THIRD INTERNATIONAL MATHEMATICS AND SCIENCE STUDY
National Reports
Appendices

Wendy Keys, Sue Harris and Cres Fernandes

nfer

THIRD INTERNATIONAL MATHEMATICS AND SCIENCE STUDY

National Reports
Appendices

Additional information relating to the study in England

Wendy Keys, Sue Harris and Cres Fernandes

nfer

Published in November 1996
by the National Foundation for Educational Research,
The Mere, Upton Park, Slough, Berkshire SL1 2DQ

Registered Charity No. 313392
ISBN 0 7005 1446 5

CONTENTS

Appendix 1	National steering committee	1
Appendix 2	Countries taking part in different components of TIMSS	2
Appendix 3	The design and administration of the study	4
Appendix 4	The development of the tests and questionnaires	20
Appendix 5	Statistical details	37
References		49

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 *(Flemish)*		●			
Belgium *(French)*		●			
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.*

APPENDIX III
The design and administration of the study

III.1 Introduction

An international study such as TIMSS necessarily requires formal procedures which are implemented in all countries participating in the study. These procedures ensure that the data gathered from countries taking part are comparable, since the tests and questionnaires had the same content, structure and time allowance, and were used with similar samples of students and teachers. This appendix provides details of the national arrangements for the administration of the study, which were, of course, set within the context of the international requirements.

III.2 The design of the research

The TIMSS study was the third international study for both mathematics and science organised by the IEA (International Association for the Evaluation of Educational Achievement). However, it was the first time that the IEA had surveyed both subjects within one study, rather than as separate studies.

The study was intended to provide opportunities to make cross-national comparisons, in terms of:

- students' knowledge and understanding of mathematics and science
- students' and teachers' attitudes towards these subjects
- mathematics and science curricula
- teaching conditions and practices.

The research focused on three different stages of education: middle primary (Population 1), lower secondary (Population 2) and students in their final year of schooling (Population 3). Internationally, Population 2 was deemed a 'core' population: each of the 46 countries taking part in TIMSS (see Appendix II) was required to survey at least this age group; in practice, countries surveyed one, two or all three of the target populations. In England, the study focused on Populations 1 and 2 only.

For each population, data were collected at school level by means of written achievement tests for students (eight rotated versions were used: see Appendix IV, section IV.3), and questionnaires for students, their teachers of mathematics and science and headteachers (see Appendix IV, sections

IV.9, IV.10 and IV.11). In addition, information about the mathematics and science curricula, and particularly in relation to the age groups on which the study focused, was collected at national level from a panel of experts (see section III.6). The content of the written achievement tests was the same for all participating countries, whereas some national amendments to the questionnaires were permitted. National translations of the tests and questionnaires had to be approved by the International Study Centre prior to administration in schools, so as to ensure that the translation had not resulted in any change of content or emphasis in the materials. The research instruments completed by schools are shown below.

Population 1 (primary)

Student achievement tests (eight rotated versions)
Student questionnaire
Primary teacher questionnaire
Headteacher questionnaire

Population 2 (secondary)

Student achievement tests (eight rotated versions)
Student questionnaire
Mathematics teacher questionnaire
Science teacher questionnaire
Headteacher questionnaire

To supplement the data concerning students' levels of achievement gathered from the written tests, a number of practical tasks formed a further component of the study. These tasks (known as the Performance Assessment) focused on students' practical, investigative and analytical skills in mathematics and science. This aspect was designed for administration to Populations 1 and 2, although countries participating in this part of the study were permitted to carry out the tasks with only one population if they preferred. In England, the Performance Assessment component was used only with the secondary age students (Population 2).

III.3 Populations and samples

The study design involved focusing on specific age groups, or populations, of students in primary and secondary schools. A number of the countries taking part in TIMSS had participated in earlier IEA studies of mathematics and science; by selecting similar age groups to those used in the previous studies, these countries could compare their students' performance with the levels achieved by their own students in the earlier studies. The age groups surveyed in previous mathematics and science studies, as well as in TIMSS, are shown in Table III.3.

Table III.3: Age groups surveyed in mathematics and science studies

Year	Subject(s)		Organiser	Ages tested
1964	Mathematics	(FIMS)	IEA	13, 18
1970	Science	(FISS)	IEA	10, 14, 18
1981	Mathematics	(SIMS)	IEA	13, 18
1984	Science	(SISS)	IEA	10, 14, 18
1988	Mathematics and Science	(IEAP1)	IAEP	13
1991	Mathematics and Science	(IAEP2)	IAEP	9, 13
1995	Mathematics and Science	(TIMSS)	IEA	9, 13

III.3.1 The choice of age groups for the international study

Three age groups were selected by the International Coordinating Centre:

- students aged about nine years (Population 1)
- students aged about 13 years (Population 2)
- students in their final year of secondary education (Population 3).

Populations 1 and 2 were defined internationally as follows:

Population 1: Students enrolled in the two adjacent grades that contained the largest proportion of nine-year-old students at the time of testing third- and fourth-grade students in most countries (Years 4 and 5 in England).

Population 2: Students enrolled in the two adjacent grades that contained the largest proportion of 13-year-old students at the time of testing seventh- and eighth-grade students in most countries (Years 8 and 9 in England) (see Beaton *et al.*, 1996a, 1996b).

For countries taking part in the research at Population 3, a single grade was surveyed.

III.3.2 The choice of age groups for the national study

England participated in the survey of two age groups: Population 1 and Population 2.

Population 1 consisted of students in all schools (excluding special schools) born between 1st September 1984 and 31st August 1986, that is students in National Curriculum Years 4 and 5. At the time of testing (early March 1995), the ages of the students surveyed in England therefore ranged from eight years seven months to ten years six months. At the time of testing, the two year groups tested included over 99 per cent of the nine-year-olds in England: 58 per cent in the lower grade and 41 per cent in the upper grade.

The students included in Population 2 in England were those whose dates of birth fell between 1st September 1980 and 31st August 1982, that is students in National Curriculum Years 8 and 9. At the time of testing in England, Population 2 students' ages ranged from 12 years seven months to 14 years six months. The two year groups selected included 99 per cent of 13-year-olds in England: 57 per cent in the lower grade and 42 per cent in the upper grade.

III.3.3 Replacement schools

The sample design required the selection of 150 first-choice schools together with two matching sets of 150 replacement schools selected at the same time which were used to replace schools that refused to take part in the study (i.e. a total of 450 schools were selected for each population, with the intention of including only 150 at each population). For each of the first- choice schools, the two replacement schools were selected to represent the same characteristics, so that, overall, the range of types of school participating would remain the same irrespective of whether any individual school was one of the first-choice schools, or one of the replacement schools.

The first-choice schools were invited to take part in the study; where schools agreed to participate, no approach was made to the replacement schools. However, when a first-choice school refused to take part, the first replacement for that school was then approached: if that school agreed, then no contact was made with the second replacement school; if the school refused, then the second (final) replacement school was approached. If all three schools refused to participate, no further replacements were permitted.

The data from replacement schools are included in the national mean scores. The response rates for England (as for other countries) were reported in two ways: firstly, on the number of first-choice schools participating in the study, and secondly, in terms of the total number of schools taking part, i.e. first-choice plus replacement schools.

III.3.4 Exclusions

Guidance was provided by the international coordinators concerning permissible exclusion of students who had been selected to take part in the study. In accordance with the international guidelines, details of the criteria for excluding particular students were prepared for English schools, so that teachers would not include these students in the testing programme. The three categories defined for excluding students were as follows:

- students with functional disabilities
- other students with special educational needs
- non-native English speakers.

The first category covered students with permanent physical disabilities (e.g. lack of fine motor control) or sensory impairment (e.g. visual or auditory impairment) which meant that they were unable to perform in the testing situation.

The second category concerned students other than those with physical disabilities who either had a statement of special educational needs, who had been referred for a statement, or who, in the professional judgement of their teachers, should have been referred for a statement. These two categories reflect the fact that although special schools were excluded from the samples, there are a number of students with statements of special educational needs who are routinely taught in mainstream schools in England, whereas in some other countries, such as The Netherlands, all such students are removed from mainstream schools.

The final category allowed schools to exclude students who were unable to read and/or speak English and who would be unable to overcome the language barrier in the testing situation. Typically, this applied to students for whom English was not their first language and who had been taught in English for less than one year.

The guidance to schools informed teachers that if they felt it would be preferable for any student(s) who fell within the exclusion categories to complete the tests (e.g. so that they did not feel that they were being treated differently from the rest of the class), they could include them in the testing programme, although this fact had to be indicated on the documents returned by the school to NFER. (In practice this only applied to schools in Population 1 where whole classes were selected, since the exclusion of one or more students may have caused those students some distress; as the Population 2 sample was randomly selected from across whole year groups, students who fell within the exclusion categories were never aware that they had been selected.) The marks of any excluded students who actually completed the tests were **not** used in calculating national mean scores. No substitute students were used at either population to replace students who fell within the exclusion categories.

III.3.5 Population 1 (nine-year-olds)

Once the schools willing to participate had been identified, they were asked to provide details of all the classes in Years 4 and 5, together with some identifier, such as the class teacher's name. For the second stage, one entire class (representing on average approximately 30 students) from each target year group was randomly selected by NFER for inclusion in the study.

Some schools indicated that students were taught in vertically grouped classes, sometimes combining Years 4 and 5, but frequently combining Years 3 and 4, and Years 5 and 6. For these schools, only students from the selected class(es) whose date of birth fell within the range defined by the study were surveyed.

If a school had only one of the year groups required, two classes were selected from that year group (this was usually the case for the middle schools that took part).

III.3.6 Population 2 (13-year-olds)

For Population 2, the procedures concerning the selection of schools were followed in the same way as for Population 1, but subsequently, the sample was selected in a different way. In contrast to the general pattern of mixed ability teaching in primary schools, it is more common for secondary schools in England to implement some form of ability grouping such as setting or streaming. Consequently, if one teaching group had been randomly selected from each of the two years with students of the target age (Years 8 and 9), the groups selected may have included students representing specific ability groups only. In order to avoid this, and with the approval of the international coordinators, a different sampling strategy was used.

Schools which had indicated that they were willing to participate in the study were asked to provide lists of all students in Years 8 and 9. From these lists, 16 students were randomly selected by NFER from each year group, making a total of 32 students. For schools with only one of the target year groups, 32 students were selected from the year group available.

III.3.7 The sub-sample for the Performance Assessment component

A sub-sample of the secondary schools that had participated in the main survey was invited to take part in the additional component which focused on students' practical and investigative skills, the Performance Assessment. This element of the study was intended to focus on only the upper age group of the population (i.e. Year 9 within Population 2). Since the administration of the Performance Assessment tasks was carried out by administrators trained by NFER, rather than teachers within the schools involved (see section III.5), the individuals concerned had to visit schools throughout England. The international guidelines permitted national research centres to exclude from the Performance Assessment sample any schools that were located in geographically isolated areas, or regions that were considerable distances from the home bases of the administrators. A maximum of 25 per cent of the schools with students in the target age group could be excluded on these criteria; in England, only 21 per cent of Population 2 schools were excluded.

As with the main survey, the sample comprised first-choice and replacement schools. A total of 100 schools with Year 9 students was selected: 50 first-choice schools, and, for each one, a single replacement school with similar characteristics. As before, replacement schools were only involved if the corresponding first-choice schools refused to take part. A total of 50 schools took part: 37 first-choice and 13 replacement schools.

From each school, nine of the Year 9 students who had completed the written achievement tests were randomly selected, together with two substitute students, who were only to be included if any of the nine selected students were absent on the day the tasks were administered. In the event of any absence, the substitute students were selected in sequence, i.e. if one of the first-choice students was absent, s/he was replaced with the first-listed substitute; the second absentee was replaced by the second-listed substitute. If more than two students were absent, or if either/both of the substitutes were absent, no further replacements were permitted. Any students who, at the time of the main survey, had been identified by teachers as falling within the exclusion categories (see section III.3.4) were excluded before the sample for the Performance Assessment was selected.

III.4 Administration of the main testing programme

The tests and questionnaires which comprised the main survey were administered in schools by teachers. Schools were free to choose when to carry out the testing within a specified two-week period (27th February to 10th March 1995). The written achievement tests were subject to precise time limits (different times applied to Populations 1 and 2): these were common to all participating countries. The student questionnaires had no time limit, but teachers were advised to allow sufficient time for all students to complete the whole questionnaire (roughly 20 minutes). Teachers and headteachers were asked to fill in the appropriate questionnaires, which were designed to take approximately 30 minutes to complete.

The content of the written tests (prepared by the international Study Centre) was rotated to provide eight different versions: these were used by every country participating in the study. The rotation was designed to provide wide curriculum content coverage, yet, at the same time, permit the inclusion of some core items in the tests for all students within a population. Each test, numbered 1-8, contained a mixture of both mathematics and science questions. In England, rather than using one thick booklet containing all the test items for both sessions (see below), the written tests were presented to students as two separate booklets, so each student completed an A and a B booklet (e.g. booklets 1A and 1B); both booklets attempted by each student included both mathematics and science questions, arranged in clusters. The student questionnaire formed a third booklet.

