<i>L12 (Pop 2) equivalent to L8 (Pop1)</i>	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	61.7	72.9
International	68.0	73.1
Median country	73.0	78.8
Highest-scoring country	90.1	93.7

* * * * *

3.5 Data representation, analysis and probability

The mean percentages of correct responses attained by the two year groups tested are shown for England and selected other countries in Table 3.5, together with the international means and the differences between year groups.

Table 3.5 Mean percentages of correct answers for *Data representation, analysis and probability* in nine countries (12 items)

COUNTRY	Mean score of international third grade (Year 4 in England) %	Mean score of international fourth grade (Year 5 in England) %	Difference between Years 4 and 5
Japan	69	79	10
Singapore	68	81	13
Netherlands	56	75	19
United States	56	73	17
Canada	52	68	16
England	50	64	14
Scotland	49	66	17
Hungary	45	60	15
Norway	37	59	22
International mean	48	62	14

Differences may be subject to rounding errors

Source: Tables 2.1 and 2.2 (Mullis et al., 1997)

3.5.1 Comparisons of total topic scores: *Data representation, analysis and probability*

This was one of two content areas for which the mean scores of pupils in England were significantly above their overall percentage of correct responses for mathematics (see Mullis *et al.*, 1997) in both grades tested (the other being *Geometry*). The mean scores of pupils in England were above the international means for both Year 4 and Year 5 by two percentage points. However, in both the lower and the upper grades, there were four countries in which pupils' mean scores were more than five percentage points higher than those of their counterparts in England.

Comparisons with previous studies

Comparisons between England and the four selected countries which took part in IAEP2 reveal few changes in relative performance; although the relative performance of the United States appears to have improved compared with England.

3.5.2 Differences in mean scores between international third grade and international fourth grade (Years 4 and 5 in England): *Data representation, analysis and probability*

The performance of pupils in all the selected countries showed an increase in mean score from the lower to the upper grade of at least ten percentage points (see Table 3.5); the largest increase in mean score was 22 percentage points (Norway).

3.5.3 Examples of items for *Data representation, analysis and probability*

Example items K4, S1 and M1 are shown below, together with national and international percentages correct. It should be noted that in the opinion of the Test-curriculum Matching Analysis panel (see Chapter 1, section 1.2.4 and Chapter 2, section 2.2.4) although both K4 and S1 were considered to be appropriate to the National Curriculum for mathematics, none of the items on probability was deemed to be appropriate for Year 4 or for Year 5; item M1 is included to indicate the type of question that pupils answered in this content area.

Item K4

K4. Keith and Bob are playing a game. The aim of the game is to get the highest total of points. This chart shows how many points they each scored.

Scorecard

Player	Keith	Bob
Round 1	125	100
Round 2	125	125
Round 3	150	100
Round 4	50	150

Who won, and by how many points?

- A. Bob won by 25 points.
- B. Bob won by 100 points.
- C. Keith won by 25 points.
- D. Keith won by 175 points.

K4	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	36.0	45.3
International	34.5	49.5
Median country	31.4	52.5
Highest-scoring country	50.9	71.4

The most popular wrong answer for pupils in England was option (D), chosen by 26.2 per cent of Year 4 and 21.6 per cent of Year 5.

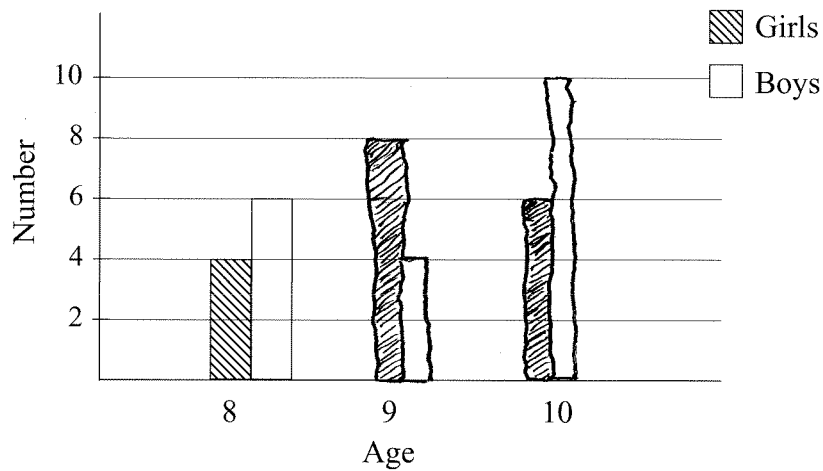
* * * * *

Item S1

S1. This table shows the ages of the girls and boys in a club.

Age	Number of Girls	Number of Boys
8	4	6
9	8	4
10	6	10

Use the information in the table to complete the graph for ages 9 and 10.



S1 fully correct	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	28.5	49.5
International	24.2	40.8
Median country	18.7	37.5
Highest-scoring country	68.5	83.2

S1 fully or partially correct*	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	46.1	67.8
International	38.1	56.1
Median country	36.3	58.5
Highest-scoring country	79.8	93.2

* For this question, pupils' responses were marked as either fully correct, partially correct or incorrect. The international data tables showed the percentages of pupils who gave fully correct answers and the percentages who gave partially correct OR fully correct responses, together with the percentages providing incorrect answers.

The international marking guidelines required all four bars to be correct in terms of height, placement and shading for a fully correct answer; however, they also allowed full marks if all four bars were of the correct height and there were 'either bars misplaced or bars shaded incorrectly in no more than one set (i.e. for age 9 or age 10)'.

Two criteria were stipulated for answers deemed partially correct, either:

- a) 'placement, shading and height all correct for one, two or three bars. (At least one bar completely correct)'; or
- b) 'all four bars of correct height, but two or more errors involving placement or shading'.

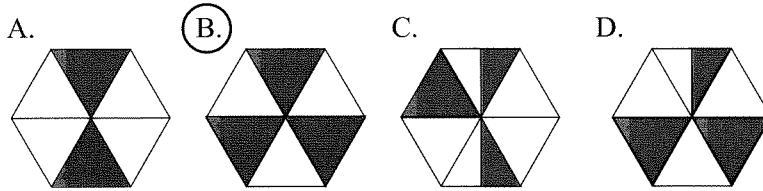
Responses which did not meet any of the above requirements were incorrect.

Of the fully correct answers offered by pupils in England, the majority showed all four bars on the graph correct in terms of height, placement and shading (24.7 per cent of Year 4 and 42.2 per cent of Year 5 pupils). The majority of the partially correct answers provided by pupils in England fulfilled criterion (a) above: 12.7 per cent of Year 4 and 12.8 per cent of Year 5 pupils.

* * * * *

Item M1

M1. Samantha drops a stone onto each of these targets. In which target does the stone have the best chance of landing on a shaded space?



<i>M1</i>	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	73.1	78.3
International	68.6	77.8
Median country	70.3	79.1
Highest-scoring country	88.8	89.0

Option (A) was the most popular wrong answer, chosen by 12.4 per cent of Year 4 and 11.3 per cent of Year 5 pupils in England.

* * * * *

3.6 Geometry

The mean percentages of correct responses achieved by the two year groups tested are shown for England and selected other countries in Table 3.6, together with the international means and the differences between the lower and upper grades.

Table 3.6 Mean percentages of correct answers for *Geometry* in nine countries (14 items)

COUNTRY	Mean score of international third grade (Year 4 in England) %	Mean score of international fourth grade (Year 5 in England) %	Difference between Years 4 and 5
Scotland	65	72	7
England	63	74	11
Japan	62	72	10
Canada	62	72	10
Netherlands	61	71	10
United States	61	71	10
Singapore	60	72	12
Hungary	52	66	14
Norway	44	58	14
International mean	56	64	8

Differences may be subject to rounding errors

Source: Tables 2.1 and 2.2 (Mullis et al., 1997)

3.6.1 Comparisons of total topic scores: *Geometry*

The mean scores of pupils in England for *Geometry* were significantly higher in both year groups than their overall score for mathematics (see Mullis *et al.*, 1997). Furthermore, the mean scores of pupils in Year 4 and Year 5 were further above the international mean than for any other reporting category: by seven percentage points for the lower grade and by ten percentage points for the upper grade. Pupils in only one of the selected countries (Scotland) achieved higher mean scores than pupils in Year 4 in England, and no country showed a higher mean score than England at Year 5. These results represent the highest mean scores for pupils in both Years 4 and 5 across all of the mathematics categories.

Comparisons with previous international studies

Comparisons between England and the four selected countries which took part in IAEP2 reveal few changes in relative performance, although the relative performance of Hungary appears to have deteriorated compared with England.

3.6.2 Differences in mean scores between international third grade and international fourth grade (Years 4 and 5 in England): *Geometry*


All the selected countries showed an increase in mean score between the two grades tested (see Table 3.6). The difference ranged from seven to 14 percentage points, with pupils in England showing an increase of 11 percentage points.

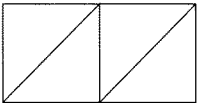
3.6.3 Examples of items for *Geometry*

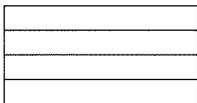
Example items K8, L3 and T5 are shown below, together with national and international percentages correct.

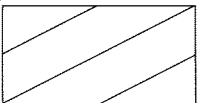
Item K8

K8. Which rectangle is NOT divided into 4 equal parts?

A. 

B. 

C. 

D. 

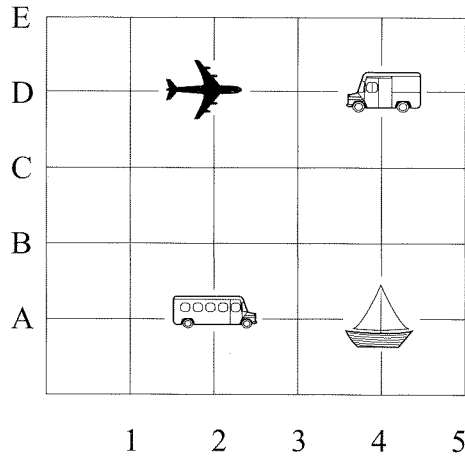
K8	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	64.4	75.2
International	59.6	72.8
Median country	61.3	77.4
Highest-scoring country	83.3	89.8

There was a different pattern of incorrect responses for pupils in Years 4 and 5 in England. For the younger age group the most popular wrong answer was option (C), selected by 13.7 per cent of pupils, whereas for the older pupils the most popular wrong answer was option (B), chosen by 11.1 per cent.





* * * * *

Item L3

L3. This is a game board.



Which object is located at (2,D)?

- (A.) The plane 
- B. The truck 
- C. The bus 
- D. The boat 

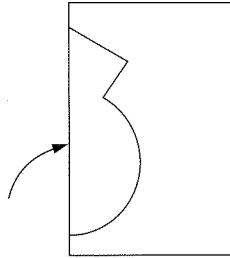
L3	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	89.4	94.4
International	79.9	87.8
Median country	82.9	92.0
Highest-scoring country	92.6	97.4

Very few pupils in either Year 4 or Year 5 in England selected a wrong answer; it would therefore be misleading to say that any one incorrect response was widespread.

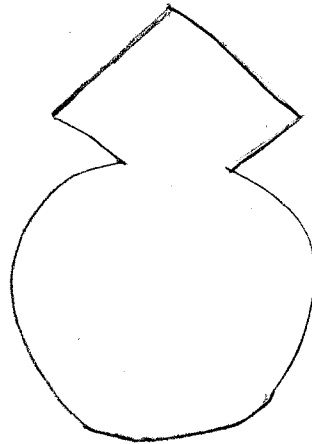
* * * * *

Item T5

T5. Craig folded a piece of paper in half and cut out a shape.

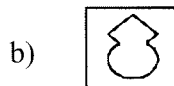


Draw a picture to show what the cut-out shape will look like when it is opened up and flattened out.



T5	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	61.1	76.5
International	45.1	58.5
Median country	48.8	63.7
Highest-scoring country	74.6	88.4

Drawings which represented either the symmetrical cut-out shape (a) or the left-over paper, opened out (b) were accepted as correct.



The majority of pupils in both Year 4 and Year 5 in England drew shape (a) as their answer (53.3 per cent of the younger pupils and 67.2 per cent of the older pupils). There was no one incorrect shape which was widespread amongst the answers of pupils in England.

* * * * *

3.7 Patterns, relations and functions

The mean percentages of correct responses achieved by the two year groups tested are shown for England and selected other countries in Table 3.7, together with the international means and the differences in mean scores between the lower and upper grades.

Table 3.7 Mean percentages of correct answers for *Patterns, relations and functions* in nine countries (ten items)

COUNTRY	Mean score of international third grade (Year 4 in England) %	Mean score of international fourth grade (Year 5 in England) %	Difference between Years 4 and 5
Singapore	65	76	11
Japan	64	76	12
Hungary	57	69	12
United States	53	66	13
Netherlands	50	65	15
Canada	48	62	14
Scotland	45	57	12
England	43	55	12
Norway	34	50	16
International mean	48	60	12

Differences may be subject to rounding errors
 Source: Tables 2.1 and 2.2 (Mullis et al., 1997)

3.7.1 Comparisons of total topic scores: *Patterns, relations and functions*

The mean scores of pupils in England were five percentage points below the international means in both years tested. These mean scores were ranked second-lowest out of the selected countries for Year 4 and Year 5. Pupils in six of the selected countries achieved mean scores which were at least five percentage points higher than those of pupils in England for both grades tested.

Comparisons with previous international studies

Comparisons between England and the four selected countries which took part in IAEP2 reveal few changes in relative performance; although the relative performance of the United States appears to have improved compared with England.

3.7.2 Differences in mean scores between international third grade and international fourth grade (Years 4 and 5 in England): *Patterns, relations and functions*

All of the selected countries showed an increase in performance between the two grades tested of more than ten percentage points: the difference for England was 12 percentage points (see Table 3.7).

3.7.3 Examples of items for *Patterns, relations and functions*

Example items J5, K3, L4 and U4 are shown below, together with national and international percentages correct. It should be noted that the Test-curriculum Matching Analysis panel (see Chapter 1, section 1.2.4 and Chapter 2, section 2.2.4) considered item L4 to be representative of the National Curriculum for Year 5 pupils, but not for Year 4 pupils.

Item J5

- J5. What do you have to do to each number in Column A to get the number next to it in Column B?

Column A	Column B
10	2
15	3
25	5
50	10

- A. Add 8 to the number in Column A.
 B. Subtract 8 from the number in Column A.
 C. Multiply the number in Column A by 5.
 D. Divide the number in Column A by 5.

J5	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	19.1	34.7
International	27.5	39.1
Median country	25.3	36.9
Highest-scoring country	49.7	70.5

Option (B) was the most popular wrong answer for pupils in both Year 4 and Year 5 in England, selected by 41.1 per cent and 31.9 per cent respectively.

* * * * *

Item K3

K3. Which pair of numbers follows the rule “Multiply the first number by 5 to get the second number”?

A. 15 → 3

B. 6 → 11

C. 11 → 6

D. 3 → 15

K3	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	30.7	41.8
International	37.3	52.7
Median country	34.4	50.2
Highest-scoring country	63.7	79.6

The pattern of pupils' incorrect responses was different in Years 4 and 5 in England. Similar percentages of the younger pupils chose options (A) and (B): 23.1 per cent and 21.1 per cent respectively. However, amongst the older pupils option (A) was clearly the most popular wrong answer, selected by 24.1 per cent, about twice as many as chose either option (B) or option (C).

* * * * *

Item L4

L4. These shapes are arranged in a pattern.

○△○○△△○○○△△△

Which set of shapes is arranged in the same pattern?

A. ★□★□★□□★□□

B. □★□□★□□□★□□□□

C. ★□★□★□□★□□□□

D. □□★□★□□★□□★□□

L4	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	60.3	74.9
International	60.8	72.4
Median country	63.1	75.3
Highest-scoring country	89.1	92.8

The majority of pupils in Years 4 and 5 in England completed this item correctly, and none of the incorrect options was notably more popular than the others at either age.

This item was also included in the tests for 13-year-olds (Population 2). Information on the percentages of correct responses achieved by students nationally and internationally in the international seventh and eighth grades (equivalent to Years 8 and 9 respectively in England) is presented for comparative purposes below, although it should be remembered that more countries took part in the Population 2 survey than the Population 1 survey (see Appendix II).

