TIMSS 2007 U.S. Technical Report and User Guide

September 2009

Trevor Williams
David Ferraro
Stephen Roey
Summer Brenwald
David Kastberg
Leslie Jocelyn
Westat

Connie Smith
Pat Stearns
Pearson

Patrick Gonzales
Project Officer
National Center for Education Statistics

NCES 2009-012
U.S. Department of Education

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Easton
Director

National Center for Education Statistics
Stuart Kerachsky
Acting Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other countries. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public. Unless specifically noted, all information contained herein is in the public domain.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

September 2009

The NCES World Wide Web Home Page address is http://nces.ed.gov.
The NCES World Wide Web Electronic Catalog is http://nces.ed.gov/pubsearch.

Suggested Citation
Williams, T., Ferraro, D., Roey, S., Brenwald, S., Kastberg, D., Jocelyn, L., Smith, C., and Stearns, P. (2009). *TIMSS 2007 U.S. Technical Report and User Guide* (NCES 2009–012). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

For ordering information on this report, write to

U.S. Department of Education
ED Pubs
P.O. Box 1398
Jessup, MD 20794-1398

or call toll free 1-877-4ED-Pubs or order online at http://www.edpubs.org.

Content Contact
Patrick Gonzales
(415) 920-9229 or patrick.gonzales@ed.gov

Contents

Contents (continued)

Contents (continued)

Contents (continued)

Contents (continued)

List of Appendixes

List of Tables

Contents (continued)

List of Tables

Contents (continued)

List of Tables

Contents (continued)

List of Tables

List of Exhibits

Contents (continued)

List of Exhibits

1. Introduction

The *U.S. TIMSS 2007 Technical Report and User Guide* provides an overview of the design and implementation of the Trends in International Mathematics and Science Study (TIMSS) 2007 in the United States, along with information designed to facilitate access to the U.S. TIMSS 2007 data.

1.1 TIMSS in brief

The Trends in International Mathematics and Science Study (TIMSS) 2007 is the fourth time since 1995 that this international comparison of student achievement has been conducted. Developed and implemented at the international level by the International Association for the Evaluation of Educational Achievement (IEA), an international organization of national research institutions and governmental research agencies, TIMSS is used to measure trends in the mathematics and science knowledge and skills of fourth- and eighth-graders.

TIMSS is designed to align broadly with mathematics and science curricula in the participating countries. The results, therefore, suggest the degree to which students have learned mathematics and science concepts and skills likely to have been taught in school. TIMSS also collects background information on students, teachers, and schools to allow cross-national comparisons of educational contexts related to student achievement. In 2007, 58 countries and educational jurisdictions participated in TIMSS at the fourth- or eighth-grade level, or both.

A detailed treatment of TIMSS 2007 from an international perspective can be found in three reports published by the IEA and available online at http://timssandpirls.bc.edu/TIMSS2007.