To assist schools in administering the survey, an administration manual, which included a script for the teacher administering the tests and questionnaires together with details of the time limits, was sent with the other materials to schools. The timetable for Population 1 was as follows:

Population 1 — Timetable for testing

Session 1	Test A	37 minutes
Session 2	Test B	27 minutes
Session 3	Student Questionnaire	20 minutes (approx.)

In addition to the actual testing time, teachers spent a further five to ten minutes on administrative matters such as going through examples printed in the front of both the test and questionnaire booklets. Teachers were asked to ensure that students had a break of about 20 minutes after completing the first test (i.e. the A booklet). A further break was permitted after the completion of the second test (B) booklet, although teachers were informed that they could move on to the student questionnaire without a break if that proved more convenient.

The arrangements for Population 2 were very similar, the only difference being the time allowed for the completion of the test booklets (46 minutes and 44 minutes for booklets A and B respectively).

Completed tests and questionnaires were returned to NFER for marking, coding and data processing.

III.5 Administration of the Performance Assessment component

Although the administration of the main survey tests and questionnaires had been carried out by teachers in participating schools, a different approach was used for the administration of the Performance Assessment. Due to their practical nature, the 12 tasks that comprised this element of the study carried additional requirements in terms of:

- the equipment necessary
- the supervision time.

So as to avoid placing additional burdens on schools, a team of ten administrators was used to set up and administer the tasks in schools. The administrators took with them all equipment needed to carry out the activities in schools. The 12 tasks were set up at nine workstations: some workstations had one 30-minute task, whilst others had two 15-minute tasks. At each workstation, students worked from booklets which provided details of the activity and asked questions related to the task; students wrote their responses within the booklets, although a few tasks required additional forms of response (see Appendix IV for further details of the tasks). The Performance Assessment component was administered in the sample of 50 schools over a six-week period (from June to July 1995) on dates and at times to suit the schools concerned.

Within each school, each of the nine Year 9 students selected to take part in the Performance Assessment visited a total of three workstations, with 30 minutes allocated at each one. Depending on whether the workstations visited had one or two tasks each, students attempted either three, four or five tasks in a total testing time of 90 minutes. In most cases, students were given a short break of about 15-20 minutes after completing the second workstation. However, in a few instances, at the school's request, students worked straight through to the third and final workstation (this was usually

because this reflected the students' normal pattern of afternoon work without a break).

The booklets completed by students at the workstations were returned to NFER for marking, coding and data processing.

III.6 Test-curriculum Matching Analysis

TIMSS used the same test items to assess students' achievement in mathematics and science in all countries that took part in the study. This ensured that comparisons could be made between different countries in terms of attainment in particular content areas of the two subjects tested. However, although this was an important element of the study, it did not recognise the fact that the curricula for mathematics and science vary from one country to another. Many countries have found that some items represent content areas that are not covered in their curriculum for the age group tested, although the selection of test items for TIMSS was carried out on the basis that the content of an item was applicable to the curriculum of at least half of the countries taking part in the study. For example, from the English point of view, the items concerned with *Earth science* did not represent part of the science curriculum in England. However, students may have covered the content within the geography curriculum. In other instances, particular aspects of item content may not have been covered at all by certain countries.

The Test-curriculum Matching Analysis was set up with the purpose of collecting information about the relevance of the TIMSS test items to the mathematics and science curricula in countries taking part in the study. For this component, each country was asked to review all the test items and indicate which ones most closely matched their national mathematics and science curricula for the populations tested. The sub-sets of items identified by each country as being covered within their curriculum provided additional comparative data, presenting each country's performance *on the items selected as being relevant for that country.* This analysis will allow countries to compare their results with other countries on a similar basis: their own self-selected sub-sets of items.

In order to identify the relevant sub-sets of items, experts in each country had to consider whether or not each item was part of the 'intended curriculum'. In practical terms this was interpreted in England as:

- whether or not the topic of the item formed part of the National Curriculum for the majority of students (at least 50 per cent) for Years 4 and/or 5 (for Population 1) and for Years 8 and /or 9 (for Population 2)
- whether or not the content of the item should have been covered by students at the time of testing (in England, early March 1995).

In England, the Test-curriculum Matching Analysis was carried out by panels of subject-matter experts during September 1995. Four panels met:

Subject	Year groups
Mathematics	Years 4 and 5
Science	Years 4 and 5
Mathematics	Years 8 and 9
Science	Years 8 and 9

Each panel consisted of:

- one HMI from the Office for Standards in Education
- one professional officer from the School Curriculum and Assessment Authority
- two or three teachers.

In each case the individuals concerned were familiar with the curriculum content for the age range under consideration.

Details of the sub-sets of items (per subject and per population) identified by individual countries were collated by the international study centre and used to produce appropriate comparative tables. The Test-curriculum Matching Analysis data have been presented in two-way tables, with each country taking part in TIMSS listed across the top and down the side. These tables provide three different perspectives on the achievement test data:

- firstly, selecting the column for a particular country (e.g. England), the reader can compare how England and all other countries performed on the English sub-set of items
- secondly, reading across the row for England, we can see how English students performed on the sub-sets of items identified by each other country
- finally, reading the diagonal cells provides a comparison of results based on the sub-sets of items identified by each country as being relevant to their own curriculum.

Test-curriculum Matching Analysis data are presented in tabular form throughout the report as appropriate. The international tables B.1 and B.2 (mathematics) and B.1 and B.2 (science) are reproduced from the international reports (Beaton *et al.*, 1996b, 1996a).

III.7 Procedures used for quality control

In accordance with the requirements of the international Study Centre, two Quality Control Monitors (QCMs) were appointed, who were independent of the research team at NFER. Between them, they carried out ten visits to monitor the administration of the tests in schools, five in Population 1 schools and five in Population 2 schools. Aspects of the administration to which the QCMs paid particular attention included:

- the suitability of the accommodation for the test administration
- whether or not the test administrator deviated from the script provided
- the accuracy of the timing of the sessions
- whether or not the students worked independently, in appropriate test conditions.

The QCMs' reports containing their observations and interviews with Test Administrators were sent direct to the Study Centre.

Three further quality control visits carried out by members of the NFER research team: the same documents were completed as were used by the QCMs, although in this instance they were retained at NFER.

III.8 Summary

As part of TIMSS, two samples of students in English schools were surveyed: Population 1 comprised approximately 6,200 students aged about nine years old (National Curriculum Years 4 and 5) and Population 2 comprised approximately 3,500 students aged about 13 years old (National Curriculum Years 8 and 9). At each population, students completed:

- tests of mathematics and science
- a questionnaire about their learning experiences and attitudes towards these subjects.

Further relevant information, including details of teaching practices, resources for learning, and numbers of mathematics and science teachers within the school, was collected by means of questionnaires directed to:

- primary teachers
- primary headteachers
- secondary mathematics teachers
- secondary science teachers
- secondary headteachers.

In addition, a sub-sample of about 450 students in 50 secondary schools participated in an additional component of the study (the Performance Assessment) which focused on students' practical and investigative skills in mathematics and science.

To supplement the comparative data based on students' responses to all mathematics and science items, the Test-curriculum Matching Analysis provided inter-national comparisons on the basis of countries' self-selected sub-sets of items identified as being most relevant to their national curricula for mathematics and science.

Table B.1 (Mathematics) Test-curriculum Matching Analysis results–international eighth grade* (Year 9 in England)

Average percentages correct based on subsets of items specially 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	Singapore	Japan	Korea	Hong Kong	Belgium (Fl)	Czech Republic	Slovak Republic	Switzerland	Austria	Hungary	France	Slovenia	Russian Federation	Netherlands	Bulgaria	Canada	Ireland	Belgium (Fr)	Australia	Israel	Sweden	Germany	New Zealand	Norway	England	United States	Denmark	Scotland	Latvia (LSS)	Spain	Iceland	Greece	Romania	Lithuania	Cyprus	Portugal	Iran, Islamic Rep.	Kuwait	Colombia	South Africa
	(Number of Score Points Included)																																								
	162**	144	153	148	150	140	150	152	133	147	162	140	151	126	116	119	147	145	138	154	159	127	155	145	150	130	162	135	125	161	158	133	76	143	155	124	152	147	140	133	129
Singapore	79 (0.9)	79	79	80	79	79	79	79	80	80	79	79	79	81	79	80	79	79	79	79	79	79	79	79	79	81	79	80	80	79	79	80	79	79	79	80	79	79	80	79
Japan	73 (0.4)	73	73	74	73	73	73	73	75	74	73	73	73	74	73	73	74	73	73	74	73	73	73	75	74	75	73	75	74	73	73	74	74	74	73	73	73	74	74	74	73
Korea	72 (0.5)	71	72	73	72	71	72	71	72	72	72	71	71	72	70	72	72	71	71	72	72	71	72	73	72	73	72	73	72	72	72	73	72	72	71	71	72	72	72	72	70
Hong Kong	70 (1.4)	70	70	71	70	70	70	70	71	71	70	70	70	71	69	71	70	70	69	70	70	69	70	71	70	71	70	71	70	70	70	70	70	70	70	70	70	70	71	70
Belgium (Fl)	66 (1.4)	65	65	67	65	65	65	65	68	66	66	66	65	67	64	65	66	65	65	66	66	67	66	67	66	68	66	67	67	66	65	67	65	66	65	65	65	66	66	66	65
Czech Republic	66 (1.1)	65	66	67	66	66	66	66	68	66	66	66	66	67	64	65	66	66	65	66	66	66	66	67	66	67	66	67	66	66	65	67	68	66	66	65	66	66	66	66	65
Slovak Republic	62 (0.8)	63	63	64	63	63	63	63	65	63	62	63	63	64	61	63	63	62	62	63	62	62	63	63	62	64	62	64	62	62	62	64	64	62	63	62	63	63	63	63	63
Switzerland	62 (0.6)	61	62	63	61	61	61	62	64	62	62	62	61	62	62	60	63	61	60	62	62	64	62	64	62	66	62	64	65	62	62	63	60	61	62	61	62	62	62	61	61
Austria	62 (0.8)	62	62	63	62	61	62	62	64	62	62	62	61	63	61	61	63	62	61	62	62	64	62	63	62	65	62	64	63	62	62	63	62	62	62	62	62	62	62	62	62
Hungary	62 (0.7)	61	61	63	61	61	61	61	63	62	62	61	61	62	60	61	62	61	61	62	62	62	62	63	62	63	62	63	61	62	61	62	63	61	61	61	61	62	62	62	61
France	61 (0.8)	61	61	62	61	60	61	61	63	61	61	61	61	62	60	60	62	61	60	62	61	62	61	63	62	63	61	63	62	61	61	62	60	61	61	60	61	61	62	62	60
Slovenia	61 (0.7)	61	61	62	61	61	61	61	63	62	61	61	61	63	59	61	62	61	61	62	61	61	61	62	61	63	61	62	62	61	61	61	62	62	61	61	61	62	62	62	61
Russian Federation	60 (1.3)	61	60	61	60	60	60	60	62	61	60	60	60	62	58	61	61	60	60	61	60	60	60	61	60	61	60	61	60	60	60	61	62	60	60	60	60	61	61	61	60
Netherlands	60 (1.6)	59	60	60	59	58	59	59	61	60	60	59	59	59	59	58	61	59	58	60	60	62	60	63	61	63	60	62	62	60	59	61	60	59	59	58	60	60	60	59	58
Bulgaria	60 (1.2)	60	59	61	59	59	59	59	61	60	60	59	59	61	57	60	60	59	59	60	60	58	60	61	60	61	60	61	59	60	60	61	61	60	59	58	60	60	60	61	59
Canada	59 (0.5)	58	58	59	58	58	58	58	60	59	59	58	58	59	58	57	60	58	57	59	59	60	59	61	59	62	59	61	60	59	58	60	58	59	58	58	59	59	59	58	58
Ireland	59 (1.2)	58	58	59	58	58	58	58	60	59	59	58	58	59	58	57	59	58	57	59	59	60	58	61	59	62	59	60	60	59	58	60	59	58	58	58	59	59	59	58	58
Belgium (Fr)	59 (0.9)	58	58	60	58	58	58	58	61	59	59	59	58	59	57	58	59	58	58	59	59	60	59	60	59	62	59	61	60	59	59	59	58	58	58	58	58	59	59	59	58
Australia	58 (0.9)	57	58	59	57	57	58	58	60	58	58	57	57	58	57	56	59	57	56	58	58	59	58	61	58	61	58	60	60	58	58	59	59	58	58	57	58	59	58	57	57
Israel	57 (1.3)	57	57	59	57	57	57	57	59	58	57	57	57	59	55	57	58	57	57	58	58	58	57	59	57	59	57	58	58	57	57	57	60	58	57	57	57	58	58	58	57
Sweden	56 (0.7)	55	55	56	55	54	55	55	57	55	56	55	54	55	56	53	57	55	54	56	56	60	55	59	57	59	56	58	59	55	55	57	53	55	55	54	56	56	56	54	54
Germany	54 (1.1)	53	54	55	53	53	54	53	56	54	54	54	53	54	53	52	55	54	52	54	54	55	54	56	54	57	54	56	55	54	54	55	54	53	54	53	54	54	54	54	53
New Zealand	54 (1.0)	53	54	54	53	52	53	53	55	54	54	53	53	53	53	51	55	53	52	54	54	56	54	57	54	57	54	56	57	54	54	55	55	53	53	52	54	55	54	53	52
Norway	54 (0.5)	53	53	54	53	52	53	53	56	54	54	53	53	53	53	51	54	53	52	54	54	56	53	56	54	57	54	56	56	53	53	55	52	53	53	52	53	54	54	52	52
England	53 (0.7)	52	53	53	52	51	52	52	54	53	53	52	52	52	52	49	54	52	50	53	53	55	53	56	54	57	53	55	56	53	53	54	55	52	52	51	53	54	53	51	51
United States	53 (1.1)	52	53	54	52	52	53	52	54	53	53	52	52	53	52	51	54	52	51	53	53	54	53	55	53	56	53	55	54	53	53	54	54	53	53	52	53	54	53	52	52
Denmark	52 (0.7)	51	52	52	52	51	51	52	54	52	52	51	51	52	52	49	53	51	50	53	52	54	52	55	53	56	52	55	55	52	52	53	53	51	52	50	52	53	52	51	50
Scotland	52 (1.3)	51	51	52	51	50	51	51	53	51	52	50	51	51	51	48	52	51	49	52	52	54	51	55	52	55	52	54	55	52	51	52	53	51	51	49	52	52	52	50	50
Latvia (LSS)	51 (0.8)	51	51	52	51	51	51	51	53	52	51	51	51	53	50	51	52	51	51	52	51	51	52	53	51	53	51	53	52	51	51	52	53	51	51	50	51	52	52	52	51
Spain	51 (0.5)	50	51	52	51	50	51	51	52	51	51	51	50	51	48	50	51	50	50	51	51	51	51	52	51	53	51	52	51	51	51	52	53	51	51	50	51	51	51	51	50
Iceland	50 (1.1)	49	50	50	49	48	49	49	52	49	50	49	49	49	49	47	50	49	48	50	50	53	49	53	50	54	50	53	53	50	49	51	49	49	49	48	49	51	50	49	48
Greece	49 (0.7)	49	49	50	49	49	49	49	51	49	49	49	49	50	48	49	50	49	49	50	49	50	49	51	50	52	49	51	50	49	49	50	50	49	49	49	49	50	50	49	48
Romania	49 (1.0)	49	49	50	49	49	49	49	50	50	49	49	49	51	47	50	49	49	49	49	49	48	49	50	49	50	49	49	48	49	49	49	50	49	49	49	49	50	49	51	49
Lithuania	48 (0.9)	49	49	49	48	49	49	49	51	49	48	48	48	50	47	48	50	48	48	49	48	49	49	50	49	50	48	50	49	48	49	49	50	48	49	47	49	49	49	49	48
Cyprus	48 (0.5)	48	47	49	47	47	48	47	49	48	48	48	47	49	46	48	48	47	47	48	48	48	48	49	48	50	48	48	48	48	48	48	48	48	47	47	48	48	48	48	47
Portugal	43 (0.7)	42	43	43	43	42	43	43	45	43	43	42	42	43	41	41	44	42	41	43	43	44	43	45	43	45	43	45	44	43	43	43	44	42	43	41	43	44	43	43	41
Iran, Islamic Rep.	38 (0.6)	38	38	39	38	38	38	38	39	38	38	38	38	39	37	38	39	37	37	38	38	37	38	39	38	40	38	39	38	38	38	39	39	38	38	37	38	38	38	39	37
Kuwait	30 (0.7)	29	30	30	29	29	29	29	31	29	30	29	30	30	29	28	30	29	29	30	30	29	30	32	30	31	30	30	30	30	30	30	32	30	30	28	30	30	30	30	28
Colombia	29 (0.8)	29	29	30	29	29	29	29	31	29	29	29	29	29	28	28	30	29	28	30	29	30	29	31	30	32	29	30	30	29	29	30	31	29	29	28	29	30	30	29	28
South Africa	24 (1.1)	23	24	25	24	24	24	24	25	24	24	23	24	24	23	24	24	24	24	24	24	24	24	25	24	25	24	24	24	24	24	25	24	24	24	23	24	24	24	24	23
International Average	55 (0.9)	55	55	56	55	54	55	55	57	55	55	55	55	56	54	54	56	55	54	55	55	56	55	57	55	58	55	57	56	55	55	56	56	55	55	54	55	56	55	55	54

