

Mullis, I.V.S., Jones, C., and Garden, R.A. (1996) "Training Sessions for Free-Response Scoring and Administration of Performance Assessment" in M.O. Martin and D.L. Kelly (eds.), *Third International Mathematics and Science Study (TIMSS) Technical Report, Volume I: Design and Development*. Chestnut Hill, MA: Boston College.

## **10. TRAINING SESSIONS FOR FREE-RESPONSE SCORING AND ADMINISTRATION OF PERFORMANCE ASSESSMENT.....10-1**

*Ina V.S. Mullis, Chancey Jones, and Robert A. Garden*

10.1	OVERVIEW.....	10-1
10.2	THE TIMSS FREE-RESPONSE CODING TRAINING TEAM.....	10-2
10.3	THE SCHEDULE OF THE REGIONAL TRAINING SESSIONS.....	10-3
10.4	DESCRIPTION OF EACH TRAINING SESSION.....	10-4
10.5	THE TRAINING MATERIALS.....	10-8
10.6	CONCLUDING REMARKS.....	10-12

# **10. Training Sessions for Free-Response Scoring and Administration of Performance Assessment**

Ina V.S. Mullis  
Chancey Jones  
Robert A. Garden

## **10.1 OVERVIEW**

For the TIMSS main survey, about one-third of the written test time is devoted to free-response items, both short-answer and extended-response. This includes the five TIMSS tests: Population 1, Population 2, Population 3 mathematics and science literacy, Population 3 physics, and Population 3 advanced mathematics. Additionally, for Populations 1 and 2, subsamples of students in approximately 20 countries participated in a performance assessment consisting of hands-on tasks for which students were expected to record results or show other products from their activities (see Chapter 6). Across the five main surveys and the performance assessment, TIMSS included approximately 300 free-response questions and tasks.

With large within-country samples of students responding to the tests, and those student samples representing widely diverse cultures from countries spanning the world's continents, ensuring reliability of scoring was a major concern for TIMSS. The scope of the effort was enormous, with 27 countries participating for Population 1, 46 countries for Population 2, and 21 countries for Population 3. The sample size was approximately 5,000

to 7,500 per population for the main survey. Although the samples for the performance assessment component were smaller (approximately 450 students per country for each of Population 1 and Population 2), the performance assessment entailed setting up equipment and conducting testing sessions involving 12 different hands-on investigations in science and mathematics.

Because of the scope of TIMSS, the training sessions were designed to assist representatives of national centers who would then be responsible for training personnel in their countries to apply the two-digit scoring codes reliably. A four-day training session was developed in which attendees were introduced to the coding system and given practice in coding example papers. In the most effective schedule for the sessions, the first three days were devoted to Populations 1, 2, and 3, respectively, and the fourth day to the administration and coding of the performance assessment. Considering that English is not the native language of many participants and that free-response scoring is a very challenging undertaking requiring subtle distinctions, four days is about the maximum length for any such training session without driving people to total exhaustion.

The four-day training period was demanding, intense, and appropriate for most participants. However, for any future study of the scope of TIMSS, more time needs to be spent on training. For example, one day each could easily have been devoted to training for the advanced mathematics and physics for Population 3. Also, even without discussing administration procedures, a full day could easily have been spent training for coding on the performance assessment. Training for administering the performance assessment ideally would include a separate training session for administering the tests.

Training sessions were conducted in seven regions to provide easy access for participants and smaller groups for the TIMSS trainers to manage. Consistency across sessions was provided by using essentially the same training team and training materials across all the sessions. All in all, this model of “training the trainers” appears to have worked relatively successfully.

## **10.2 THE TIMSS FREE-RESPONSE CODING TRAINING TEAM**

The members of the training team embodied considerable knowledge of the TIMSS tests and of procedures used in training coders to achieve high reliability. The team members are briefly described below.

*Mr. Chancey Jones, United States.* Mr. Jones was heavily involved in developing the mathematics instruments for Populations 1 and 2. As part of his work in managing development of mathematics tests at Educational Testing Service (United States), he has had extensive experience in establishing scoring criteria, training personnel in scoring procedures, and managing large-scale mathematics scoring sessions for the College Board’s Advanced Placement Program and for the U.S. National Assessment of Educational Progress. Mr. Jones was also responsible for reviewing the TIMSS mathematics training materials and conducting training for scoring

mathematics items. He also assisted the International Study Center by serving as the team leader for several training sessions.

