Chapter 1

Student Achievement on the Performance Assessment Tasks

The performance assessment tasks were chosen to sample a broad array of mathematics and science content, and to elicit from students a variety of abilities, skills, and knowledge. Each task is presented in this chapter, together with the questions (items) asked of the students, the scoring criteria applied to their responses, and a sample student response. For the majority of the tasks administered to both fourth- and eighth-grade students, the version of the task administered was different at the two grade levels. For these tasks, the full-task example with student responses is shown for the eighth-grade version, and only the modified items and scoring criteria are shown for the fourth-grade version. At both grades, the average performance of the students in each country is presented for each item, and averaged across items to provide an average task performance score for each country.

As depicted by the task averages, there is a wide range of performance across countries on each task. There is, however, also a wide range of performance across the items within each task for each country. This is a natural consequence of the way the tasks are structured, with each task containing some items that even the weakest students could attempt, as well as some quite challenging items. The variation in item difficulty within a task is a consequence also of the range of performance expectations addressed by the items, since some of those expectations make more demands on the students' abilities than others. Because of the varying difficulty of the items within each task, detailed results are presented for each item to allow a full appreciation of the performance of the students.

Although more than half the tasks require both science and mathematics knowledge and skills, each task has been classified for discussion purposes as a science or mathematics task depending on the primary content addressed. The tasks of primarily science content are Pulse, Magnets, Batteries, Rubber Band, Solutions (eighth grade only), and Containers (fourth grade only). The mathematics tasks are Dice, Calculator, Folding and Cutting, Around the Bend, and Packaging. Two tasks – Plasticine and Shadows – are referred to as "combination" tasks because they required students to integrate their mathematics and science knowledge and skills. A summary of the overall performance across tasks for each country is presented in Chapter 2.

PULSE

In the Pulse task, students were asked to find out how their pulse changed during and after exercise (5 minutes stepping up and down). They were provided with a stopwatch and a step 20-25 cm high. The task was intended to measure the ability to design and conduct an investigation (no written plan was required), that is, to collect, tabulate, analyze, and interpret data and use appropriate concepts to explain findings.

The task for eighth-grade students is shown in Figure 1.1, together with a sample student response and the criteria for a fully-correct response. Decisions about how many measurements to make, and at what intervals, were left to the individual student. Item 1 for eighth graders has two quite different aspects: the organization and representation of data in a table – a procedural skill; and the quality of the data and of the way they were collected – an aspect requiring knowledge of the content area, and of how to conduct an investigation. In common with all items involving data collection, two scores were assigned to this item – one for the quality of presentation, and the other for the quality of the data.[1] The fourth-grade version of this task did not require students to construct a data table but instead provided a practice task (Item 1) along with prepared tables (see Figure 1.2). Item 2 for fourth graders provided instructions about how often to count and record pulse beats. For these items, students' ability to organize, label, and display data in a table was not assessed, nor was the ability to decide an experimental design.

The second item for the eighth graders (Item 3 for fourth grade) required an accurate description of the trends in the data, and the third item (Item 4 for fourth grade) required conceptual knowledge of the connections among muscle work, energy needs, circulation, and heart rate in order to explain the data.

Tables 1.1 and 1.2 show the average percentage score[2] for each country on each item of the Pulse task for eighth- and fourth-grade students, respectively. The overall task average is the arithmetic mean of these average percentage scores. The 12 countries shown in the upper part of the table, in decreasing order of achievement on the overall task, were judged to have met the TIMSS requirements for testing a representative sample of students. Although all countries tried very hard to meet the TIMSS sampling requirements, several encountered difficulties in securing participation, and did not have participation rates for both schools and students of 85% (or a combined rate for the product of school and student participation of 75%) as specified in the TIMSS guidelines (i.e., Australia, England, the Netherlands, and the United States at the eighth grade). To provide a better curricular match, Colombia, Romania, and Slovenia elected to test their eighth-grade students even though these students were somewhat older than those in other countries (of these three countries, only Slovenia participated at the fourth grade). Because Hong Kong had low school participation at eighth grade, and consequently a small sample size,

[1] These two aspects of data collection, together with organization and interpretation of data, comprise what the Curriculum Frameworks call "Conducting an Investigation." For the full curriculum frameworks see Robitaille, D.F., McKnight, C.C., Schmidt, W.H., Britton, E.D., Raizen, S.A., and Nicol, C. (1993). *TIMSS Monograph No. 1: Curriculum Frameworks for Mathematics and Science.* Vancouver, B.C.: Pacific Educational Press.

[2] The percentage score on an item is the score achieved by a student expressed as a percentage of the maximum points available on that item. A country's average percentage score is the average of its students' percentage scores.

FIGURE 1.1 - PULSE

FULL-TASK EXAMPLE AND SCORING CRITERIA – EIGHTH GRADE

ITEM 1

PULSE

At this station you should have

A watch with a second hand
A step on the floor to climb on

Read ALL directions carefully.

Your task:

Find out how your pulse changes when you climb up and down on a step for 5 minutes.

This is what you should do:

- Find your pulse and be sure you know how to count it. IF YOU CANNOT FIND YOUR PULSE ASK A TEACHER FOR HELP.
- Decide how often you will take measurements starting from when you are at rest.
- Climb the step for about 5 minutes and measure your pulse at regular intervals.

1. Make a table and write down the times at which you measured your pulse and the measurements you made.

Time	Pulse (beats per min)
Rest	90
1 min	110
2 min	120
3 min	125
4 min	128
5 min	125

page 1 TASK S1-P2

ITEMS 2 AND 3

2. How did your pulse change during this exercise?

My pulse became faster during this exercise. At first, it became faster quite quickly but near the end it became only a little faster with each minute until on the 5th minute it dropped a little.

3. Why do you think your pulse changed in this way?

My pulse changed this way because I was using up oxegen faster when I was excercising therefore I needed more oxegen. My heart had to pump the blood around my body faster so I would get all the oxegen I needed, as blood carries oxegen around the body.

PUT EVERYTHING BACK THE WAY YOU FOUND IT SO THAT SOMEONE ELSE CAN USE THE STATION.

TASK S1-P2 page 2

Criteria for Fully-Correct Response

Item 1 - Measure pulse rates and record in table.
Response is scored for both the quality of the presentation and the quality of data collection.

> **Quality of presentation.** i) Presents at least 2 sets of measurements in table. ii) Measurements are paired: time and number of pulse beats. iii) Labels table appropriately: data entries in columns identified by headings and/or units; units incorporated into headings or placed beside each measurement; headings or units for the number of pulse beats include the time interval.
> *Total Possible Points: 2*
>
> **Quality of data**. i) Makes at least 5 measurements (at rest, and 4 or more during exercise). ii) Pulse rates are plausible: 7 to 25 counts per 10 seconds (40-150 pulse beats per minute). iii) Pulse rate increases with exercise (may level off or slow near the end).
> *Total Possible Points: 3*

Item 2 - Describe how pulse changes during exercise. i) Description consistent with data presented. ii) Description includes identification of the trend or pattern in the data.
Total Possible Points: 2

Item 3 - Explain why pulse changes. Includes the following three elements relating to physiological needs during exercise: i) role of muscle action (exercise results in need for more energy and oxygen in the muscles); ii) role of blood (more oxygen or food supplied by an increase in blood flow); iii) connection with heart action or pulse rate, (heart is pumping faster to supply more blood).
Total Possible Points: 3

these results are presented in Appendix B. Israel did not follow within-school sampling guidelines at the fourth grade or eighth grade and at the eighth grade it had a small sample size; its results are also presented in Appendix B. A full discussion of the sampling procedures and outcomes for each country can be found in Appendix A.

To facilitate comparisons across grades, the results for eighth and fourth grades are presented together. However, not all of the countries that tested at eighth grade also tested at fourth grade, and consequently the countries making up the international averages at eighth grade are not the same as those making up those averages at fourth grade. Comparisons across grades on the basis of the international averages should therefore be made with caution. Within individual countries, however, the relative performance between the two grades is directly comparable where the task included identical items at both grade levels.

Average country performance on the Pulse task at the eighth grade varied considerably around the international average of 44% (see Table 1.1). Despite the substantial difference between the highest- and lowest-performing countries, differences between countries with adjacent score levels may not be statistically significant because of sampling variability. Internationally, eighth-grade students had most success on the item that required them to describe the trend in their data. They found this easier than the rather straightforward task of recording and presenting their data. Explaining the causes of the changes in pulse rate observed was a challenge for most students, with relatively few students able to make the connection between muscle action, blood supply, and heart rate. Fourth-grade students were generally able to measure their pulse at rest, but found the other items demanding, particularly explaining the data.

FIGURE 1.2 - PULSE ITEMS AND SCORING CRITERIA – FOURTH GRADE

PULSE

At this station you should have

A watch with a second hand
A step on the floor to climb on

Read ALL directions carefully.

Your task:

Find out how your pulse changes when you climb up and down on a step for 5 minutes.

This is what you should do:

- Find your pulse and be sure you know how to count it. IF YOU CANNOT FIND YOUR PULSE ASK A TEACHER FOR HELP.
- Count your pulse for 10 seconds.

1. Write this number of counts in the table below on the line marked 0 minutes.
2. Now climb up and down on the step for about 5 minutes. Stop after each minute and write your pulse in the table below.