*Eighth grade in most countries; see Table 2 for more information about the grades tested in each country.

**Of the 151 items in the mathematics test, some items had two parts and some extended-response items were scored on a two- or three-point scale, resulting in 162 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: Beaton et al., 1996b

Table B.2 (Mathematics) Test-curriculum Matching Analysis results–international seventh grade* (Year 8 in England)

Average percentages correct based on subsets of items specially 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	Singapore	Japan	Korea	Hong Kong	Belgium (Fl)	Czech Republic	Austria	Bulgaria	Netherlands	Belgium (Fr)	Slovak Republic	Hungary	Ireland	Switzerland	Russian Federation	Slovenia	Australia	Canada	France	Germany	United States	England	Sweden	New Zealand	Scotland	Norway	Latvia (LSS)	Denmark	Romania	Iceland	Spain	Cyprus	Greece	Lithuania	Portugal	Iran, Islamic Rep.	Colombia	South Africa
	(Number of Score Points Included)																																						
	162**	126	145	144	139	104	145	131	119	81	104	152	158	113	91	121	144	114	87	128	132	162	93	100	115	76	118	150	59	87	102	150	101	108	146	148	128	89	81
Singapore	73 (1.3)	74	73	74	74	77	73	76	74	76	74	73	73	75	77	75	74	75	78	73	74	73	76	73	75	78	74	73	78	75	74	72	75	76	73	74	73	77	77
Japan	67 (0.4)	66	68	68	68	70	67	68	67	69	66	67	68	67	70	68	67	71	74	67	68	67	72	68	71	71	69	67	74	67	70	67	67	68	67	67	67	71	70
Korea	67 (0.6)	66	67	68	68	69	67	68	67	67	65	67	68	66	70	68	67	70	73	66	68	67	70	67	70	70	68	67	75	69	69	67	66	68	67	67	66	70	69
Hong Kong	65 (1.8)	65	66	66	66	68	65	67	66	67	64	65	66	65	67	67	65	68	69	65	65	65	69	66	68	70	66	65	72	66	66	65	65	67	65	65	65	69	68
Belgium (Fl)	65 (0.8)	65	65	66	65	69	65	67	65	67	66	64	65	66	71	66	65	70	73	66	66	65	72	67	70	74	68	65	76	66	69	65	65	68	65	65	64	69	70
Czech Republic	57 (1.2)	57	58	59	59	61	58	59	56	58	58	57	58	57	63	58	57	61	65	58	58	57	63	59	61	64	60	57	69	60	61	57	56	59	57	58	57	62	62
Austria	56 (0.7)	55	56	56	56	59	56	57	54	57	56	55	56	56	62	56	56	60	64	56	57	56	62	59	61	64	59	56	68	57	59	55	55	57	55	56	54	61	60
Bulgaria	55 (1.7)	55	55	57	57	59	55	57	57	53	55	55	56	55	57	57	55	58	60	55	55	55	58	54	58	60	55	55	63	58	57	55	53	56	55	56	56	60	57
Netherlands	55 (1.0)	53	55	56	54	57	55	56	53	59	56	54	55	54	62	55	55	60	63	55	58	55	65	60	62	65	62	55	69	54	60	56	54	56	54	55	53	59	59
Belgium (Fr)	54 (0.9)	53	55	55	55	57	54	56	53	56	56	54	55	55	61	55	54	60	63	55	57	54	62	58	60	64	59	54	68	55	59	54	54	56	54	54	54	60	59
Slovak Republic	54 (0.8)	54	55	56	56	58	55	56	54	54	54	54	55	54	59	56	54	58	62	54	55	54	59	55	58	60	56	54	65	57	57	54	53	56	54	55	54	59	59
Hungary	54 (0.8)	53	54	55	54	57	54	54	53	54	53	53	54	53	58	55	53	58	61	53	55	54	60	55	58	60	56	54	64	55	57	53	52	55	53	54	53	58	58
Ireland	53 (1.0)	53	53	54	54	56	53	54	52	55	54	53	54	55	60	54	53	58	61	53	55	53	60	56	58	61	56	54	64	54	56	53	53	55	53	53	52	56	58
Switzerland	53 (0.5)	52	53	53	53	55	53	54	50	57	55	52	53	54	62	53	53	60	63	54	56	53	62	59	60	65	60	54	68	52	58	53	53	55	53	53	52	58	59
Russian Federation	53 (0.9)	53	53	54	54	56	53	54	53	52	53	53	53	53	57	55	53	56	59	54	53	53	57	53	55	59	54	53	64	56	55	53	52	55	53	54	53	57	56
Slovenia	53 (0.7)	51	53	53	53	55	53	53	51	53	52	52	53	52	58	54	52	57	60	52	54	52	58	54	57	59	55	53	65	55	56	53	51	54	52	53	51	58	56
Australia	52 (0.8)	51	52	53	53	55	52	53	50	54	52	52	53	52	59	52	52	57	60	53	54	52	61	57	59	61	57	53	65	51	57	52	51	54	52	52	51	56	56
Canada	52 (0.5)	50	52	52	52	54	51	52	49	54	52	51	52	51	59	51	51	57	61	51	54	52	60	55	58	60	56	52	65	51	56	51	51	53	51	51	50	55	56
France	51 (0.8)	49	51	51	51	52	51	51	49	52	51	50	51	49	56	51	51	55	60	51	53	51	58	54	57	60	55	50	65	51	56	52	49	51	50	51	49	56	54
Germany	49 (1.0)	48	49	49	49	52	49	50	47	51	50	48	49	49	56	49	49	55	59	49	51	49	57	53	56	58	53	49	64	50	54	49	48	51	49	49	47	54	54
United States	48 (1.2)	46	48	48	48	50	48	48	46	50	47	47	48	47	54	47	47	52	56	48	49	48	55	50	54	55	52	48	59	48	52	47	46	49	47	48	46	51	51
England	47 (0.9)	45	47	47	47	48	47	47	44	47	46	46	47	45	53	45	46	52	55	47	49	47	57	52	54	56	53	47	60	45	52	48	44	47	46	47	45	49	49
Sweden	47 (0.6)	45	47	47	47	48	47	47	44	49	48	46	47	47	55	46	46	53	56	47	49	47	57	54	55	57	54	47	62	45	53	48	45	47	47	47	44	51	50
New Zealand	46 (0.9)	44	46	46	46	48	46	47	43	48	47	45	47	45	53	45	46	51	54	46	48	46	56	51	53	55	52	46	59	44	51	47	44	47	46	46	45	49	49
Scotland	44 (0.9)	42	44	44	44	46	44	45	41	46	45	44	45	43	51	43	44	50	53	45	46	44	55	50	51	54	51	44	58	42	50	45	42	44	43	44	43	47	47
Norway	44 (0.7)	43	45	44	45	46	45	45	42	47	46	43	44	44	52	44	44	50	53	45	47	44	54	50	52	54	51	45	59	43	50	45	43	45	44	44	42	48	48
Latvia (LSS)	44 (0.7)	43	45	45	45	46	44	45	44	43	44	44	44	43	49	45	44	48	50	45	45	44	50	46	48	50	46	44	56	47	47	44	42	45	44	44	43	49	47
Denmark	44 (0.5)	42	44	44	44	45	44	44	41	45	44	43	44	43	50	43	43	49	53	44	46	44	54	49	51	54	50	44	60	44	49	44	41	44	43	44	42	48	46
Romania	43 (0.8)	44	44	45	44	46	44	45	44	41	43	43	44	43	45	45	44	45	47	44	43	43	46	43	45	48	44	43	51	47	45	43	42	44	44	44	44	48	44
Iceland	43 (0.7)	42	44	44	44	45	43	44	40	46	44	42	44	42	51	42	43	49	53	43	46	43	54	50	51	54	50	43	60	41	50	44	41	44	42	43	42	47	47
Spain	42 (0.6)	41	42	43	43	44	42	42	41	41	42	42	43	41	47	43	42	46	48	43	43	42	49	44	47	49	46	42	55	43	46	43	40	43	42	42	41	47	44
Cyprus	42 (0.4)	41	42	43	42	44	42	43	42	41	41	41	42	42	44	43	42	45	46	41	43	42	47	42	45	47	43	42	51	44	43	42	42	43	42	42	41	46	44
Greece	40 (0.6)	40	41	41	41	43	41	41	40	42	42	40	41	41	46	41	41	44	47	41	42	40	46	42	45	47	44	41	51	42	44	41	40	42	40	41	39	45	45
Lithuania	38 (0.8)	37	39	39	39	41	38	39	37	37	38	38	38	37	43	40	38	42	45	39	39	38	44	39	42	44	40	38	50	40	41	38	36	39	38	38	37	42	41
Portugal	37 (0.6)	35	37	37	37	38	37	37	35	36	37	36	37	36	42	37	36	41	44	37	38	37	43	39	42	44	40	37	49	38	41	37	34	37	37	37	35	41	39
Iran, Islamic Rep.	32 (0.5)	32	32	33	33	33	33	33	32	32	32	32	33	32	35	33	33	35	36	33	34	32	37	32	35	37	34	32	41	33	34	32	31	33	32	32	32	37	33
Colombia	26 (0.6)	25	26	26	27	27	26	26	25	26	26	26	26	26	30	26	26	30	32	26	27	26	31	28	30	32	29	26	35	27	29	26	24	27	26	26	25	30	28
South Africa	23 (0.9)	22	22	23	23	24	22	23	22	23	23	22	23	23	25	23	22	24	26	22	23	23	25	23	24	26	24	23	29	23	25	22	22	24	22	23	22	24	24
International Average	49 (0.8)	48	49	50	50	52	49	50	48	50	49	49	49	49	55	50	49	53	56	49	51	49	56	52	54	56	53	49	61	50	53	49	48	50	49	49	48	53	52

*Seventh grade in most countries; see Table 2 for more information about the grades tested in each country.