L13 (Pop 2) equivalent to L4 (Pop 1)	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	93.6	94.6
International	87.3	90.1
Median country	90.4	93.1
Highest-scoring country	97.3	98.3

* * * * *

Item U4

U4. These numbers are part of a pattern.

50 , 46 , 42 , 38 , 34 , **30**

What do you have to do to get the next number?

Answer: Take away four

U4	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	40.7	56.6
International	41.3	56.9
Median country	41.4	63.5
Highest-scoring country	72.5	79.1

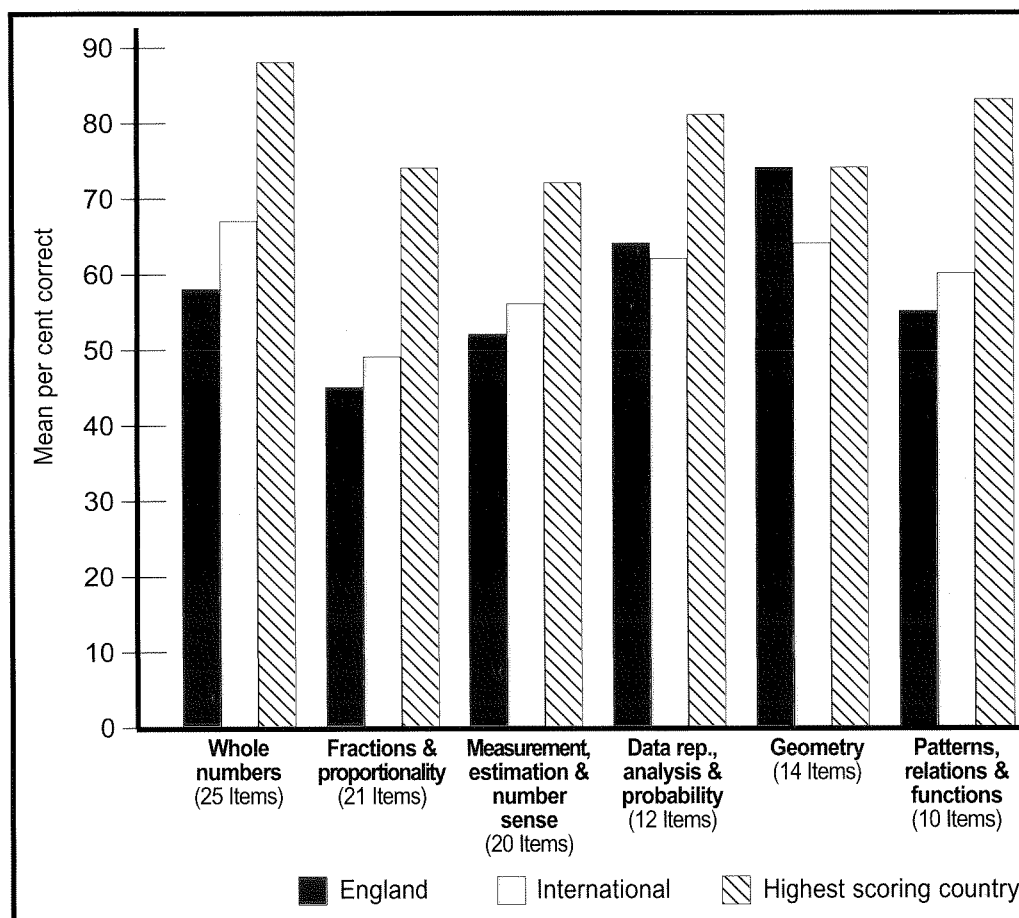
Answers which indicated the process for determining the next number in the pattern (subtract four) or gave one or more correct numbers to continue the pattern were correct. In England, 26.3 per cent of Year 4 pupils and 39.7 per cent of Year 5 pupils gave the correct process. Eight per cent of the pupils in both Years 4 and 5 indicated an increase by four as their answer.

* * * * *

3.8 Overview of results for each mathematics reporting category

The mean percentages of correct responses achieved by pupils in Year 5 in England for each of the six mathematics reporting categories are shown in Figure 3.8 below, together with the international means for the equivalent year groups and the mean scores of the highest-scoring country for each category.

Figure 3.8 Overview of mean percentages for mathematics reporting categories: *international fourth grade (Year 5 in England)*



Source: Tables 2.1 and 2.2 (Mullis et al., 1997)

The results for pupils in Year 4 in England were broadly similar to those for Year 5.

3.9 Gender differences

The preceding sections have presented the mean scores of pupils in Years 4 and 5 in England for each of the mathematics reporting categories and drawn comparisons with the mean scores of pupils in the equivalent years within eight other selected countries. This section identifies where there are statistically significant gender differences in performance in those categories of mathematics.

Two countries (Hungary and Scotland) showed no significant gender differences in mean scores for any of the six reporting categories in mathematics: this was the case for both the lower and upper grades tested (equivalent to Years 4 and 5 in England).

Most instances of gender differences were found in two content areas:

- ◆ **Whole numbers**
- ◆ **Measurement, estimation and number sense.**

There were no significant gender differences in mean scores in any of the selected countries in either the lower or the upper grade for *Fractions and proportionality* and *Data representation, analysis and probability*. For the two remaining reporting categories, only one of the selected countries showed a significant gender difference in performance in each case, as follows:

- ◆ **Geometry:** a difference favouring boys in the upper grade only in the Netherlands.
- ◆ **Patterns, relations and functions:** a difference in favour of boys in Year 4 only in England.

All of the gender differences identified in the selected countries were in favour of boys, with one exception: the mean score of the upper grade girls in Singapore was higher than that of the boys.

The gender differences in pupils' mean scores for England and other selected countries where gender differences were apparent are shown in Table 3.9.1.

Table 3.9.1 Gender differences for mathematics reporting categories in nine countries

COUNTRY	WHOLE NUMBERS		MEASUREMENT etc.	
	International third grade (Year 4 in England)	International fourth grade (Year 5 in England)	International third grade (Year 4 in England)	International fourth grade (Year 5 in England)
Canada	*		*	
England	*	*	*	*
Hungary				
Japan	*	*	*	*
Netherlands	*			*
Norway	*		*	*
Scotland				
Singapore		G		
United States				*

* Gender difference in favour of boys, statistically significant at the 0.05 level.

G Gender difference in favour of girls, statistically significant at the 0.05 level.

Gender differences in the mean scores achieved by pupils in England for each of the six mathematics reporting categories are shown in Table 3.9.2 below. Where boys' mean scores were significantly higher than those of the girls (i.e. statistically significant at the 0.05 level), they are shown in bold type.

Table 3.9.2 Gender differences in mean percentages of correct answers for mathematics reporting categories for pupils in England

Reporting category	Whole numbers	Fractions & proportionality	Measurement, estimation & number sense	Data rep., analysis & probability	Geometry	Patterns, relations & functions
	(25 items) %	(21 items) %	(20 items) %	(12 items) %	(14 items) %	(10 items) %
Boys Year 4	47	35	43	50	63	45
Girls Year 4	44	34	40	50	63	40
Boys Year 5	60	46	54	64	74	56
Girls Year 5	57	45	50	65	74	54

Source: Tables 2.4 and 2.5 (Mullis et al., 1997)

CHAPTER 4

The pupils' overall performance on the science tests

Summary of main points

- ◆ The mean overall science scores of pupils in Year 5 in England were:
 - significantly *lower* than those of pupils in only three of the 25 other countries in the TIMSS survey of nine-year-olds
 - not significantly different from those of pupils in nine other countries
 - significantly *higher* than those of pupils in 13 of the 25 other countries that took part.
- ◆ The pattern of achievement of Year 4 pupils was virtually the same.
- ◆ Compared with pupils in other countries, the relative performance of pupils in England was much better in science than in mathematics.
- ◆ The relative position of England in terms of the science scores of nine-year-olds was similar to its position based on the science scores of 13-year-olds.
- ◆ Thirteen per cent of the pupils in England were in the international top ten per cent in both year groups.
- ◆ Although the boys' overall mean science scores were higher than the girls' in both Year 4 and Year 5 in England, the differences were not statistically significant. There were, however, statistically significant differences in favour of boys in several of the other selected countries.

4.1 Preface

This chapter presents and compares the mean scores on the TIMSS science tests of primary school pupils in England and pupils of the same age in other countries. As with the mathematics results, separate mean scores are provided for the two year groups tested (international third and fourth grades, equivalent to Years 4 and 5 in England), together with mean scores for nine-year-olds (i.e. pupils aged 9.0–9.11 at the time of testing).

Comparisons with other countries are made in terms of:

- ◆ pupils' overall mean scores on the TIMSS science tests
- ◆ pupils' mean scores on the sub-set of items identified by science curriculum experts in England as being covered by the science curriculum for each year group (the Test-curriculum Matching Analysis)
- ◆ ranges of scores in different countries (percentages of pupils reaching the following international marker levels: top ten per cent, top 25 per cent, top half and bottom 25 per cent of the international distribution)
- ◆ differences in mean scores of pupils in the international third and fourth grades (equivalent to Years 4 and 5 in England)
- ◆ differences in the mean scores of boys and girls.

Where appropriate, the results of TIMSS are compared with the results of the following previous large-scale international studies of science.

Date of testing	Study	References to reports
1970-71	First International Science Study (FISS)	Comber and Keeves (1973)
1984	Second International Science Study (SISS)	Postlethwaite and Wiley (1992) IEA (1988) Keys (1987)
1991	The second study carried out by the International Association for the Evaluation of Educational Progress (IAEP2)	Lapointe <i>et al.</i> (1992b) Foxman (1992)

In the following discussions, previous studies will be referred to by their abbreviations (e.g. FISS, SISS and IAEP2). In comparing the results of previous studies with TIMSS, it should be remembered that:

- ◆ some of the countries taking part in TIMSS have not participated in any previous international comparative studies in science

- ◆ different sets of countries took part in different studies
- ◆ only three countries (England¹, Hungary and the United States) have participated in all three previous large-scale studies
- ◆ because of curriculum changes, different tests were used in each study and there are only small numbers of common items between any two studies.

4.2 Overall mean scores on the science tests

In this chapter, the overall mean science scores of pupils in Years 4 and 5 England are compared with those of pupils of similar ages in four distinct groups of other countries, as follows:

continental Western European countries	Table 4.2.1
English-speaking countries	Table 4.2.2
Eastern European countries	Table 4.2.3
Asian Pacific Rim countries	Table 4.2.4.

The information presented in these tables is derived from Tables 1.1 (Science) and 1.2 (Science) and Figures 1.1 (Science) and 1.2 (Science) from the international report on TIMSS (Population 1) (Martin *et al.*, 1997); these tables and figures, which give detailed information on pupils' mean science scores in all of the countries taking part in TIMSS together with the average age and numbers of years of formal schooling for each sample, have been reproduced at the end of this chapter. In the international tables, countries which did not meet the international sampling requirements are shown 'below the line', i.e. the mean scores of their pupils are not included in the rank orders of results. In the figures, however, these countries are included in the rank ordering of results, but are listed in italics to differentiate them. The reasons why each country did not meet the sampling requirements are explained in the footnotes to the international tables.

Overall mean science scores are expressed in terms of a scale with an international mean of 500 and a standard deviation of about 100, based on the performance of pupils in both grades (year groups) in all participating countries. The international means were 473 for the lower grade and 524 for the upper grade.

It is apparent from the international tables and figures that the relative performance of pupils in England, compared with pupils in most of the other countries taking part in the TIMSS survey of nine-year-olds, was much higher in science than in mathematics. The mean overall science scores of pupils in both Year 4 and Year 5 in England were significantly

¹ In TIMSS, FISS and SISS, samples were selected from schools in England. In IAEP2, samples were selected from schools in England and Wales.

higher than those of pupils in half of the other countries participating in the study (Figures 1.1 (Science) and 1.2 (Science)), whereas in mathematics the mean scores of pupils in both year groups were significantly *lower* than those of pupils in more than half of the other countries that took part in the survey. The mean overall science scores of pupils in Years 4 and 5 in England were more than 25 scale points above the international means for the equivalent years (26 scale points for the younger pupils and 27 scale points for the older pupils). There were only two countries (Korea and Japan) in which pupils at both grades tested achieved mean scores that were significantly higher than those of pupils in England. Detailed comparisons for science are given below.

The relative position of England in terms of the science scores of nine-year-olds was similar to its position based on the scores of 13-year-olds (Keys *et al.*, 1996a).

4.2.1 Overall mean science scores: comparisons between England and continental Western European countries

In none of the Western European countries did pupils achieve higher overall mean science scores than pupils in England (see Table 4.2.1). The mean science scores of pupils in Years 4 and 5 in England were significantly higher than those of four of the six Western European countries taking part in the TIMSS survey of nine-year-olds: the differences in mean scores, favouring pupils in England, ranged from 49–76 scale points for the younger pupils and from 21–71 scale points for the older pupils.

Table 4.2.1 Overall mean science scores: comparisons between England and continental Western European countries

Compared with England	International third grade (Year 4 in England)	International fourth grade (Year 5 in England)
Significantly higher	<i>None</i>	<i>None</i>
Not significantly different	Austria; Netherlands	Austria; Netherlands
Significantly lower	Greece; Iceland; Norway; Portugal	Greece; Iceland; Norway; Portugal

Source: Figures 1.1 and 1.2 (Martin *et al.*, 1997)

Comparisons with previous international studies

Only three of the continental Western European countries taking part in TIMSS at the Population 1 level have taken part in previous studies.

- ◆ The relative position of England, compared with Norway appears to have improved slightly since the early 1980s (SISS) when pupils in both countries scored at the same level.

- ◆ The relative position of England compared with the other two countries, the Netherlands and Portugal, has not changed.

4.2.2 Overall mean science scores: comparisons between England and other English-speaking countries

The performance of pupils in Years 4 and 5 in England is compared with that of pupils in the six other English-speaking countries in Table 4.2.2. In both year groups, there was no significant difference in the mean scores of pupils in England and four of the other countries. The mean score of pupils in the upper grade in the United States was significantly higher than that of pupils in Year 5 in England by 14 scale points.

Table 4.2.2 Overall mean science scores: comparisons between England and other English-speaking countries

Compared with England	International third grade (Year 4 in England)	International fourth grade (Year 5 in England)
Significantly higher	None	United States
Not significantly different	Australia; Canada; United States; Scotland	Australia; Canada; Ireland; Scotland
Significantly lower	Ireland; New Zealand	New Zealand

Source: Figures 1.1 and 1.2 (Martin et al., 1997)

Comparisons with previous international studies

The results of TIMSS suggest that, compared with other English speaking countries, the relative position of England in science remains unchanged since previous studies took place.

In TIMSS, as in most previous studies, pupils in England achieved mean scores which were similar to those of pupils in Australia, Canada and Scotland.

Two countries, however, appear to have improved their relative position compared with England.

- ◆ In IAEP2, the mean science score of pupils in Ireland was six percentage points below that of England, whereas in TIMSS, pupils in Ireland scored at the same level as pupils in England.
- ◆ In previous studies (FISS and IAEP2), pupils in the United States had performed at about the same level as those in England, whereas in TIMSS United States pupils in one of the year groups tested achieved a significantly higher mean score than pupils in the equivalent grade in England.

4.2.3 Overall mean science scores: comparisons between England and Eastern European countries

There were no Eastern European countries in which pupils' overall mean science scores were significantly higher than those of pupils in Years 4 and 5 in England (see Table 4.2.3). Pupils in both year groups in England outperformed their counterparts in Hungary and Latvia: by 35 scale points and 34 scale points respectively for the younger pupils, and by 19 scale points and 39 scale points respectively for the older pupils.

Table 4.2.3 Overall mean science scores: comparisons between England and Eastern European countries

Compared with England	International third grade (Year 4 in England)	International fourth grade (Year 5 in England)
Significantly higher	<i>None</i>	<i>None</i>
Not significantly different	Czech Republic; Slovenia	Czech Republic; Slovenia
Significantly lower	Hungary; Latvia	Hungary; Latvia

Source: Figures 1.1 and 1.2 (Martin et al., 1997)

Comparisons with previous international studies

The relative position of England, compared with Hungary, appears to have improved since previous studies were carried out.