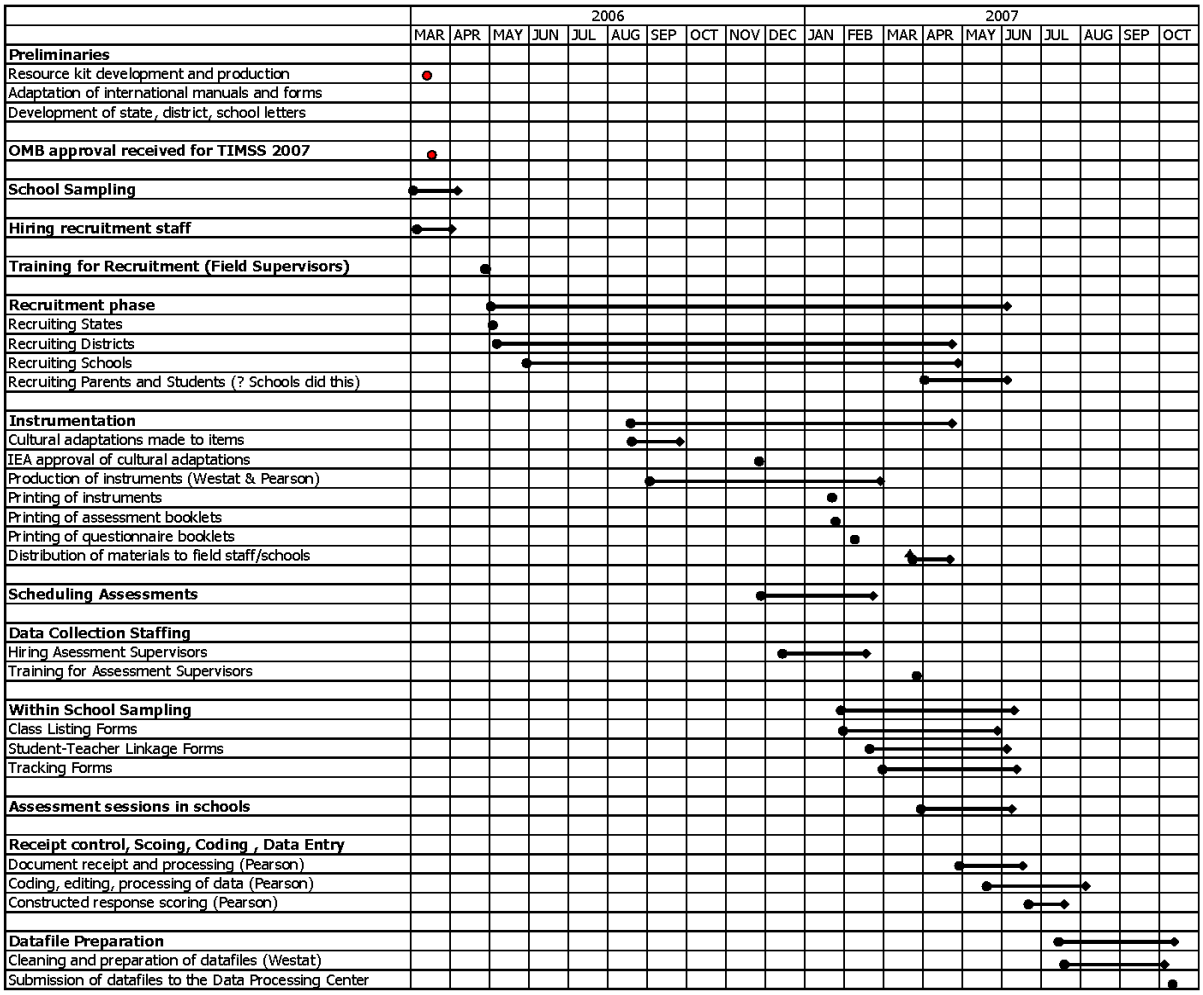
- *TIMSS 2007 International Mathematics Report: Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades (Mullis et al. 2008);*
- *TIMSS 2007 International Science Report: Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades (Martin et al. 2008); and*
- *TIMSS 2007 Technical Report (Olson, Martin, and Mullis 2008).*

A U.S. national report is available as well. It is titled *Highlights From TIMSS 2007: Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context* (NCES 2009–001). (Gonzales et al. 2008). This report describes the performance of U.S. students relative to their peers in other countries, changes in mathematics and science achievement since 1995, and additional details about the achievement of U.S. students that are not available in the international reports. The report is available online at http://nces.ed.gov/Pubsearch/pubsinfo.asp?pubid=2009001.

1.2 TIMSS 2007 U.S. data collection activities and schedule

Descriptions of data collection activities and their timing within the U.S. provide a foundation for researchers seeking to understand the detail of the data. These activities are listed in exhibit 1.1 along with the timing of their implementation. The activities are described in detail in chapter 4 of this report.

Exhibit 1.1. Schedule for U.S. TIMSS 2007 data collection activities



1.3 Overview of the design and administration of TIMSS 2007

The basic parameters of the design and administration of TIMSS 2007 in the United States are outlined below. A more detailed treatment is provided in subsequent chapters of this report. The *TIMSS 2007 Technical Report* (Olson, Martin, and Mullis 2008) provides a basic reference for details of the international design.

Sampling. Participating countries administered TIMSS to national probability samples of students and schools. Countries were required to draw samples of students who were nearing the end of their fourth year or eighth year of formal schooling. Formal schooling was defined as beginning with Level 1 as defined in the International Standard Classification of Education—ISCED (OECD 1999). In most countries, including the United States, these students were in the fourth and eighth grades. The U.S. sample included both public and private schools, randomly selected and weighted to be representative of the nation. In total, 257 fourth-grade U.S. schools and 7,896 fourth-grade students, along with 239 eighth-grade schools and 7,377 eighth-grade students participated in TIMSS 2007 at grade eight. Detailed information on sampling, administration, response rates, and other technical issues is provided in the chapters which follow.