Number of minutes climbing	Number of pulse counts in 10 seconds
0 Minutes	
1 Minutes	
2 Minutes	
3 Minutes	
4 Minute	
5 Minutes	

3. How did your pulse change during this exercise?

4. Why do you think your pulse changed in this way?

CRITERIA FOR FULLY-CORRECT RESPONSE

Item 1 - Measure "at rest" pulse rate and record in table. Pulse beats are plausible: 7 to 25 counts per 10 seconds (40 to 150 counts per minute).
Total Possible Points: 1

Item 2 - Measure "after exercise" pulse rates and record in table. i) Records pulse at least 4 different times during the exercise (in addition to "at rest" measurement). ii) Pulse rates are plausible: 7 to 25 counts per 10 seconds at the beginning (40 to 150 counts per minute). iii) Pulse rate increases with exercise (may level off or slow near the end).
Total Possible Points: 3

Item 3 - Describe how pulse changes during exercise. i) Description consistent with data presented. ii) Description includes identification of the trend or pattern in the data.
Total Possible Points: 2

Item 4 - Explain why pulse changes. Includes the following three elements relating to physiological needs during exercise: i) role of muscle action (exercise results in need for more energy and oxygen in the muscles); ii) role of blood (more oxygen or food supplied by an increase in blood flow); iii) connection with heart action or pulse rate, (heart is pumping faster to supply more blood).
Total Possible Points: 3

Task layout condensed for display

Table 1.1 Pulse Task: Average Percentage Score on Items – Eighth Grade*

Country	Overall Task Average▼	Average Percentage Scores on Items●			
		Item 1 Measure Pulse		Item 2 Describe Trend	Item 3 Explain Results
		Presentation	Data Quality		
		2 Points	3 Points	2 Points	3 Points
Singapore	60 (2.7)	59 (4.0)	56 (2.7)	82 (3.8)	42 (3.5)
Iran, Islamic Rep.	55 (4.5)	76 (4.8)	58 (4.8)	53 (9.0)	33 (6.6)
† Scotland	55 (2.9)	61 (3.7)	56 (3.4)	67 (4.0)	34 (3.3)
†1 Switzerland	51 (1.9)	58 (3.2)	43 (3.6)	75 (3.7)	27 (3.2)
Norway	48 (1.6)	44 (2.9)	48 (3.4)	72 (2.7)	29 (3.7)
Canada	46 (2.4)	53 (3.0)	44 (3.0)	60 (3.3)	26 (2.5)
Czech Republic	46 (2.9)	45 (5.5)	38 (4.8)	72 (3.8)	27 (2.6)
Sweden	45 (2.6)	45 (3.6)	50 (2.7)	62 (4.3)	22 (4.3)
New Zealand	44 (2.0)	51 (3.5)	37 (3.0)	61 (3.2)	28 (2.6)
Spain	36 (2.1)	36 (3.0)	30 (2.7)	52 (5.3)	26 (3.0)
Cyprus	33 (2.1)	31 (3.6)	32 (3.5)	55 (3.8)	15 (1.9)
Portugal	24 (2.5)	31 (3.2)	24 (3.2)	26 (4.1)	17 (2.8)
Countries Not Satisfying Guidelines for Sample Participation Rates (See Appendix A for Details):					
Australia	54 (2.6)	68 (3.7)	46 (3.5)	71 (3.6)	31 (3.5)
2 England	59 (2.2)	65 (2.4)	59 (3.6)	75 (3.0)	39 (2.7)
Netherlands	45 (2.6)	50 (4.1)	44 (3.9)	56 (3.6)	29 (2.9)
United States	50 (2.0)	54 (2.9)	43 (2.6)	72 (2.6)	33 (3.4)
Countries Not Meeting Age/Grade Specifications (See Appendix A for Details):					
Colombia	11 (1.0)	10 (1.9)	4 (1.3)	20 (3.1)	11 (2.1)
3 Romania	41 (3.6)	45 (4.5)	29 (5.1)	63 (5.7)	27 (4.1)
Slovenia	40 (3.2)	54 (3.7)	33 (5.0)	53 (4.6)	19 (3.3)
International Average	44 (0.6)	49 (0.8)	41 (0.8)	60 (1.0)	27 (0.8)

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

* Eighth grade in most countries; see Table 2 for information about the grades tested in each country.
● Percent of total possible points on each item averaged over students.
▼ Average of percentage scores across items; all items weighted equally.
† Met guidelines for sample participation rates only after replacement schools were included (see Appendix A for details)
1 National Desired Population does not cover all of International Desired Population (see Table A.2) - German-speaking cantons only.
2 National Defined Population covers less than 90 percent of National Desired Population for the main assessment (see Table A.2).
3 School-level exclusions for performance assessment exceed 25% of the National Desired Population (see Table A.2).
() Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

Pulse Task: Average Percentage Score on Items – Fourth Grade* Table 1.2

Country	Overall Task Average▼	Average Percentage Scores on Items●			
		Item 1 Measure Pulse at Rest	Item 2 Measure Pulse During Exercise	Item 3 Describe Trend	Item 4 Explain Results
		1 Point	3 Points	2 Points	3 Points
Iran, Islamic Rep.	41 (3.3)	77 (5.7)	44 (5.7)	32 (4.3)	9 (1.9)
Cyprus	38 (3.0)	76 (5.4)	47 (4.3)	23 (4.7)	3 (1.3)
Canada	36 (1.5)	73 (3.2)	34 (2.6)	33 (3.9)	7 (1.4)
†1 New Zealand	27 (2.1)	66 (4.7)	19 (3.2)	19 (2.5)	4 (1.2)
Portugal	22 (1.8)	61 (4.5)	23 (3.3)	5 (1.9)	1 (0.5)
Countries Not Satisfying Guidelines for Sample Participation Rates (See Appendix A for Details):					
Australia	38 (2.3)	74 (4.7)	30 (4.8)	44 (4.1)	5 (1.3)
Hong Kong	39 (2.1)	46 (4.7)	46 (4.1)	55 (3.5)	7 (1.6)
United States	42 (1.7)	77 (2.9)	32 (3.3)	48 (3.5)	9 (1.4)
Countries Not Meeting Age/Grade Specifications (See Appendix A for Details):					
Slovenia	39 (2.7)	69 (5.2)	46 (3.7)	38 (5.1)	3 (1.1)
International Average	36 (0.8)	69 (1.5)	36 (1.3)	33 (1.3)	6 (0.5)

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

* Fourth grade in most countries; see Table 2 for information about the grades tested in each country.
● Percent of total possible points on each item averaged over students.
▼ Average of percentage scores across items; all items weighted equally.
† Met guidelines for sample participation rates only after replacement schools were included (see Appendix A for details)
1 School-level exclusions for performance assessment exceed 25% of the National Desired Population (see Table A.3).
() Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

MAGNETS

Students performing the Magnets task were given two magnets, similar in appearance but of different magnetic strengths, and a number of magnetizable and non-magnetizable items such as steel balls, hairpins, and poker chips. The task was to conduct experiments to find which of the two magnets was the stronger, and to describe the experiments. Although on the surface the task appears to be a simple exercise in problem solving, because of its open nature it also required investigatory skills. The task was intended to measure problem solving in both strategy development and its implementation (Item 1), and the ability to support the conclusion with evidence (Item 2). Eighth-grade students were asked to experiment with the magnets and materials, without any directive as to the number or type of experiments, whereas fourth-grade students were asked to test magnets in two different the ways (see Figures 1.3 and 1.4 for descriptions of the tasks, scoring criteria for fully-correct responses, and a sample response). At either grade, only the first correct experiment was used to compute a student's score.[3]

For both grades, Item 1 was coded simply for the correctness of the solution to the problem. Students were given credit provided that they identified the correct magnet and that the test described could indeed have led to that conclusion. More than seven different approaches to the problem were recorded under Item 2, the two most popular being comparing the number of objects the magnet could hold and comparing the relative weights the magnets could lift.

Students at both grade levels found this task relatively easy. In most countries almost all eighth-grade students were able to identify the stronger magnet and to explain their strategy, and in the fourth grade the majority of students in most countries also were successful (see Tables 1.3 and 1.4). It is noteworthy that both the ability to apply a strategy to solve this problem and the ability to describe that strategy seem well developed among eighth-grade students, whereas among fourth graders there was a substantial performance difference between solving the problem and describing the strategy used for the solution.

[3] Because an explicit number of tests was not required at the eighth grade, there was no penalty if students performed only one, and no extra credit if they performed several.

FIGURE 1.3 - MAGNETS FULL-TASK EXAMPLE AND SCORING CRITERIA – EIGHTH GRADE

ITEM 1

MAGNETS

At this station you should have:

6 steel balls
10 hair pins or paper clips
6 poker chips
2 steel bars
10 washers
2 magnets
A 30 cm ruler

Read ALL directions carefully.

Your task:

Use the things in the bag to find which magnet, A or B, is stronger.

This is what you should do:

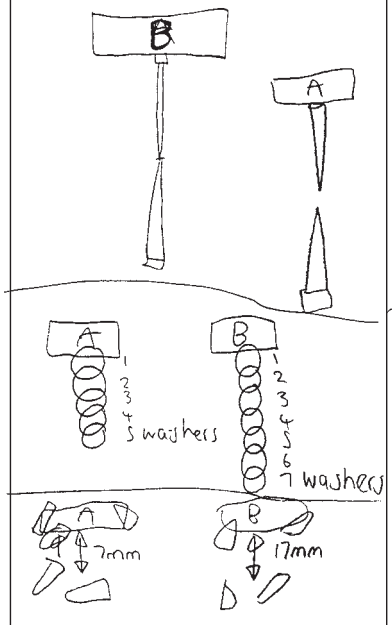
- Experiment with the things in the bag to complete the sentence below.