- ◆ In TIMSS, pupils in England achieved higher mean science scores than those in Hungary whereas in SISS, the mean scores of pupils in Hungary had been significantly higher than those in England, and in IAEP2 the mean science scores of pupils in England and Hungary were similar.

4.2.4 Overall mean science scores: comparisons between England and Asian Pacific Rim countries

The performance of pupils in Years 4 and 5 in England is compared with that of pupils in the Asian Pacific Rim countries in Table 4.2.4. In both year groups, pupils in Korea and Japan achieved significantly higher overall mean science scores than pupils in England, by 54 scale points and 23 scale points respectively for the younger pupils, and by 46 scale points and 23 scale points respectively for the older pupils. However, these differences in performance were considerably smaller than were evident in the comparisons of mathematics scores reported in Chapter 2. Whilst the mean scores of pupils in both grades in Singapore were not significantly different from those of pupils in England, the latter outperformed pupils in

Hong Kong and Thailand by 17 scale points and 66 scale points respectively in the lower grade (Year 4), and by 18 scale points and 78 scale points respectively at the upper grade (Year 5).

Table 4.2.4 Overall mean science scores: comparisons between England and Asian Pacific Rim countries

Compared with England	International third grade (Year 4 in England)	International fourth grade (Year 5 in England)
Significantly higher	Japan; Korea	Japan; Korea
Not significantly different	Singapore	Singapore
Significantly lower	Hong Kong; Thailand	Hong Kong; Thailand

Source: Figures 1.1 and 1.2 (Martin et al., 1997)

Comparisons with previous international studies

The relative position of England, compared with Hong Kong, appears to have improved.

- ◆ In TIMSS, pupils in England achieved higher mean scores than those in Hong Kong, whereas in SISS, pupils in England and Hong Kong scored at a similar level.

The relative position of England compared with the other Asian Pacific Rim countries is unchanged since previous studies.

- ◆ There were no significant differences between the science scores of pupils in England and Singapore in TIMSS or in SISS.
- ◆ As in previous studies, pupils in Korea (SISS, IAEP2) and Japan (FISS, SISS) achieved higher mean science scores than those in England in TIMSS.

4.3 Comparisons of nine-year-olds

The preceding sections have compared the performance of pupils in England with that of pupils in the other countries that participated in TIMSS. In the rest of this chapter, comparisons will be limited to the nine countries identified in Chapter 1.

The rank order of median scores of nine-year-olds in England and the selected eight countries is shown in Table 4.3, together with the median scores for the international third grade (Year 4 in England). The median

science scores for nine-year-olds were estimated in the same way as the median mathematics scores. A full explanation of how the medians were estimated is given in Chapter 2, section 2.3.

Table 4.3 Median science achievement in nine countries: *the scores of nine-year-olds compared with the scores of pupils in international third grade (Year 4 in England)*

COUNTRY	nine-year-olds		International third grade (Year 4 in England)		Mean age (years)
	Median score	Rank	Median score	Rank	
United States	535	1	515	2	9.2
Japan	529	2	524	1	9.4
Scotland	523	3	486	7	8.7
Canada	522	4	494	5	9.1
England	521	5	501	3	9.1
Netherlands	515	6	500	4	9.3
Norway	505	7	451	9	8.8
Singapore	500	8	489	6	9.3
Hungary	477	9	468	8	9.4

Source: Tables 1.2 and 1.8 and C.2 (Martin et al., 1997)

It is apparent from Table 4.3 that there are some similarities in the rank orders of the two sets of results.

4.4 Test-curriculum Matching Analysis

The Test-curriculum Matching Analysis was designed and carried out to investigate the extent of the match between the curriculum covered within the TIMSS science tests for pupils in the two grades tested (Years 4 and 5 in England) and the science curriculum within participating countries for the same two grades, and to show how pupils' performance varied when based only on test items that were judged to be relevant to their own curriculum.² It was carried out in the same way as the Test-curriculum Matching Analysis for mathematics described in Chapter 2, section 2.4.

The scores of pupils in the lower grade (Year 4 in England) on the sub-set of items selected by the science curriculum experts in England as being relevant to the National Curriculum for this age are shown in Table 4.4.1

² Because there may also be curriculum areas covered in some countries that are not covered by the TIMSS tests, the Test-curriculum Matching Analysis does not provide complete information about how well the TIMSS tests cover the curricula of the participating countries. The purpose of the Test-curriculum Matching Analysis was to investigate whether or not the content covered by the actual test items was in the intended curricula of participating countries. (Beaton et al., 1996b)

for the nine selected countries, together with their respective scores on the entire science test. Similar information for the upper grade (Year 5 in England) is shown in Table 4.4.2. The complete matrices, showing how pupils in every country performed on their own and other countries' selected sub-sets of items are reproduced from the international report in Annex B at the end of this report.

Table 4.4.1 Science Test-curriculum Matching Analysis results for international third grade (Year 4 in England): the scores of pupils in nine selected countries on the sub-set of items selected by curriculum experts in England as addressing the National Curriculum compared with their scores on the entire science test

COUNTRY	Mean percent correct: for all 105 items %	Rank	Mean percent correct: for 32 items selected by England %	Rank	Difference in mean percent correct %
Japan	61	1	68	1	7
United States	56	2	62	2	6
Netherlands	56	2	62	2	6
England	55	4	60	5	5
Singapore	53	5	62	2	9
Canada	53	5	58	6	5
Scotland	51	7	57	7	6
Hungary	50	8	55	8	5
Norway	46	9	49	9	3
International mean	51		56		5

Some of the 97 items in the TIMSS science tests had two parts. The total number of score points available for analysis was therefore 105.

Differences may be subject to rounding errors.

Source: Table B.2 (Martin et al., 1997)

Table 4.4.2 Science Test-curriculum Matching Analysis results for international fourth grade (Year 5 in England): the scores of pupils in nine selected countries on the sub-set of items selected by curriculum experts in England as addressing the National Curriculum compared with their scores on the entire science test

COUNTRY	Mean percent correct: for all 105 items	Rank	Mean percent correct: for 53 items selected by England	Rank	Difference in mean percent correct
	%		%		%
Japan	70	1	71	1	1
Netherlands	67	2	68	3	1
United States	66	3	69	2	3
Singapore	64	4	66	4	2
Canada	64	4	64	6	0
England	63	6	65	5	2
Hungary	62	7	62	7	0
Scotland	60	8	62	7	2
Norway	60	8	61	9	1
International mean	60		60		0

Some of the 97 items in the TIMSS science tests had two parts. The total number of score points available for analysis was therefore 105.

Differences may be subject to rounding errors.

Source: Table B.1 (Martin et al., 1997)

As with mathematics, pupils in England achieved higher mean percentages of correct responses on the sub-sets of science items selected as being relevant to the National Curriculum for Years 4 and 5 than they achieved on the entire TIMSS science test of 105 items. However, the differences in performance were not considerable: five percentage points for the younger pupils and only two percentage points for the older pupils. The mean percentages of correct responses achieved by both lower and upper grade pupils in other countries also improved when calculated on the sub-sets of items selected by science experts in England, so that, when the rank orders are compared (i.e. based on the full set of science items, and based on the sub-set chosen as appropriate to pupils in England), the two orders are very similar. These results suggest that the TIMSS science test results provide a reasonable basis for comparing achievement in the participating countries: this is not surprising as one of the main considerations during the development of the tests was to make them as fair as possible to pupils in all participating countries.

The number of items selected by each of the nine countries as being relevant to their own curriculum are shown below for both grades tested (equivalent to Years 4 and 5 in England). To date, there is no more detailed information about the specific items selected and rejected by different countries.

	International third grade (Year 4 in England)	International fourth grade (Year 5 in England)
United States	105	105
Canada	64	93
Norway	49	62
Hungary	41	49
Netherlands	34	68
Scotland	33	45
England	32	53
Singapore	32	52
Japan	17	29
International test	105	105

Source: Tables B.1 and B.2 (Martin et al., 1997)

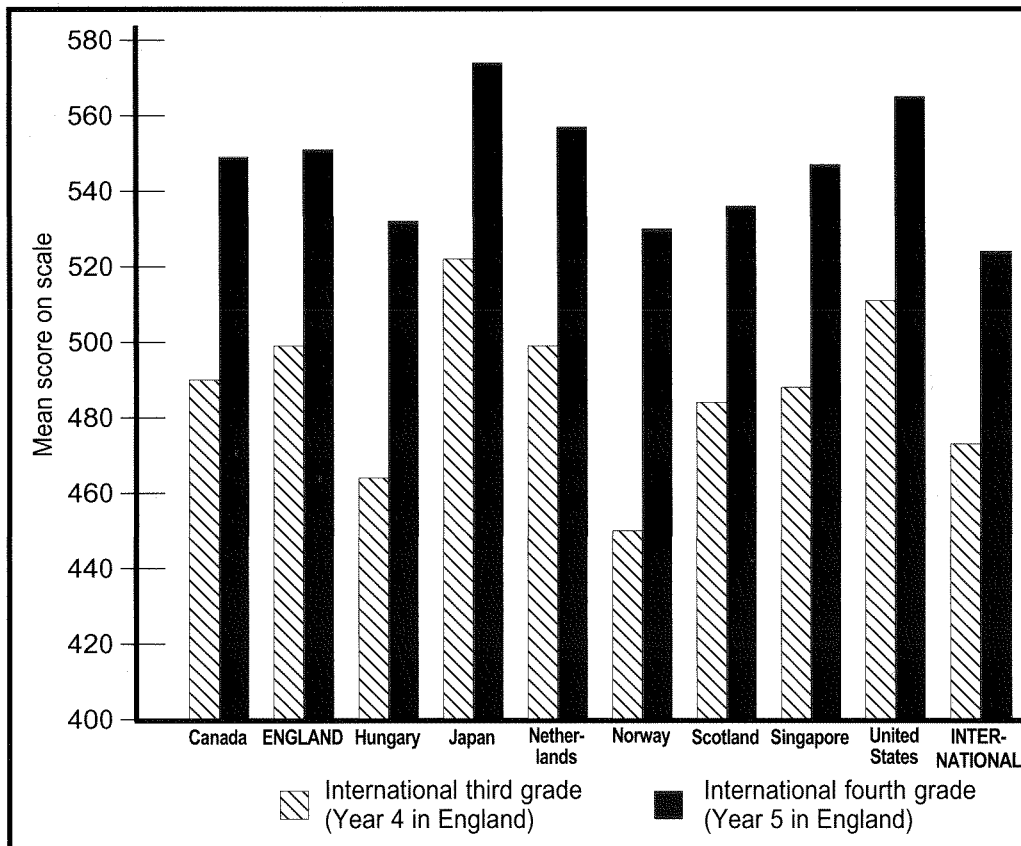
It is apparent that, at both grades tested, there was considerable variation in the numbers of items deemed relevant to countries' respective science curricula: only two of the nine countries selected more than half of the total number of science items for the lower grade tested, although five countries selected at least half of the items for the upper grade. This suggests that the content of participating countries' science curricula varied even more than their mathematics curricula (see Chapter 2, section 2.4). However, it is difficult to determine whether or not there is a positive association between curriculum coverage and achievement. The highest numbers of science items were selected by the United States and Canada: the mean overall science scores of pupils in both the lower and upper grades in these two countries were above the international means by at least 20 scale points. However, the mean overall science scores of pupils in Years 4 and 5 in England also exceeded the international means (by 29 scale points and by 28 scale points respectively) despite far fewer items having been selected by science experts as being relevant to the curriculum for these two year groups. Furthermore, the numbers of science items selected for each grade by curriculum experts in Japan were the lowest within the nine countries, yet the overall mean science scores of pupils in Japan (both grades tested) were the highest of the nine countries.

4.5 Comparisons between nine countries: differences in overall mean science scores between international third and fourth grades (Years 4 and 5 in England)

The difference in the mean science scores of lower and upper grade pupils in each country was used as a proxy measure for the progress achieved by pupils between the third and fourth grades (the progress score).

Figure 4.5 shows the differences in the mean science scores for pupils in both grades tested in each of the selected countries. In each country the mean score of pupils in the upper grade was higher than that of pupils in the lower grade; differences in mean scores ranged from 52 to 80 scale points. The increase in mean score shown by pupils in England was 52 scale points; this was close to the increase in mean score from the international third grade to the international fourth grade (51 scale points), but somewhat lower than the median increase in mean score for the nine selected countries (58 scale points).

Figure 4.5 Comparisons between nine countries: difference in overall mean science score of international third and fourth grade pupils (Years 4 and 5 in England)



Differences may be subject to rounding errors.

Source: Tables 1.1, 1.2 and 1.3 (Martin et al., 1997)

As with the mathematics analyses, it is difficult to find a consistent pattern in these results. The two countries in which lower grade pupils achieved mean scores which were below the international mean (Norway and Hungary) showed the highest progress scores of the selected countries. Pupils in Japan and the United States (the two highest-scoring countries at both the lower and the upper grades) showed relatively low progress scores. On the other hand, the mean scores of pupils in Scotland were relatively low in the ranking for both the lower and upper grades, and their progress score was also relatively low. Analyses of the science curricula in these countries may suggest some of the reasons for differences in progress scores.

4.6 International marker levels

One of the questions which TIMSS was designed to answer was: how do the proportions of very high-scoring pupils and very low-scoring pupils vary from country to country? In order to answer this question, international marker levels were calculated, which correspond to the percentiles computed from the combined data from all participating countries. The marker levels are shown below.

	International third grade (Year 4 in England)	International fourth grade (Year 5 in England)
• Top 10 per cent benchmark (90th percentile)	610	660
• Top 25 per cent benchmark (75th percentile)	554	607
• Top half benchmark (50th percentile or median)	488	541
• Bottom 25 per cent benchmark (25th percentile)	413	466

Source: Tables 1.4 and 1.5 (Martin et al., 1997) and additional analysis provided by the International Study Center

Table 4.6 presents information from the international report showing the percentages of international fourth grade pupils (Year 5 in England) in each of the selected countries in the top ten per cent and the top half of the international distribution. The pattern was generally similar for the international third grade (Year 4 in England).

The proportion of pupils in England that reached the international top ten per cent marker level was higher than in seven of the other eight selected countries (Table 4.6). However, when comparisons are made in terms of the proportions of pupils reaching the top half marker level, pupils in four of the selected countries equalled or exceeded the performance of pupils in England (Japan, Netherlands, United States and Canada). The pattern was quite similar for the international third grade (Year 4 in England).