**Of the 151 items in the mathematics test, some items had two parts and some extended-response items were scored on a two- or three-point scale, resulting in 162 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: Beaton et al., 1996b

Table B.1 (Science) Test-curriculum Matching Analysis results–international eighth grade* (Year 9 in England)

Average percentages correct based on subsets of items specially 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	Singapore	Korea	Japan	Czech Republic	Netherlands	Bulgaria	Slovenia	Austria	England	Hungary	Belgium (Fl)	Australia	Slovak Republic	Sweden	Canada	Ireland	United States	Russian Federation	Germany	New Zealand	Norway	Hong Kong	Israel	Switzerland	Spain	Scotland	France	Iceland	Greece	Denmark	Belgium (Fr)	Latvia (LSS)	Portugal	Romania	Lithuania	Iran, Islamic Rep.	Cyprus	Kuwait	Colombia	South Africa
	(Number of Score Points Included)																																								
	146**	109	59	86	136	102	112	140	131	124	129	98	133	129	125	121	90	146	96	129	126	111	68	102	105	146	97	73	146	111	70	58	113	133	99	120	87	78	131	112	74
Singapore	70 (1.0)	72	72	73	69	72	70	69	70	71	69	69	70	70	71	71	72	70	70	70	72	70	71	71	72	70	71	71	70	71	72	70	68	70	71	71	69	72	70	72	73
Korea	66 (0.3)	66	67	68	66	67	66	66	66	66	65	67	65	66	66	67	67	66	68	66	67	67	65	66	68	66	67	67	66	65	67	67	65	65	65	68	65	66	66	67	65
Japan	65 (0.3)	66	65	70	65	68	66	66	67	66	65	66	65	66	66	67	66	65	68	66	67	66	66	66	68	65	67	65	65	65	68	67	65	65	65	67	65	66	66	65	66
Czech Republic	64 (0.8)	65	66	67	64	66	65	64	65	63	65	64	64	66	66	66	67	64	67	64	65	65	65	65	67	64	64	65	64	65	66	68	62	64	64	66	64	66	64	66	65
Netherlands	62 (1.0)	63	62	66	63	66	63	62	63	63	63	64	62	63	64	65	64	62	64	63	65	64	64	64	66	62	63	64	62	63	66	67	61	62	63	64	61	65	62	63	61
Bulgaria	62 (1.0)	64	65	65	62	64	64	62	63	61	62	61	61	63	63	62	63	62	66	62	62	63	62	62	65	62	61	64	62	62	64	62	60	62	63	64	61	64	61	64	64
Slovenia	62 (0.5)	62	63	64	62	63	62	62	63	61	62	63	61	63	63	63	63	62	64	62	63	63	62	62	65	62	62	64	62	62	63	66	60	61	60	64	61	63	61	63	63
Austria	61 (0.7)	63	65	65	61	63	62	62	63	61	62	61	61	63	63	63	64	61	64	62	63	63	63	62	64	61	61	63	61	63	64	64	59	61	62	63	62	64	61	63	62
England	61 (0.6)	62	62	63	62	64	61	61	62	62	62	63	61	62	62	64	63	61	62	61	64	63	63	62	64	61	62	61	61	63	64	63	59	61	61	63	61	64	61	63	61
Hungary	61 (0.6)	62	63	64	60	62	61	61	61	60	62	61	61	63	62	62	61	61	64	61	62	62	60	61	63	63	59	62	61	62	61	63	59	61	60	63	61	64	60	62	61
Belgium (Fl)	60 (1.1)	61	62	64	61	64	60	60	61	60	61	62	60	62	62	63	63	60	62	61	62	63	63	62	64	60	61	62	60	61	65	65	59	60	60	62	59	63	60	62	59
Australia	60 (0.7)	61	61	62	60	62	60	60	60	60	60	60	60	61	61	62	62	60	60	60	62	61	60	60	62	60	60	60	60	61	62	64	58	59	59	61	59	62	60	61	60
Slovak Republic	59 (0.6)	60	59	63	60	61	60	59	61	58	60	59	59	61	60	61	62	59	63	60	60	61	62	60	62	59	59	61	59	60	61	62	58	59	60	62	60	63	59	60	61
Sweden	59 (0.6)	58	60	63	59	61	59	59	59	58	59	58	59	60	61	60	60	59	61	59	61	61	60	60	63	59	60	61	59	59	62	63	57	58	59	61	58	61	59	60	59
Canada	59 (0.5)	59	58	61	59	61	59	59	59	59	59	59	59	59	60	61	61	59	60	59	61	60	60	60	62	59	59	59	59	60	61	62	58	58	58	61	57	61	58	60	59
Ireland	58 (0.9)	59	60	59	59	60	58	58	59	59	59	59	58	59	60	60	62	58	59	58	60	60	60	60	61	58	59	59	58	60	61	60	57	58	58	60	57	61	58	60	57
United States	58 (1.0)	59	60	60	58	60	58	58	59	59	59	59	58	59	60	60	61	58	59	58	60	60	59	59	61	58	58	60	58	59	62	63	57	58	57	60	56	61	58	60	58
Russian Federation	58 (0.8)	59	61	62	59	61	59	58	59	58	59	58	58	60	60	60	61	58	65	59	59	60	60	60	61	58	59	60	58	59	61	61	57	58	59	61	59	61	59	60	61
Germany	58 (1.0)	59	60	62	58	60	58	58	59	58	58	58	58	59	59	60	60	58	61	59	60	59	59	58	61	58	58	59	58	59	60	61	56	57	57	59	57	61	58	59	59
New Zealand	58 (0.8)	58	58	60	58	60	58	57	58	58	58	58	58	58	59	59	59	58	58	58	60	59	59	59	60	58	58	58	58	59	60	60	56	57	57	59	56	60	58	59	57
Norway	58 (0.4)	56	56	61	58	60	57	58	58	58	58	58	58	58	59	59	59	58	58	58	60	60	60	58	62	58	59	57	58	58	62	61	57	57	58	60	56	60	58	59	58
Hong Kong	58 (1.0)	59	58	61	58	60	59	58	59	58	57	57	57	58	59	59	59	58	61	58	60	58	59	58	61	58	58	60	58	58	59	62	56	57	57	59	57	59	57	60	60
Israel	57 (1.1)	57	57	60	57	59	57	57	58	56	58	58	57	58	58	58	59	57	58	57	58	58	58	58	59	57	57	57	57	58	60	60	55	57	57	59	56	61	57	58	56
Switzerland	56 (0.5)	57	56	60	56	59	56	57	57	56	57	57	56	58	58	59	57	56	58	57	58	58	58	57	60	56	56	57	56	57	59	60	55	56	57	58	56	59	56	58	56
Spain	56 (0.4)	56	57	58	56	57	56	56	57	55	56	56	55	57	56	57	57	56	57	55	57	58	56	56	57	56	55	55	56	57	58	58	53	55	55	57	55	59	55	57	55
Scotland	55 (1.0)	56	55	58	56	58	56	55	56	56	56	56	55	56	56	57	57	55	56	55	58	57	57	56	58	55	56	55	55	57	57	57	53	55	56	57	55	59	55	57	55
France	54 (0.6)	54	54	57	54	56	53	54	54	54	55	54	53	55	55	56	55	54	56	54	55	55	54	55	57	54	54	57	54	54	55	59	53	53	53	57	52	56	53	54	52
Iceland	52 (0.9)	51	51	55	52	55	53	52	53	52	52	53	51	53	53	54	52	52	54	53	55	55	53	53	55	52	52	52	52	52	55	57	51	51	52	54	51	55	52	53	51
Greece	52 (0.5)	53	53	54	52	53	53	52	53	52	53	52	52	53	53	53	53	52	53	52	53	53	53	53	54	52	52	52	52	54	52	53	50	52	52	54	52	55	52	54	54
Denmark	51 (0.6)	50	49	54	51	53	53	51	52	51	51	52	50	52	52	53	51	51	53	51	53	53	51	52	54	51	52	52	51	51	55	57	49	50	50	53	50	53	51	51	50
Belgium (Fr)	50 (0.7)	50	50	54	50	53	51	50	51	49	51	51	50	51	52	52	51	50	53	51	52	52	51	51	54	50	50	52	50	50	52	56	49	50	50	53	49	53	50	51	49
Latvia (LSS)	50 (0.6)	51	52	55	50	52	51	50	51	51	51	51	50	52	51	52	51	50	54	51	51	51	52	52	53	50	50	52	50	50	51	54	49	50	51	52	50	52	50	51	51
Portugal	50 (0.6)	51	53	53	50	52	50	50	51	50	51	49	50	51	51	51	52	50	52	50	51	51	49	50	52	50	49	52	50	51	52	54	48	50	49	52	49	54	50	52	49
Romania	50 (0.8)	51	52	52	50	51	50	50	51	49	50	51	50	52	50	51	52	50	54	50	50	51	49	50	52	50	48	52	50	49	52	52	47	50	48	51	50	51	49	52	51
Lithuania	49 (0.7)	50	51	53	49	51	50	49	50	48	50	49	49	51	50	50	51	49	55	49	49	50	50	50	51	49	49	51	49	49	50	53	47	49	50	52	49	50	49	50	51
Iran, Islamic Rep.	47 (0.6)	49	51	50	48	48	49	47	49	48	48	48	48	49	48	49	49	47	51	48	48	49	49	47	49	47	47	48	47	49	47	47	47	48	50	49	50	52	48	50	48
Cyprus	47 (0.4)	48	47	50	47	48	47	47	48	47	47	47	46	48	48	48	50	47	49	47	48	48	49	48	50	47	47	48	47	48	48	50	45	47	47	49	47	51	47	49	48
Kuwait	43 (0.9)	43	42	45	43	44	44	43	44	43	42	43	42	43	43	44	44	43	44	43	44	45	43	42	44	43	42	44	43	44	44	43	42	43	44	44	43	47	43	45	43
Colombia	39 (0.8)	39	40	41	39	40	39	39	40	39	39	40	39	39	39	40	41	39	40	38	40	40	39	40	41	39	39	39	39	40	42	41	36	39	38	40	38	41	38	41	39
South Africa	27 (1.3)	28	27	29	27	28	27	27	28	27	26	27	26	27	27	27	29	27	30	27	27	27	28	27	28	27	26	28	27	27	28	29	26	27	27	28	27	28	26	28	28
International Averag	**55 (0.7)**	**56**	**56**	**59**	**56**	**58**	**56**	**56**	**56**	**56**	**56**	**56**	**55**	**57**	**57**	**57**	**57**	**55**	**58**	**56**	**57**	**57**	**57**	**56**	**58**	**55**	**56**	**57**	**55**	**56**	**58**	**58**	**54**	**55**	**56**	**57**	**55**	**58**	**55**	**57**	**56**

*Eighth grade in most countries; see Table 2 for more information about the grades tested in each country.

**Of the 135 items in the science test, some items had two parts and some extended-response items were scored on a two- or three-point scale, resulting in 146 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: Beaton et al., 1996a.