1. I found that magnet B is stronger.

page 1 TASK S2-P2

ITEM 2

2. Describe all the different ways you used to find which magnet was stronger. You may draw pictures or diagrams as part of your answer if it helps you to explain.

What I did	What happened
	Finest tip to tip stayed together on magnet B. But on magnet A It did not. magnet A held less washers than magnet B. Paper clips. at 7mm A magnet picked all of them up. At 17mm magnet B picked all of the paper clips up.

PUT ALL THE MATERIALS BACK IN THE BAG AND LEAVE THE STATION AS YOU FOUND IT.

TASK S2-P2 page 2

Criteria for Fully-Correct Response

Item 1 - Identify stronger magnet. Correct magnet identified according to administrator's notes.
Total Possible Points: 1

Item 2 - Describe all tests used to identify stronger magnet. Includes at least one correct test that: i) includes description or clearly interpretable diagram; ii) shows how results of test were interpreted.
Total Possible Points: 1

FIGURE 1.4 - MAGNETS ITEMS AND SCORING CRITERIA – FOURTH GRADE

MAGNETS

At this station you should have:

6 steel balls
10 hair pins or paper clips
6 poker chips
2 steel bars
10 washers
2 magnets
A 30 cm ruler

Read ALL directions carefully.

Your task:

Use the things in the bag to find which magnet, A or B, is stronger.

This is what you should do:

- Test the magnets in at least two different ways.

1. I found that magnet _____ is stronger.

2. Describe 2 different ways you used to find which magnet was stronger. You can draw pictures as part of your answer if it helps you to explain.

	What I did	What happened
Test One Magnet A Magnet B		
Test Two Magnet A Magnet B		

CRITERIA FOR FULLY-CORRECT RESPONSE

Item 1 - Identify stronger magnet. Correct magnet identified according to administrator's notes.
Total Possible Points: 1

Item 2 - Describe two tests used to identify stronger magnet.
i) Records what he or she did with each magnet in both tests.
ii) Relates results of each test to the identification of the stronger magnet. (Note: Student score reflects that at least one correct test is described.)
Total Possible Points: 1

Task layout condensed for display

Table 1.3 Magnets Task: Average Percentage Scores on Items – Eighth Grade*

Country	Overall Task Average▼	Average Percentage Scores on Items●	
		Item 1 Identify Stronger Magnet	Item 2 Describe Strategy
		1 Point	1 Point
† Scotland	98 (0.9)	99 (0.6)	96 (1.5)
†1 Switzerland	97 (1.2)	98 (1.3)	97 (1.7)
Spain	96 (1.4)	97 (1.5)	96 (1.9)
Sweden	95 (1.6)	95 (1.7)	95 (2.2)
Singapore	95 (0.9)	98 (1.0)	92 (1.7)
Portugal	94 (1.6)	97 (1.4)	90 (2.4)
New Zealand	93 (1.6)	92 (1.6)	94 (1.9)
Canada	92 (1.5)	95 (1.6)	89 (2.4)
Norway	91 (2.0)	88 (3.0)	94 (2.3)
Czech Republic	86 (2.3)	86 (3.3)	86 (2.2)
Cyprus	86 (2.3)	93 (1.9)	78 (3.5)
Iran, Islamic Rep.	45 (4.9)	52 (8.0)	39 (5.6)
Countries Not Satisfying Guidelines for Sample Participation Rates (See Appendix A for Details):			
Australia	92 (1.4)	97 (1.2)	86 (2.1)
2 England	99 (0.6)	99 (0.9)	99 (0.7)
Netherlands	94 (2.1)	96 (1.9)	93 (2.6)
United States	85 (2.5)	90 (2.6)	81 (3.4)
Countries Not Meeting Age/Grade Specifications (See Appendix A for Details):			
Colombia	96 (1.3)	96 (1.7)	95 (1.8)
3 Romania	83 (3.5)	78 (5.2)	89 (3.6)
Slovenia	92 (1.9)	94 (2.1)	91 (2.1)
International Average	90 (0.5)	92 (0.6)	88 (0.6)

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

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

● Percent of total possible points on each item averaged over students.

▼ Average of percentage scores across items; all items weighted equally.

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

1 National Desired Population does not cover all of International Desired Population (see Table A.2) - German-speaking cantons only.

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

3 School-level exclusions for performance assessment exceed 25% of the National Desired Population (see Table A.2).

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

Magnets Task: Average Percentage Score on Items – Fourth Grade* Table 1.4

Country	Overall Task Average▼	Average Percentage Scores on Items●	
		Item 1 Identify Stronger Magnet	Item 2 Describe Strategy
		1 Point	1 Point
Canada	84 (2.3)	92 (2.1)	76 (3.1)
†1 New Zealand	84 (2.2)	86 (2.3)	83 (2.9)
Portugal	74 (3.1)	83 (3.3)	65 (4.2)
Cyprus	68 (3.9)	82 (4.1)	54 (4.6)
Iran, Islamic Rep.	42 (5.1)	49 (5.2)	35 (6.1)
Countries Not Satisfying Guidelines for Sample Participation Rates (See Appendix A for Details):			
Australia	77 (3.1)	80 (3.8)	74 (3.6)
Hong Kong	74 (3.8)	82 (3.9)	65 (5.5)
United States	73 (3.0)	82 (3.2)	65 (4.2)
Countries Not Meeting Age/Grade Specifications (See Appendix A for Details):			
Slovenia	74 (3.8)	84 (3.8)	63 (4.6)
International Average	72 (1.2)	80 (1.2)	64 (1.5)

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

* Fourth grade in most countries; see Table 2 for information about the grades tested in each country.
● Percent of total possible points on each item averaged over students.
▼ Average of percentage scores across items; all items weighted equally.
† Met guidelines for sample participation rates only after replacement schools were included (see Appendix A for details)
1 School-level exclusions for performance assessment exceed 25% of the National Desired Population (see Table A.3).
() Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

BATTERIES

In the Batteries task, students were provided with four unmarked batteries and a flashlight. To begin, they were asked to find out which of the batteries were good and which were worn out. The task was intended to measure students' ability to develop and implement problem-solving strategies and use experimental evidence to support a conclusion, but it also sampled specific knowledge about electricity to solve a routine problem and to develop a concept-based explanation for the solution. Item 1 required students to identify the good batteries, which could be achieved by a systematic process of trial and error. Item 2 called for a description of the strategy used to identify the good batteries. Item 3 in this task required selection of the correct arrangement of batteries in a flashlight. Item 4 asked students to explain why their solution was correct, which requires knowledge of the concept of a complete circuit and an understanding of the direction of flow of electrical current.

Scoring criteria and a sample response for the Batteries task are presented in Figure 1.5. This task was exactly the same for eighth- and fourth-grade students.

Eighth-grade students in most countries had no difficulty in identifying the correct alignment of batteries in the flashlight (see Table 1.5, Item 3 – average percentage score: 91%), and were well able to identify the good batteries (Item 1 – average percentage score: 74%). They were somewhat less successful in describing the strategy used to identify the good batteries (Item 2 – average percentage score: 59%), and in explaining why their choice of battery alignment was the best one (Item 4 – average percentage score: 42%). Fourth-grade students also did quite well in identifying the correct battery alignment (see Table 1.6, Item 3 – average percentage score: 72%) and moderately well in finding the good batteries (Item 1 – average percentage score: 51%), but found describing their strategy (Item 2 – average percentage score: 23%) and explaining their choice (Item 4 – average percentage score: 19%) much more difficult.

The most typical partial-credit responses stated that the "positive pole must touch the negative" (without mentioning the reason), or "I tried all the combinations one after the other and this is what I got." A number of students merely repeated their strategy descriptions, perhaps not understanding the difference between describing what happened and explaining why it happened.

FIGURE 1.5 - BATTERIES

FULL-TASK EXAMPLE AND SCORING CRITERIA – EIGHTH AND FOURTH GRADES

ITEMS 1 AND 2

BATTERIES

At this station you should have:

A flashlight (or torch)
Four batteries in a plastic bag: Batteries A, B, C, D

Read ALL directions carefully.

Your task:

Find out which of the batteries are good and which are worn-out.

This is what you should do:

- Think about how you could solve this problem.
- Then work out which batteries are good and which are worn-out.

1. Based on your investigation which of the batteries are good and which are worn-out? Write the letters of the batteries in the spaces below.

   Good batteries AD

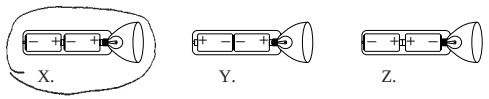
   Worn-out batteries BC

2. Write down how you decided which batteries were worn-out.

I put different batteries together and switch the Torch on. The combinations I used were 1 CD 2 AB 3 DB 4 CA 5 AD 6 CB and A and D batteries together were the brightest so A and D are the got batteries and BC are the worn-out ones.

ITEMS 3 AND 4

3. How should the batteries be put in the flashlight to give the brightest light? Here are 3 different ways of putting the batteries in the flashlight. Draw a circle around the picture that you think shows the correct way.