Table 4.6 Comparisons between nine countries: percentages of pupils reaching international benchmarks (top half and top ten per cent) of science achievement in international fourth grade (Year 5 in England)

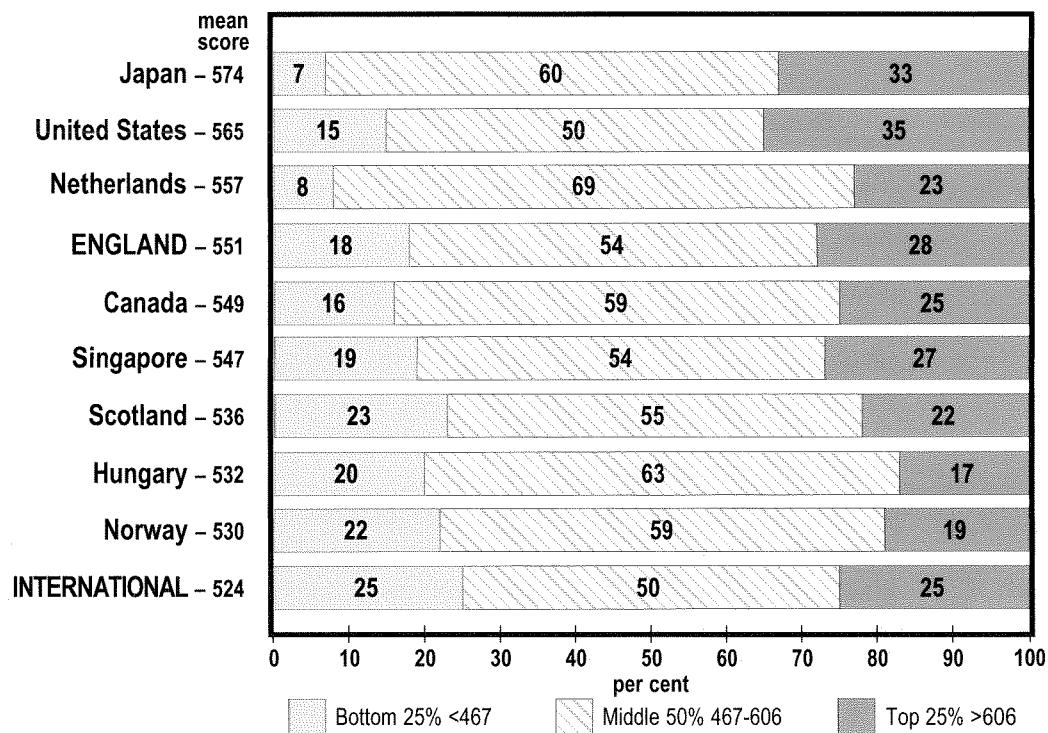
COUNTRY	Top half		Top ten per cent	
	%	Rank	%	Rank
Japan	68	1	11	3
United States	63	2	16	1
Netherlands	60	3	5	8
England	55	4	13	2
Canada	55	4	9	5
Singapore	54	6	11	3
Scotland	50	7	9	5
Hungary	48	8	5	8
Norway	47	9	6	7

Differences may be subject to rounding errors

Source: Table 1.4 (Martin et al., 1997)

Figure 4.6 shows the percentages of pupils in the upper grade (Year 5 in England) in each of the selected countries that were in the top 25 per cent, middle 50 per cent and bottom 25 per cent of the international distribution. The percentages for the lower grade (Year 4 in England) showed the same general pattern.

Figure 4.6 Percentages in the bottom 25%, middle 50% and top 25% of the international distribution for science (international fourth grade : Year 5 in England): comparisons between nine countries



Source: Table 1.4 (Martin et al., 1997) and additional analyses provided by the International Study Center

It can be seen that, in Figure 4.6, there is a very close correlation between the proportion of pupils in the international top 25 per cent and a country's mean science score. Several points are illustrated by Figure 4.6:

- ◆ England had 28 per cent of pupils in the top 25 per cent and 18 per cent in the bottom 25 per cent, reflecting the fact that the mean science score of pupils in England was above the international mean.
- ◆ Comparisons between England and the Netherlands show that the proportion of pupils in the top and bottom 25 per cent were higher in England illustrating that there are more high scoring and low scoring pupils in England than the Netherlands.
- ◆ The mean score of pupils in England was higher than the mean score of those in Hungary (551 compared with 532), yet the proportions of pupils in the bottom 25 per cent were similar (18 and 20 per cent). This is because the proportion of pupils in the international top 25 per cent was higher in England than in Hungary (28 per cent compared with 17 per cent).

4.7 Comparisons between nine countries: gender differences in overall mean science scores

Gender differences in science performance are shown in Table 4.7 for pupils in the upper grade in the selected countries. The results for the lower grade were generally similar, although none of the differences was greater than 13 scale points.

Table 4.7 Comparisons between nine countries: gender differences in overall mean science score for international fourth grade (Year 5 in England)

COUNTRY	Boys' Mean	Girls' Mean	Difference in favour of boys
Netherlands	570	544	26*
Hungary	539	525	14*
Japan	580	567	14*
United States	571	560	12*
Canada	553	545	8
Norway	534	526	8
England	555	548	7
Scotland	538	533	4
Singapore	549	545	4

*Gender difference statistically significant at the five per cent level.

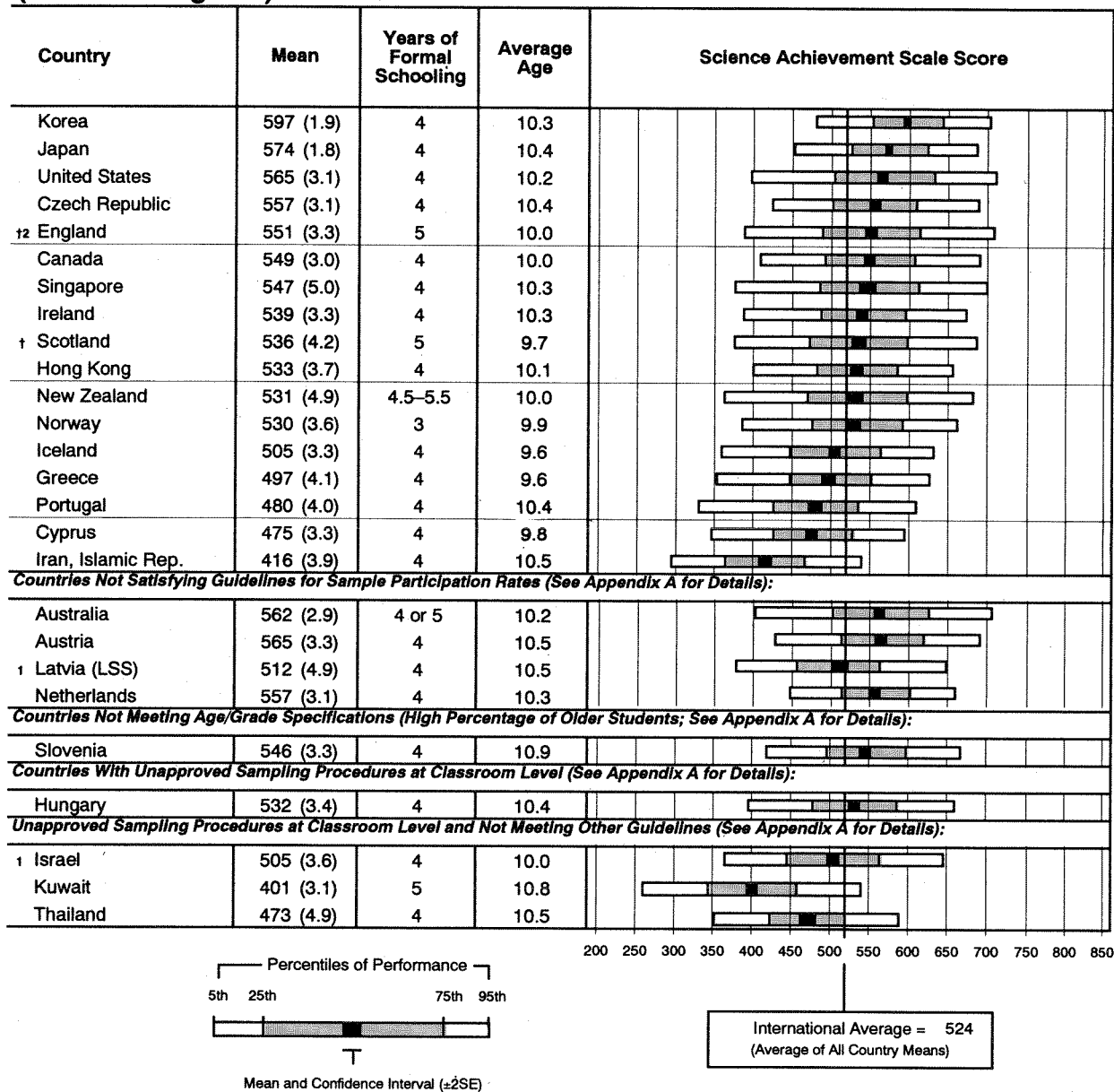
Differences may be subject to rounding errors.

Source: Table 1.6 (Martin et al., 1997)

In each of the selected countries, the mean scores of the boys were higher than the girls' in both the lower and the upper grades. However, the differences in favour of the boys were not statistically significant in most of the countries in both grades tested. The differences between the boys' and girls' mean scores in England were eight scale points in Year 4 and seven scale points in Year 5, neither of which was statistically significant. Statistically significant differences favouring boys were found in the Netherlands, Hungary, Japan and the United States.

The results for both grades presented in the international report (Martin *et al.*, 1997) also show the majority of gender differences in favour of boys, although at both grades tested, these differences were statistically significant in fewer than half of the countries.

Table 1.1 (Science)
Distributions of achievement in the sciences—international fourth grade*
(Year 5 in England)



*Fourth grade in most countries.

¹Met guidelines for sample participation rates only after replacement schools were included (see Appendix A for details).

[†]National Desired Population does not cover all of International Desired Population (see Table A.2). Because coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.

²National Defined Population covers less than 90 percent of National Desired Population (see Table A.2).

() Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

SOURCE: Martin et al., 1997




Figure 1.1 (Science)

**Multiple comparisons of achievement in the sciences—international fourth grade*
(Year 5 in England)**

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 two countries.¹

Country	Korea	Japan	United States	Austria	Australia	Netherlands	Czech Republic	England	Canada	Singapore	Slovenia	Ireland	Scotland	Hong Kong	Hungary	New Zealand	Norway	Latvia (LSS)	Israel	Iceland	Greece	Portugal	Cyprus	Thailand	Iran, Islamic Rep.	Kuwait
Korea		▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Japan	▼		●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
United States	▼	●		●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Austria	▼	●	●		●	●	●	●	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Australia	▼	▼	●	●		●	●	●	▲	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Netherlands	▼	▼	●	●	●		●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Czech Republic	▼	▼	●	●	●	●		●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
England	▼	▼	▼	●	●	●	●		●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Canada	▼	▼	▼	▼	●	●	●	●		●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Singapore	▼	▼	▼	●	●	●	●	●	●		●	●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲
Slovenia	▼	▼	▼	▼	▼	●	●	●	●	●		●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Ireland	▼	▼	▼	▼	▼	▼	▼	●	●	●	●		●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲
Scotland	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●		●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲
Hong Kong	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●		●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲
Hungary	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●		●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲
New Zealand	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●		●	●	▲	▲	▲	▲	▲	▲	▲	▲
Norway	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	▼	●	●	●	●	●		●	▲	▲	▲	▲	▲	▲	▲	▲
Latvia (LSS)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●		●	●	●	▲	▲	▲	▲	▲
Israel	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●		●	●	▲	▲	▲	▲	▲
Iceland	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●		●	▲	▲	▲	▲	▲
Greece	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●		●	▲	▲	▲	▲
Portugal	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●		●	▲	▲	▲
Cyprus	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●		●	▲	▲
Thailand	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●		▲	▲
Iran, Islamic Rep.	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼		●
Kuwait	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●

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

 Mean achievement significantly higher than comparison country
  No statistically significant difference from comparison country
  Mean achievement significantly lower than comparison country

*Fourth grade in most countries.

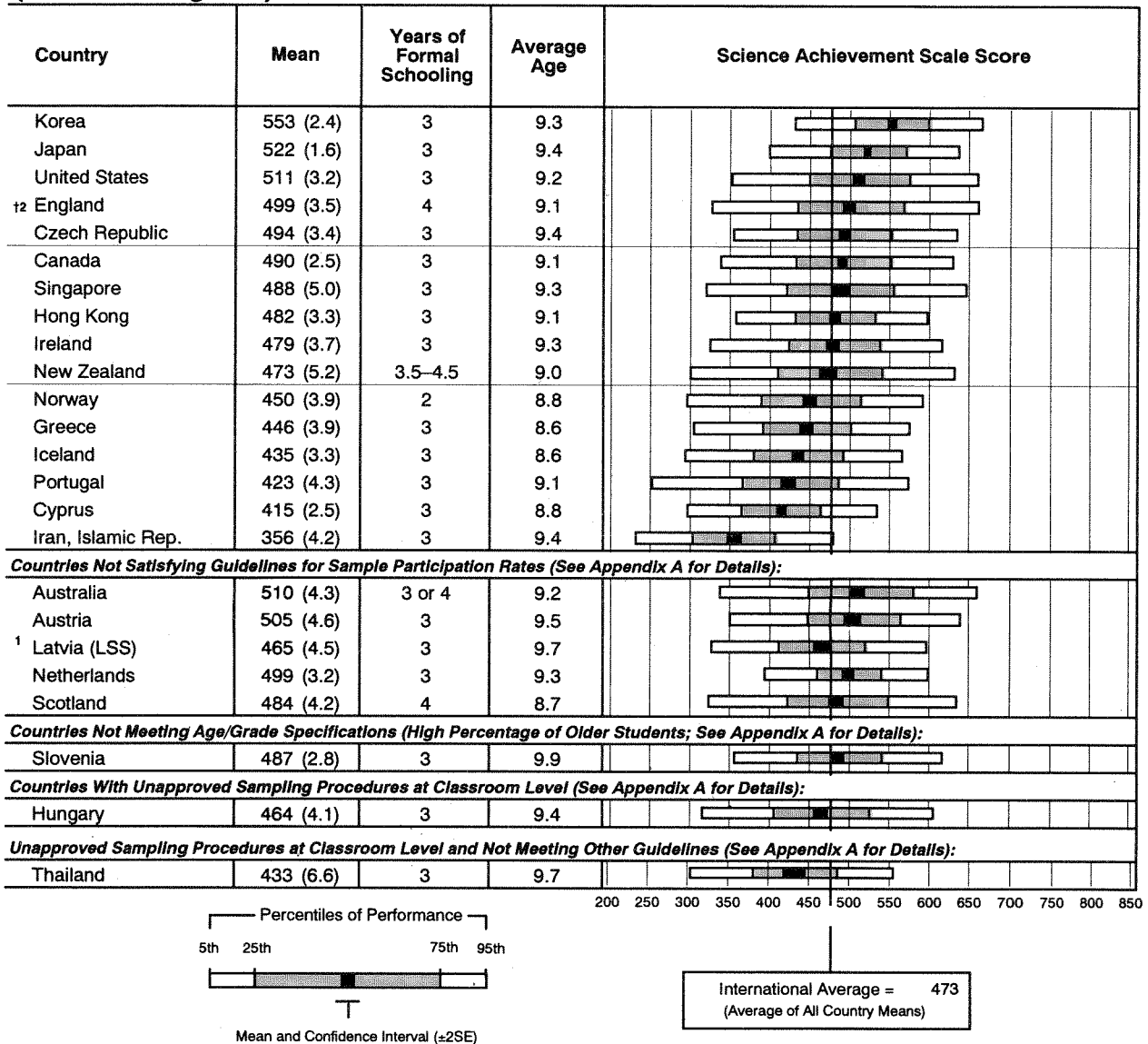
¹Statistically significant at .05 level, adjusted for multiple comparisons.

Because coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools 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: Martin et al., 1997

Table 1.2 (Science)
Distributions of achievement in the sciences—international third grade*
(Year 4 in England)



*Third grade in most countries.

¹Met guidelines for sample participation rates only after replacement schools were included (see Appendix A for details).

¹²National Desired Population does not cover all of International Desired Population (see Table A.2). Because coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.

²National Defined Population covers less than 90 percent of National Desired Population (see Table A.2).

() Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

SOURCE: Martin et al., 1997




Figure 1.2 (Science)

**Multiple comparisons of achievement in the sciences—international third grade*
(Year 4 in England)**

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 two countries.[†]

Country	Korea	Japan	United States	Australia	Austria	England	Netherlands	Czech Republic	Canada	Singapore	Slovenia	Scotland	Hong Kong	Ireland	New Zealand	Latvia (LSS)	Hungary	Norway	Greece	Iceland	Thailand	Portugal	Cyprus	Iran, Islamic Rep.
Korea		▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Japan	▼		●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
United States	▼	●		●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Australia	▼	●	●		●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Austria	▼	▼	●	●		●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
England	▼	▼	●	●	●		●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Netherlands	▼	▼	●	●	●	●		●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Czech Republic	▼	▼	▼	●	●	●	●		●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Canada	▼	▼	▼	▼	●	●	●	●		●	●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲
Singapore	▼	▼	▼	▼	●	●	●	●	●		●	●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲
Slovenia	▼	▼	▼	▼	▼	●	●	●	●	●		●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲	▲
Scotland	▼	▼	▼	▼	▼	●	●	●	●	●	●		●	●	●	●	▲	▲	▲	▲	▲	▲	▲	▲
Hong Kong	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●		●	●	●	▲	▲	▲	▲	▲	▲	▲	▲
Ireland	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●		●	●	▲	▲	▲	▲	▲	▲	▲	▲
New Zealand	▼	▼	▼	▼	▼	▼	▼	●	●	●	●	●	●	●		●	●	▲	▲	▲	▲	▲	▲	▲
Latvia (LSS)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●	●		●	●	▲	▲	▲	▲	▲	▲
Hungary	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●		●	▲	▲	▲	▲	▲	▲
Norway	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●		●	●	●	▲	▲	▲
Greece	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●		●	●	▲	▲	▲
Iceland	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●		●	●	▲	▲
Thailand	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●	●		●	●	▲
Portugal	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●		●	▲
Cyprus	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	●	●		▲
Iran, Islamic Rep.	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	

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

 Mean achievement significantly higher than comparison country
  No statistically significant difference from comparison country
  Mean achievement significantly lower than comparison country

*Third grade in most countries.