Assessment design. Data for TIMSS 2007 were collected through pencil-and-paper assessments administered to the students and through questionnaires completed by principals, teachers, and students. To keep the testing burden to a minimum, and to ensure broad subject-matter coverage, TIMSS used a rotated block design that included both mathematics and science items in each assessment booklet. The booklets were constructed in such a way that no student would have the opportunity to respond to all of the items. This is consistent with other large-scale assessments, such as the National Assessment of Educational Progress (NAEP).

Test administration. Test administration in the United States occurred over the period March through June 2007. The administration was carried out by professional staff trained according to the international guidelines. School personnel were asked only to assist with listings of students, the identification of space for testing in the school, and the specification of any parental consent procedures required. The International Study Center monitored compliance with the standardized procedures. The *TIMSS and PIRLS International Study Center*, located at Boston College is the arm of IEA responsible for the conceptual development and implementation of studies in the areas of mathematics, science, and reading.

Scoring. The TIMSS assessment items included both multiple-choice and constructed-response items. A scoring rubric (guide) was provided for every constructed response item. The national research coordinator in each country was responsible for the scoring and coding of data in that country, following established guidelines.

Scaling. Total scores for mathematics and science, along with scores that reflect performance in specific domains of mathematics and science, were estimated using an item response theory (IRT) model. For example, the TIMSS 2007 eighth-grade assessment had four scales describing four mathematics content areas and four science content areas, as well as three cognitive domains in each of mathematics and science. Benchmark scores were also derived. IRT estimation procedures were also used to place scores from the four TIMSS assessments conducted in 1995, 1999, 2003 and 2007 on the same scale (the scale of the 1995 administration). This allows for the calculation of trends in achievement even though the makeup of the countries participating in TIMSS changed over time. Details are provided in Olson, Martin, and Mullis (2008).

Plausible values. The rotated block design used in the TIMSS assessment meant that no student responded to all of the items. To accommodate the missing data generated by this design, during the scaling process plausible values were estimated to characterize students participating in the assessment. Plausible values are imputed values and not test scores for individuals in the usual sense. They represent what the true performance of an individual might have been, had it been observed. They are estimated as random draws (five, in the case of TIMSS) from an empirically derived distribution of score values based on the student's observed responses to assessment items and on background variables. A more technical treatment can be found in Olson, Martin, and Mullis (2008). From the point of view of analysis this means that special statistical procedures need to be applied to estimate results and uncertainty around results (see chapter 5).

Weighting. Responses from the groups of students were assigned sampling weights to adjust for the complex sample design which resulted in students having an unequal, but known, probability of selection. As well, an adjustment for school and student nonresponse was built into the weighting. The estimation of sampling weights was carried out by Statistics Canada. A detailed description is provided in Olson, Martin, and Mullis (2008). In analyses of the TIMSS data it is necessary to use sampling weights to obtain accurate population estimates.

1.4 Reporting TIMSS results

Achievement results from TIMSS are reported on a scale from 0 to 1,000, with a TIMSS scale average of 500 and standard deviation of 100. Even though the countries participating in TIMSS have changed across the four assessments between 1995 and 2007, comparisons between the 2007 results and prior results are still possible because the achievement scores in each of the TIMSS assessments are placed on a scale which is not dependent on the list of participating countries in any particular year. A brief description of the assessment equating and scaling is presented in appendix A to this volume. A more detailed presentation can be found in Olson, Martin, and Mullis (2008).

In addition to numerical scale results, TIMSS also includes international benchmarks. The TIMSS international benchmarks provide a way to interpret the scale scores and to understand how students' proficiency in mathematics and science varies along the TIMSS scale. The TIMSS benchmarks describe four levels of student achievement in each subject, based on the kinds of skills and knowledge students at each score cutpoint would need to successfully answer the mathematics and science items. More information on the development of the benchmarks and the procedures used to set the score cutpoints can be found in Olson, Martin, and Mullis (2008).