4. Why is the way you chose the best way to put in the batteries?

Because electricity flows in a circle or circuit and to get a circuit with batteries the pole have to be the opposite of each other like a magnet, and to complete the circuit the top and bottom poles have to be joint with a electrical conducting metal

Criteria for Fully-Correct Response

Item 1 - Identify which batteries are good and which are worn out.
All batteries correctly identified (per administrator notes).
Total Possible Points: 2

Item 2 - Describe how worn-out batteries were identified.
i) Shows evidence of systematic and definitive testing of different combinations of batteries. ii) "Systematic" is evidenced by trying all combinations of batteries or trying selected combinations using reasoning and scientific knowledge to eliminate some trials.
Total Possible Points: 2

Item 3 - Identify which arrangement of batteries inside flashlight will produce the brightest light. Correct arrangement identified (X).
Total Possible Points: 1

Item 4 - Explain why chosen arrangement is the best one.
i) Identifies correct arrangement. ii) Includes concepts of complete circuit and/or current flowing in one direction.
Total Possible Points: 2

Table 1.5 Batteries Task: Average Percentage Score on Items – Eighth Grade*

Country	Overall Task Average▼	Average Percentage Scores on Items●			
		Item 1 Identify Good/Bad Batteries	Item 2 Describe Tests	Item 3 Identify Arrangement	Item 4 Explain Arrangement
		2 Points	2 Points	1 Points	2 Points
Singapore	79 (2.1)	83 (3.4)	72 (3.7)	98 (1.4)	63 (2.7)
†1 Switzerland	75 (2.1)	87 (3.3)	77 (3.7)	94 (2.2)	41 (3.1)
Spain	73 (1.7)	84 (3.0)	75 (3.6)	93 (2.2)	41 (2.3)
Sweden	71 (2.9)	77 (3.7)	61 (4.5)	90 (2.7)	57 (4.3)
† Scotland	68 (2.4)	72 (4.1)	59 (3.5)	94 (2.1)	47 (3.2)
New Zealand	68 (1.6)	78 (3.0)	49 (3.0)	97 (1.2)	47 (2.1)
Norway	67 (1.7)	92 (2.1)	56 (3.6)	91 (2.4)	29 (2.3)
Czech Republic	66 (2.8)	76 (4.2)	63 (4.3)	87 (4.3)	39 (2.7)
Cyprus	66 (2.2)	81 (4.1)	72 (3.3)	87 (3.1)	25 (2.6)
Canada	62 (2.1)	66 (3.4)	52 (4.0)	92 (1.8)	38 (1.5)
Iran, Islamic Rep.	52 (4.0)	78 (5.0)	52 (7.7)	64 (3.5)	15 (3.0)
Portugal	50 (2.2)	39 (4.3)	29 (4.1)	92 (2.1)	41 (2.4)
Countries Not Satisfying Guidelines for Sample Participation Rates (See Appendix A for Details):					
Australia	71 (1.8)	81 (2.4)	71 (3.5)	93 (3.1)	40 (2.5)
2 England	77 (2.0)	89 (2.7)	71 (3.3)	91 (2.3)	56 (3.2)
Netherlands	63 (2.9)	68 (3.4)	42 (5.8)	93 (1.7)	47 (3.4)
United States	56 (1.9)	59 (4.1)	35 (3.7)	97 (1.1)	34 (2.3)
Countries Not Meeting Age/Grade Specifications (See Appendix A for Details):					
Colombia	55 (2.2)	61 (5.5)	39 (4.0)	80 (5.5)	40 (3.8)
3 Romania	75 (2.2)	73 (4.4)	75 (4.3)	96 (1.6)	56 (3.2)
Slovenia	71 (1.8)	69 (3.4)	64 (3.3)	97 (1.4)	52 (3.0)
International Average	67 (0.5)	74 (0.9)	59 (1.0)	91 (0.6)	42 (0.7)

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

* Eighth grade in most countries; see Table 2 for information about the grades tested in each country.
● Percent of total possible points on each item averaged over students.
▼ Average of percentage scores across items; all items weighted equally.
† Met guidelines for sample participation rates only after replacement schools were included (see Appendix A for details)
1 National Desired Population does not cover all of International Desired Population (see Table A.2) - German-speaking cantons only.
2 National Defined Population covers less than 90 percent of National Desired Population for the main assessment (see Table A.2).
3 School-level exclusions for performance assessment exceed 25% of the National Desired Population (see Table A.2).
() Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

Batteries Task: Average Percentage Score on Items – Fourth Grade* Table 1.6

Country	Overall Task Average▼	Average Percentage Scores on Items●			
		Item 1 Identify Good/Bad Batteries	Item 2 Describe Tests	Item 3 Identify Arrangement	Item 4 Explain Arrangement
		2 Points	2 Points	1 Points	2 Points
Canada	48 (2.0)	60 (3.7)	27 (2.6)	82 (3.0)	23 (2.2)
Cyprus	41 (2.2)	66 (5.0)	27 (3.9)	61 (3.5)	11 (2.1)
Iran, Islamic Rep.	40 (3.2)	73 (4.2)	27 (4.4)	48 (5.5)	13 (2.6)
†1 New Zealand	37 (1.4)	38 (3.7)	8 (1.5)	80 (2.6)	21 (2.2)
Portugal	31 (2.5)	32 (5.3)	11 (2.5)	62 (5.2)	19 (2.1)
Countries Not Satisfying Guidelines for Sample Participation Rates (See Appendix A for Details):					
Australia	40 (1.9)	48 (5.2)	25 (2.7)	71 (4.4)	17 (1.9)
Hong Kong	42 (2.0)	49 (5.0)	17 (3.1)	88 (2.9)	15 (3.5)
United States	38 (2.2)	38 (3.4)	21 (3.4)	76 (3.0)	19 (1.7)
Countries Not Meeting Age/Grade Specifications (See Appendix A for Details):					
Slovenia	54 (2.0)	58 (4.4)	44 (2.6)	83 (2.7)	30 (3.0)
International Average	41 (0.7)	51 (1.5)	23 (1.0)	72 (1.3)	19 (0.8)

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

* Fourth grade in most countries; see Table 2 for information about the grades tested in each country.
● Percent of total possible points on each item averaged over students.
▼ Average of percentage scores across items; all items weighted equally.
† Met guidelines for sample participation rates only after replacement schools were included (see Appendix A for details)
1 School-level exclusions for performance assessment exceed 25% of the National Desired Population (see Table A.3).
() Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

RUBBER BAND

The Rubber Band task asked students to investigate what would happen to the length of a rubber band as more and more rings were hung on it. They were provided with an experimental set-up that included a clipboard with a sheet of white paper and a length of rubber band suspended from the clip. A paper clip bent into the shape of a hook was attached to the other end of the rubber band, and students were given a set of metal rings to be hung onto the hook as weights (see Figures 1.6 and 1.7).

The instructions were to add weights to the band one at a time and record the length of the rubber band each time. The task was intended as an investigation into elasticity, with specific items within the task assessing particular skills, i.e., the ability to follow a procedure and measure and record data accurately (Item 1); to make a graph of the data (Item 2 – eighth-grade students only); to extract information from a table or graph students have constructed (Item 3 for eighth grade, Item 2 for fourth grade); to describe a trend in the data they have recorded (Item 4 for eighth grade, Item 3 for fourth grade); to extrapolate beyond the data they have recorded (Item 5 for eighth grade, Item 4 for fourth grade); and to explain the trend in the data that justified the extrapolation (Item 6 – eighth grade only).

Fourth-grade students were provided with a table and explicit instructions on how to collect and record data, whereas eighth-grade students had to construct and label the table themselves. Therefore, the quality of the data organization was not assessed for fourth graders, nor were fourth graders required to graph their data. Criteria for a fully-correct response to each item and a sample response are provided in Figures 1.6 and 1.7.

Eighth-grade students found the data collection and display easy in this task, probably due to the prescriptive directions that guided them step by step in what and how to measure and record (Table 1.7, Item 1 – average percentage scores: 85% and 88%). However, graphing the resulting data (Item 2 – average percentage score: 50%) proved more difficult. Calculating the increase in length, which required a combination of two routine procedures – reading the data table and computing a designated difference – was also difficult (Item 3 – average percentage score: 47%). Eighth-grade students were generally able to describe the trend in their data (Item 4 – average percentage score: 64%) and make an extrapolation on the basis of that trend (Item 5 – average percentage score: 59%), but were less successful in explaining the trend and justifying their extrapolation (Item 6 – average percentage score: 49%).

Fourth-grade students were very successful in measuring and recording the data from the task (Table 1.8, Item 1 – average percentage score: 84%), and in some countries could extrapolate from their data quite well (Item 4 – average percentage score: 50%), but generally found the other items very difficult.

FIGURE 1.6 - RUBBER BAND FULL-TASK EXAMPLE AND SCORING CRITERIA – EIGHTH GRADE

INTRODUCTION TO TASK

RUBBER BAND

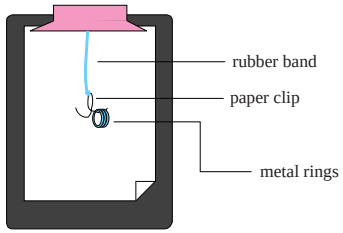
At this station you should have:

A clipboard with a rubber band
A large paper clip attached to one end of the rubber band
Metal rings to hang on the large paper clip
A 30 cm ruler
Some sheets of plain paper.
2 sheets of graph or squared paper

Read ALL directions carefully.

Your task:

Find out how the length of the rubber band changes as more and more rings are hung on it.