[†]Statistically significant at .05 level, adjusted for multiple comparisons.

Because coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools 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: Martin et al., 1997

CHAPTER 5

The pupils' performance on the science topics

Summary of main points

Compared with the international means for each of the four reporting categories for science, pupils in England showed:

- ◆ higher scores both in Year 4 and in Year 5 in all four content areas: *Earth science; Life science; Physical science and Science and environment.*

These results are in line with the results for 13-year-old students in England, in which the mean scores of students in Years 8 and 9 were above the international means for all five science content areas for 13-year-olds.

With regard to gender differences in achievement:

- ◆ the boys' mean scores were equal to or higher than those of the girls in both Years 4 and 5 for each reporting category with only one exception: Year 5 girls achieved a higher score in *Science and environment*
- ◆ only two gender differences were statistically significant: the differences in favour of boys in both Year 4 and Year 5 in *Earth science*.

5.1 Preface

The science items used in the TIMSS tests were matched to the curriculum frameworks which were drawn up to ensure that the tests covered a wide range of the science curriculum relevant to pupils in the international third and fourth grades (Years 4 and 5 in England). The curriculum frameworks for TIMSS are described in detail in Robitaille (1993). The total set of science items has been organised into four reporting categories, each of which represents a different aspect of the science curriculum. These categories are listed below, together with the number of items within each category and the nearest corresponding attainment targets from the 1995 National Curriculum Order for science¹.

Reporting category ²	No. of items	Attainment target(s)
Earth science	17	Sc3: Materials and their Properties Sc4: Physical Processes
Life science	41	Sc2: Life Processes and Living Things
Physical science	30	Sc4: Physical Processes Sc3: Materials and their Properties
Science and environment	9	Sc4: Physical Processes Sc1: Experimental and Investigative Science
Total 97		

The majority of items focused on areas of science knowledge that are linked to Sc2–4, although a few of the items within the *Science and environment* category addressed skills and concepts related to Sc1, such as planning a fair test. It should be acknowledged, however, that pupils' activities in connection with this part of the National Curriculum for science would usually involve more practical and investigative work than is possible within the format of written tests.

¹ At the time of testing (February–March 1995), the 1991 Statutory Orders for science were still in effect. The attainment targets at that time were as follows:

Sc1: Scientific investigation

Sc2: Life and living processes

Sc3: Materials and their properties

Sc4: Physical processes.

The expert panels convened for the Test-curriculum Matching Analysis (which included professional officers from the School Curriculum and Assessment Authority and inspectors from the Office for Standards in Education as well as practising teachers) considered that at the time of testing schools were more likely to be working to the 1995 Statutory Order than the 1991 Order.

² The reporting categories were broadly similar to those for Population 2 (13-year-olds) which were as follows:

Earth science

Life science

Physics

Chemistry

Science and environment.

This chapter presents information about the mean scores of pupils in England for each of the four science reporting categories and draws comparisons with the performance of pupils in the eight other countries listed in Chapter 1. Separate comparisons are made for each of the following for each reporting category:

- ◆ comparisons of pupils' mean percentages of correct answers
- ◆ differences in mean scores between Year 4 and Year 5 (a proxy measure for progress in achievement)
- ◆ gender differences in mean scores for both Year 4 and Year 5: these are presented for all four science content areas in the final section of this chapter.

Within each of the following sections relating to science reporting categories, illustrative examples of items are presented after the discussion of mean scores. For each item shown, the percentages of pupils making correct responses are given for England and the countries with the median and highest scores (see Martin *et al.*, 1997), together with the international percentage correct. Details of the international coding guidelines for free response items are included where appropriate. Additional information, for example, relating to incorrect answers that were frequently given, is provided for each illustrative item. Each of the examples was one of those selected by the national panel of science experts as being representative of the National Curriculum for science for both years tested (Years 4 and 5 in England) unless otherwise stated.

Within this chapter, comparisons are made using the mean percentages of correct responses³ attained by pupils in the nine selected countries and differences are expressed in percentage points. This is in contrast to Chapter 4, where scores on a scale were quoted, and differences were expressed in scale points. Differences of five percentage points in the mean percentages of correct responses have been taken as evidence of real differences in performance.

³ Pupils who omitted to answer specific items were included in the percentage of pupils offering incorrect responses. Full details are provided in the international report (Beaton *et al.*, 1996b).

5.2 Earth science

The mean percentages of correct answers achieved by pupils in the international third and fourth grades (equivalent to Years 4 and 5 respectively in England) are shown for England and selected other countries in Table 5.2, together with the international means and the differences between the mean scores of pupils in the two grades.

Table 5.2 Mean percentages of correct answers for *Earth science* in nine countries (17 items)

COUNTRY	Mean score of international third grade (Year 4 in England) %	Mean score of international fourth grade (Year 5 in England) %	Difference between Years 4 and 5
Japan	58	66	8
United States	55	64	9
England	53	61	8
Canada	52	62	10
Singapore	51	58	7
Scotland	50	58	8
Netherlands	49	61	12
Hungary	47	62	15
Norway	47	60	13
International mean	49	57	8

Source: Tables 2.1 and 2.2 (Martin et al., 1997)

Comparisons with previous studies

Four of the other selected countries took part in IAEP2. There has been very little change in the relative position of England compared with these countries for the upper grade (the grade tested in IAEP2).

Five of the other countries took part in SISS. The relative position of England as compared with Canada and Hungary appears to have improved slightly; students in both of these countries performed at the same level as those in England in TIMSS, whereas in SISS they had outperformed students in England.

5.2.1 Comparisons of total topic scores: *Earth science*

The mean scores of pupils in Years 4 and 5 in England were both above the international means. For both the lower and the upper grades, in only one of the selected countries (Japan) were pupils' mean scores at least five percentage points higher than those of pupils in England.

5.2.2 Differences in mean scores between international third grade and international fourth grade (Years 4 and 5 in England): *Earth science*

In each of the selected countries, pupils in the upper grade achieved higher mean scores than pupils in the lower grade. The differences ranged from seven to 15 percentage points (see Table 5.2).

5.2.3 Examples of items for *Earth science*

Example items N1, Z1 and Y1 are shown below, together with national and international percentages correct. In the opinion of the Test-curriculum Matching Analysis panel (see Chapter 1, section 1.2.4 and Chapter 2, section 2.2.4) none of the released items was considered to be representative of the National Curriculum for Year 4 pupils; each of these examples was, however, considered to be representative of the National Curriculum for Year 5 pupils.

Item N1

N1. This table shows the temperature and amount of rain or snow in four different towns on the same day.

	Town A	Town B	Town C	Town D
Lowest Temperature	13°C	-9°C	22°C	-12°C
Highest Temperature	25°C	-1°C	30°C	-4°C
Rain or Snow	0 cm	5 cm	2.5 cm	0 cm

Where did it snow?

- A. Town A
- B. Town B
- C. Town C
- D. Town D

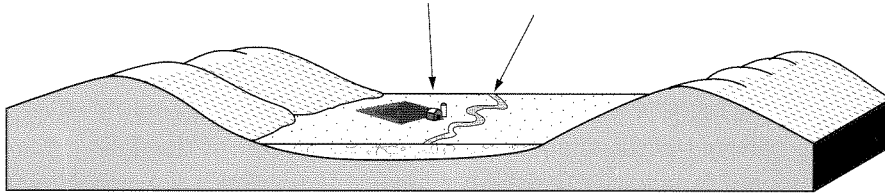
N1	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	27.1	31.9
International	31.7	43.7
Median country	30.6	43.9
Highest-scoring country	49.4	68.8

Approximately one-third of pupils in both Years 4 and 5 in England selected Option (C): 35.6 per cent of the younger pupils and 31.2 per cent of the older pupils.

* * * * *

Item Z1

- Z1. The diagram shows a river flowing through a wide plain. The plain is covered with several layers of soil and sediment.



- a. Write down one reason why this plain is a good place for farming.

The ~~can~~ animals can get a drink from the river.

- b. Write down one reason why this plain is NOT a good place for farming.

If the river floods it will flood the farm.

Z1A good place	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	62.7	73.6
International	47.8	62.2
Median country	49.2	63.4
Highest-scoring country	81.0	91.2

Z1B bad place	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	31.7	40.1
International	15.8	22.7
Median country	17.0	22.6
Highest-scoring country	31.7	44.6

The international marking guidelines stated that acceptable reasons for part (a) should include references to either:

- i. plenty of good/fertile soil
- ii. the river providing water for animals and crops, or
- iii. plenty of level ground to make farming crops easier.

Other reasons which seemed plausible were also acceptable.

The correct answer most frequently mentioned was (ii), which was mentioned in the responses of 39.6 per cent of Year 4 and 47.6 per cent of Year 5 pupils in England.

Acceptable reasons for part (b) included references to either:

- i. the possibility of the river flooding, or the soil being too wet; or
- ii. the possibility of erosion due to the action of the wind or water.

Again, additional reasons suggested by pupils were accepted as correct if they seemed plausible (e.g. the difficulties of using farm machinery on steep slopes).

As in the first part of the question, most of the correct answers offered by pupils in England focused on the river: 20.2 per cent of Year 4 and 26.8 per cent of Year 5 pupils mentioned the possibility of flooding.

* * * * *

Item Y1

Y1. The Sun is bigger than the Moon, but they look about the same size when you look at them from the Earth. Why is this?

is smaller than the sun but it looks bigger because the moon is closer to the earth so it

Y1	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	26.2	35.2
International	21.1	30.5
Median country	21.9	33.1
Highest-scoring country	33.4	45.6

The international marking guidelines stated that, to be correct, answers should mention that the Moon is closer to the Earth than the Sun or that the Sun is further away. An incorrect answer that was widespread in both Years 4 and 5 in England was a general reference to ‘a long distance away’ but without any comparison between the two distances. This type of answer was given by 24 per cent of the pupils in both year groups.

This item was also included in the tests for 13-year-olds (Population 2). Information on the percentages of correct responses achieved by students nationally and internationally in the international seventh and eighth grades (equivalent to Years 8 and 9 respectively in England) is presented for comparative purposes below, although it should be remembered that more countries took part in the survey of 13-year-olds than surveyed nine-year-olds (see Appendix II).

014 (Pop 2) equivalent to Y1 (Pop 1)	Percentage correct for international seventh grade (Year 8 in England) %	Percentage correct for international eighth grade (Year 9 in England) %
England	56.4	59.7
International	51.4	56.9
Median country	53.8	61.2
Highest-scoring country	78.2	79.0

* * * * *

5.3 Life science

The mean percentages of correct answers achieved by pupils in the two age groups tested (equivalent to Years 4 and 5 in England) are shown for England and selected other countries in Table 5.3, together with the international means and the differences between the lower and upper grades.

Table 5.3 Mean percentages of correct answers for *Life science* in nine countries (41 items)

COUNTRY	Mean score of international third grade (Year 4 in England) %	Mean score of international fourth grade (Year 5 in England) %	Difference between Years 4 and 5
Japan	65	73	8
Netherlands	63	73	10
United States	62	71	9
England	60	68	8
Canada	59	68	9
Singapore	58	70	12
Scotland	57	65	8
Hungary	55	66	11
Norway	52	67	15
International mean	55	64	9

Source: Tables 2.1 and 2.2 (Martin et al, 1997)

Comparisons with previous studies

Four of the other selected countries took part in IAEP2. There has been very little change in the relative position of England compared with these countries for the upper grade (the grade tested in IAEP2).

Five of the other countries took part in SISS. The relative position of England compared with Canada and Hungary appears to have improved slightly; pupils in both these countries performed at the same level as those in England in TIMSS, whereas in SISS they had outperformed pupils in England.

5.3.1 Comparisons of total topic scores: *Life science*

The mean scores of pupils in England were above the international means for both age groups tested. The difference between the lowest and the highest mean scores for the selected countries was greater at the younger age group (13 percentage points) than for the older pupils (eight percentage points). Only pupils in Japan achieved scores which were five percentage points higher than those of pupils in Year 4 in England, although pupils in both Japan and the Netherlands achieved scores which were five percentage points higher than those of pupils in Year 5 in England.

5.3.2 Differences in mean scores between international third grade and international fourth grade (Years 4 and 5 in England): *Life science*

There was an increase in the performance achieved by pupils in the upper grade as compared with the lower grade in each of the selected countries. The difference ranged from eight to 15 percentage points (see Table 5.3).

5.3.3 Examples of items for *Life science*

Example items N2, Y2, P9, X4 and Z2 are shown below, together with the national and international percentages correct.

Item N2

<p>N2. Which one of these includes only living things?</p> <p>A. clouds, fire, rivers</p> <p>B. fire, rivers, trees</p> <p>C. rivers, birds, trees</p> <p><input checked="" type="radio"/> D. birds, trees, worms</p> <p>E. trees, worms, clouds</p>
--

N2	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	70.8	79.2
International	63.4	74.4
Median country	66.0	76.2
Highest-scoring country	85.7	94.0

The incorrect answer most frequently chosen by pupils in Year 4 and 5 in England was option (C), selected by 11.5 per cent of the younger pupils and 12.4 per cent of the older pupils.

* * * * *

Item Y2

Y2. Write down two changes that take place in children's bodies as they become adults.

When children get bigger
 there small teeth fall
 out and bigger ones form
 As children get older
 they get taller.

Y2 first change	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	60.1	70.2
International	52.5	63.9
Median country	53.8	66.2
Highest-scoring country	83.4	91.0

Y2 second change	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	29.0	42.2
International	27.5	39.1
Median country	26.1	42.3
Highest-scoring country	51.3	60.4

The international marking guidelines indicated a range of acceptable answers, such as:

- ◆ references to growth, such as increased height, weight
- ◆ references to sexual maturity
- ◆ references to other appropriate changes, such as the growth of wisdom teeth.

In both Years 4 and 5 in England, most pupils who gave correct answers focused on growth-related changes as both the first and the second change that they identified (the marking guidelines allowed responses mentioning ‘height’ and ‘weight’ to count as two separate acceptable changes). As the first change, 45.4 per cent of the younger pupils and 51.1 per cent of the older pupils referred to growth; 15.9 per cent of Year 4 and 21.3 per cent of Year 5 pupils referred to growth as their second change.

More of the older pupils referred to changes related to sexual maturity: 9.2 per cent of Year 5 pupils mentioned this as their first change and 12.0 per cent as their second change, whereas 5.0 per cent of Year 4 pupils mentioned sexual maturity as their first change and 4.9 per cent gave this as their second change.

* * * * *

Item P9

P9. Seeds develop from which part of a plant?

(A) Flower

B. Leaf

C. Root

D. Stem

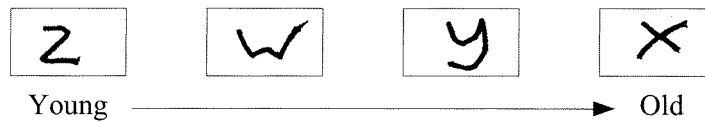
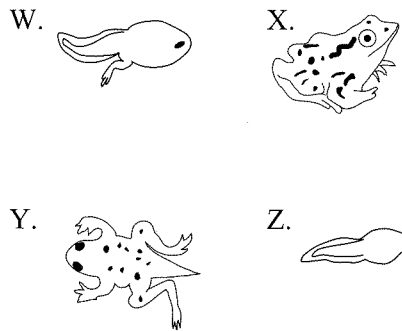
P9	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	28.6	34.7
International	38.6	45.8
Median country	31.8	41.1
Highest-scoring country	67.0	78.8

The majority of pupils in both Year 4 and Year 5 in England chose option (C) as their answer (52.6 per cent and 50.6 per cent respectively).