1.5 U.S. international and national datasets

U.S. data are available through three avenues:

1. The "U.S. international dataset" that is available as part of the international database released by the TIMSS and PIRLS International Study Center. This file can be downloaded at http://timssandpirls.bc.edu/TIMSS2007/idb_ug.html. This U.S. datafile conforms to the international specifications common to the datafiles from all countries. However, it does not include the U.S.-specific adaptations made to a few questions in the questionnaires or the additional questions added to the school and student questionnaires; for example, the question on race/ethnicity added to the student questionnaire.

2. The "U.S. national public-use dataset" that is available through the National Center for Education Statistics. This file can be downloaded from http://nces.ed.gov/timss or can be obtained on a CD-ROM along with an electronic codebook by ordering online at http://nces.ed.gov/timss. This U.S. datafile includes the U.S.-specific adaptations made to the questionnaire item and the additional questions added to the school and student questionnaires.

3. The "U.S. national restricted-use dataset" that is only available from the National Center for Education Statistics as a CD-ROM. This CD-ROM contains all of the information described above plus supplemental link-files which link TIMSS school ID numbers to the ID numbers of schools as these appear in the publicly available Common Core of Data (CCD) or the Private School Universe Survey (PSS). The supplemental link files allow a user to merge school-level data from the CDD and PSS with TIMSS school data and, subsequently, with other TIMSS teacher and student data. Since school names are revealed in this process the data require a restricted-use license (for information on how to obtain a license, see http://nces.ed.gov/pubsearch/licenses.asp).

The most comprehensive treatment of the TIMSS international data and hence, of the U.S. international dataset, is provided in the various TIMSS 2007 publications produced by the International Association for the Evaluation of Educational Achievement (IEA), particularly the *TIMSS 2007 User Guide for the International Database* (Foy and Olson 2009). This publication should be seen as the primary reference. The *TIMSS 2007 U.S. Technical Report and User Guide* draws heavily on the international user guide for much of its datafile-related content. This content is supplemented with detail on those aspects of the TIMSS 2007 data that were unique to the United States.

2. Sampling Schools

The 2007 TIMSS school samples were drawn for the United States in November, 2005. The sample design for this school sample was developed to retain most of the properties of the 2003 TIMSS U.S. school sample, and to follow international requirements as given in the TIMSS sampling manual. Like the 2003 TIMSS sample, the U.S. sample followed a two-stage sampling process with the first stage a sample of schools, and the second stage a sample of classrooms within schools. All students in sampled classrooms were selected for assessment. As in 2003, the school sample was not clustered for the 2007 TIMSS.

Two samples were selected for the 2007 TIMSS, one for fourth grade and one for eighth grade. The sample designs for both grades were the same. Unlike the 2003 TIMSS sample, in which low-income schools were oversampled in the fourth grade sample, there was no oversampling for either grade in 2007. The overlap with the 2007 National Assessment of Educational Progress (NAEP) school sample was not minimized for the 2007 TIMSS as it was in 2003. This is because the TIMSS sample was selected before the NAEP sample because of TIMSS scheduling constraints.[1] Thus the overlap between the samples was minimized when the 2007 NAEP sample was selected.

The student population for the fourth grade 2007 TIMSS is the set of all fourth graders in the United States in both public and private schools. The TIMSS school sample consists of 300 schools containing a fourth grade class. The schools were selected with probability proportionate to the school's estimated grade enrollment of fourth graders from the 2006 NAEP school frame using 2003-04 school data. A total of two fourth grade mathematics classes were selected within each school in an equal probability sample (unless there were only one or two classes, in which case all fourth grade classes were taken with certainty). The overall sample design was intended to approximate a self-weighting sample of students as much as possible in accordance with the international guidelines, with each fourth grade student in the United States having an equal probability of being selected.

The student population for the eighth grade 2007 TIMSS is the set of all eighth graders in the United States in both public and private schools. The TIMSS school sample consists of 300 schools containing an eighth grade class. The schools were selected with probability proportionate to the school's estimated grade enrollment of eighth graders from the 2006 NAEP school frame using 2003-04 school data. A total of two eighth grade mathematics classes were selected within each school in an equal probability sample (unless there were only one or two classes, in which all eighth grade classes were taken with certainty). The overall sample design was intended to approximate a self-weighting sample of students as much as possible in accordance with the international guidelines, with each eighth grade student in the U.S. having an equal probability of being selected.