This is what you should do:

- Hang the metal rings onto the paper clip one by one
- Measure the length for each new ring.
- Record your measurements in the table.

ITEMS 1, 2, 3, AND 4

1. Write your measurements in the table. Remember to write a heading for each column.

Number of Rings	Length in Cm
1	17.5
2	18.0
3	18.5
4	19.0
5	19.5
6	20.0
7	20.5
8	21.0
9	21.5
10	22.0

2. Graph your results on the paper provided. You may use a graph or a bar chart.

ANSWER QUESTIONS 3 TO 6, USING YOUR TABLE, GRAPH, OR BAR CHART.

3. When there are 2 rings on the paper clip and 3 more are then added, how much longer does the rubber band become?

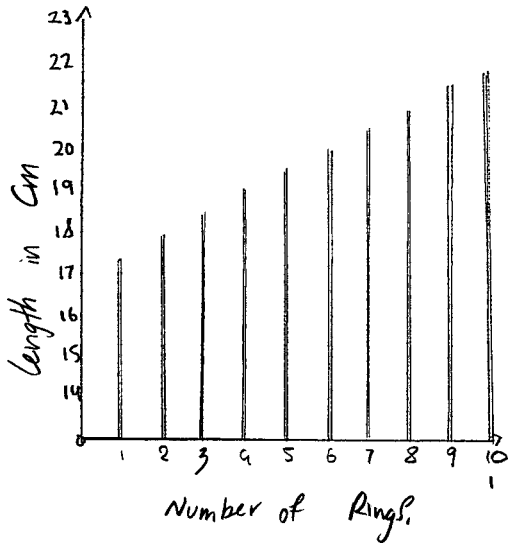
The rubber band becomes 1.5 cm longer.

4. Describe how the rubber band changed in length as more and more rings were added.

Each time I added a ring the length became 0.5 of a cm longer

Please turn the page.

ITEM 2 RESPONSE



ITEMS 5 AND 6

5. What do you think would be the length of the rubber band if you could add 2 more rings than you have been given?

I think the total length of the rubber band might be 23 cm.

6. Why do you think this would happen? This would happen because for every Ring put on the length becomes 0.5 of a centimeter longer.

PUT EVERYTHING BACK THE WAY YOU FOUND IT SO THAT SOMEONE ELSE CAN USE THE STATION.

FIGURE 1.6 (CONT.) RUBBER BAND – EIGHTH GRADE

CRITERIA FOR FULLY-CORRECT RESPONSE

Item 1 - Record rubber band length as rings are added. Response is scored for both the quality of the presentation and the quality of data collection.

Quality of presentation. i) Presents at least 2 sets of measurements in table. ii) Measurements are paired: number of rings and length of rubber band. iii) Labels table appropriately: data entries in columns identified by headings and/or units; units incorporated into headings or placed beside each measurement.
Total Possible Points: 2

Quality of data. i) Records length of rubber band for five or more different numbers of rings. ii) Shows reasonable trend in data: rubber band length increases with number of rings (at least for first few measurements); length *may* increase steadily at first and then stabilize or level off; elastic limit of rubber band *may* be exceeded and measurements toward the end show very large or erratic increases.
Total Possible Points: 3

Item 2 - Graph results (graph or bar chart). i) Axes correctly scaled. ii) Axes correctly labeled, including units where appropriate. iii) Measurements recorded in graph are consistent with data table. iv) Graph reflects trend in data.
Total Possible Points: 3

Item 3 - Calculate increase in length of rubber band when rings are added. i) Records amounts consistent with data in table, graph, or bar chart. ii) Calculates increase correctly.
Total Possible Points: 2

Item 4 - Describe how rubber band length changes as more rings are added. i) Description corresponds to data in table or graph. ii) Identifies trend in data. Trend may show that rubber band length increases consistently with each added ring; initially rubber band length increases consistently, then begins to level off; increases become larger or erratic with more rings (elastic limit of band exceeded); no change in length occurs (rubber band too strong for weights, per administrator notes).
Total Possible Points: 2

Item 5 - Predict increase in length of rubber band. Makes reasonable prediction, based on the data presented in the table or graph.
Total Possible Points: 1

Item 6 - Explain reason for prediction. i) Refers to the increase in length as read from the table or extrapolated from graph. ii) Attempts to relate weight or number of rings to elasticity of the rubber band. iii) Response is consistent with data in table or graph.
Total Possible Points:2

FIGURE 1.7 - RUBBER BAND ITEMS AND SCORING CRITERIA – FOURTH GRADE

RUBBER BAND

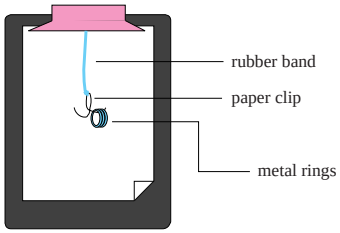
At this station you should have:

A board with a rubber band
A paper clip attached to one end of the rubber band
Metal rings to hang on the paper clip
A 30 cm ruler
Some sheets of plain paper

Read ALL directions carefully.

Your task:

Find out how the length of the rubber band changes as more and more rings are hung on it.



This is what you should do:

- Measure the length of the rubber band and write it in the table on the line marked "0 - rubber band with no ring."
- Hang one ring on the paper clip and measure the new length of the rubber band. Write it in the table opposite "1 ring."
- Keep adding rings one by one. Measure each new length and write it in the table.

1. Table of Measurements

NUMBER OF MASSES	LENGTH OF RUBBER BAND (in centimeters)
0 - rubber band with no ring	
1 ring	
2 rings	
3 rings	
4 rings	
5 rings	
6 rings	
7 rings	
8 rings	
9 rings	
10 rings	

USE YOUR TABLE TO ANSWER QUESTIONS 2 TO 5.

2. When there are 2 rings on the paper clip and 3 more are then added, how much longer does the rubber band become?

   The rubber band becomes _____ cm longer.

3. How did the rubber band change in length as more and more rings were added?

4. What do you think the length of the rubber band would be if you could add 2 more rings than you have been given?

   I think the total length of the rubber band might be _____ cm.

5. Why do you think this would happen?

Task layout condensed for display

FIGURE 1.7 (CONT.) RUBBER BAND – FOURTH GRADE

CRITERIA FOR FULLY-CORRECT RESPONSE

Item 1 - Record rubber band length as rings are hung from it. i) Records length of rubber band for at least five different numbers of rings. ii) Shows reasonable trend in data: rubber band lengths increase with numbers of rings (at least for first few sets of measurements); length *may* increase steadily at first, and then stabilize or level off; elastic limit of rubber band *may* be exceeded and measurements toward the end show very large or erratic increases.
Total Possible Points: 3

Item 2 - Calculate increase in length of rubber band when rings are added. i) Records amounts consistent with data in table.
ii) Calculates increase correctly.
Total Possible Points: 2

Item 3 - Describe how rubber band length changes as more rings are added. i) Description corresponds to data in table or graph.
ii) Identifies trend in data. Trend may show that rubber band length increases consistently with each added ring; initially rubber band length increases consistently, then begins to level off; increases become larger or erratic with more rings (elastic limit of band exceeded); no change in length occurs (rubber band too strong for weights, per administrator notes).
Total Possible Points: 2

Item 4 - Predict increase in length of rubber band. Makes reasonable prediction, based on the data presented in the table.
Total Possible Points: 1

Item 5- Explain reason for prediction. i) Refers to the increase in length as read from the table. ii) Attempts to relate weight or number of rings to elasticity of the rubber band. iii) Response is consistent with data in table.
Total Possible Points:2