Table B.2 (Science) Test-curriculum Matching Analysis results–international seventh grade* (Year 8 in England)

Average percentages correct based on subsets of items specially 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	Singapore	Korea	Japan	Czech Republic	Slovenia	Belgium (Fl)	Bulgaria	Netherlands	England	Hungary	Austria	Slovak Republic	United States	Canada	Australia	Hong Kong	Germany	Ireland	Sweden	New Zealand	Norway	Switzerland	Russian Federation	Spain	Scotland	Iceland	France	Belgium (Fr)	Romania	Greece	Denmark	Iran, Islamic Rep.	Latvia (LSS)	Portugal	Cyprus	Lithuania	Colombia	South Africa
	(Number of Score Points Included)																																					
	146**	83	43	45	108	132	46	105	34	104	98	52	111	146	78	94	32	88	60	86	110	92	36	49	132	49	146	26	23	91	72	20	48	46	81	29	79	108	26
Singapore	61 (1.2)	66	65	65	62	62	62	63	63	63	61	65	62	61	65	66	62	63	65	67	65	64	66	65	62	66	61	68	61	64	63	68	63	62	62	66	64	64	66
Korea	61 (0.4)	63	64	67	62	62	63	64	65	63	61	67	63	61	65	63	60	65	63	66	65	63	68	67	62	65	61	69	61	62	64	68	61	62	63	63	65	62	68
Japan	59 (0.3)	60	61	67	60	60	60	61	63	61	58	63	62	59	62	61	59	64	59	66	62	61	64	66	60	63	59	65	63	61	62	66	61	59	60	65	62	59	65
Czech Republic	58 (0.8)	60	61	64	61	59	60	61	63	59	59	64	62	58	60	60	55	63	62	65	62	61	63	64	59	59	58	65	65	62	61	62	62	57	59	59	63	61	61
Slovenia	57 (0.5)	59	60	61	58	58	59	58	60	57	58	61	60	57	58	59	55	61	62	62	59	60	63	62	57	59	57	65	65	59	59	60	61	57	59	58	61	59	61
Belgium (Fl)	57 (0.5)	58	58	60	59	58	60	58	65	58	59	61	60	57	61	59	59	62	60	66	61	63	66	63	59	61	57	61	64	59	62	66	63	57	59	58	64	59	58
Bulgaria	56 (1.0)	57	60	60	57	57	58	60	60	57	56	61	60	56	58	57	54	59	59	63	59	59	60	62	57	56	56	62	61	58	57	62	57	53	57	55	60	58	59
Netherlands	56 (0.7)	58	59	58	58	56	58	57	63	57	56	59	59	56	59	58	57	60	59	64	60	60	65	60	58	61	56	62	62	58	61	62	59	57	59	60	62	58	59
England	56 (0.6)	57	56	57	56	56	56	57	60	58	55	56	57	56	58	58	55	58	57	62	60	59	62	58	57	60	56	56	59	57	58	58	59	54	57	55	59	58	59
Hungary	56 (0.6)	57	57	57	57	56	56	57	61	55	56	59	58	56	56	58	52	59	59	60	58	58	60	62	56	53	56	60	60	58	57	61	59	52	56	51	60	58	57
Austria	55 (0.6)	58	56	58	55	56	57	56	60	55	56	60	59	55	57	57	53	59	58	62	59	59	62	59	56	55	55	60	63	59	58	59	59	54	57	57	60	58	58
Slovak Republic	54 (0.6)	56	58	61	57	54	55	56	58	55	55	60	58	54	55	56	53	59	59	60	57	58	60	61	55	56	54	61	58	57	56	59	59	54	55	54	58	56	57
United States	54 (1.1)	55	54	54	55	54	56	55	59	55	54	54	56	54	56	56	54	55	57	60	58	57	61	56	55	58	54	59	59	56	58	57	56	54	57	54	58	56	58
Canada	54 (0.5)	55	56	55	54	54	55	55	59	55	54	56	56	54	57	56	53	55	55	61	58	57	62	57	55	58	54	58	57	56	57	57	57	53	56	54	58	56	57
Australia	54 (0.7)	56	57	56	54	54	54	56	58	55	53	55	56	54	55	56	52	57	56	60	57	56	61	56	55	57	54	57	59	56	57	54	57	52	55	53	57	56	58
Hong Kong	53 (1.2)	56	57	57	54	54	56	56	60	55	52	58	54	53	56	55	54	56	55	60	57	55	61	59	54	58	53	60	60	55	55	54	55	52	54	51	56	54	61
Germany	53 (0.8)	55	55	57	53	54	54	55	58	53	53	57	56	53	55	55	51	57	55	60	57	56	60	57	54	52	53	59	60	55	56	58	56	51	54	53	58	55	56
Ireland	52 (0.7)	54	52	51	52	52	54	52	53	53	53	52	54	52	53	55	53	53	56	57	55	55	59	53	53	57	52	52	53	54	55	55	56	54	54	52	56	54	53
Sweden	51 (0.5)	53	53	56	53	52	52	53	57	53	52	54	53	51	54	53	50	56	54	59	56	56	60	56	53	53	51	55	59	53	55	58	53	52	53	54	57	53	54
New Zealand	50 (0.7)	52	50	51	51	50	50	52	54	52	50	51	52	50	52	53	50	53	51	56	55	54	58	51	52	53	50	53	54	52	54	52	53	50	52	51	55	52	52
Norway	50 (0.6)	51	53	54	52	51	50	52	57	51	50	52	53	50	53	52	49	54	53	58	54	55	59	54	52	52	50	52	55	52	55	58	53	50	52	52	56	52	52
Switzerland	50 (0.4)	53	52	54	50	51	52	51	55	51	51	55	53	50	52	52	49	55	53	57	54	54	60	55	52	52	50	56	56	52	53	55	54	50	52	51	56	51	52
Russian Federation	50 (0.8)	52	51	55	51	51	52	51	57	51	51	53	53	50	52	51	50	53	55	57	52	53	59	61	51	53	50	52	56	52	54	58	53	52	53	54	57	52	53
Spain	49 (0.4)	50	49	50	50	49	50	50	52	50	50	50	52	49	50	50	47	51	51	55	52	53	54	51	50	51	49	53	54	51	52	52	52	49	51	48	53	52	49
Scotland	48 (0.8)	50	50	50	49	48	49	50	51	50	48	49	50	48	51	51	49	50	50	54	52	52	55	50	49	53	48	50	51	50	50	49	51	46	49	46	52	50	52
Iceland	46 (0.6)	47	48	49	48	46	47	49	53	47	46	46	49	46	47	47	44	49	49	53	50	50	51	50	48	47	46	49	53	48	49	48	49	44	48	43	51	48	51
France	46 (0.6)	48	49	50	46	47	47	47	48	48	47	51	48	46	48	47	45	49	46	53	49	50	55	50	47	50	46	57	50	48	48	51	47	44	47	44	50	47	48
Belgium (Fr)	45 (0.7)	47	49	48	45	46	47	47	49	46	45	50	47	45	47	46	44	48	47	52	48	49	53	50	46	48	45	54	52	47	47	46	49	42	46	40	49	46	47
Romania	45 (0.7)	46	45	47	46	45	47	47	50	45	45	49	48	45	45	46	42	47	48	50	46	47	48	50	45	44	45	52	51	47	47	46	48	43	46	43	48	47	50
Greece	45 (0.5)	45	45	47	46	44	44	46	48	46	45	45	46	45	45	48	43	47	45	49	48	47	48	46	45	47	45	50	46	47	48	48	47	43	46	45	48	47	51
Denmark	44 (0.4)	45	46	49	45	44	45	46	50	45	44	48	47	44	47	45	42	49	45	53	47	48	53	51	45	46	44	48	52	46	48	52	47	41	46	44	49	45	47
Iran, Islamic Rep.	42 (0.6)	46	41	44	43	43	44	45	44	43	42	41	43	42	44	45	42	45	46	46	44	44	44	44	41	42	42	43	44	44	43	48	45	40	42	40	45	45	47
Latvia (LSS)	42 (0.5)	44	43	46	42	42	45	43	47	43	42	45	44	42	44	43	41	45	46	48	44	45	48	48	42	43	42	47	49	44	44	50	45	41	44	41	47	43	44
Portugal	41 (0.5)	43	41	42	41	41	42	42	43	41	41	44	44	41	42	42	37	43	45	47	44	44	45	45	42	43	41	47	44	42	44	47	45	42	45	42	46	44	42
Cyprus	40 (0.4)	43	42	43	40	40	42	41	42	42	41	44	42	40	42	42	40	42	43	45	42	43	45	43	40	45	40	47	43	43	42	48	44	40	41	44	43	42	45
Lithuania	38 (0.7)	40	39	44	37	38	41	39	42	39	38	42	40	38	40	38	36	41	42	44	40	40	45	46	38	40	38	41	44	40	40	45	40	38	40	38	44	39	43
Colombia	35 (0.7)	37	36	35	35	35	36	37	39	35	35	37	37	35	34	36	33	35	38	40	37	37	34	36	35	38	35	39	40	37	38	36	39	33	37	34	38	38	39
South Africa	26 (1.0)	27	27	28	26	25	27	26	28	26	25	27	27	26	25	26	26	27	29	29	27	27	27	30	26	26	26	28	27	27	27	26	27	25	26	25	27	27	30
International Average	50 (0.7)	52	52	53	51	50	51	51	54	51	50	53	52	50	52	52	49	53	52	56	53	53	56	54	51	52	50	55	54	52	52	54	53	49	51	50	54	52	53

*Seventh grade in most countries; see Table 2 for more information about the grades tested in each country.

**Of the 135 items in the science test, some items had two parts and some extended-response items were scored on a two- or three-point scale, resulting in 146 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: Beaton et al., 1996a

APPENDIX IV

The development of the tests and questionnaires

IV.1 Introduction

This appendix provides details of the development of the research instruments used in the TIMSS study:

- the written achievement tests
- the questionnaires
- the practical tasks.

In addition, information about the procedures used for marking the students' test and task booklets is given. However, we begin with essential background information concerning the intended curriculum coverage of the tests, the test blueprints.

IV.2 The test blueprints and their development

Due to the different curricula in different countries, a common framework was essential for two purposes, to provide:

- reference points for the comparison of the mathematics and science curricula in place in participating countries
- guidelines for the development of tests so that they would be relevant for all the countries taking part in terms of the curriculum content.

Full details of the curriculum frameworks used for mathematics and science in the TIMSS study are described in Robitaille (1993) but some of the most pertinent points are summarised here. The framework for the purposes listed above (for both subjects) involves three distinct aspects:

- subject matter content
- performance expectations
- perspectives (or the context of the item).

The first area, content, relates to the content of the schools' mathematics and science curricula within the countries taking part in TIMSS. The performance expectations describe the types of response that may be expected from students given the specific subject matter items; these are reflected in the coding system used in the main survey. (The coding system used was a refinement of the codes used in the international field trial of

1994, and includes codes for particular types of response that were found to be relatively widespread; see section IV.6 for further details of the coding for the achievement tests and section IV.8 for details of the performance assessment coding.) The perspectives element is particularly concerned with the analysis of the curricula and teaching resources in individual countries, but, in terms of the test items, relates to the context within which the question is set. All three aspects are, therefore, relevant with regard to the test items.

The content element for each subject is divided into main categories which define major aspects of the curriculum; each of these is then divided into sub-categories, many of which are further divided into more specific areas; finally, there is elaboration of each of the specific areas within the sub-categories, providing further detail of the content covered. There are ten major categories in the mathematics framework; these are listed below, together with the number of sub-categories for each one.

Major category	Number of sub-categories
Numbers	5
Measurement	3
Geometry: position, visualisation, and shape	5
Geometry: symmetry, congruence, and similarity	3
Proportionality	4
Functions, relations, and equations	2
Data representation, probability, and statistics	2
Elementary analysis	2
Validation and structure	2
Other content	1

The science framework has eight major categories, most of which are split into sub-categories, as shown below.

Major category	Number of sub-categories
Earth sciences	3
Life sciences	5
Physical sciences	6
Science, technology, and mathematics	3
History of science and technology	0
Environmental and resource issues	6
Nature of science	2
Science and other disciplines	2

So as to illustrate how each of the major categories is broken down into further sub-categories, and the subsequent elaboration into more detail for each of these, let us take as an example the major mathematics category of *Measurement*. This has three sub-categories, as follows:

- units
- perimeter, area, and volume
- estimation and error.

Looking for more detail, we find that the sub-category *Perimeter, area, and volume* is elaborated thus: concepts of perimeter, area, surface area, volume; formulae for perimeters, areas, surface areas, and volume. Other major areas of subject matter are broken down into greater detail in a similar way.

The tests themselves comprised a mixture of three different types of item:

- multiple-choice questions
- items requiring a short answer
- questions requiring an extended response.

The multiple-choice format had already been used in the previous IEA studies of mathematics and science. Items of this type accounted for 77 per cent of all the assessment items for Population 1 and about 80 per cent for Population 2. Some of the multiple-choice questions selected for TIMSS had been used in either the Second International Mathematics Study, or the Second International Science Study, as appropriate; these were deemed 'link' items and provided opportunities for countries that had participated in the previous studies to compare their students' performance on these specific items.

Two types of item were defined as 'free response', that is questions for which the student had to construct his/her own response rather than selecting one of the choices offered, as was the case for the multiple-choice items. One type, the short answer, required a very brief response from the student, frequently a single word, number, time or measurement. The second type required an extended response, where, typically, students were asked to show all their working out to reach an answer to a mathematical question, or to explain the scientific principles relating to a particular phenomenon.

The totals of each type of item for each subject and each population are shown below. (It should be borne in mind that these figures represent the **total** number of items for each group; the actual percentages of each type of item varied between different versions of the achievement test, i.e. booklets numbered from one to eight.)

Population 1

	NUMBER			PERCENTAGE			
	multiple-choice	short answer	extended response	multiple-choice	short answer	extended response	Total
Mathematics	79	15	8	40%	8%	4%	52%
Science	74	13	10	37%	6%	5%	48%
Total	153	28	18	77%	14%	9%	100%

Population 2

	NUMBER			PERCENTAGE			
	multiple-choice	short answer	extended response	multiple-choice	short answer	extended response	Total
Mathematics	125	19	7	44%	7%	2%	53%
Science	102	22	11	36%	7%	4%	47%
Total	227	41	18	80%	14%	6%	100%

Each of the different versions of the achievement tests included multiple-choice items and some free-response questions. In addition to the multiple-choice items which were link items relating to the previous mathematics and science studies, some items (both multiple-choice and free-response) were included in the tests for both Populations 1 and 2, so that there were link items between populations. This latter category enabled comparisons to be made concerning the achievement of students of different ages on the same items.

The final selection of test items for inclusion in the main survey was made on the basis of the following:

- item statistics based on the responses of students who participated in the international field trial (in which England took part) in 1994
- relevance to the test blueprints for each subject, outlined above
- appropriateness to the curricula in participating countries, based on the feedback from expert review panels convened in each country taking part in the field trial (a panel of mathematics experts and one of science experts reviewed each of the assessment items for each population and rated them according to specific criteria, such as appropriateness to the national curriculum, typical content and context for both the subject and the target age range).