*Mr. Robert Garden, New Zealand.* Mr. Garden coordinated the development of the TIMSS mathematics instruments. As International Coordinator of the IEA's Second International Mathematics Study, he is experienced in conducting international studies. Mr. Garden was director of research and statistics at the Education Ministry in New Zealand, and more recently became a private consultant. He too was responsible for reviewing mathematics training materials and conducting training for scoring mathematics items.

*Dr. Graham Orpwood, Canada.* Dr. Orpwood coordinated the development of the TIMSS science instruments. He is a professor of science education at the Faculty of Education of York University in Ontario. Dr. Orpwood had responsibility for reviewing the science training materials and conducting training for scoring the science items. He also was involved in developing the TIMSS performance assessment and had responsibility for training related to the administration of the performance assessment tasks.

*Dr. Jan Lokan, Australia.* Dr. Lokan is the National Research Coordinator for TIMSS in Australia. A senior researcher at the Australian Council for Educational Research, Dr. Lokan contributed substantially to developing the coding guides for the science items and TIMSS performance assessment. She shared responsibility for conducting training for scoring the science items. She also had a central role in training related to the administration of the performance assessment tasks.

*Dr. Ina Mullis, United States.* Dr. Mullis, codeputy director of the TIMSS International Study Center, coordinated the activities of the training team. Before joining TIMSS, she was director of the National Assessment of Education Progress in the United States, where she gained extensive experience in the evaluation of students' answers to free-response questions in large-scale assessments. She coordinated preparation of the TIMSS manuals containing the coding guides and example responses, and of the materials used at the training sessions.

### **10.3 THE SCHEDULE OF THE REGIONAL TRAINING SESSIONS**

As shown in Table 10.1, the regional training sessions for free-response coding and administering the performance assessment were held across a one-year period from October 1994 through September 1995. This time period was established to accommodate the different school schedules in the countries in terms of the TIMSS schedule. For example, the school schedule for Southern Hemisphere countries is such that the TIMSS tests for Populations 1 and 2, including the performance assessment, were administered in late 1994, while the Population 3 instruments were given in mid- to late 1995. The countries in South America and South Africa administered their tests for multiple populations on a schedule similar to that for Population 3 in the Southern Hemisphere. One training session, focusing

solely on administering the performance assessment, was held in Slovenia in December 1994 for countries that were doing the assessment before their scheduled regional training session.

In general, resources for TIMSS, both within countries and overall, precluded having training sessions devoted only to test administration. Yet separating training for administration and scoring activities would benefit future international assessments. First, it would enable more rigor in training for test administration and could include procedures for the main survey as well as for special components like the performance assessment. Perhaps even more important, the training for scoring could be conducted at a time that would improve those procedures. It is best to conduct scoring training after data collection has begun. The training materials are thus based on responses to the final test items incorporating all of the revisions. Also, training closer in time to the actual scoring process means that the information is fresh in the minds of the scorers.

**TABLE 10.1 TIMSS Training Sessions: Free-Response Item Coding and Performance Assessment Administration**

<b>Location</b>	<b>Dates</b>
Wellington, New Zealand (Populations 1 and 2)	October 10-12, 1994
Ljubljana, Slovenia (Only PA Administration)	December 18-19, 1994
Hong Kong	January 18-21, 1995
Boston, United States	January 25-28, 1995
Enschede, Netherlands	March 7-10, 1995
Budapest, Hungary	March 13-16, 1995
Pretoria, South Africa	July 18-19, 1995
Miami, United States	July 17-18, 1995
Wellington, New Zealand (Population 3)	September 6, 1995
Melbourne, Australia (Population 3)	September 28-29, 1995

#### **10.4 DESCRIPTION OF EACH TRAINING SESSION**

**Wellington, New Zealand.** This first session was attended by 11 representatives, from Australia (3), Korea (1), New Zealand (6), and Singapore (1). The training team included Ina Mullis, Robert Garden, and Graham Orpwood. The session was designed for countries on a school schedule necessitating the administration of Population 1 and 2 instruments in late 1994 with Population 3 administration to follow in 1995. Therefore, it did not include training for Population 3 items and was three days long rather than four. These countries either had administered the Population 1 and 2 tests, including the performance assessment, or were about to do so. The exception was Korea, which did not participate in the performance assessment.