2.1 School sampling frame

The school frame for both samples was developed from the 2006 NAEP school frame. The data for public schools was from the 2003-04 Common Core of Data (CCD), and the data for private schools was from the 2003-04 Private School Survey (PSS). For more information on the NAEP school frame, see the NAEP technical reports on the NAEP website at http://nces.ed.gov/nationsreportcard/tdw/. Additional NAEP reports are scheduled for release in the coming months.

[1] In order to maximize response rates from both districts and schools it was necessary to begin the recruitment of both prior to the end of the 2005-06 school year. Since the 2007 NAEP sampling frame was not available until March 2006, it was necessary to base the TIMSS samples on the 2006 NAEP sampling frame.

2.1.1 Fourth grade frame

Any school having a fourth grade as of the school year 2003-04 was included on the fourth grade school sampling frame. Tables 2-1 and 2-2 present frame tabulations of the number of schools by the school grade span (lowest to highest grade level of the school) and school control, respectively.

Table 2-1. Number and percentage of students and schools included in the U.S. TIMSS fourth-grade school sampling frame, by grade span: 2007

Grade span	Students	Percent	Schools	Percent
Total	4,049,655	100.0	72,670	100.0
Grades 1-5	1,915,302	47.3	25,251	34.7
Grades 1-6	808,665	20.0	14,554	20.0
Grades 1-8	443,111	10.9	14,727	20.3
Grades 1-12	124,689	3.1	6,354	8.7
Other	757,888	18.7	11,784	16.2

NOTE: Detail may not sum to totals because of rounding.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 2-2. Number and percentage of students and schools included in the U.S. TIMSS fourth-grade school sampling frame, by school control: 2007

School control	Students	Percent	Schools	Percent
Total	4,049,655	100.0	72,670	100.0
Private	419,553	10.4	21,490	29.6
Public	3,630,102	89.6	51,180	70.4

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

2.1.2 Eighth grade frame

Any school having an eighth grade as of the school year 2003-04 was included on the eighth grade school sampling frame. Tables 2-3 and 2-4 present frame tabulations of the number of schools by the school grade span (lowest to highest grade level of the school) and school control, respectively.

Table 2-3. Number and percentage of students and schools included in the U.S. TIMSS eighth-grade school sampling frame, by grade span: 2007

Grade span	Students	Percent	Schools	Percent
Total	4,219,262	100.0	46,112	100.0
Grades 1 to 8	436,151	10.3	14,727	31.9
Grades 6 to 8	2,150,637	51.0	9,353	20.3
Grades 1 to 12	145,209	3.4	6,354	13.8
Grades 7 to 12	703,286	16.7	2,780	6.0
Grades 7 to 8	174,450	4.1	3,582	7.8
Other	609,529	14.4	9,316	20.2

NOTE: Detail may not sum to totals because of rounding.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 2-4. Number and percentage of students and schools included in the U.S. TIMSS eighth-grade school sampling frame, by school control: 2007

School control	Students	Percent	Schools	Percent
Total	4,219,262	100.0	46,112	100.0
Private	404,861	9.6	18,965	41.1
Public	3,814,401	90.4	27,147	58.9

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

2.2 School sample

2.2.1 Measures of size for school selection

The goal for the TIMSS sample was to attain a self-weighting student sample. To achieve this, schools' probability of selection was related to their measure of size (MOS), which is proportional to its share of the target population, that is, the fourth or eighth grade enrollments. This method also reduces the chance of selection for smaller schools. This improves cost efficiency by increasing the number of students per school. However, students in schools with enrollments of only a few students would have very large weights if selected. To minimize the impact of these small schools on variances and estimates, the minimum measure of size was set to 5.

The following is a summary of the steps for assigning measures of size to the schools on the TIMSS frame.