The content of the achievement tests was the same for all countries taking part in TIMSS, although, obviously, National Research Coordinators in each country had to arrange for the translation of the tests into the national language.

IV.3 The item clusters and compilation of the test booklets

In order to provide broad coverage of mathematics and science, with sufficient items to reflect adequately the curricula of all countries taking part in TIMSS, the test blueprints had identified a wide body of knowledge to be covered. This resulted in a considerable number of test items for each population, and the issue of how to achieve broad content coverage, whilst at the same time avoiding excessive burdens on individual students, was raised. The chosen solution involved collating questions into groups or 'clusters' of items for each population. Using the clusters of items, eight different versions of the test were prepared for each population.

Each version of the test was obtained by putting together a different combination of clusters; each had an alphabetical identifier, from A to Z; within the cluster, items were numbered consecutively, beginning with 1. The 26 clusters (A to Z) for Population 1 were different from those for Population 2. The number of items within clusters varied considerably: some contained as few as two items, whereas others comprised as many as 19. Each cluster contained items focusing on either mathematics, science, or both.

The test design involved the production of eight different versions of the tests for Populations 1 and 2; these were numbered from 1 to 8, and were allocated to the students selected for the study in rotation. This meant that in a Population 1 class of about 30 students, only three or four attempted the same version of the test; with 16 students sampled from Years 8 and 9 for Population 2, only two in each year group completed the same version.

Each version of the test was obtained by combining seven different clusters: the particular clusters for each version, together with the order in which they had to be presented, were specified by the Study Centre. For both Populations 1 and 2, the design required students to attempt four clusters in the first part of the test, then, after a 20-minute break, attempt a further three clusters; in England these two parts of the test were presented to students as two separate booklets, A and B. The clusters used in each version of the test are shown below.

Population 1

		Test version							
Booklet	**Cluster order**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
A	1st	B	C	D	E	F	G	H	B
	2nd	A	A	A	A	A	A	A	A
	3rd	C	D	E	F	G	H	B	R
	4th	S	W	T	X	U	Y	V	Z
		BREAK							
B	5th	E	F	G	H	B	C	D	I
	6th	J	N	K	O	L	P	M	Q
	7th	T	X	U	Y	V	Z	W	S

Population 2

		Test version							
Booklet	**Cluster order**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
A	1st	B	C	D	E	F	G	H	B
	2nd	A	A	A	A	A	A	A	A
	3rd	C	D	E	F	G	H	B	Q
	4th	S	W	T	X	U	Y	V	*
					BREAK				
B	5th	E	F	G	H	B	C	D	R
	6th	I	J	K	L	M	N	O	P
	7th	T	X	U	Y	V	Z	W	*

* *No cluster.*

Although there were fewer clusters in Booklet version 8, the total testing time, based on the number of items, was comparable since there were more items in clusters P, Q, and R than in some of the other clusters.

Some clusters were included in only one version of the test, others in two, three, or four versions; in both populations, cluster A was common to all eight versions. It was always placed as the second cluster in the first part of the test (booklet A for English students). This cluster comprised only multiple-choice items for each population; it was included so as to provide a core of items (both mathematics and science) attempted by all the students in a population which could be used for comparative purposes; it was positioned as the second cluster in order to reduce any 'warm-up' effects which may have affected students' performance.

To summarise, for Populations 1 and 2, each version of the achievement tests showed the following features:

- a mixture of mathematics and science items
- both multiple-choice and free-response questions
- cluster A was common to all versions, and was in the same position in each test.

The total testing times for Populations 1 and 2 (54 minutes and 90 minutes respectively) were calculated on the basis of allowing specific periods of time for the completion of different types of item, as shown below.

Item type	Population 1	Population 2
multiple-choice	1 minute	1 minute
short answer	1 minute	2 minutes
extended response	3 minutes	5 minutes

IV.4 The tests for nine-year-olds (Years 4 and 5)

The tests for Population 1 were a mixture of both mathematics and science items. In the international field trial, conducted in 1994, the test booklets had focused on either one subject or the other, and students taking part had attempted *either* one of the four different versions of the mathematics test, *or* one of four different science tests. Clearly, whilst this arrangement was satisfactory for trialling the items, it did not provide achievement data for students on each subject; a number of teachers in England whose students had been involved in the field trial commented on this and expressed their surprise that each student had not had opportunities to demonstrate his/her abilities in both subjects.

The subject content for each of the eight versions of the tests represented areas selected from the main categories defined in the curriculum frameworks (see section IV.2). Clearly, with the breadth of content identified, it was impossible to cover all major categories for either mathematics or science within any one version, hence the particular content varied from one booklet to another, although all versions addressed the curriculum content specified in the frameworks. In practice, this meant that the mathematics items in a particular test may have included items on numbers, measurement and (for geometry) symmetry, with no items concerning the other major categories; on the other hand, a different version of the test may have included numbers, proportionality, data representation and elementary analysis.

One cluster of test items, the A cluster, was common to all eight version of the tests: it comprised five mathematics and five science items, all of which were multiple-choice format. Each of the other six clusters in other versions of the test contained either mathematics or science items, or a mixture of both.

As previously stated, a number of items which had been included in either the Second International Mathematics Study or the Second International Science Study were selected for TIMSS (see section IV.2). Further items were selected from the item banks maintained by the IEA, and a small number of new items were generated specifically with the test blueprints for TIMSS in mind. These new items were subject to international trialling in a pilot study carried out in 1993 and a larger field trial, which involved about 15 countries, including England, in 1994. Twice as many items as were needed were used in the field trial, so as to provide a large bank from which to make the final selection. In many cases, amendments were made to the items in order to make them more acceptable to as many countries as possible. The revisions were prompted by:

- comments made by National Research Coordinators
- the views expressed by expert review panels convened for each subject in each country taking part in the field trial
- the item statistics as a result of the field trial.

Some items for Population 1 were also included in the tests for Population 2. The number of items for each subject that were identified as link items, either to a previous study or to Population 2, are shown below.

	Number of items
Link items from SIMS	NA[1]
Link items from SISS	12[2]
	Number of items
Mathematics link items to Population 2	15
Science link items to Population 2	17

IV.5 The tests for 13-year-olds (Years 8 and 9)

The tests for Population 2, like those for the younger age group, consisted of a mixture of mathematics and science questions in each of the two booklets that each student attempted. Again, the content for each of the eight versions of the test was determined by reference to the curriculum frameworks, so that, whilst each individual version did not cover all the major categories for either mathematics or science, taken collectively, the eight versions covered all the major categories for both subjects.

As with Population 1, one cluster of items (cluster A) was included in all eight versions of the test, always presented as the second cluster within the first booklet attempted. This cluster comprised 12 multiple-choice items in total, six mathematics and six science; none of these questions was a link item with Population 1.

Some of the items were selected from those used in the IEA's previous mathematics and science studies, as shown below.

	Number of items
Link items from SIMS	22
Link items from SISS	14[3]

Some items were link items with Population 1 (see above). Other questions were either chosen from the item banks maintained by the IEA, or were developed specifically for TIMSS, in accordance with the test blueprints.

---

[1] *SIMS did not include a primary school age population.*

[2] *In SISS, Population 1 consisted of ten-year-olds, i.e. the students were, on average, about a year older than those taking part in TIMSS.*

[3] *In SISS, Population 2 consisted of 14-year-olds. They were selected from a single year group equivalent to Year 9, whereas TIMSS tested students in Years 8 and 9.*

IV.6 Marking the free-response items in the main study

Whereas all the multiple-choice items in the achievement tests were pre-coded, and therefore needed no further attention before being processed and analysed, all those questions classified as free-response items were coded manually before the details were processed electronically. There were two types of free-response item that required coding: questions prompting a short answer, and those requiring an extended response. The international study centre used two strategies to ensure that research teams working in different countries would approach the coding of the free-response items in a similar way:

- Firstly, they provided guidelines for coding students' responses to the relevant questions in the form of rubrics for each item. These rubrics indicated various responses that students might have made: these were based on data collated from trials, so that, typically, the responses described in the rubrics had been offered by at least ten per cent of students.

- Secondly, international training sessions were held in a number of venues in order to train representatives from each country in the coding procedures, and to provide opportunities for these representatives to familiarise themselves with the structure and content of the rubrics.

The fact that all countries taking part in TIMSS used the same rubrics (translated into their native language) to code free-response questions and had undergone the same training process ensured that there were few, if any, opportunities for differences of interpretation concerning which code would be allocated to a specific response.

A particular feature of the coding used in TIMSS was the two-digit system which provided more information than whether or not a student had supplied the 'correct' answer. This approach was different from the single-digit coding used in the field trial, and was adopted so as to retain more information about the quality of the students' responses. The system was developed by a team of researchers representing several countries on the basis of the field trial data. The team drafted a new set of rubrics for a sample of free-response questions and re-coded a number of students' booklets from different countries to assess the validity of the approach. This showed that the two-digit system of coding was feasible, and had the advantage of retaining information about the quality of the response, in terms of the content included or approach used. As a result, this approach to coding was adopted for all the free-response questions in the main survey.

The concept behind the two-digit coding was that each digit conveyed a separate piece of information as follows:

- **The first digit** represented the 'correctness' code: this ranged from four to one, with a higher number representing a more accurate or fully

complete answer (the maximum correctness code varied from item to item, so that some questions had a maximum correctness code of four, whereas for others, the maximum was only three, two or even one). The number seven as a first digit (or prefix) indicated that the response as a whole was incorrect.

- **The second digit**, used as a 'suffix' to the first (correctness) code, represented the content included in the response, or the approach used to reach the answer. *(The numbers used ranged from zero to five, hence, where a question offered a maximum correctness code of two, the following codes may have been allocated to a particular response, depending on the content: 20, 21, 22, 23. For a partially correct answer, the possible codes may have been 10, 11, 12, 13. In addition, the suffix '9' was used where a particular response did not fit into any of the pre-specified categories with another suffix, hence code 29 for a fully correct answer which used a different approach from those already specified in the '2' prefix codes, code 19 for a partially correct answer using a different approach, and code 79 for an incorrect answer where the content had not already been specified under any of the preceding '7' codes.)*

Finally, specific codes were used to indicate particular types of response, where:

i. the student reiterated the information provided in the question (used for some science items only)

ii. the response was illegible, unintelligible, erased or crossed-out

iii. no response was made (i.e. completely blank).

These specific codes were used for both populations and for both subjects, with the exception of (i), which was not used for mathematics items.

Figure 1 shows the possible codes for one question.

In England, the coding exercise was carried out by two teams of coders, one working with Population 1 booklets, and the other with Population 2. The coders were chosen specifically for their suitability for the task: most were experienced teachers or lecturers and/or had higher qualifications in mathematics and science. All coders were trained and supervised by the NFER research team.

In order to ensure that the coding was accurate, all countries were required to implement the following procedures to monitor, and provide a measure of, the reliability of the exercise:

- Ten per cent of each batch of booklets was coded by two coders independently. The first one to examine the booklets recorded the codes allocated on a separate sheet, and the second marked the codes directly on the students' booklets; the codes from both the separate

sheets and the actual booklets were processed electronically. The data from the two sets of codes concerning the booklets dealt with in this way provided a measure of the inter-rater reliability (i.e. the level of agreement between different coders as to how to code a particular response).

- Once a batch of booklets had been coded, a random sample of ten per cent was reviewed by a member of the research team, to check the coding; any inaccuracies were rectified and discussed with the individual concerned.

Figure 1 Codes for *Item W2.*

W2 (ER): Diagram: Rain from Another Place

W2. Draw a diagram to show how the water that falls as rain in one place may come from another place that is far away.

Code	Response
	Correct Response
20	Response includes the three following steps: i. Evaporation of water from a source. ii. Transportation of water as vapour/clouds to another place. iii. Precipitation in other places.
	Partial Response
10	As in code 20, but response does not mention evaporation.
11	As in code 20, but response does not mention transportation.
12	As in code 20, but response does not mention precipitation.
19	Other partially correct.
	Incorrect Response
70	Responses indicates precipitation only; it may use vertical or diagonal lines.
79	Other incorrect.
	Non-response
90	Crossed out/erased, illegible, or impossible to interpret.
99	BLANK

IV.7 The Performance Assessment tasks

The Performance Assessment component of TIMSS was designed to assess students' practical skills in mathematics and science, which was not possible within the context of the written achievement tests used in the main survey. This aspect of the study was intended to set out a number of specific tasks that would allow students to demonstrate their practical, investigative, recording and analytical skills in controlled situations.

Considerable effort was exerted in the development of suitable activities. Representatives from a number of different countries put forward a total of 22 tasks which provided opportunities for practical activities in mathematics and science. These tasks were subjected to a field trial in 15 countries, including England, during 1994. In addition to considerations such as the cost and complexity of equipment necessary, the clarity of the instructions, and the time required to complete each task, the field trial showed that local features such as climate affected some tasks, and therefore meant that the conditions for administration of certain tasks in different countries were not identical.

As a result of the field trial data and comments provided by National Research Coordinators concerning the proposed tasks, some activities were deemed unsuitable and some of those selected for the main survey were subject to minor amendment. A total of 12 tasks was chosen:

- five focusing on mathematics
- five focusing on science
- two containing both mathematical and scientific elements.