Because the coding schemes had not yet been applied in countries, participants at the New Zealand training session were able to make an important contribution to determining how they were organized. All representatives had participated extensively in the TIMSS field tests and were familiar with the materials, approaches to free-response coding, and how to administer the performance assessment tasks.

---

For this initial session, training materials were prepared for all but the most straightforward items. That is, for each item, participants were given a coding guide and from about 10 to 20 example student responses, depending on the complexity of the question and the number of codes involved. The papers had been given preliminary codes to insure a range of example answers. The participants at the New Zealand session worked through the guide for each item and scored the example responses. In striving for reliable coding for all the guides, they made many clarifications and refinements in the guides for Populations 1 and 2, including both the main survey and the performance assessment.

Robin Caygill from New Zealand presented the performance assessment equipment being used in New Zealand and the group reviewed the *Performance Assessment Administration Manual for the Main Survey* (TIMSS, 1994b). Because the group was so familiar with the performance assessment materials, there was no real need to “train” participants in administrative procedures. However, their review of the materials was enormously productive. The TIMSS International Study Center is very grateful for the thoughtful work accomplished at the New Zealand session.

**Ljubljana, Slovenia.** The session in Slovenia dealt only with training for administering the performance assessment. The session was attended by 11 representatives, from Norway (1), Austria (1), Iceland (1), Czech Republic (2), and Slovenia (6). It was designed particularly for countries that were beginning performance assessment administration before the main survey. Graham Orpwood served as the trainer, and the representatives from Norway and Slovenia both had their performance assessment equipment available for the group to use. The participants at this session, especially those not involved in developing and field testing the performance assessment tasks, found two days of discussion about these complex administration procedures to be very helpful.

**Hong Kong.** Designed for countries in the Asian region, the Hong Kong session was attended by 16 representatives, from Hong Kong (12), Japan (1), the Philippines (1), and Thailand (2). The training team included Chancey Jones, Robert Garden, Graham Orpwood, and Jan Lokan.

The session began with an orientation covering the importance of coding the free-response questions and performance tasks. Topics included the need to maintain high reliability in coding, the importance of conducting similar training in the participants’ own countries, and the necessity of finding exemplars within their countries to use in the training process. The remainder of the first day was devoted to the performance assessment. The Australian materials and equipment for each of the performance assessment tasks were set up for demonstration and discussion purposes. Jan Lokan described the equipment necessary for each task and gave advice about how to conduct the administration. Also, the training team worked with participants on coding approaches and practiced coding for several of the performance assessment tasks.

The second day began with a review of questions raised by the participants concerning coding procedures in general, and the significance of the first and second digits used to code

the free-response questions and performance tasks. The rest of the second day was devoted to training on Population 1 mathematics and science free-response items. Day 3 was spent primarily on training for the Population 2 free-response items, although at the conclusion of the day there was a discussion of the procedures to be followed for planning, organizing, and implementing a successful coding endeavor. This session covered the crucial nature of training materials, including exemplar student responses, the importance of subject-matter expertise in coding the Population 3 specialist items, and effective ways to organize staff to do the scoring (including information about table leaders and backreading procedures). Procedures for implementing the within-country reliability studies were discussed, and the vital need to maintain high reliability was again emphasized. Participants were told that the most important factor in coding student responses is that codes be applied accurately and consistently. Although speed is desirable, accuracy and consistency should not be sacrificed. Coders must be encouraged to follow the manual at all times. The fourth day was dedicated to training for Population 3, although Hong Kong was the only country at the training session with plans to participate in Population 3 testing.

**Boston, United States.** The session in Boston was attended by 12 representatives, from the United States (5), Canada (4), Mexico (1), Norway (1), and Kuwait (1). The Boston session tended to parallel that in Hong Kong. However, it was decided that beginning with this session, it was preferable to devote the last rather than the first day to the performance assessment. All countries needed to participate in the training for Population 2, but only some in the training for Populations 1 and 3 and the performance assessment. In an attempt to arrange the most convenient schedule for the most countries, the performance assessment had been placed first. This had been convenient, but it was a difficult initiation into TIMSS scoring procedures. Therefore, it was decided to begin with Population 1, follow with Populations 2 and 3, and conclude with the performance assessment on Day 4.