- Determine the estimated target population size for the school. This is the estimated enrollment per grade (4 or 8) from the school frame. If the grade 4/8 enrollment is not available, it is calculated by dividing the school's total enrollment by the number of grades in the school; and

- Calculate measures of size according to the estimated enrollment per grade as shown:

$$\text{MOS} = \begin{cases} 5 & \text{if grade 4/8 enrollment } <= 5 \\ \text{grade 4/8 enrollment} & \text{otherwise} \end{cases}$$

2.2.2 Fourth grade school sample

The sample design for the fourth grade TIMSS sample was a stratified systematic sample, with sampling probabilities proportional to the measure of size (PPS). The fourth grade sample had no explicit stratification. A sample of 300 schools was drawn from the frame as a single stratum. The frame was implicitly stratified by four categorical stratification variables. They are listed in table 2-5 shown in order of priority. There were a total of 128 implicit strata. The frame was sorted in alternating (serpentine) sort order according to these school characteristics, implicitly stratifying the frame. The last sort within the implicit stratification was by grade enrollment (*MOS*) in descending order. Alternating the sort order sorts the frame from lowest to highest value with respect to the first sort variable, then within each level of the first sort variable, the second sort variable alternates its sort order, from lowest to highest for the first level of the first sort variable, then from highest to lowest for the second level of the first sort variable, then, again, from lowest to highest for the third level of the first sort variable, etc. Each of the variables will alternate the sort order within each level of the preceding sort variable. This means that schools adjacent on the list are not substantially different or at most different by one sorting characteristic.

Table 2-5. U.S. TIMSS implicit stratification variables: 2007

Variable	Variable definition	Number of levels
School control	Public or private	2
Region	Northeast, Southeast, Central, West	4
Location of school relative to populous areas	1 = large city 250,000+ 2 = mid-size city <250,000 3 = urban fringe of large city 4 = urban fringe of mid-size city 5 = large town 25,000+ 6 = small town 2,500-25,000 7 = rural outside CBSA 8 = rural inside CBSA	8
Race/ethnicity status	Above or below 15 percent of Black, non-Hispanic; Hispanic; and American Indian or Alaska Native students	2

NOTE: Region is the state-based region of the country (see section 3.5 for state listing). CBSA stands for Core Based Statistical Area. Black includes African American, and Hispanic includes Latino. Racial categories exclude Hispanic origin.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

2.2.3 Fourth grade tabulations within subgroups for frame and sample

This section provides an overview of the frame and sample for the implicit strata used in the sample process. The PPS sampling and implicit stratification worked effectively: the sample percentage of schools is close to the measure-of-size percentage of the frame for all the implicit strata. For these strata-defining subgroups, tables 2-6 through 2-9 present the following summary tabulations in these subgroups:

- **Total measure of size.** This is the summation of MOS_{ij} over the subgroup. Note that this is larger than the national population student size because the minimum MOS_{ij} is set to 5 for small schools; and

- **Sample size.** This is the final realized sample size of schools in the subgroup for the U.S. TIMSS fourth-grade sample.

Table 2-6. Number and percentage of schools included in the U.S. TIMSS fourth-grade school sampling frame and sample, by school control: 2007

	Frame		Sample	
School control	Measure of size	Percent	Number of schools	Percent
Total	4,049,655	100.0	300	100.0
Private	419,553	10.4	31	10.3
Public	3,630,102	89.6	269	89.7

NOTE: Measure of size is the number of students enrolled in the target grade.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 2-7. Number and percentage of schools included in the U.S. TIMSS fourth-grade school sampling frame and sample, by region: 2007

	Frame		Sample	
Region	Measure of size	Percent	Number of schools	Percent
Total	4,049,655	100.0	300	100.0
Northeast	805,194	19.9	60	20.0
Southeast	967,875	23.9	72	24.0
Central	909,611	22.5	67	22.3
West	1,366,975	33.8	101	33.7

NOTE: Detail may not sum to totals because of rounding. Region is the state-based region of the country (see section 3.5 for state listing). Measure of size is the number of students enrolled in the target grade.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 2-8. Number and percentage of schools included in the U.S. TIMSS fourth-grade school sampling frame and sample, by location of school relative to populous areas: 2007

	Frame		Sample	
Location of school	Measure of size	Percent	Number of schools	Percent
Total	4,049,655	100.0	300	100.0
Large city	688,383	17.0	51	17.0
Mid-size city	608,250	15.0	45	15.0
Urban fringe of large city	1,184,346	29.2	89	29.7
Urban fringe of mid-size city	495,187	12.2	36	12.0
Large town	41,783	1.0	2	0.7
Small town	263,299	6.5	20	6.7
Rural outside CBSA	351,778	8.7	26	8.7
Rural inside CBSA	416,629	10.3	31	10.3