Some of the activities were short, 15-minute tasks requiring routine skills such as constructing tables to show results, and summarising observations, whereas others were long, 30-minute tasks involving skills such as planning and carrying out an investigation, graphing results and drawing conclusions.

The 12 tasks were set out at nine workstations within schools, each one having either one 30-minute task, or two 15-minute tasks. During the administration of the Performance Assessment, each student visited three workstations, attempting three, four or five tasks, depending on their length (see Appendix III, section III.5 for further details of the administration of the Performance Assessment component).

IV.8 Marking the Performance Assessment tasks

The booklets completed by the Population 2 students taking part in the Performance Assessment contained students' own records of the activities they had completed. Just as the free-response items in the written achievement tests had to be marked and coded, students' responses to the Performance Assessment tasks were dealt with in a similar way.

As with the written tests, the Study Centre supplied coding guides for each of the 12 different tasks, which were used by each country taking part in this aspect of the study. The guides (or rubrics) provided used the same principles as had been applied for the marking and coding of the written tests, whereby two-digit codes were allocated to students' responses, the first digit representing the correctness of the answer, and the second digit representing the content or approach used.

In order to retain detailed information about the quality and content of students' responses, in many cases more than one code was allocated to one question. For example, several tasks required students to construct a table and present data they had collected within the table; in these instances, one code was allocated for the *presentation* of the table (an appropriate number of columns and suitable labels for each one), and a further code for the *quality* of the data included within the table (appropriate increments and data within the expected range). By marking different aspects of the response separately, it was possible to collate information about students' abilities in relation to particular skills.

Figure 2 shows the range of codes that applied to part of the *Pulse* activity.

The arrangements for monitoring the accuracy of markers' work that were used in coding the written tests were also implemented for the Performance Assessment (i.e. a sample of booklets was independently coded by two markers and the supervisor reviewed a further sample). The procedures adopted varied in one respect, however: in view of the relatively small number of booklets for each task (approximately 150), and bearing in mind the level of interpretation required, it was felt to be desirable to select a sample larger than ten per cent for the inter-rater reliability check. Consequently, a 20 per cent sample of booklets was coded by two markers for each of the 12 tasks; as with the written tests, the codes allocated by the two markers were processed so as to provide a measure of inter-rater reliability. A further ten per cent was reviewed by a member of the research team

Figure 2 Codes for *Pulse*

Q2. How did your pulse change during this exercise?

Criteria for a complete response:

i) Description must be consistent with data presented in preceding question.
ii) Description includes identification of the trend or pattern in the data if there is one.

Code	Response
	Correct Response
20	Number of pulse beat increases with exercise.
21	Number of pulse beats increases at first then stabilizes or slows.
29	Other complete.
	Partial Response
11	Describes pulse at specific time intervals instead of summarizing trend. *Examples: At 2 minutes slow, at 4 min faster.* *At 2 min 60/min at 4 min 70/min.*
12	Pulse does not increase perceptibly with exercise but student presents a plausible reason. *Example: I counted each 2 minutes and I rested after each time because I was tired.*
19	Other partially correct.
	Incorrect Response
70	Description not consistent with the data.
76	Merely repeats information in stem or in a previous question.
79	Other incorrect.
	Non-response
90	Crossed out/erased, illegible, or impossible to interpret.
99	BLANK

IV.9 The student questionnaires

The student questionnaire was designed to obtain background information from students about themselves, their home background, their attitudes towards mathematics and science and their perceptions of lessons in these subjects. Information collected from the students in England included:

- **the students themselves**: age, gender, country of birth, language spoken at home, out-of-school activities; perceived ability in mathematics and science; educational aspirations; and time spent on homework.
- **home background**: family composition; country of birth of parents; perceptions of parental interest; and surrogate measures to provide an indication of the socio-educational status of the family, such as the approximate number of books in the home.
- **attitudes**: liking for mathematics and science; views about mathematics and science and the importance of doing well in these subjects.
- **perceptions of mathematics and science lessons**: teaching approaches used by teachers; activities undertaken by students including practical work and the use of calculators and computers; and behaviour in class.

IV.10 The teacher questionnaires

The teacher questionnaire was designed to collect biographical details and information on teaching and learning approaches from the students' mathematics and science teachers. Information collected included:

- **biographical details**: age, gender, educational background, teaching experience;
- **how they spent their time**: teaching time for different subjects; lesson preparation time; other timetabled activities; frequency of collaborative planning with colleagues
- **teaching approaches and resources**: size of classes; extent of whole class, group and individual work; questioning and correcting wrong responses; amount and nature of homework set; assessing students' work; resources used for lesson planning; use of textbooks and schemes of work; use of calculators and computers; and factors, if any, which limited how they taught mathematics or science.
- **attitudes**: views on mathematics or science as subjects; views on teaching and learning mathematics or science; views on teaching as a career.

IV.11 The school questionnaires

The school questionnaire sought general background information on the schools taking part in the study and on the level of resources devoted to mathematics and science. Information provided by schools included:

- **general background information**: location; length of school week and teaching week; length of each teaching period; length of school year; admissions criteria.
- **teaching staff**: number and stability of teaching staff; numbers of teachers teaching mathematics and science and proportions teaching these subjects for three-quarters or more of their teaching-time; average non-contact time for mainscale mathematics and science teachers.
- **students**: number of boys and girls on roll; number of students eligible for free school meals, from ethnic minorities, needing ESL support and with statements of special educational needs; stability of student population; numbers of students in the year groups being tested.
- **resources**: number of computers available to students; shortages or inadequacies, if any, of other resources.
- **organisational features**: existence of written curriculum plans for mathematics and science; extent of streaming, setting or banding for mathematics and science; criteria for allocation to streams, sets or bands; teaching time per week for mathematics and science; remedial and/or enrichment provision in mathematics and science.

IV.12 Summary

The TIMSS project began with the development of the test blueprints which identified major categories within mathematics and science curricula which were to be addressed by the tests. Each of the major content areas was broken down into more detailed categories, which were the focus of specific test items. The written tests comprised items in three different formats:

- multiple-choice
- short answer
- extended response.

Groups of items were combined to form 26 'clusters' for each population. Eight different versions of the achievement tests were prepared using different combinations of clusters, although one cluster was common to all versions for each population.

The tests for nine-year-olds and 13-year-olds contained both mathematics and science items in the formats described above. Whereas the multiple-choice items were pre-coded, the short answer and extended response items were manually marked and coded according to the guidelines provided by the International Study Centre before data processing.

The 12 tasks which comprised the Performance Assessment component of the study focused on students' practical and investigative skills. These were also manually marked and coded in accordance with the international guidelines.

Specific procedures were followed to monitor the accuracy of the coding of both the written tests and the Performance Assesment booklets by markers.

Questionnaires collected information from students, their mathematics and science teachers and school headteachers. The questionnaires obtained data about:

- students' and teachers' attitudes towards mathematics and science
- teaching conditions and practices.

APPENDIX V
Statistical details

V.1 Introduction

This appendix provides details of the samples of schools and students taking part in TIMSS in England, the participation rates achieved, the weighting procedures used, scaling and data analysis.

V.2 Population 2: samples, participation rates, representativeness and weighting

V.2.1 The sample of schools

Using the Register of Schools provided by the Department for Education and Employment (DfEE), schools with students in Years 8 and 9 (normally those born between 1 September 1980 and 31 August 1982) were stratified by:

- size (in terms of numbers of students of the relevant ages);
- type (comprehensive – 18; comprehensive – 16; other maintained; and independent)
- region (metropolitan; non-metropolitan)
- GCSE results (percentage of Year 11 students achieving five or more GCSE grades A*-C in 1993.

The sample design required 150 first-choice schools and two matching sets of 150 replacement schools selected at the same time to replace schools refusing to take part (see Appendix III, section III. 3, for full details of the replacement sampling procedure).

The national sample consisted of first-choice and replacement schools (Table V.2.4). Two participation rates were calculated by the International Study Centre, one excluding replacement schools (i.e. first-choice schools only) and one including first-choice and replacement schools (Table V.2.4).

V.2.2 The samples of students within schools

Two random samples, each consisting of 16 students, were selected from each school: one from all the students in Year 8 and one from all the students in Year 9. If a school contained only one of the target year groups, a random sample of 32 students was selected from that year group (five schools contained no Year 8 students and six included no Year 9 students). Day of birth was used to select students randomly.

Guidance was prepared for all schools taking part in TIMSS in England concerning the exclusion of students from the survey. The guidance was drafted using the guidelines provided by the Study Centre, but taking into

account national procedures concerning students with special educational needs. Full details of the guidance given to schools in England on exclusions is given in Appendix III, section III. 4.

V.2.3 Selecting the mathematics and science teachers

Schools were asked to provide the names of all the mathematics and science teachers teaching the selected students. Since each sample of 16 students was selected randomly from across an age group, the selected students could be drawn from any of the mathematics (or science) sets/classes in that year group. In a large school with ten mathematics sets, say, up to ten mathematics teachers could be selected, some of whom would teach only one or two of the selected students. In order to reduce the burden on schools, only those teachers who taught three or more of the selected students were sent a questionnaire (this more than met the requirements of the International Study Centre, which stipulated that countries could, if they wished, limit the teachers surveyed to those who taught five or more of the selected students). In accordance with the international guidelines, teachers who taught selected students in both year groups were sent one questionnaire only and asked to complete it in relation to their teaching approaches with the older age group. These decisions meant that it would not be possible to link every student with a mathematics (or science) teacher. In the event, 65 per cent of the students were linked with a mathematics teacher and 65 per cent with a science teacher.

It should be noted that the teachers selected for TIMSS were not true probability samples of teachers but *the teachers of a probability sample of students*. Thus, the data did not describe the characteristics of a sample of teachers but instead described the teacher characteristics experienced by a probability sample of students.

V.2.4 Participation rates from schools

The Population 2 achieved sample consisted of 127 schools, 19 (15 per cent) of which were single-sex schools; ten of these schools were for boys only and nine for girls only. Table V.2.4 below provides details of the numbers of schools invited to take part in the study and those which actually participated.

Table V.2.4 Population 2: school participation rates and sample size

	Year 8	Year 9
Number of schools in original sample	150	150
Number of eligible schools in original sample	*145	*144
Number of original schools that participated	*81	*80
Number of replacement schools that participated	*41	*41
Total number of schools that participated	*122	*121
Weighted school participation rate before replacement	*57*	*56*
Weighted school participation rate after replacement	*85*	*85*

* *Five of the selected schools did not include Year 8 students; six of the selected schools did not include Year 9 students. The total number of schools taking part was 127 (85 first-choice and 42 replacement schools).*

V.2.5 Participation rates within schools

As noted in Appendix III, students taking part in the study were required to complete two test booklets and a questionnaire. Questionnaires were also completed by their mathematics and science teachers and a school questionnaire was completed by the headteacher. Details of the Population 2 datafile, which consisted of 3,579 students (in Years 8 and 9 combined) in 127 schools, are given in Table V.2.5 below, together with information about the participation rates achieved from students, teachers and schools.

Table V.2.5 Population 2: participation rates within schools

Booklet/questionnaire	Sent to participating schools	Completed	Participation rate from schools taking part
Test Booklet A and/or B	3851	3579	93%
Student Questionnaires	3851	3535	92%
Mathematics Teacher Questionnaires	558	485	87%
Science Teacher Questionnaires	690	599	87%
School Questionnaires	127	110	87%

V.2.6 Coverage of the TIMSS target population

The international desired population for Population 2 was defined as: all students enrolled in the two adjacent grades with the largest proportion of 13-year-old students at the time of testing. The two grades/year groups (Years 8 and 9) selected in England included 99 per cent of the 13-year-olds (57 per cent in Year 8 and 42 per cent in Year 9).

The sampling frame for England included all county, grant-maintained, special and independent schools in every LEA in England. Special schools and schools with less than 20 students in a year group were excluded from sampling. These represented 1.8 per cent of the target student population. In addition, researchers were not able to approach a random sample of schools (containing about 6.6 per cent of students) which were taking part in trials for National Curriculum assessment. Of the schools taking part, about 2.9 per cent of the students within schools were excluded either on educational grounds (see Appendix III, section III.4) or had left the school between the time the sample was selected and the time of testing. Details of school-level and within-sample exclusion rates are given below.

Table V.2.6 Population 2: School-level and within-sample exclusion rates

School-level exclusions	8.4%
Within-school exclusions	2.9%
Overall exclusions	11.3%

V.2.7 Representativeness

The table below shows the distribution of these schools by the GCSE results, school type and area. Of the first preference schools, those with low GCSE results (< 25 per cent with five or more grades A–C) were as likely to respond as those with high GCSE grades (> 55 per cent). Those with average results and schools in metropolitan areas were less likely to participate. Again, the use of replacement schools reduces the bias and in general the characteristics of all the responding schools reflect the target sample.

Table V.2.7 Population 2: representativeness of sample schools

	First-preference schools responding		First-preference and replacements responding		National student population
	Number	%	*Number*	%	%
Total	***85***	***100***	***127***	***100***	***100***
GCSE results					
<25% 5+ grades A-C	21	25	29	23	23
26-35%	14	17	22	18	18
36-45%	12	14	22	17	19
46-55%	16	19	20	16	16
>55%	17	20	25	20	18
Missing or N/A	5	6	8	6	6
School type					
Independent	9	11	13	10	7
Comp. to 16	27	32	41	33	35
Comp. to 18	39	46	56	44	45
Others	10	12	16	13	13
Area					
Metropolitan	24	28	43	34	36
Non-metropolitan	61	72	83	66	64

Notes: The sample distribution of schools should be compared with the national sample of pupils as schools were chosen with probability proportional to the numbers of students in the year groups.