Chancey Jones opened the session by providing an orientation to the TIMSS scoring approach and the training session itself. During the next three days, he and Robert Garden conducted training for the mathematics items, and Graham Orpwood for the science items; and Ina Mullis discussed procedures for doing the actual coding (as described in the *Guide to Checking, Coding, and Entering the TIMSS Data* (TIMSS, 1995c). In general, Day 1 was devoted to Population 1, Day 2 to Population 2, and Day 3 to Population 3. The performance assessment training took place on the fourth day using the equipment and materials from the United States. Maryellen Harmon, who coordinated development of the performance assessment tasks for the International Study Center, presented and discussed techniques for administering the tasks. Graham Orpwood and Robert Garden provided training on the science and mathematics performance tasks, respectively.

**Enschede, Netherlands.** With 28 participants, the session in Enschede was the largest. It was attended by representatives from Belgium (Flemish) (1), Denmark (1), England (1), France (2), Germany (1), Greece (1), Indonesia (2), Iran (1), Ireland (1), the Netherlands (4), Portugal (2), Scotland (1), Spain (2), Sweden (3), and Switzerland (5). The complete

---

training team was in attendance: Chancey Jones, Robert Garden, Graham Orpwood, Jan Lokan, and Ina Mullis.

Since beginning with the free-response scoring for Populations 1, 2, and 3 and then moving to the performance assessment training worked well during the Boston session, this order was followed also in the Enschede and Budapest sessions. Thus, the Enschede session began with an orientation to the TIMSS approach to coding the free-response items and the importance of coding reliably. This was followed by training for Population 1. The second day was devoted to Population 2 training and some discussion of procedures for coding and conducting the within-country reliability study. Day 3 was dedicated to training for Population 3, both the literacy and specialist components. On Day 4, Jan Lokan led a demonstration on administering the performance assessment tasks using the Australian equipment. This was followed by training in free-response coding for the performance assessment.

**Budapest, Hungary.** Representatives from the following 16 countries took part in the training session held in Budapest: Austria (1), Bulgaria (1), Canada (2), Cyprus (1), Czech Republic (2), Hungary (3), Iceland (1), Israel (1), Latvia (1), Lithuania (1), Norway (1), Romania (1), Russia (1), Slovak Republic (2), Slovenia (1), and the Ukraine (1). The training for the 21 participants was conducted by Chancey Jones, Robert Garden, Graham Orpwood, and Jan Lokan.

The first day followed the agenda of the Boston and Enschede sessions. After a brief orientation to free-response coding for TIMSS, the team reviewed the goals of the training session: to instruct the participants in the nature and volume of coding, to model procedures for training staff to apply the free-response codes reliably and efficiently, and to discuss staff requirements and facilities needed for successful free-response coding. The greater part of Day 1 was spent in training for Population 1.

On Day 2, it was decided to include the coding of practice examples of both mathematics literacy and science literacy for Population 3. This provided time on Day 3 to cope with the complexity of Population 3 coding for the advanced mathematics and physics items. Since some items are part of both Population 2 testing and the literacy assessment for Population 3, this change in schedule worked well at the Budapest session, where most countries were participating in both Population 2 and 3 testing. The science training for Population 2 and Population 3 literacy was followed by the mathematics training for Population 2 and Population 3 literacy. These were followed (as in earlier sessions) by the discussion of guidelines for successful coding within countries. Day 3 was devoted to training for the advanced mathematics and physics items. The extra time gained permitted discussion of additional mathematics questions that were not part of the subset used for practice coding during the training. In response to requests from the participants, training for the performance assessment was begun in an early morning session on Day 3 and concluded on Day 4.

**Pretoria, South Africa.** South Africa participated in TIMSS on the schedule for Southern Hemisphere countries testing Population 3, but also tested Population 2. Therefore, a special training session was held to provide training for Population 2 testing and for the Population 3 literacy free-response items. (South Africa did not participate in the specialist testing for Population 3.) There were 24 participants, all from South Africa. Robert Garden led the training session, which covered most of the mathematics and science free-response items for Population 2 and the literacy portion for Population 3. Because the Population 3 specialist tests did not need to be covered, there was additional time for covering the items relevant to South Africa. South Africa provided financial support for this training session.

**Miami, United States.** Like the session in South Africa, this training session was for the South American countries—Colombia and Argentina (2 and 3 representatives respectively)—that also participated on the Southern Hemisphere schedule for Population 3. Both of these countries participated only at Population 2, but for both the main survey and the performance assessment. Ina Mullis and Eugenio Gonzalez from the TIMSS International Study Center led the training session. One day was devoted to coding training for Population 2 mathematics and science for the main survey, and the second day to the performance assessment. Although some discussion was held about administering the latter, both countries had participated in the pilot, already had arranged for their equipment, and felt comfortable about administration procedures. Thus, training on the second day focused mainly on procedures for coding the performance assessment responses.