* * * * *

Item X4

X4. The pictures show different stages in a frog's growth.



Write the letters in the boxes to show the order of how a frog grows.

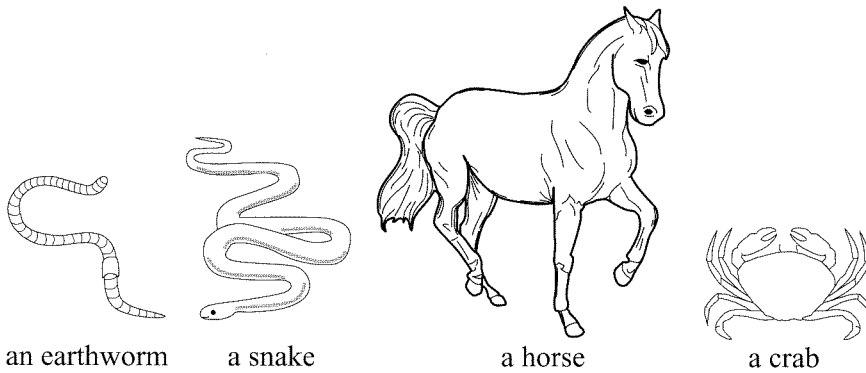
X4	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	81.5	90.6
International	71.0	82.6
Median country	76.0	88.0
Highest-scoring country	93.1	96.8

No distinctions were made between incorrect responses listing different orders.

* * * * *

Item Z2

Z2. Use the pictures to answer the questions. (Use each animal once only.)



An animal that has a hard outside skeleton is

a crab

An animal without a backbone that has many segments to its body is

an earthworm

An animal that has hair and an inside skeleton is

a horse

An animal that has a scaly skin and an inside skeleton is

a snake

Z2	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	60.4	72.3
International	51.8	62.2
Median country	52.6	66.8
Highest-scoring country	91.2	94.0

The majority of pupils in Years 4 and 5 in England completed this item correctly. However, 11.1 per cent of the younger pupils and 8.0 per cent of the older pupils had only two animals correct: the crab and the horse.

* * * * *

5.4 Physical science

The mean percentages of correct answers achieved by pupils in England and selected other countries are shown in Table 5.4, together with the international means for the two grades tested and the differences in mean scores between the lower and upper grades.

Table 5.4 Mean percentages of correct answers for *Physical science* in nine countries (30 items)

COUNTRY	Mean score of international third grade (Year 4 in England) %	Mean score of international fourth grade (Year 5 in England) %	Difference between Years 4 and 5
Japan	61	70	9
Netherlands	53	65	12
Singapore	52	64	12
England	52	60	8
United States	51	60	9
Canada	50	61	11
Hungary	48	59	11
Scotland	48	57	9
Norway	41	55	14
International mean	48	57	9

Source: Tables 2.1 and 2.2 (Martin et al., 1997)

Comparisons with previous studies

Four of the other selected countries took part in IAEP2. There has been very little change in the relative position of England compared with these countries for the upper grade (the grade tested in IAEP2).

Five of the other countries took part in SISS. The relative position of England compared with Canada and Norway appears to have improved slightly; pupils in both these countries performed at the same level as those in England in TIMSS, whereas in SISS they had outperformed pupils in England. It should be noted, however, that these comparisons were made between the categories *Physical sciences* in TIMSS and *Physics* in SISS.

5.4.1 Comparisons of total topic scores: *Physical science*

The mean scores of pupils in Year 4 and 5 in England were above the international means. In the lower grade, only pupils in Japan and the Netherlands achieved mean scores which were above the mean scores of Year 4 pupils in England. Pupils in four of the selected countries achieved mean scores which were higher than those of Year 5 pupils in England, although there were only two countries (Japan and the Netherlands) in which the differences were five percentage points or more.

5.4.2 Differences in mean scores between international third grade and international fourth grade (Years 4 and 5 in England): *Physical science*

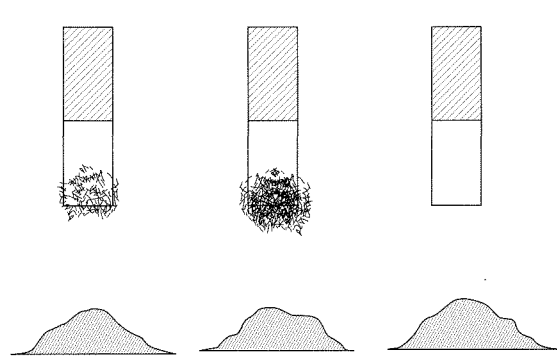
In all of the selected countries the scores of pupils in the upper grade were higher than those of pupils in the lower grade. The difference in pupils' performance in the two grades ranged from eight to 14 percentage points (see Table 5.4).

5.4.3 Examples of items for *Physical science*

Example items P5, Q3 and Q8 are shown below, together with the national and international percentages correct.

Item P5

P5. Each of the three magnets shown has been dipped into the substance below it. Which of the substances could be coffee?



Substance A Substance B Substance C

A. A only
 B. B only
 C. C only
 D. A and B only

P5	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	48.7	59.5
International	41.4	49.9
Median country	40.1	51.9
Highest-scoring country	74.9	83.6

The incorrect options which were most popular with pupils in England were (B) and (D). Option (B) was selected by 23.0 per cent of Year 4 and 16.2 per cent of Year 5 pupils, and option (D) was chosen by 22.2 per cent of Year 4 and 20.6 per cent of Year 5 pupils.

* * * * *

Item Q3

Q3. Some things were buried in wet ground. Several years later they were dug up. Which thing is MOST likely to have stayed the same?

A. An egg shell

B. A plastic cup

C. A paper plate

D. An orange peel

Q3	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	50.6	59.5
International	45.6	53.7
Median country	44.7	52.8
Highest-scoring country	69.4	76.2

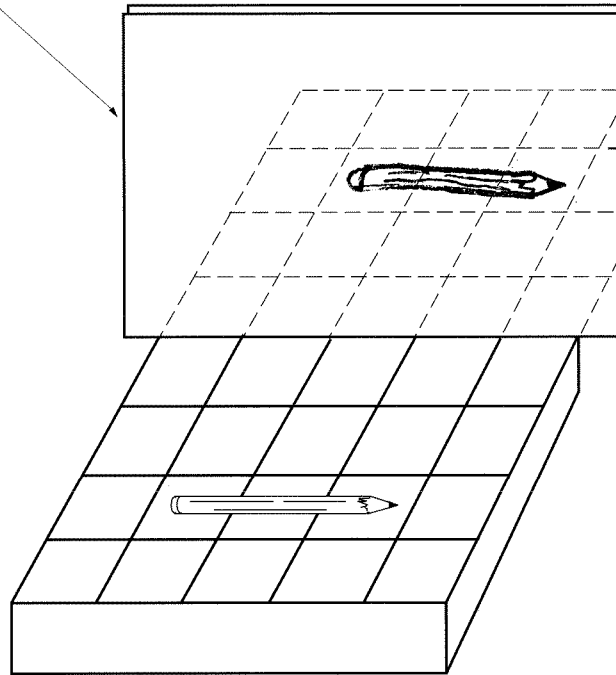
Option (A) was the most popular wrong answer selected by pupils in Year 4 and 5 in England, chosen by one-fifth of pupils in each year group: 19.7 per cent of the younger pupils and 19.6 per cent of the older pupils.

* * * * *

Item Q8

- Q8. The picture shows a pencil that is lying on a shelf in front of a mirror. Draw a picture of the pencil as you would see it in the mirror. Use the patterns of lines on the shelf to help you.

Mirror



Q8	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	43.7	61.3
International	36.9	47.1
Median country	41.0	51.5
Highest-scoring country	61.4	66.0

The most common incorrect response for Year 4 and Year 5 pupils in England placed the reflection of the pencil one row closer to the mirror (26.2 per cent of the younger pupils and 16.3 per cent of the older pupils).

This item was also included in the tests for 13-year-olds (Population 2). Information on the percentages of correct responses achieved by students nationally and internationally in the international seventh and eighth grades (equivalent to Years 8 and 9 respectively in England) is presented for comparative purposes below, although it should be remembered that more countries took part in the survey of 13-year-olds than surveyed nine-year-olds (see Appendix II).

PUPILS' PERFORMANCE ON SCIENCE TOPICS

<i>M14 (Pop 2) equivalent to Q8 (Pop 1)</i>	Percentage correct for international seventh grade (Year 8 in England) %	Percentage correct for international eighth grade (Year 9 in England) %
England	80.1	84.1
International	66.4	68.9
Median country	68.7	73.1
Highest-scoring country	83.4	85.2

* * * * *

5.5 Science and environment

The mean percentages of correct answers achieved by pupils in the two grades tested are shown for England and selected other countries in Table 5.5, together with the international means and the differences between the mean scores of pupils in the lower and upper grades.

Table 5.5 Mean percentages of correct answers for *Science and environment* in nine countries (nine items)

COUNTRY	Mean score of international third grade (Year 4 in England) %	Mean score of international fourth grade (Year 5 in England) %	Difference between Years 4 and 5
Japan	52	62	10
United States	52	65	13
Netherlands	48	61	13
England	45	56	11
Canada	44	56	12
Scotland	42	53	11
Singapore	41	53	12
Hungary	39	50	11
Norway	34	53	19
International mean	40	51	11

Source: Tables 2.1 and 2.2 (Martin et al., 1997)

5.5.1 Comparisons of total topic scores: *Science and environment*

Pupils in both Year 4 and Year 5 in England achieved mean scores which were above the international means for this topic. At the lower grade, only pupils in Japan and the United States achieved mean scores which were more than five percentage points higher than their counterparts in England (by seven percentage points in both cases). At the upper grade, there were three countries (United States, Japan and the Netherlands) in which pupils' performance was five percentage points or more above the mean score of pupils in Year 5 in England.

5.5.2 Differences in mean scores between international third grade and international fourth grade (Years 4 and 5 in England): *Science and environment*

There was an increase in performance of at least ten percentage points between the lower and the upper grades in each of the selected countries. The differences in achievement ranged from ten to 19 percentage points (see Table 5.5).

5.5.3 Examples of items for Science and environment

Example items R2, Q7 and X3 are shown below, together with the national and international percentages correct.

Item R2

R2. Some children were trying to find out which of three light bulbs was brightest. Which one of these gives the best START toward finding the answer?

A. "One bulb looks brightest to me, so I already know the answer."

B. "All the bulbs look bright to me, so there cannot be an answer."

C. "It would help if we had a way to measure the brightness of a light bulb."

D. "We can take a vote and each person will vote for the bulb he or she thinks is the brightest."

R2	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	37.2	48.2
International	32.6	43.3
Median country	33.9	43.4
Highest-scoring country	48.1	60.6

The most popular wrong answer chosen by pupils in Years 4 and 5 in England was option (D), selected by 26.7 per cent of Year 4 and 28.1 per cent of Year 5 pupils.

* * * * *

Item Q7

- Q7. Four children can feel and smell an object inside a bag, but they cannot see it. Which of the following is NOT an observation about the object?
- A. "It is flat at one end and round at the other."
 - B. "It smells like peppermint."
 - C. "It has a bump on it."
 - D. "I hope it is chocolate."

Q7	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	41.1	52.6
International	34.3	43.1
Median country	35.8	45.3
Highest-scoring country	48.6	59.2

Two options, (A) and (B), were popular wrong answers chosen by pupils in England. Option (A) was selected by 20.0 per cent of Year 4 and 17.2 per cent of Year 5 pupils, and option (B) was selected by 22.5 per cent of Year 4 and 17.3 per cent of Year 5.

* * * * *

Item X3

X3. Write as completely as possible why large oil spills in rivers and seas are harmful to the environment.

Because they pollute the seas and rivers they poison and kill the animals. The oil produces slick which means when fish and other water creatures come up to breathe they get stuck in the slick and inhale the oil and die and seagulls and other fish eating birds land on the polluted water to fish and they can't lift their wings properly to fly away. They eat the poisoned fish and are then themselves poisoned.

The oil also washes up on the shore and beaches become dangerous and are closed and with huge spills sometimes millions of pounds are spent on the clean up operation alone.

When spill in rivers the oil washes upstream and the entire river can be devastated and if young children get in they could drown with strong currents and danger with the oil.

X3 fully correct	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	18.8	28.4
International	16.4	27.0
Median country	14.2	24.5
Highest-scoring country	51.1	63.8

X3 fully or partially correct*	Percentage correct for international third grade (Year 4 in England) %	Percentage correct for international fourth grade (Year 5 in England) %
England	56.7	69.4
International	48.9	62.9
Median country	47.9	66.5
Highest-scoring country	83.2	90.4

The international marking guidelines stated that fully correct answers should either:

- identify at least one effect and provide an explanation or elaboration (e.g. *Because of the oil in the water, birds get oil in their feathers and then they cool down and die*), or
- list two or more effects but without necessarily giving an explanation (e.g. *It can kill animals and pollute rivers*).

Answers which identified one effect without elaboration were partially correct (e.g. *Fish get killed*).

In England, 10.7 per cent of responses from Year 4 pupils satisfied (b) above, as did 14.1 per cent of responses from Year 5 pupils.

The most frequently offered partially correct response identified effects on animals (e.g. *The oil kills the fish*): 29.9 per cent of the younger pupils and 28.9 per cent of the older pupils in England gave such responses.

More than ten per cent of pupils in both years in England did not attempt the question (17.7 per cent of Year 4 and 13.1 per cent of Year 5).

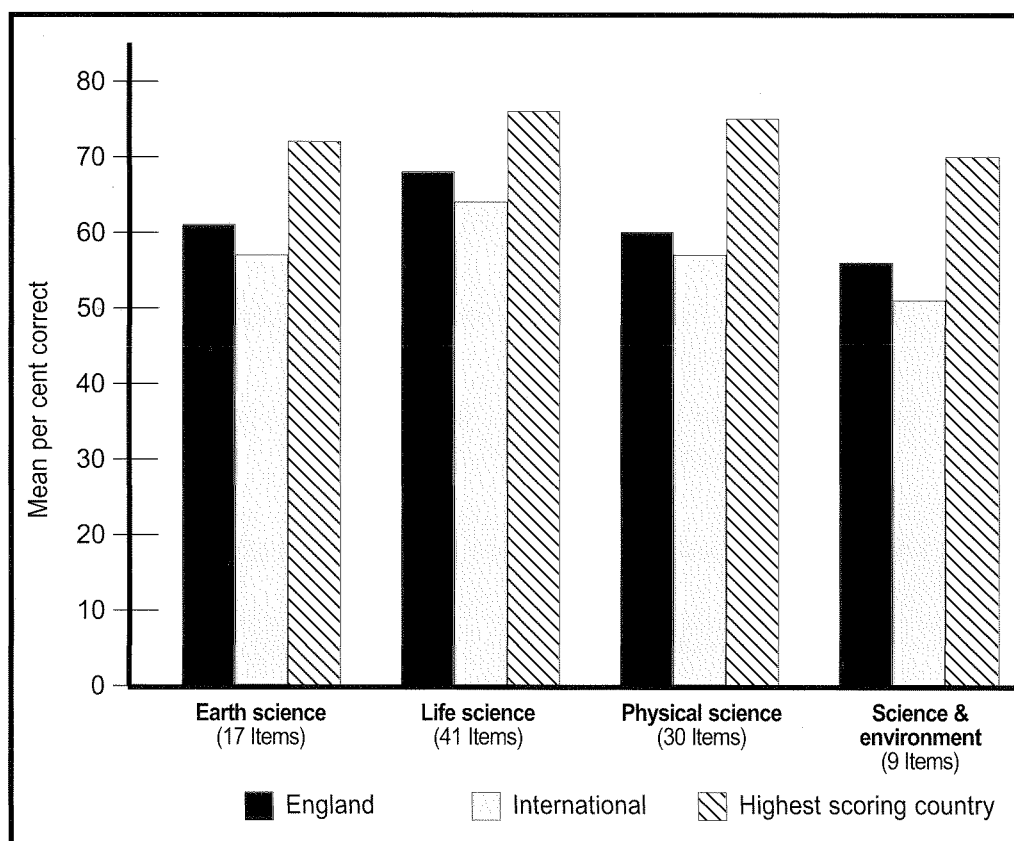
* * * * *

* For this question, pupils' responses were marked as either fully correct, partially correct or incorrect. The international data tables showed the percentages of pupils who gave fully correct answers and the percentages who gave partially correct OR fully correct responses, together with the percentages providing incorrect answers.