Percentages may not add up to 100% due to rounding.

V.2.8 Weighting

In order to ensure that the samples of students selected for the study were as representative as possible of all students in England and, thus, would provide unbiased estimates for the whole population, weights were applied to the data. The weighting of every sample from every country participating in TIMSS was carried out by the International Sampling Centre (Statistics Canada).

The weighting factors for Population 2 in England were the product of four components:

- **a school factor**: the inverse of the school selection probability; computed at the stratum level
- **a school participation adjustment** (which redistributed the school factor of non-participating schools to participating schools): based on originally selected schools and replacements; computed at the stratum level
- **a student factor**: the inverse of the student selection probability level; computed at the school level
- **a student participation adjustment** (which redistributed the student factor of non-participating students to participating students); based on students participating in the original test session; computed at the school level.

V.3 Population 1: samples, participation rates and weighting

V.3.1 The sample of schools

Using the Register of Schools provided by the Department for Education and Employment (DfEE), schools with students born between 1 September 1984 and 31 August 1986 (normally those in Years 4 and 5) were stratified by:

- size (in terms of numbers of students of the relevant ages)
- type (junior and infant, junior, other maintained; and independent)
- region (metropolitan, non-metropolitan)
- percentage of students eligible for free school meals in (year).

The sample design was similar to the design used for Population 2 (see above): 150 first-choice schools and two matching sets of 150 replacement schools selected at the same time to replace schools refusing to take part.

V.3.2 The sample of students within schools

Two intact classes were selected from each school. In schools which contained both Year 4 and Year 5 students, one Year 4 class and one Year 5 class were randomly selected by NFER. In schools which contained mixed-age classes including Year 4 and Year 5 students, two such classes were selected. In schools which contained only one of the target year groups, two classes were randomly selected from the same year group. Rules for exclusion were the same as those used for Population 2 (see Section V.2.2).

V.3.4 Selecting the teachers

Schools were asked to provide a list of the teacher(s) of mathematics and science to the selected classes. In most cases the students were taught by their class teacher for both subjects; one teacher questionnaire, covering both mathematics and science, was completed for each class. If a class was taught by another teacher for either subject, an additional questionnaire was provided for that teacher, who was asked to complete the background questions and those relating to the relevant subject.

V.3.5 Participation rate from schools

The achieved sample consisted of 134 schools. Table V.3.5 provides details of the numbers of schools invited to take part in the study and those which actually participated.

Table V.3.5 Population 1: school participation rates and sample size

	Year 4	Year 5
Number of schools in original sample	150	150
Number of eligible schools in original sample	*145	*145
Number of original schools that participated	*93	*92
Number of replacement schools that participated	*36	*35
Total number of schools that participated	*129	*127
Weighted school participation rate before replacement	*64*	*63*
Weighted school participation rate after replacement	*88*	*88*

* *Five of the selected schools did not include Year 4 students; five of the selected schools did not include Year 5 students. The total number of schools taking part was 134.*

V.3.6 Participation rates within schools

As noted in Appendix III, students taking part in the study were required to complete two test booklets and a questionnaire. Questionnaires were also completed by the teacher(s) who taught the selected classes for mathematics and science (in most cases the students' class teacher taught them for both subjects) and a school questionnaire was completed by the head teacher. Details of the Population 1 datafile, which consisted of 6,142 students (Years 4 and 5 combined) in 134 schools are given in Table V.3.6 below, together with information about the participation rates achieved from students, teachers and schools.

Table V.3.6 Population 1: participation rates within schools

Booklet/questionnaire	Sent to participating schools	Completed	Participation rate from schools taking part
Test Booklet A and/or B	6534	6142	94%
Student Questionnaires	6534	6044	94%
Teacher Questionnaires	297	259	87%
School Questionnaires	134	126	94%

V.3.7 Coverage of the TIMSS target population

The international desired population for Population 1 was defined as: all students enrolled in the two adjacent grades with the largest proportion of nine-year-old students at the time of testing. The two grades/year groups (Years 4 and 5) selected in England included 99 per cent of the nine-year-olds (58 per cent in Year 4 and 41 per cent in Year 5).

The sampling frame for England included all county, grant-maintained, special and independent schools in every LEA in England. Special schools and schools with less than 12 students in a year group were excluded from sampling. These exclusions represented 2.8 per cent of the target student population. In addition, researchers were not able to approach a random sample of schools (containing about 5.8 per cent of students) which were taking part in trials for National Curriculum assessment. About 3.8 per cent of the students within the schools taking part were excluded on educational grounds (see Appendix III, section III. 4). Details of school-level and within-sample exclusion rates are given below.

Table V.3.7 Population 2: School-level and within-sample exclusion rates

School-level exclusions	8.6%
Within-school exclusions	3.8%
Overall exclusions	12.4%

V.3.8 Representativeness

The table below shows the distribution of these schools by the free school meal indicator, KS2 results, school type and area. Of the first-preference schools, those with low average KS2 scores or high percentages receiving free school meals or in the metropolitan areas were less likely to participate. For example, 26 per cent of the first-preference schools responding were from the metropolitan area. With replacement schools this increases to 33 per cent as against 38 per cent in the target sample. The use of replacement schools reduces the bias and in general the characteristics of all the responding schools closely reflect the population.

Table V.3.8 Population 1: representativeness of sample of schools

	First-preference schools responding		First-preference and replacements responding		National student population
	Number	*%*	*Number*	*%*	*%*
Total	***96***	***100***	***134***	***100***	***100***
KS2 results					
Lowest band 1	*11*	*12*	*18*	*13*	*16*
2	*22*	*23*	*29*	*22*	*20*
3	*16*	*17*	*24*	*19*	*19*
4	*14*	*15*	*21*	*16*	*18*
Top band 5	*25*	*26*	*31*	*23*	*18*
Missing/NA	*8*	*8*	*11*	*8*	*9*
Free school meals					
Highest % FSM	18	19	30	22	22
2nd highest	18	19	29	22	21
Middle	18	19	26	19	19
2nd lowest	20	21	23	17	18
Lowest	22	22	26	19	20
School type					
Independent	5	5	7	5	5
Junior & infant	50	52	68	51	57
Junior	27	28	37	28	24
Others	14	15	22	16	14
Area					
Metropolitan	25	26	44	33	38
Non-metropolitan	71	74	90	67	62

Notes: *The sample distribution of schools should be compared with the national sample of pupils as schools were chosen with probability proportional to the numbers of students in the year groups.*

V.3.9 Weighting

In order to ensure that the samples of students selected for the study were as representative as possible of all students in England and, thus, would provide unbiased estimates for the whole population, weights were applied to the data. The weighting of every sample from every country participating in TIMSS was carried out by the International Sampling Centre (Statistics Canada).

The weighting factors for Population 1 in England were the product of five components:

- **a school factor**: the inverse of the school selection probability; computed at the stratum level
- **a school participation adjustment** (which redistributed the school factor of non-participating schools to participating schools): based on originally selected schools and replacements; computed at the stratum level
- **a classroom factor**: the inverse of the classroom selection probability; computed at the school level
- **a classroom participation adjustment** (which redistributed the classroom factor of non-participating classrooms to participating classrooms); computed at the school level
- **a student factor**; the inverse of the student selection probability level; computed at the school level
- **a student participation adjustment** (which redistributed the student factor of non-participating students to participating students); based on students participating in the original test session; computed at the school level.

V.4 Performance Assessment: sampling, participation rates and weighting

The international Performance Assessment study consisted of parallel assessments at Populations 1 and 2. In each case, students in the older age group only (Year 5 in Population 1 and Year 9 in Population 2) were assessed. England took part in the Population 2 assessment only.

V.4.1 The sample of schools

The sub-sample of schools invited to take part in the Performance Assessment was selected from those schools that were involved in the main survey achievement tests and questionnaires. The international guidelines on sampling permitted up to 25 per cent of schools to be omitted from the sub-sample if they were deemed to be in remote areas. In fact, 21 per cent (26 schools in 14 local education authorities (LEAs)) of the schools in the secondary school sample for the main study in England were excluded on this basis. The sub-sample was structured in a similar way to the main survey, with 50 first-choice schools and, for each one, a replacement school selected from the same stratum.

V.4.2 The sample of students within schools

A random sub-sample of nine students (plus two replacement students) was selected from the 16 Year 9 students who had taken part in the main study (excluding those students who had been excluded from the main study (see Appendix III section III.4).

V.4.3 Participation rates from schools and within schools

A total of 50 schools took part in the Performance Assessment: 37 first-choice schools and 13 replacement schools. A total of 443 students in these schools completed the Performance Assessment tasks.

V.5 Scaling and data analysis

The mathematics and science scores were derived in the same way. The description below,[1] therefore, which relates to the mathematics scores is also relevant to the science scores.

Two general analysis methods were used for this report — item response scaling methods and mean percentage correct.

V.5.1 Item Response Theory (IRT) Scaling

The overall mathematics results were summarised using an item response theory (IRT) scaling method—Rasch model. This scaling method produces a mathematics score by averaging the responses of each student to the items which they took in a way that takes into account the difficulty of each item. The methodology used in TIMSS includes refinements which enable reliable scores to be produced even though individual students responded to relatively small subsets of the total mathematics item pool. Analyses of the response patterns of students from participating countries indicated that, although the items in the test address a wide range of mathematical content, the performance of the students across the items was sufficiently consistent that it could be usefully summarised in a single mathematics score.

The IRT methodology was preferred for developing comparable estimates of performance for all students, since students answered different test items depending upon which of the eight test booklets they received. The IRT analysis provides a common scale on which performance can be compared across countries. In addition to providing a basis for estimating mean achievement, scale scores permit estimates of how students within countries vary and provide information on percentiles of performance. The scale was

[1] *Section V.5 has been adapted slightly from Appendix A in Beaton et al. ,1996b.*

standardised using students from both the grades tested. When all participating countries and grades are treated equally, the TIMSS scale average is 500 and the standard deviation is 100. Since the countries varied in size, each country was re-weighted to contribute equally to the mean and standard deviation of the scale. The average of the scale scores was constructed to be the average of the 41 means of countries that were available at the eighth grade and the 39 means at the seventh grade. The average and standard deviation of the scale scores are arbitrary and do not affect scale interpretations.

V.5.2 Mean percentage correct

The analytic approach underlying the mean percentage correct scores, which were used mainly in Chapters 3 and 5 of the First National Report Part 1 (Keys *et al.*, 1996) involved calculating the percentage of correct answers for each item for each participating country (as well as the percentages of different types of incorrect responses). The percentages correct were averaged to summarise mathematics performance overall and in each of the content areas for each country as a whole and by gender. For items with more than one part, each part was analysed separately in calculating the mean percentages correct. Also, for items with more than one point awarded for full credit, the mean percentages correct reflect an average of the points received by students in each country. This was achieved by including the percentage of students receiving one score point as well as the percentage receiving two score points and three score points in the calculations. Thus, the mean percentages correct are based on the number of score points rather than the number of items *per se*. An exception to this is the international mean percentages correct reported for example items, where the values reflect the percentage of students receiving full credit.

V.5.3 Estimating sampling error

Because the statistics presented in this report are estimates of national performance based on samples of students, rather than the values that could be calculated if every student in every country would have answered every question, it is important to have measures of the degree of uncertainty of the estimates. The jack-knife procedure was used to estimate the standard error associated with each statistic presented in this report. The use of confidence intervals, based on the standard errors, provided a way to make inferences about the population means and proportions in a manner that reflects the uncertainty associated with the sample estimates. An estimated sample statistic plus or minus two standard errors represents a 95 per cent confidence interval for the corresponding population result.

V.6 Summary

Appendix V provided the following information:

- Selection of the samples of schools, students and teachers for Populations 1 and 2 in England.
- Participation rates from schools and within schools for Populations 1 and 2 in England.
- Coverage of TIMSS target Populations 1 and 2 in England.
- Representativeness of the TIMSS samples for Populations 1 and 2 in England.
- Selection of the sub-samples of schools and students for the Performance Assessment in England.
- Scaling and data analysis.

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).* Boston: 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).* Boston: Boston College.

KEYS, W., HARRIS, S. and FERNANDES, C. (1996). *Third International Mathematics and Science Study: First National Report, Part 1.* Slough: NFER.

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

nfer

THIRD INTERNATIONAL MATHEMATICS AND SCIENCE STUDY
National Reports
Appendices

This volume contains the Appendices relating to the National Reports for England on the Third International Mathematics and Science Study (TIMSS). Subjects covered include:

- research design, populations and samples
- development and structure of the mathematics and science tests and their relationship with the National Curriculum
- questionnaires for students, teachers and headteachers
- Performance Assessment tasks
- administration and quality control procedures
- marking procedures for free-response items in the written tests and in the Performance Assessment
- statistical details relating to samples, participation rates, representativeness and weighting for nine- and 13-year-olds and for the Performance Assessment
- scaling and data analysis.

This information supplements the data presented in the National Reports.

ISBN 0 7005 1446 5

£6.00