Rubber Band Task: Average Percentage Score on Items - Eighth Grade* Table 1.7

Country	Overall Task Average▼	Average Percentage Scores on Items ●						
		Item 1 Measure Lengths		Item 2 Graph Results	Item 3 Calculate Increase	Item 4 Describe Trend	Item 5 Predict Length	Item 6 Explain Prediction
		Presentation	Data Quality					
		2 Points	3 Points	3 Points	2 Points	2 Points	1 Points	2 Points
Singapore	80 (1.5)	95 (1.3)	99 (0.7)	67 (2.3)	67 (4.2)	87 (1.7)	84 (3.0)	61 (2.3)
† Scotland	75 (1.8)	95 (1.7)	96 (1.6)	69 (2.8)	57 (3.7)	73 (2.6)	78 (4.2)	54 (3.8)
Canada	71 (2.0)	87 (2.0)	95 (1.4)	66 (3.2)	55 (6.1)	59 (3.9)	73 (3.1)	59 (3.2)
Sweden	70 (2.4)	83 (2.8)	93 (1.8)	55 (3.4)	64 (3.8)	65 (4.6)	72 (4.1)	59 (4.3)
†1 Switzerland	67 (1.9)	93 (2.2)	93 (1.9)	31 (4.2)	58 (5.0)	73 (3.4)	68 (4.5)	53 (4.4)
New Zealand	65 (1.8)	89 (1.8)	93 (1.4)	67 (3.1)	56 (3.6)	68 (2.8)	51 (3.4)	33 (2.8)
Czech Republic	65 (3.6)	81 (2.3)	86 (2.6)	44 (4.1)	54 (5.8)	70 (4.3)	66 (6.6)	54 (5.9)
Norway	63 (1.9)	80 (2.6)	96 (1.2)	49 (3.1)	56 (3.8)	60 (4.0)	53 (4.1)	46 (3.7)
Cyprus	59 (2.3)	83 (3.1)	87 (2.9)	41 (4.4)	46 (5.2)	53 (3.7)	58 (3.7)	46 (4.7)
Iran, Islamic Rep.	56 (5.4)	83 (5.4)	80 (6.4)	26 (6.5)	20 (5.2)	62 (4.1)	56 (8.6)	63 (8.9)
Spain	51 (2.0)	68 (3.3)	75 (3.0)	33 (3.3)	33 (4.0)	58 (3.9)	44 (4.4)	48 (2.6)
Portugal	51 (2.3)	78 (3.0)	83 (3.5)	41 (3.4)	32 (3.8)	44 (4.2)	47 (4.6)	34 (4.5)
Countries Not Satisfying Guidelines for Sample Participation Rates (See Appendix A for Details):								
Australia	64 (2.4)	92 (2.2)	92 (1.8)	55 (3.1)	49 (4.1)	65 (5.1)	57 (4.3)	41 (4.1)
2 England	79 (1.4)	95 (1.5)	98 (0.9)	76 (2.4)	55 (3.3)	84 (2.9)	80 (3.1)	68 (3.6)
Netherlands	70 (1.9)	89 (2.1)	95 (1.3)	71 (2.9)	62 (3.5)	63 (4.0)	53 (5.8)	61 (3.8)
United States	63 (2.4)	83 (3.1)	88 (2.3)	54 (3.6)	45 (4.1)	68 (3.2)	59 (4.0)	41 (3.1)
Countries Not Meeting Age/Grade Specifications (See Appendix A for Details):								
Colombia	40 (3.7)	58 (7.4)	67 (5.8)	14 (3.0)	17 (3.9)	55 (5.4)	28 (4.3)	39 (3.9)
3 Romania	45 (3.0)	87 (3.1)	60 (4.1)	39 (4.2)	26 (5.4)	47 (3.0)	30 (4.8)	25 (3.6)
Slovenia	64 (1.7)	93 (1.6)	91 (1.8)	58 (3.3)	35 (4.1)	57 (3.6)	65 (5.1)	47 (3.8)
International Average	63 (0.6)	85 (0.7)	88 (0.7)	50 (0.8)	47 (1.0)	64 (0.9)	59 (1.1)	49 (1.0)

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

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

● Percent of total possible points on each item averaged over students.

▼ Average of percentage scores across items; all items weighted equally.

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

1 National Desired Population does not cover all of International Desired Population (see Table A.2) - German-speaking cantons only.

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

3 School-level exclusions for performance assessment exceed 25% of the National Desired Population (see Table A.2).

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

Table 1.8 Rubber Band Task: Average Percentage Score on Items – Fourth Grade*

Country	Overall Task Average▼	Average Percentage Scores on Items●				
		Item 1 Record Lengths	Item 2 Calculate Increase	Item 3 Describe Trend	Item 4 Predict Length	Item 5 Explain Prediction
		3 Points	2 Points	2 Points	1 Point	2 Points
Canada	55 (1.5)	96 (0.8)	23 (2.5)	49 (2.6)	69 (2.6)	39 (2.6)
Cyprus	45 (3.2)	89 (3.2)	26 (4.0)	41 (3.0)	44 (5.0)	26 (4.1)
†1 New Zealand	44 (2.3)	89 (2.9)	27 (3.2)	36 (2.9)	53 (3.4)	13 (2.8)
Iran, Islamic Rep.	36 (3.3)	76 (4.8)	13 (2.6)	36 (2.8)	31 (6.3)	26 (4.4)
Portugal	27 (2.2)	55 (4.4)	16 (3.3)	23 (2.9)	25 (3.5)	15 (2.9)
Countries Not Satisfying Guidelines for Sample Participation Rates (See Appendix A for Details):						
Australia	52 (2.9)	84 (2.7)	30 (3.4)	41 (5.4)	71 (5.3)	33 (2.7)
Hong Kong	43 (2.5)	91 (2.5)	26 (4.2)	43 (2.9)	37 (5.4)	18 (3.4)
United States	45 (1.8)	77 (1.9)	30 (3.5)	41 (3.0)	54 (3.4)	22 (2.5)
Countries Not Meeting Age/Grade Specifications (See Appendix A for Details):						
Slovenia	51 (1.7)	96 (1.5)	20 (3.5)	43 (2.5)	64 (4.6)	34 (3.1)
International Average	44 (0.8)	84 (1.0)	23 (1.1)	39 (1.1)	50 (1.5)	25 (1.1)

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

* Fourth grade in most countries; see Table 2 for information about the grades tested in each country.
● Percent of total possible points on each item averaged over students.
▼ Average of percentage scores across items; all items weighted equally.
† Met guidelines for sample participation rates only after replacement schools were included (see Appendix A for details)
1 School-level exclusions for performance assessment exceed 25% of the National Desired Population (see Table A.3).
() Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

SOLUTIONS

In the Solutions task, students were asked to investigate the effect of temperature on the speed with which tablets dissolved in water. The students were provided with several beakers, a supply of hot and cold water, tablets that would fizz as they dissolved, a stirrer, a thermometer, a 30cm ruler, and a stopwatch or wall clock with a second hand. This task was intended to measure students' ability to plan an investigation; use a thermometer correctly and accurately; collect, tabulate, analyze, and interpret data; invoke concept knowledge to explain findings; and evaluate the entire investigation. This task was administered to eighth-grade students only.

Unique among all items in the performance assessment, the last item of this task was intended to measure students' ability to evaluate the experiment by identifying the need and reasons for changes. Possible changes could be in design, materials, method, use of equipment, number of repeated measurements, or they could be intended to verify that the variables ignored were indeed irrelevant. Full credit was awarded only if the reasons for the changes were included. A description of the task, along with a sample response and scoring criteria for a fully-correct response for each item, are presented in Figure 1.8.

Students internationally found it somewhat difficult to describe their plan for the investigation in writing (Table 1.9, Item 1 – average percentage score: 44%). Carrying out the measurements, recording, and presenting the data were generally easier (Item 2 – average percentage scores: 62% and 59%). Students did best at providing conclusions consistent with their data (Item 3 – average percentage score: 77%). Presenting an explanation for the observed phenomena was much more difficult (Item 4 – average percentage score: 22%). To receive full credit, the explanation needed to demonstrate knowledge both of the relationship between higher temperature and greater energy and of the effect of this energy on the dissolving process.

Students also found it difficult to evaluate their plan and describe modifications they would make (Item 5 – average percentage score: 30%). Almost half of the students claimed, erroneously, that no changes were needed in their original plans. On the other hand, about one quarter of the students internationally stated, correctly, "no change required," i.e., their original plan was found to be complete and correct. To be able to evaluate one's work – design, data collection, and results – is a sophisticated form of scientific thinking, and one that many eighth-grade students in these countries have yet to acquire.

FIGURE 1.8 - SOLUTIONS

FULL-TASK EXAMPLE AND SCORING CRITERIA – EIGHTH GRADE

ITEM 1

SOLUTIONS

At this station you should have:

Hot and cold water
Several beakers
Some tablets
A stirrer
A clock or watch with a second hand
A thermometer
A 30 cm ruler

Read ALL directions carefully.

Your task:

Investigate what effect different water temperatures have on the speed with which the tablet dissolves.

This is what you should do:

- Plan an experiment to find out what effect different water temperatures have on the speed with which the tablet dissolves.

1. Write your plan here. Your plan should include
   - what you will measure.
   - how many measurements you will make.
   - how you will present your measurements in a table.

Take a measurement of the time it takes for the tablet to completely dissolve in 200ml
Take 1 measurement in cold, one in hot and one in warm (½ cold, ½ hot)
Make table with time it takes for each ~~tablet~~ to dissolve in each temperature

page 1 TASK S5-P2

ITEMS 2 AND 3

2. Carry out your tests on the tablets. Make a table and record all your measurements.

Temperature °C	time (s)
cold 13°C	62
warm 42°C	32
Hot 79°C	30
1part Hot - 3part cold 28°C	41
1part cold - 3part hot 54°C	30