NOTE: Detail may not sum to totals because of rounding. Measure of size is the number of students enrolled in the target grade. CBSA stands for Core Based Statistical Area.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 2-9. Number and percentage of schools included in the U.S. TIMSS fourth-grade school sampling frame and sample, by race/ethnicity status: 2007

	Frame		Sample	
Race/ethnicity status	Measure of size	Percent	Number of schools	Percent
Total	4,049,655	100.0	300	100.0
Above 15 percent	2,377,265	58.7	175	58.3
Below 15 percent	1,672,390	41.3	125	41.7

NOTE: Measure of size is the number of students enrolled in the target grade. Race/ethnicity status is above or below 15 percent of Black, non-Hispanic; Hispanic; and American Indian or Alaska Native students.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

2.2.4 Eighth grade school sample

The sample design for the eighth grade TIMSS sample was a stratified systematic sample, with PPS sampling. The eighth grade sample had no explicit stratification. A sample of 300 schools was drawn from the frame as a single stratum. The same four categorical implicit stratification variables were used as in the fourth grade sample (see table 2-5). The frame was sorted in alternating sort order according to these school characteristics, implicitly stratifying the frame. There were a total of 128 implicit strata. The last sort within the implicit stratification was by grade enrollment (*MOS*) in descending order.

2.2.5 Eighth grade tabulations within subgroups for frame and sample

This section provides an overview of the frame and sample for the implicit strata used in the sampling process. The PPS sampling and implicit stratification worked effectively: the sample percentage of schools is close to the measure-of-size percentage of the frame for all the implicit strata. For these strata-

defining subgroups, tables 2-10 through 2-13 present the following summary tabulations in these subgroups:

- **Total measure of size.** This is the summation of MOS_{ij} over the subgroup. Note that this is larger than the national population size because the minimum *MOS* is 5; and
- **Sample size.** This is the final realized sample size of schools in the subgroup for the U.S. TIMSS eighth-grade sample.

Table 2-10. Number and percentage of schools included in the U.S. TIMSS eighth-grade school sampling frame and sample, by school control: 2007

	Frame		Sample	
School control	Measure of size	Percent	Number of schools	Percent
Total	4,219,262	100.0	300	100.0
Private	404,861	9.6	29	9.7
Public	3,814,401	90.4	271	90.3

NOTE: Measure of size is the number of students enrolled in the target grade.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 2-11. Number and percentage of schools included in the U.S. TIMSS eighth-grade school sampling frame and sample, by region: 2007

	Frame		Sample	
Region	Measure of size	Percent	Number of schools	Percent
Total	4,219,262	100.0	300	100.0
Northeast	852,511	20.2	61	20.3
Southeast	1,015,672	24.1	72	24.0
Central	957,762	22.7	68	22.7
West	1,393,317	33.0	99	33.0

NOTE: Region is the state-based region of the country (see section 3.5 for state listing). Measure of size is the number of students enrolled in the target grade.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 2-12. Number and percentage of schools included in the U.S. TIMSS eighth-grade school sampling frame and sample, by location of school relative to populous areas: 2007

	Frame		Sample	
Location of school	Measure of size	Percent	Number of schools	Percent
Total	4,219,262	100.0	300	100.0
Large city	659,334	15.6	47	15.7
Mid-size city	640,931	15.2	45	15.0
Urban fringe of large city	1,230,633	29.2	88	29.3
Urban fringe of mid-size city	517,138	12.3	37	12.3
Large town	44,187	1.0	2	0.7
Small town	304,770	7.2	22	7.3
Rural outside CBSA	367,823	8.7	27	9.0
Rural inside CBSA	454,446	10.8	32	10.7

NOTE: Measure of size is the number of students enrolled in the target grade. CBSA stands for Core Based Statistical Area.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 2-13. Number and percentage of schools included in the U.S. TIMSS eighth-grade school sampling frame and sample, by race/ethnicity status: 2007

	Frame		Sample	
Race/ethnicity status	Measure of size	Percent	Number of schools	Percent
Total	4,219,262	100.0	300	100.0
Above 15 percent	2,426,797	57.5	176	58.7
Below 15 percent	1,792,465	42.5	124	41.3