5.6 Overview of results for each science reporting category

The mean percentages of correct responses achieved by pupils in Year 5 in England for each of the four science reporting categories are shown in Figure 5.6 below, together with the international means and the means achieved by pupils in the highest-scoring country for the equivalent year group.

Figure 5.6 Overview of mean percentages for science reporting categories: international fourth grade (Year 5 in England)



Source: Tables 2.1 and 2.2 (Martin et al., 1997)

The mean scores of pupils in Year 4 showed a similar pattern.

5.7 Gender differences

The preceding sections have presented the mean percentages of correct responses achieved by pupils in Years 4 and 5 in England and drawn comparisons with the performance of pupils in eight selected other countries. This section identifies where there are statistically significant gender differences in performance in those science content areas.

Gender differences were widespread within the *Earth science* category: eight of the selected countries (including England) showed a difference in performance in favour of boys in both the lower and the upper grades. Fewer gender differences were found within the *Physical science* category: most of these were at the upper grade. Statistically significant gender differences in performance in *Earth science* and *Physical science* are shown in Table 5.7.1.

For the remaining two science reporting categories, only one of the selected countries showed a significant gender difference in performance in each case, as follows:

- ◆ **Life science:** a difference favouring boys in the upper grade only in the Netherlands
- ◆ **Science and environment:** a difference favouring girls in the lower grade only in Canada.

Table 5.7.1 Gender differences for science reporting categories in nine countries

COUNTRY	EARTH SCIENCE		PHYSICAL SCIENCE	
	International third grade (Year 4 in England)	International fourth grade (Year 5 in England)	International third grade (Year 4 in England)	International fourth grade (Year 5 in England)
Canada	*	*		
England	*	*		
Hungary	*	*		*
Japan	*	*		*
Netherlands	*	*	*	*
Norway	*	*		
Scotland	*	*		
Singapore				
United States	*	*		*

Source: Tables 2.4 and 2.5 (Martin et al., 1997)

* Gender difference in favour of boys, statistically significant at the 0.05 level.

Gender differences in the mean percentages of correct responses achieved by pupils in England for each of four science reporting categories are shown in Table 5.7.2 below. Where boys' mean scores were significantly higher than those of the girls (i.e. statistically significant at the 0.05 level), they are shown in bold type.

Table 5.7.2 Overview of gender differences in mean percentages of correct answers for science reporting categories for students in England

Reporting category	Earth science (17 items) %	Life science (41 items) %	Physical science (30 items) %	Science & environment (9 items) %
Boys Year 4	56	60	53	45
Girls Year 4	51	59	51	44
Boys Year 5	63	68	62	55
Girls Year 5	60	68	59	58

Source: Tables 2.4 and 2.5 (Martin et al., 1997)

CHAPTER 6

Preview of further analyses

In this report, the performance of nine-year-old pupils in England on the TIMSS mathematics and science tests has been compared with that of pupils of a similar age in other countries. A previous report, which was published in November 1996, presented the results for 13-year-old students in England.

In addition to collecting pupil achievement data for mathematics and science, TIMSS also collected a wide range of background information from the pupils, teachers and schools that took part in the study. Two further reports (one for nine-year-olds and one for 13-year-olds) will draw upon the background data to describe the contexts of mathematics and science education at the primary and lower secondary levels, and to explore possible reasons for the key findings in this study. Each of these additional reports will compare the responses to the TIMSS questionnaires of students, teachers and headteachers in England with those of their counterparts in the selected countries in order to identify any factors or patterns of behaviour which might differentiate between high- and low-scoring countries.

These further reports (which will each be identified as Part 2 of the national report for one age group) will both be published later this year.

Another component of TIMSS (the Performance Assessment) focused on skills and attainment in practical mathematics and science tasks. England took part in this aspect of the study for the 13-year-olds, but not the nine-year-olds: a report to be published in autumn 1997 will present the results for lower secondary students on this aspect of TIMSS.

REFERENCES

BEATON, A.E., MARTIN, M.O., MULLIS, I.V.S., GONZALEZ, E.J., SMITH, T.A. and KELLY, D.L. (1996a). *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.

BEATON, A.E., MULLIS, I.V.S., MARTIN, M.O., GONZALEZ, E.J., KELLY, D.L. and SMITH, T.A. (1996b). *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.

COMBER, L.C. and KEEVES, J.P. (1973). *Science Education in Nineteen Countries: an Empirical Study*. New York, NY: John Wiley.

FOXMAN, D. (1992). *Learning Mathematics and Science: the Second International Assessment of Educational Progress in England*. Slough: NFER.

INTERNATIONAL ASSOCIATION FOR THE EVALUATION OF EDUCATIONAL ACHIEVEMENT (IEA) (1988). *Science Achievement in Seventeen Countries: a Preliminary Report*. Oxford: Pergamon Press.

KEYS, W. (1987). *Aspects of Science Education in English Schools* (International Studies in Pupil Performance Series). Windsor: NFER-NELSON.

KEYS, W., HARRIS, S. and FERNANDES, C. (1996a). *Third International Mathematics and Science Study, First National Report. Part 1: Achievement in Mathematics and Science at Age 13 in England*. Slough: NFER.

KEYS, W., HARRIS, S. and FERNANDES, C. (1996b). *Third International Mathematics and Science Study, National Reports. Appendices: Additional Information Relating to the Study in England*. Slough: NFER.

LAPOINTE, A.E., ASKEW, J. and MEAD, N.A. (1992a). *Learning Mathematics*. New Jersey: Educational Testing Service, Center for the Assessment of Educational Progress.

LAPOINTE, A.E., ASKEW, J. and MEAD, N.A. (1992b). *Learning Science*. New Jersey: Educational Testing Service, Center for the Assessment of Educational Progress.

MARTIN, M.O., MULLIS, I.V.S., BEATON, A.E., GONZALEZ, E.J., SMITH, T.A. and KELLY, D.L. (1997). *Science Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.

MULLIS, I.V.S., MARTIN, M.O., BEATON, A.E., GONZALEZ, E.J., KELLY, D.L. and SMITH, T.A. (1997). *Mathematics Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.

POSTLETHWAITE, T.N. and WILEY, D.E. (1992). *The IEA Study of Science 2: Science Achievement in Twenty-three Countries*. Oxford: Pergamon Press.

ROBITAILLE, D.F. (Ed) (1993). *Curriculum Frameworks for Mathematics and Science (TIMSS Monograph No. 1)*. Vancouver: Pacific Educational Press.

TIMSS INTERNATIONAL STUDY CENTER (1997a). *TIMSS Mathematics Items: Released Set for Population 2 (Seventh and Eighth Grades) (IEA's Third International Mathematics and Science Study)*. Chestnut Hill, MA: Boston College.

TIMSS INTERNATIONAL STUDY CENTER (1997b). *TIMSS Science Items: Released Set for Population 2 (Seventh and Eighth Grades) (IEA's Third International Mathematics and Science Study)*. Chestnut Hill, MA: Boston College.

APPENDIX I

National steering committee

Member	Association
Mark Neale (Chairman, 1993–1995)	Department for Education and Employment
Michael Richardson (Chairman 1995–present)	Department for Education and Employment
Robert Wood	Department for Education and Employment
John Gardner	Department for Education and Employment
Dr Seamus Hegarty	Director, National Foundation for Educational Research
Dr Wendy Keys	National Foundation for Educational Research
Sue Harris	National Foundation for Educational Research
Cres Fernandes	National Foundation for Educational Research
George Smith (1993–1996)	OFSTED
Christine Agambar (1996–present)	OFSTED
Richard Browne	School Curriculum and Assessment Authority
Miranda Simond (1993–1995)	School Curriculum and Assessment Authority
Carolyn Swain	School Curriculum and Assessment Authority
Dr John Marks	Consultant
Dr Hilary Steedman	Centre for Economic Performance, LSE
Prof. Geoffrey Howson	University of Southampton
Dr Barbara Jaworski (1993–1995)	Mathematical Association (secondary)
Roy Ashley (1995–present)	Mathematical Association (secondary)
Susan Sanders	Mathematical Association (primary)
Dr Alan Eales	Association of Teachers of Mathematics (secondary)
Marjorie Gorman	Association of Teachers of Mathematics (primary)
Judith Lee	Association for Science Education (secondary)
Mick Revell	Association for Science Education (primary)
Brian Semple	Scottish Office Education Department
Hywel Jones	Welsh Office Education Department
Mike Richards (1993–1996)	Welsh Office Education Department
Andrew George (1996–present)	Welsh Office Education Department

APPENDIX II

Countries taking part in different components of TIMSS

Continental Western Europe

			PERFORMANCE ASSESSMENT		
	Population 1	Population 2	Population 1	Population 2	Population 3
Austria	●	●			●
Belgium (<i>Flemish</i>)		●			
Belgium (<i>French</i>)		●			
Cyprus	●	●	●	●	●
Denmark		●		●	●
France		●			●
Germany		●			●
Greece	●	●			●
Iceland	●	●			●
¹ Italy		●			
Netherlands	●	●			●
Norway	●	●	●	●	●
Portugal	●	●	●	●	
Spain		●		●	
Sweden		●			●
Switzerland		●		●	●

English-speaking

Australia	●	●	●	●	●
Canada	●	●	●	●	●
England	●	●		●	
Ireland	●	●			
New Zealand	●	●	●	●	●
Scotland	●	●		●	
United States	●	●	●	●	●

¹ Argentina, Italy and Indonesia were unable to complete the steps necessary for their data to appear in this report. Because the characteristics of its school sample are not completely known, achievement results for the Philippines are not included in the main tables of the international report. Mexico chose not to release its results for the international report.

Eastern Europe

			PERFORMANCE ASSESSMENT		
	Population 1	Population 2	Population 1	Population 2	Population 3
Bulgaria		●			
Czech Republic	●	●	●	●	●
Hungary	●	●	●	●	●
Latvia	●	●			●
Lithuania		●			●
Romania		●		●	
Russian Federation		●			●
Slovak Republic		●			
Slovenia	●	●		●	●
Ukraine		●			

Asia and Pacific Region

Hong Kong	●	●	●	●	
¹ Indonesia	●	●			
Japan	●	●			
Korea	●	●			
¹ Philippines		●			
Singapore	●	●	●	●	
Thailand	●	●			

Other countries

¹ Argentina		●			
Colombia		●		●	
Iran	●	●	●	●	
Israel	●	●	●	●	●
Kuwait	●	●			
¹ Mexico	●	●			●
South Africa		●			●

¹ Argentina, Italy and Indonesia were unable to complete the steps necessary for their data to appear in this report. Because the characteristics of its school sample are not completely known, achievement results for the Philippines are not included in the main tables of the international report. Mexico chose not to release its results for the international report.

ANNEX A

Test-curriculum Matching Analysis
results: mathematics

Table B.1 (Mathematics) Test-curriculum Matching Analysis results—international fourth grade*

(Year 5 in England)

Average percentages correct based on subsets of items especially identified by each country as addressing its curriculum.

Instructions: Read *across* the row to compare that country's performance based on the test items included by each of the countries across the top. Read *down* the column under a country name to compare the performance of the country down the left on the items included by the country listed on the top. Read along the *diagonal* to compare performance for each different country based on its own decisions about the test items to include.

Country	Average Percent Correct on All Items		(Number of Score Points Included)																									
	113	113	49	84	101	91	59	84	89	96	86	111	113	99	83	105	92	86	99	100	98	92	73	102	97	66		
Korea	76 (0.4)	77	79	77	77	78	77	78	75	77	77	76	76	77	78	76	77	79	78	76	76	77	79	78	78	77	76	76
Singapore	76 (0.8)	75	79	76	77	78	76	77	74	75	77	75	76	77	77	75	76	78	78	75	75	77	78	77	76	78	77	78
Japan	74 (0.4)	72	75	75	74	76	75	73	75	73	75	74	74	74	75	74	74	77	74	75	74	75	77	75	74	74	74	74
Hong Kong	73 (0.8)	69	74	72	73	73	73	71	74	74	74	72	73	74	73	73	74	74	73	72	72	74	75	73	73	74	74	74
Netherlands	69 (0.7)	68	70	69	69	75	72	70	71	70	69	69	69	69	71	71	70	72	69	71	71	70	74	70	70	70	69	69
Czech Republic	66 (0.7)	66	67	66	67	71	69	68	68	68	67	66	66	66	69	68	66	68	66	67	68	67	71	67	67	67	66	66
Slovenia	64 (0.7)	65	65	64	65	68	68	67	67	65	64	64	64	65	67	66	64	66	64	66	67	66	69	65	65	65	64	64
Hungary	64 (0.8)	65	65	63	64	69	68	66	66	65	65	63	64	65	67	66	64	66	64	65	65	66	69	65	65	65	65	65
Ireland	63 (0.8)	61	65	64	65	67	65	63	65	63	64	63	65	66	64	64	65	66	64	64	64	64	67	64	64	64	64	64
Australia	63 (0.6)	59	63	62	64	65	64	64	64	65	62	63	63	63	64	64	64	65	63	63	63	64	67	63	63	63	63	61
United States	63 (0.7)	61	64	62	64	64	64	62	65	63	62	63	64	63	63	63	64	65	64	63	63	63	67	63	63	62	62	62
Canada	60 (1.0)	58	61	60	62	63	62	60	63	61	60	60	61	61	61	62	62	63	61	61	62	61	65	61	61	61	59	59
Israel	59 (1.0)	60	61	59	60	63	62	61	62	60	59	60	60	62	61	60	62	60	62	60	61	61	60	65	60	60	60	60
Latvia (LSS)	59 (1.0)	59	60	58	59	62	62	62	61	59	58	59	59	61	60	58	60	61	58	60	61	60	64	60	60	60	59	59
Scotland	58 (0.8)	55	58	57	59	60	59	58	60	58	58	58	58	59	59	60	60	61	59	58	60	58	63	59	58	58	55	55
England	57 (0.7)	52	56	56	57	58	58	57	59	56	56	56	57	57	57	58	59	59	57	57	59	56	61	57	56	56	55	55
Cyprus	54 (0.6)	55	55	54	55	57	56	54	55	55	53	54	55	55	54	54	54	56	55	55	55	56	57	55	56	57	57	57
Norway	53 (0.7)	53	53	53	53	57	56	55	55	55	53	53	53	53	55	55	53	55	53	55	55	54	57	54	54	54	52	52
New Zealand	53 (1.0)	49	53	52	54	55	55	54	56	53	53	53	54	54	55	54	55	55	54	54	55	54	58	54	54	53	51	51
Greece	51 (0.9)	50	51	50	52	55	53	52	52	52	50	51	51	52	52	52	50	52	51	52	52	52	54	52	52	52	53	53
Iceland	50 (0.8)	47	49	49	51	53	52	51	52	50	49	50	50	51	51	51	50	52	50	50	52	50	55	50	50	50	47	47
Portugal	48 (0.7)	47	49	47	48	50	49	49	49	49	48	48	48	49	48	47	47	49	48	48	48	48	50	49	49	50	50	50
Iran, Islamic Rep.	38 (0.9)	39	39	37	39	39	39	40	39	39	38	38	39	38	38	37	37	39	38	39	38	40	40	39	40	40	43	43
Kuwait	32 (0.5)	32	32	31	31	32	33	33	33	33	32	31	32	32	31	32	30	32	31	32	31	32	33	32	32	32	33	33
International Average	59 (0.7)	58	60	59	60	62	61	60	61	60	59	59	60	61	60	60	60	61	60	60	60	60	63	60	60	60	59	59

*Fourth grade in most countries.

**Of the 102 items in the mathematics test, some items had two parts and some extended-response items were scored on a two-point scale, resulting in 113 total score points.