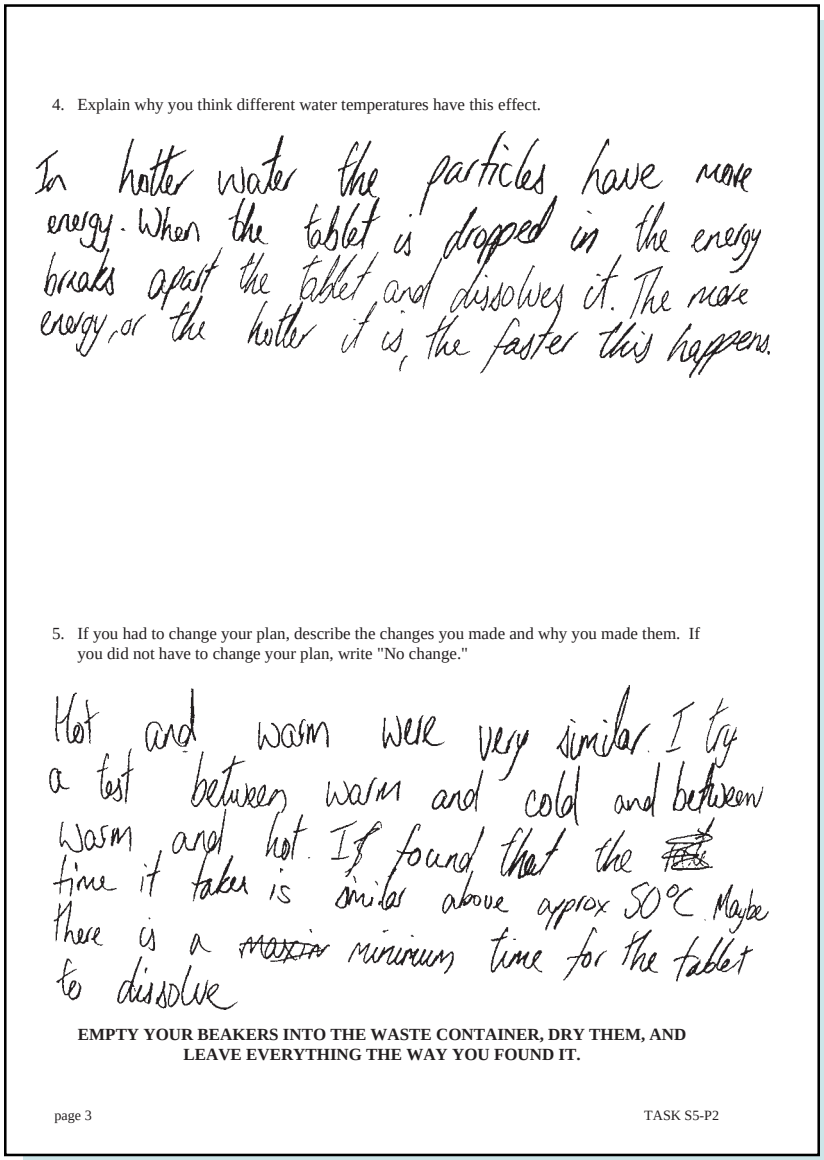
3. According to your investigation, what effect do different water temperatures have on the speed with which a tablet dissolves?

In higher water temperatures the tablets dissolve faster and more thoroughly than cold water

Please turn the page.

TASK S5-P2 page 2

ITEMS 4 AND 5



4. Explain why you think different water temperatures have this effect.

In hotter water the particles have more energy. When the tablet is dropped in the energy breaks apart the tablet and dissolves it. The more energy, or the hotter it is, the faster this happens.

5. If you had to change your plan, describe the changes you made and why you made them. If you did not have to change your plan, write "No change."

Hot and warm were very similar. I try a test between warm and cold and between warm and hot. If found that the time it takes is similar above approx 50°C. Maybe there is a ~~maxim~~ minimum time for the tablet to dissolve

EMPTY YOUR BEAKERS INTO THE WASTE CONTAINER, DRY THEM, AND LEAVE EVERYTHING THE WAY YOU FOUND IT.

page 3 TASK S5-P2

CRITERIA FOR FULLY-CORRECT RESPONSE

Item 1 - Plan investigation . i) Describes how the investigation will be conducted. ii) States what variables will be measured or observed; includes both solution time and temperature. iii) Provides control for other variables, or renders other variables irrelevant by design.
Total Possible Points: 2

Item 2 - Conduct investigation and record measurements in table. Response is scored for both the quality of the presentation and the quality of the data collection.

> **Quality of presentation.** i) Presents at least 2 sets of measurements in table. ii) Measurements are paired: dissolution time and temperature. iii) Labels table appropriately: data entries in columns identified by headings and/or units; units incorporated into headings or placed beside each measurement.
> *Total Possible Points: 2*

> **Quality of data.** i) Records solution time for at least three temperature points. ii) Measurements are plausible: time and temperature (10^o to 100^o C) iii) Records solution times that decline as temperature increases.
> *Total Possible Points: 3*

Item 3 - Draw conclusions about effect of temperature.
i) Conclusion is consistent with data table or other presentation of data (graph or text). ii) Describes relationship presented in the data.
Total Possible Points: 2

Item 4 - Explain conclusions. i) Relates higher temperature to greater energy or speed of particles (atoms, molecules, etc.). ii) Makes connection between greater speed or energy of water molecules and the effect on the tablet (may be implicit).
Total Possible Points: 2

Item 5 - Evaluate design and experiment; describe changes.
i) Response is consistent with the way student recorded and described data ("no change" is acceptable if student plan was complete).
ii) Changes may be made in method, use of equipment, number of measurements taken, etc; reason for change must be included.
Total Possible Points:2

Table 1.9 Solutions Task: Average Percentage Score on Items - Eighth Grade*

Country	Overall Task Average▼	Average Percentage Scores on Items●					
		Item 1 Plan Investigation	Item 2 Conduct Investigation		Item 3 Draw Conclusion	Item 4 Explain Conclusion	Item 5 Evaluate Design
			Presentation	Data Quality			
		2 Points	2 Points	3 Points	2 Points	2 Points	2 Points
Singapore	68 (2.7)	53 (5.0)	91 (1.8)	81 (2.1)	93 (2.2)	42 (4.0)	46 (5.4)
Czech Republic	59 (2.3)	60 (3.6)	71 (2.8)	63 (3.0)	86 (3.1)	28 (3.4)	48 (4.5)
†1 Switzerland	57 (1.9)	55 (4.8)	79 (2.7)	79 (2.9)	85 (3.0)	11 (1.5)	36 (5.2)
† Scotland	51 (2.3)	44 (3.8)	72 (3.9)	70 (3.3)	81 (3.7)	18 (2.4)	23 (3.9)
Iran, Islamic Rep.	50 (3.5)	52 (4.3)	52 (6.7)	47 (5.5)	86 (4.5)	36 (7.7)	26 (3.2)
Sweden	50 (2.2)	51 (3.9)	55 (3.3)	65 (3.0)	77 (2.9)	18 (2.5)	34 (4.7)
Canada	48 (2.1)	39 (4.0)	64 (3.1)	59 (2.6)	76 (1.9)	26 (2.4)	27 (4.1)
New Zealand	48 (2.1)	46 (3.2)	61 (3.4)	54 (2.0)	70 (2.5)	33 (2.8)	25 (3.5)
Norway	42 (1.8)	43 (2.4)	55 (3.1)	57 (3.8)	68 (4.5)	6 (1.3)	21 (3.4)
Spain	41 (2.3)	44 (4.2)	51 (3.9)	43 (3.2)	74 (3.3)	17 (2.2)	19 (3.2)
Portugal	36 (2.4)	27 (3.9)	39 (3.9)	36 (2.9)	74 (4.3)	25 (2.8)	13 (3.2)
Cyprus	29 (2.9)	14 (3.1)	42 (4.9)	44 (4.6)	47 (5.1)	18 (3.0)	10 (2.8)
Countries Not Satisfying Guidelines for Sample Participation Rates (See Appendix A for Details):							
Australia	59 (2.2)	55 (4.7)	79 (3.4)	79 (2.3)	89 (2.6)	23 (3.5)	30 (3.6)
2 England	68 (2.1)	66 (3.8)	82 (2.5)	75 (1.6)	89 (2.6)	36 (4.5)	59 (4.1)
Netherlands	43 (2.7)	45 (2.7)	46 (5.2)	52 (4.2)	77 (5.0)	12 (2.6)	23 (3.0)
United States	48 (2.2)	33 (2.6)	64 (3.7)	59 (3.2)	82 (3.1)	27 (3.3)	24 (2.6)
Countries Not Meeting Age/Grade Specifications (See Appendix A for Details):							
Colombia	26 (2.3)	14 (3.7)	43 (5.2)	43 (4.4)	41 (4.8)	8 (1.6)	6 (1.6)
3 Romania	63 (2.6)	63 (4.9)	59 (3.7)	68 (5.6)	82 (3.2)	30 (4.2)	73 (4.9)
Slovenia	49 (2.0)	37 (4.0)	75 (2.9)	57 (2.8)	81 (2.8)	12 (2.8)	34 (3.6)
International Average	49 (0.5)	44 (0.9)	62 (0.9)	59 (0.8)	77 (0.8)	22 (0.8)	30 (0.9)

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

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

● Percent of total possible points on each item averaged over students.

▼ Average of percentage scores across items; all items weighted equally.

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

1 National Desired Population does not cover all of International Desired Population (see Table A.2) - German-speaking cantons only.

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

3 School-level exclusions for performance assessment exceed 25% of the National Desired Population (see Table A.2).

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

CONTAINERS

In the Containers task, students were given three containers of different insulating capacity, for example, metal, ceramic, and plastic, and were asked to find out which one would keep a hot drink warm for the longest time. They also received thermometers, a clock, a piece of card to use as a fan, and a supply of hot water. The students were instructed to pour a measure of hot water into each of the containers, and to take the temperature in each one over a ten-minute interval. They were provided with a pre-designed data table in which to record their observations. This task assessed students' ability to make and record measurements of temperature and probed their understanding of the concept of insulation. Figure 1.9 presents the task with sample student responses and scoring criteria for a fully-correct response. This task was administered to fourth-grade students only.

In general, this was a difficult task for fourth graders. Although most students in most countries were able to use a laboratory thermometer, in many cases, the data gathered were incomplete or contained small inaccuracies in measurement (Table 1.10, Item 1 – average percentage scores: 91% and 56%). Students did reasonably well in identifying the container that kept water hottest (Item 2 – average percentage score: 48%), but almost none could explain insulating capacity in terms of the materials from which the containers were made.