**Wellington, New Zealand, and Melbourne, Australia.** These two training sessions were for the two Southern Hemisphere countries — Australia and New Zealand — testing Population 3. Both sessions were led by Robert Garden. Because New Zealand participated only in the literacy testing for Population 3, that training took only one day. It was held on September 6, 1995. As Australia administered both the literacy and specialist tests, that training was held across two days with the assistance of Dr. Jan Lokan and Dr. John Lindsey, both of the Australian Council for Educational Research. It was held September 28-29, 1995. During the two days, the time for coding training was divided about equally across physics, advanced mathematics, and literacy. In contrast to the usual approach, for both the New Zealand and Australian sessions the training was held for the actual coders.

## 10.5 THE TRAINING MATERIALS

Each participant in the training sessions needed a considerable amount of material, including the relevant manuals and packets of example papers for practice. The participants were asked to bring their own copies of the following manuals as pertinent to their participation status:

- *Coding Guide for Performance Assessment (TIMSS, 1994a)*

- 
- *Coding Guide for Free-Response Items–Populations 1 and 2 (TIMSS, 1995a)*
  - *Coding Guide for Free-Response Items–Population 3 (in three sections: Literacy Guide, Physics Guide, Mathematics Guide) (TIMSS, 1995b)*
  - *Guide to Checking, Coding, and Entering the TIMSS Data (TIMSS, 1995c)*
  - *Performance Assessment Administration Manual (TIMSS, 1994b).*

Each coding guide contained the rubrics developed for each of the TIMSS free-response items. For the main survey, each coding category within a rubric also contained some example student responses—as part of the rubric itself, or by following the rubric with some actual student responses, or both. For the performance assessment, a separate document containing examples of coded student responses, entitled the *Supplement to the Performance Assessment Coding Guide with Student Examples (TIMSS, 1995d)*, was sent to the countries after training, but before the actual coding effort began.

For the initial training session in New Zealand, the training materials were by necessity based on field-test materials. For the remaining sessions, however, the training materials for Populations 1 and 2 were based on actual test papers from the Southern Hemisphere countries that administered the tests in English: Australia, Hong Kong, and New Zealand. For the literacy and specialist tests for Population 3, again by necessity, the training materials were based on field-test materials. This problem was somewhat alleviated because several countries held a late second round of field testing of revisions to the specialist materials. Still, everything considered, trying to assemble training materials before actual testing was an enormous undertaking and is not recommended. It is better to train for scoring after testing has begun, so that the training materials can be based on actual test papers reflecting the final wording of the test items.

Training materials were prepared for the subset of items shown in Tables 10.2 and 10.3 for mathematics and science, respectively. The purpose was not to conduct the actual training for the coders, but to present a model for use in each country and an opportunity to practice with the most difficult items.

**Table 10.2 Mathematics Items For Free-Response Training Sessions**

<b>Population 1 Mathematics</b>	
S1	Graph of Numbers of Boys and Girls
T4	Girl Boy Ratio
V4	Game with Cards
<b>Population 2 Mathematics</b>	
T1	Apples in Box
U1	Estimate Time Songs (also, Population 3 literacy)
U2	Draw Rectangle, Explain Ratio
<b>Population 3 Mathematics Literacy</b>	
A12	Which Apartment Cheaper (also, Population 2)
A8	Graph of CD's
<b>Population 3 Mathematics Specialist</b>	
J19	Quadrilateral - Prove E Midpoint
K12	Coordinates of B'
K13	Bacteria in Colony
L15	Crickets (Template)
L16	Real Values of X Satisfy Equation
<b>Performance Task</b>	
M2	Calculator