() Standard errors for the average percent of correct responses on all items appear in parentheses.

Because results are rounded to the nearest whole number, some totals may appear inconsistent.

Countries shown in italics did not satisfy one or more guidelines for sample participation rates, age/grade specifications, or classroom sampling procedures.

Because population coverage falls below 65% Latvia is annotated LSS for Latvian Speaking Schools only.

SOURCE: Mullis et al., 1997.

**Table B.2 (Mathematics) Test-curriculum Matching Analysis results—international third grade*
(Year 4 in England)**

Average percentages correct based on subsets of items especially identified by each country as addressing its curriculum.

Instructions: Read **across** the row to compare that country's performance based on the test items included by each of the countries across the top.

Read **down** the column under a country name to compare the performance of the country down the left on the items included by the country listed on the top.

Read along the **diagonal** to compare performance for each different country based on its own decisions about the test items to include.

Country	Average Percent Correct on All Items											Iran, Islamic Rep.												
	(Number of Score Points Included)																							
	113	89	84	58	45	26	61	80	83	113	74	41	63	46	54	51	71	77	102	51	81	34	65	
Korea	67 (0.5)	65	67	74	76	71	68	66	69	67	67	72	73	67	72	68	72	71	69	70	51	69	68	71
Japan	63 (0.3)	62	64	68	70	68	64	63	65	63	65	68	67	65	68	65	67	65	65	67	65	65	65	68
Singapore	62 (0.9)	61	64	72	72	69	63	62	65	62	63	67	66	64	69	61	68	68	64	68	65	65	65	67
Hong Kong	59 (0.7)	56	59	65	69	63	60	59	61	59	61	62	64	64	67	60	63	63	61	64	61	61	63	65
Netherlands	52 (0.6)	51	51	57	60	63	56	54	56	52	57	58	55	59	60	57	57	56	53	58	56	56	60	56
Czech Republic	52 (0.7)	50	50	58	61	60	55	55	53	52	56	59	55	58	60	57	57	55	53	57	55	57	61	58
Slovenia	51 (0.7)	49	49	56	61	58	55	55	53	51	56	58	55	59	58	56	57	55	52	56	54	59	56	56
Australia	50 (0.9)	47	48	54	58	58	52	51	54	50	54	54	55	58	57	54	55	54	50	55	53	57	53	53
United States	49 (0.7)	47	48	55	59	55	51	50	52	49	52	55	54	56	56	52	54	53	50	54	52	52	53	52
Hungary	49 (0.9)	47	48	56	61	58	52	52	51	49	53	58	53	55	55	54	55	53	50	54	52	52	55	55
Ireland	48 (0.8)	46	47	52	57	55	49	49	50	48	50	55	52	55	54	50	52	52	49	52	49	52	50	51
Canada	47 (0.7)	44	45	52	56	55	50	48	50	47	51	51	53	55	54	50	53	51	48	53	50	53	50	53
Scotland	45 (0.8)	42	43	48	53	53	48	46	49	45	49	48	50	54	53	48	50	49	46	50	47	51	47	47
England	45 (0.6)	43	43	48	52	52	47	46	48	45	49	47	48	53	52	48	49	49	45	49	45	49	51	47
Latvia (LSS)	45 (0.8)	43	43	51	53	50	48	47	47	45	49	51	49	52	51	51	51	49	46	50	47	54	50	50
New Zealand	41 (0.8)	39	39	45	49	47	43	42	44	41	45	44	45	49	47	45	45	45	41	46	43	48	44	44
Cyprus	38 (0.6)	36	39	46	46	43	39	39	39	38	39	45	42	43	43	40	43	43	39	44	40	42	43	43
Portugal	37 (0.8)	35	36	41	45	40	36	38	37	37	38	41	41	41	42	39	40	40	38	41	37	38	40	40
Greece	37 (0.8)	35	37	42	45	46	38	39	38	37	39	43	40	42	41	40	41	40	38	44	40	41	40	41
Norway	36 (0.7)	34	34	39	43	44	38	37	38	36	40	42	39	41	40	42	40	38	36	41	38	41	38	41
Iceland	35 (0.6)	34	34	38	42	43	37	37	38	35	39	40	40	43	40	39	40	39	36	40	38	42	38	38
Iran, Islamic Rep.	28 (0.7)	25	27	33	35	27	27	30	27	28	28	34	30	31	32	31	31	31	29	32	29	31	33	33
International Average	47 (0.7)	45	46	52	56	54	49	48	49	47	50	52	51	53	50	52	51	48	48	52	49	52	51	51

*Third grade in most countries.

**Of the 102 items in the mathematics test, some items had two parts and some extended-response items were scored on a two-scale, resulting in 113 total score points.

() Standard errors for the average percent of correct responses on all items appear in parentheses.

Because results are rounded to the nearest whole number, some totals may appear inconsistent.

Countries shown in italics did not satisfy one or more guidelines for sample participation rates, age/grade specifications, or classroom sampling procedures.

Because population coverage falls below 65% Latvia is annotated LSS for Latvian Speaking Schools only.

SOURCE: Mullis et al., 1997

ANNEX B

Test-curriculum Matching Analysis

results: science

**Table B.1 (Science) Test-curriculum Matching Analysis results—international fourth grade*
(Year 5 in England)**

Average percentages correct based on subsets of items especially identified by each country as addressing its curriculum.

Instructions: Read across the row to compare that country's performance based on the test items included by each of the countries across the top.

Read down the column under a country name to compare the performance of the country down the left on the items included by the country listed on the top.

Read along the *diagonal* to compare performance for each different country based on its own decisions about the test items to include.

Country	Average Percent Correct on All Items		(Number of Score Points Included)																								
	105		Korea	Japan	Netherlands	Australia	United States	Czech Republic	Singapore	Slovenia	Canada	England	Hong Kong	Hungary	Ireland	Norway	New Zealand	Scotland	Israel	Latvia (LSS)	Iceland	Greece	Cyprus	Portugal	Iran, Islamic Rep.	Kuwait	
Korea	74 (0.4)	105	80	74	74	74	74	75	52	98	83	53	42	49	27	62	90	45	33	104	94	71	61	84	85	85	85
Japan	70 (0.3)	77	79	78	74	74	75	78	74	75	76	76	81	78	83	77	76	81	76	71	70	72	72	75	75	75	75
Netherlands	67 (0.5)	68	79	74	71	70	73	70	70	70	71	71	76	72	80	73	72	72	76	69	67	66	70	71	70	68	68
Australia	66 (0.5)	65	70	72	68	67	69	68	67	68	68	68	73	69	77	72	68	75	69	67	66	69	70	68	67	67	67
United States	66 (0.5)	66	67	70	68	66	67	66	66	66	66	66	73	67	78	70	68	75	67	66	66	69	69	68	67	67	67
Czech Republic	65 (0.5)	65	67	70	67	65	66	65	66	65	66	66	73	67	78	70	67	75	69	66	66	69	69	68	67	66	66
Singapore	64 (0.8)	62	70	69	65	64	65	65	65	65	66	66	72	68	79	70	68	72	67	65	64	68	68	66	66	66	66
Slovenia	64 (0.7)	61	64	68	64	64	66	64	64	64	64	64	74	65	76	68	66	71	65	64	64	65	67	66	65	65	64
Canada	64 (0.6)	64	64	67	65	64	66	64	63	63	63	63	71	64	76	67	65	71	65	64	63	66	67	65	64	64	64
England	63 (0.6)	63	64	67	64	63	63	66	64	63	65	65	71	65	76	68	65	71	64	63	63	66	67	65	64	64	64
Hong Kong	62 (0.7)	59	68	66	64	62	63	66	63	62	63	62	72	65	76	68	64	67	61	62	62	62	65	64	64	63	63
Hungary	62 (0.6)	60	62	66	62	62	62	64	63	61	62	62	69	65	73	67	63	66	62	62	62	62	65	63	63	62	62
Ireland	61 (0.6)	60	61	64	62	61	61	64	61	61	61	62	69	62	74	65	62	69	64	61	61	63	66	63	62	62	62
Norway	60 (0.6)	58	61	65	61	60	61	63	61	60	61	61	68	62	71	67	62	68	59	61	61	63	64	63	61	61	61
New Zealand	60 (0.9)	59	60	64	61	60	61	64	61	60	61	61	67	61	72	64	62	68	61	60	60	63	65	62	61	61	61
Scotland	60 (0.8)	59	61	64	61	60	60	64	61	60	62	62	69	62	73	65	62	69	61	60	60	63	65	62	61	61	61
Israel	57 (0.8)	56	58	61	57	57	57	61	57	57	59	59	65	59	69	61	58	65	58	57	57	59	60	60	58	57	57
Latvia (LSS)	56 (0.8)	55	56	60	57	56	56	59	57	56	56	63	63	59	66	61	58	62	56	56	57	58	59	58	58	57	56
Iceland	55 (0.7)	55	55	59	56	55	56	59	56	55	56	61	58	58	65	61	57	64	55	55	56	57	57	57	57	56	56
Greece	54 (0.8)	53	54	58	54	54	54	58	54	54	54	62	56	65	65	59	55	62	55	54	54	57	58	56	55	55	55
Cyprus	51 (0.5)	52	52	55	51	51	50	55	51	50	49	59	53	61	56	52	52	58	53	50	51	53	54	52	52	52	52
Portugal	50 (0.7)	50	50	54	51	50	50	53	51	49	50	57	52	59	55	52	52	57	52	50	50	52	53	53	51	50	50
Iran, Islamic Rep.	40 (0.7)	38	39	44	40	40	40	45	40	39	38	48	42	48	44	41	44	44	37	40	40	42	41	41	41	41	41
Kuwait	39 (0.5)	42	42	43	39	39	40	43	39	38	37	47	41	48	43	41	45	45	40	39	39	42	41	40	40	40	41
International Average	60 (0.6)	59	61	64	60	60	60	63	60	60	60	67	62	71	64	61	67	61	60	60	62	63	61	61	61	60	60

*Fourth grade in most countries.

**Of the 97 items in the science test, some items had two parts and some extended-response items were scored on a two-point scale, resulting in 105 total score points.

() Standard errors for the average percent of correct responses on all items appear in parentheses.

Because results are rounded to the nearest whole number, some totals may appear inconsistent.

Countries shown in italics did not satisfy one or more guidelines for sample participation rates, age/grade specifications, or classroom sampling procedures.

Because population coverage falls below 65% Latvia is annotated LSS for Latvian Speaking Schools only.

SOURCE: Martin et al., 1997

**Table B.2 (Science) Test-curriculum Matching Analysis results—international third grade*
(Year 4 in England)**

Average percentages correct based on subsets of items especially identified by each country as addressing its curriculum.

Instructions: Read across the row to compare that country's performance based on the test items included by each of the countries across the top.

Read down the column under a country name to compare the performance of the country down the left on the items included by the country listed on the top.

Read along the diagonal to compare performance for each different country based on its own decisions about the test items to include.

Country	Average Percent Correct on All Items										(Number of Score Points Included)													
	105	11	17	49	105	34	40	32	64	32	82	34	33	14	68	41	74	49	30	93	84	44	39	
Korea	67 (0.5)	66	76	71	67	76	78	74	71	71	67	76	77	83	71	74	68	74	78	67	68	70	69	69
Japan	61 (0.3)	62	76	67	61	68	74	68	66	65	61	70	66	77	65	66	61	68	69	61	63	64	64	64
Australia	57 (0.8)	55	60	61	57	62	65	62	61	58	57	65	66	76	59	59	58	64	69	56	58	58	58	58
United States	56 (0.6)	57	60	62	56	62	66	62	61	58	57	66	66	75	58	59	58	66	69	56	58	58	59	59
Netherlands	56 (0.7)	57	60	61	56	66	67	62	60	55	57	67	65	72	58	61	59	66	68	55	58	59	59	59
Czech Republic	55 (0.6)	52	61	58	55	61	66	60	58	56	55	64	63	74	58	59	56	64	70	55	57	56	59	59
England	55 (0.7)	51	57	59	55	61	64	60	58	56	55	64	64	73	57	58	56	63	66	54	56	55	57	57
Canada	53 (0.5)	52	56	57	53	59	62	58	57	54	53	62	62	70	55	56	55	61	65	53	55	54	55	55
Singapore	53 (0.9)	54	65	58	53	61	66	62	58	61	52	65	62	74	56	60	54	61	64	53	55	57	57	57
Slovenia	53 (0.5)	46	55	55	53	58	65	58	56	54	53	62	62	72	54	57	54	62	65	54	55	54	54	54
Hong Kong	53 (0.6)	53	59	57	53	59	64	60	57	54	52	63	61	74	56	56	54	61	62	52	54	52	57	57
Scotland	51 (0.7)	48	55	56	51	58	62	57	55	53	52	61	61	71	53	56	53	60	64	51	53	53	54	54
Ireland	51 (0.7)	47	54	56	51	58	62	57	56	54	51	62	62	72	53	55	53	60	64	51	53	52	54	54
New Zealand	51 (0.9)	51	54	54	51	57	61	56	54	53	51	60	60	69	53	55	52	58	63	50	52	53	53	53
Hungary	50 (0.8)	44	55	52	50	55	61	55	53	52	51	61	57	67	52	55	50	58	65	50	51	52	52	53
Latvia (LSS)	48 (0.9)	43	50	50	48	54	60	53	51	51	49	58	56	64	51	54	50	57	62	49	51	52	52	52
Norway	46 (0.7)	41	48	49	46	50	56	49	49	45	46	55	55	61	48	49	48	56	60	46	48	48	49	49
Greece	44 (0.7)	35	43	46	44	49	54	48	49	45	45	55	56	63	46	49	46	54	59	45	46	46	45	45
Iceland	42 (0.6)	33	40	44	42	45	51	45	46	42	42	50	52	54	44	46	44	52	56	43	44	42	44	44
Portugal	41 (0.8)	31	45	43	41	45	51	45	44	44	41	51	48	55	43	45	43	50	53	41	43	44	44	44
Cyprus	39 (0.5)	33	43	41	39	45	52	44	43	43	39	50	47	57	42	43	41	48	52	40	41	44	40	40
Iran, Islamic Rep.	30 (0.7)	23	33	31	30	31	39	31	33	33	30	40	33	42	32	33	31	36	39	31	31	33	34	34
International Average	51 (0.7)	47	55	54	51	56	61	56	54	53	51	60	59	68	53	55	52	59	63	51	52	53	53	53

*Third grade in most countries.

**Of the 97 items in the science test, some items had two parts and some extended-response items were scored on a two-scale, resulting in 105 total score points.

() Standard errors for the average percent of correct responses on all items appear in parentheses.

Because results are rounded to the nearest whole number, some totals may appear inconsistent.

Countries shown in italics did not satisfy one or more guidelines for sample participation rates, age/grade specifications, or classroom sampling procedures.

Because population coverage falls below 65% Latvia is annotated LSS for Latvian Speaking Schools only.

SOURCE: Martin et al., 1997



THIRD INTERNATIONAL MATHEMATICS AND SCIENCE STUDY

Second National Report

Part 1

England is one of 26 countries that have surveyed nine-year-old pupils as part of the Third International Mathematics and Science Study (TIMSS). This report presents the national results for nine-year-olds in England: information on pupils' performance is presented, together with comparisons with other countries, with previous international studies and with the results for 13-year-olds reported in the First National Report (Keys *et al.*, 1996).

Comparisons are made in terms of:

- pupils' overall scores in mathematics and science;
- pupils' scores on each of the following mathematics and science reporting categories:

Mathematics

Whole numbers

Fractions and proportionality

Measurement, estimation and number sense

Data representation, analysis and probability

Geometry

Patterns, relations and functions

Science

Earth science

Life science

Physical science

Science and environment

Examples of items in each of the reporting categories are provided, together with information on the responses made by pupils in England and international data relevant to the illustrative items.

This report is based on a national survey of more than 6,000 pupils in 134 schools, which was part of a world-wide sample of over 175,000 pupils in more than 4,000 schools in 26 countries. It is essential reading for all those concerned with the teaching and learning of mathematics and science in primary schools: teachers, governors, LEA advisory teams, policy makers and researchers.

ISBN 0 7005 1464 3
£15.00