An interesting misconception appeared when students were asked to apply their findings to a different situation – that of keeping ice cream cold. While 15% of students internationally (Item 4) recognized that the container that was best for keeping a hot drink warm would also be best for keeping ice cream cold, almost none could explain why (Item 5). About one-quarter of the students seemed to see the ice cream as an opposite case, explaining that the container in which the temperature of a hot drink declined most rapidly would be the one to keep ice cream cold the longest.

FIGURE 1.9 - CONTAINERS FULL-TASK EXAMPLE AND SCORING CRITERIA – FOURTH GRADE

INTRODUCTION TO TASK

CONTAINERS

At this station you should have:

Three containers (or cups) marked A, B, C
Three thermometers
A clock or watch
A container with very hot water. BE CAREFUL NOT TO SPILL HOT WATER.
Pieces of card to use as a fan if you wish
A roll of paper to wipe up spills
A measuring cup

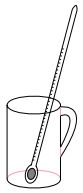
Read ALL directions carefully.

Your task:

Find out which of the containers will keep a hot drink warm for the longest time.

This is what you should do:

- Place a thermometer in each of the containers BEFORE the hot water is poured in. Your teacher will pour the hot water when you are ready. BE CAREFUL. THE WATER IS VERY HOT.



- Measure the temperature on each thermometer as soon as the hot water is poured in.
- Write these measurements and the time in the table on the opposite page.
- Now you will take measurements over a total of 10 minutes.
  - → Decide how often to read each thermometer.
  - → Write your measurements in the table on the opposite page.

ITEMS 1, 2, AND 3

1. Table of Measurements:

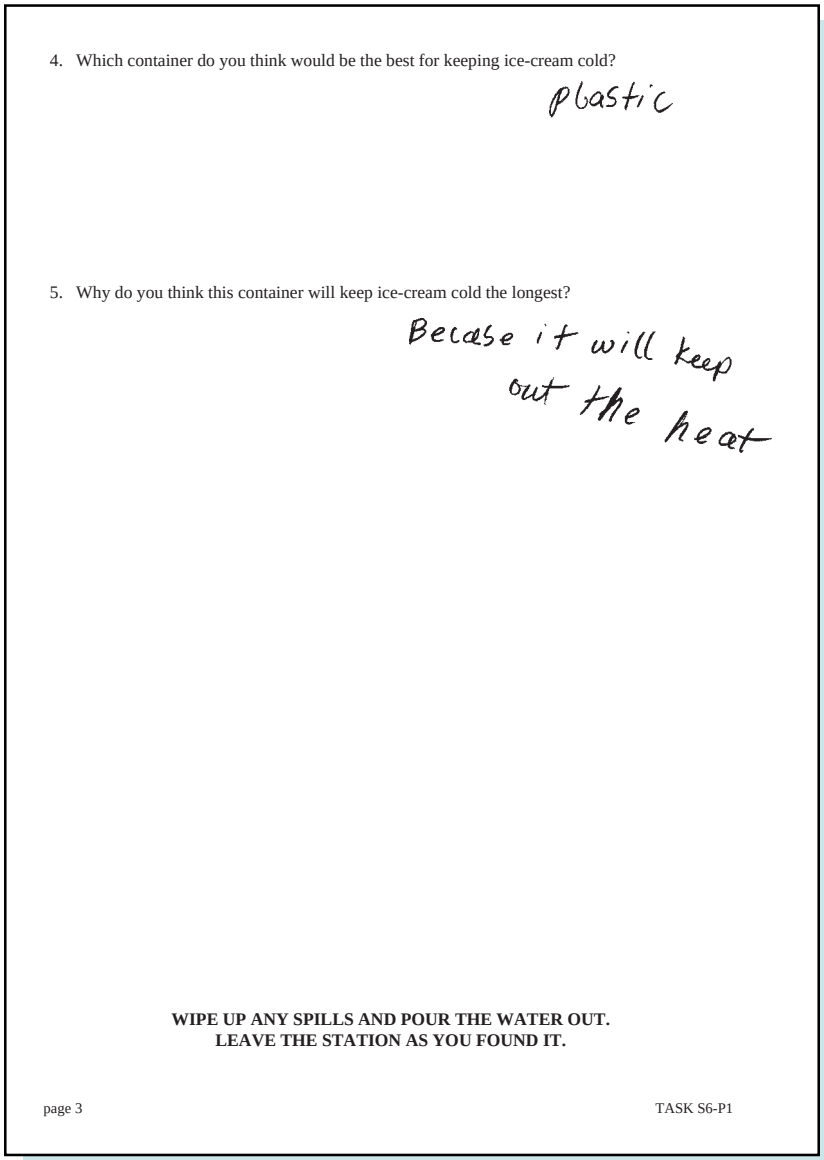
Time	Temperature of Container A	Temperature of Container B	Temperature of Container C
2:28	0	0	0
2:32	50	60	50
2:35	50	60	50
2:36	50	60	50
2:38	50	60	40
2:39	50	50½	40

2. Look at the table. Which container keeps a hot drink warm for the longest time? cupB plastic

3. Why do you think this container was best for keeping a hot drink warm? Because it holds the heat

Please turn the page.

ITEMS 4 AND 5



4. Which container do you think would be the best for keeping ice-cream cold?

plastic

5. Why do you think this container will keep ice-cream cold the longest?

Becase it will keep out the heat

WIPE UP ANY SPILLS AND POUR THE WATER OUT.
LEAVE THE STATION AS YOU FOUND IT.

page 3 TASK S6-P1

CRITERIA FOR FULLY-CORRECT RESPONSE

Item 1 - Measure temperatures and record data in table. Student is scored both on proper use of the thermometer and on the quality of data gathering.

Ability to use thermometer. Does not require assistance in proper use of the thermometer (Based on adminstrator notes on any special assitance provided.)
Total Possible Points: 1

Quality of data gathering. i) Records times and temperatures for 5 or more temperature points per container. ii) Times cover full 10-minute range. iii) Trend in the temperature is reasonable: temperature declines with time in one or more of the cups. (One cup may be too well insulated to give measurable declines in 10 minutes.)
Total Possible Points: 3

Item 2 - Identify container that keeps hot drink warm longest.
i) Identifies correct container (based on administrator notes). ii) Container identified is consistent with the data in table.
Total Possible Points: 2

Item 3 - Explain why container retains heat. i) Relates material of containers to their ability to retain or transfer heat. ii) Includes comparison of different containers based on heat transfer.
iii) Logically applies any additional relevant information (stirring, thickness of container, size differences, etc.).
Total Possible Points: 2

Item 4 - Predict best container for keeping ice cream cold. Identifies the same container that best keeps hot drink warm.
Total Possible Points: 1

Item 5 - Explain why container keeps ice cream cold. i) Relates material of containers to their ability to retain or transfer heat.
ii) Includes comparison of different containers based on heat transfer.
iii) Logically applies any additional relevant information provided (stirring, thickness of container, size difference, etc.).
Total Possible Points: 2

Table 1.10 Containers Task: Average Percentage Score on Items – Fourth Grade*

Country	Overall Task Average▼	Average Percentage Scores on Items●					
		Item 1 Measure Temperatures and Record in Table		Item 2 Identify Best Insulator	Item 3 Explain Best Insulator	Item 4 Apply to Ice Cream	Item 5 Explain Application
		Ability to Use Thermometer	Quality of Data Gathering				
		1 Points	3 Points	2 Points	2 Points	1 Point	2 Points
Cyprus	42 (1.3)**	- -	82 (3.8)	60 (4.5)	6 (1.5)	4 (1.9)	1 (0.6)
Canada	40 (1.1)	94 (2.1)	69 (2.8)	56 (2.7)	7 (1.1)	14 (2.4)	3 (0.8)
†1 New Zealand	33 (1.4)	95 (1.8)	38 (3.3)	50 (4.4)	3 (0.8)	10 (1.9)	1 (0.6)
Iran, Islamic Rep.	30 (3.5)	63 (7.6)	49 (5.3)	39 (5.3)	9 (1.8)	17 (4.2)	2 (1.4)
Portugal	26 (1.9)	78 (5.5)	31 (4.7)	33 (5.0)	3 (1.3)	11 (2.3)	2 (1.0)
Countries Not Satisfying Guidelines for Sample Participation Rates (See Appendix A for Details):							
Australia	39 (0.8)	93 (2.8)	57 (3.9)	65 (2.5)	5 (1.4)	14 (2.7)	3 (0.9)
Hong Kong	41 (1.3)	99 (0.6)	58 (2.9)	56 (3.8)	11 (2.8)	18 (3.0)	5 (1.2)
United States	40 (1.1)	98 (0.7)	64 (3.1)	32 (3.5)	8 (1.2)	31 (3.5)	4 (1.0)
Countries Not Meeting Age/Grade Specifications (See Appendix A for Details):							
Slovenia	38 (1.3)	100 (0.0)	54 (4.0)	45 (3.2)	4 (1.5)	18 (3.1)	7 (1.5)
International Average	37 (0.6)	91 (1.1)	56 (1.3)	48 (1.3)	6 (0.5)	15 (1.0)	3 (0.3)

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

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

● Percent of total possible points on each item averaged over students.

▼ Average of percentage scores across items; all items weighted equally.

A dash (-) indicates data are not available. Ability to use thermometer was not recorded in Cyprus.

**Overall task average includes an estimated average percentage score of 97% for the missing item based on overall relative country performance and international item difficulty.

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

1 School-level exclusions for performance assessment exceed 25% of the National Desired Population (see Table A.3).

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