**Table 10.3 Science Items for Free-Response Training Sessions**

<b>Population 1 Science</b>	
Q4	Glass Jar Over Lighted Candle (also, Population 2)
R1	Watering Can (also, Population 2)
W5	Reducing Air Pollution
X1	Soup Cooling
X3	Oil Spills
Y1	Sun and Moon (also, Population 2)
Z3	Weights of Blocks
<b>Population 2 Science</b>	
L18	Juanita's Experiment
K10	How Air Exists
O16	Thirsty on a Hot Day (also, Populations 1 and 3 literacy)
O17	Jose's Influenza (also, Population 3 literacy)
P2	Flashlight on the Wall
R4	Ozone Level
W2	Rain from Another Place (also, Population 3 literacy)
<b>Population 3 Science Literacy</b>	
A7	High Heels
A11	Painting the Bridge (also, Population 2)
<b>Population 3 Physics Specialist</b>	
F17	Value of Gravity and Uncertainty
G12	Collision Railway Trucks
G15	Acceleration Arrows Bouncing Ball
G18	Alpha Particles through Gold Sheet
H16	Expression Speed of Electron
<b>Performance Tasks (Population 2 Version)</b>	
SM1	Shadows
S1	Pulse

For each item selected for training, a packet of materials was prepared for each participant in the training session. This packet began with coded responses illustrating each of the categories in the rubric or guide for that item. These served as a basis of discussion to familiarize the participants with the rubric. The trainers presented the reasons for each of the assigned codes and answered any questions.

The packet also contained about 15 to 20 precoded student responses, with the codes known to the training team but not to the session participants. The trainer for the item would first invite participants to code five or six of these student responses. After the

coding had been completed, the trainer would read the scores and answer any questions from the group. This procedure was iterated until the group had scored all the responses. For variety, sometimes the participants took turns in reading out their scores. Although generally there was insufficient time at the training sessions to achieve a high degree of agreement on all items, the procedure provided some practice for participants and an example for how training might be conducted in each of their countries. The trainers emphasized the need for each country to prepare training materials for each item rather than for only a sample of items, and pointed out that for more difficult items more responses might be needed to help coders reach a high degree of reliability. The trainers also recommended that the training materials used in each country be based largely on student responses from that country.

## **10.6 CONCLUDING REMARKS**

The participants in the training sessions exhibited enthusiasm, patience, understanding, and humor in successfully completing the intense and demanding training. Most of them took part in the training for two populations, while those involved in all three populations and the performance assessment attended all the sessions during the four-day training. The training activity highlighted the complexity of the TIMSS coding process, especially for Population 3 and for the performance assessment tasks. In general, future studies should consider a more rigorous process both for deciding which codes to apply internationally and for assigning the codes to the example responses used in the coding guides and training materials. More specifically, the coding guides should be developed as an integral part of item development and modified as necessary throughout the process, particularly in light of actual student responses. The example student responses should be considered to be part of each coding guide. Particular attention should be paid to the suitability of an item for such elaborate coding.

Although demanding and intense, the four-day training period was appropriate for most participants. The difficulty was trying to fit so much material into the four days. Considering the many aspects of TIMSS, perhaps extra sessions should have been held for participants who were to be trained in how to code responses to the advanced mathematics and physics items. Or perhaps other configurations of the training sessions might have helped to ease the burden for countries participating in all aspects of TIMSS.

All in all, however, the participants in the sessions, the host countries, the staff at the International Study Center at Boston College, and the training team are to be commended. Their planning coordination, good will, patience, and support were instrumental to the success of the TIMSS training endeavor.

---

## REFERENCES

- Third International Mathematics and Science Study (TIMSS). (1994a).** *Coding Guide for Performance Assessment* (Doc. Ref.: ICC885/NRC422). Chestnut Hill, MA: Boston College.
- Third International Mathematics and Science Study (TIMSS). (1994b).** *Performance Assessment Administration Manual for the Main Survey* (Doc. Ref.: ICC884/NRC421). Chestnut Hill, MA: Boston College.
- Third International Mathematics and Science Study (TIMSS). (1995a).** *Coding Guide for Free-Response Items—Populations 1 and 2* (Doc. Ref.: ICC897/NRC433). Chestnut Hill, MA: Boston College.
- Third International Mathematics and Science Study (TIMSS). (1995b).** *Coding Guide for Free-Response Items—Population 3* (Doc. Ref.: ICC913/NRC446). Chestnut Hill, MA: Boston College.
- Third International Mathematics and Science Study (TIMSS). (1995c).** *Guide to Checking, Coding, and Entering the TIMSS Data* (Doc. Ref.: ICC918/NRC449). Chestnut Hill, MA: Boston College.
- Third International Mathematics and Science Study (TIMSS). (1995d).** *Supplement to the Performance Assessment Coding Guide with Student Examples* (Doc. Ref.: ICC933/NRC456). Chestnut Hill, MA: Boston College.