NOTE: Measure of size is the number of students enrolled in the target grade. Race/ethnicity status is above or below 15 percent of Black, non-Hispanic; Hispanic; and American Indian or Alaska Native students.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

2.2.6 TIMSS school selection

The sample was systematically selected from the ordered frame. Within each stratum a sampling interval was calculated by dividing the total measure of size by the sample size. A random number between 0 and the sampling interval was generated, and a sequence of numbers was in turn generated by adding integer multiples of the sampling interval to the random number, until the total measure of size was exceeded. For each number in the sequence, the first school with a cumulative measure of size (the summation over the preceding schools in the sequence, inclusive) that met or exceeded that number was selected.

2.3 Selecting substitute schools

Though efforts were made to secure the participation of all schools selected, it was anticipated that not all schools would choose to participate. Therefore, as each school was selected in the sample, the two

neighboring schools in the sampling frame were designated as substitute schools. The first school following the sample school was the first substitute and the first school preceding it was the second substitute. If an original school refused to participate, the first substitute was then contacted. If that school also refused to participate, the second substitute was then contacted. There were several constraints on the assignment of substitutes. One sampled school was not allowed to be a substitute for another, and a given school could not be assigned to be a substitute for more than one sampled school. Furthermore, substitutes were required to be in the implicit stratum as the sampled school. If the sampled school was the first or last school in the stratum, then the second school following or preceding the sampled school was identified as the substitute. There were no restrictions for identifying substitute schools that were also in the NAEP sample. If the first substitute was a NAEP school, the second substitute was to be contacted first to again reduce the burden on the schools. Under these rules, it was possible to identify two substitutes for all sampled schools.

2.4 Selecting classrooms

The final stage of selection was of classrooms within schools. Within each sampled school that agreed to cooperate with 2007 TIMSS all mathematics classrooms in the school were listed on the classroom sampling frame. Classrooms with fewer than 15 students were collapsed into pseudo-classrooms, so that each classroom on the school's classroom sampling frame had at least 20 students. An equal probability sample of two classrooms or pseudo-classrooms was sampled from the classroom frame for each school. All students in sampled classrooms (pseudo-classrooms) were selected for assessment.

Page intentionally left blank.

3. Participation Rates and Nonresponse Bias

In order to minimize the potential for response biases, IEA developed participation or response rate standards that apply to all participating jurisdictions and govern whether or not data are included in the TIMSS 2007 international dataset and the way in which aggregate statistics are presented in the international reports. These standards were set using composites of participation rates at the school, classroom, and student levels, and were calculated with and without the inclusion of substitute schools that were selected to replace schools refusing to participate.

The standards take the following two forms, distinguished primarily by whether or not meeting the school participation rate of 85 percent requires the counting of substitute schools.

Category 1: Met requirements. Jurisdictions that meet all of the following conditions are considered to have fulfilled the IEA requirements: (a) a minimum school participation rate of 85 percent, based on original sampled schools only; (b) a minimum classroom participation rate of 95 percent, from both original and substitute schools; and (c) a minimum student participation rate of 85 percent, from both original and substitute schools. Participation rates are weighted rates.

Category 2: Met requirements after substitutes. Jurisdictions are considered to have met requirements if the following requirements are met and provided that at least 50 percent of schools in the original sample participate: a minimum combined school, classroom and student (weighted) participation rate of 75 percent, based on the product of the participation rates described above. That is, the product of (a), (b) and (c), as defined in the category 1 standard, must be greater than or equal to 75 percent. Substitute schools are included in both the numerator and denominator of the school participation rate calculation.

Countries satisfying the category 1 standard are included in the international tabular presentations without annotation. Those only able to satisfy the category 2 standard are included as well but are annotated to indicate their participation rate status. In general, the data from jurisdictions failing to meet either standard are usually presented separately in the international tabular displays. Sometimes, though, an additional category is used to describe countries which “nearly satisfied guidelines for sample participation rates only after replacement schools were included”; see, for example, Mullis, Martin, and Foy (2008, p. 34).

3.1 Participation rates of U.S. schools, classrooms, and students

The raw numbers on which the various participation rates are based, along with the participation rates themselves, are shown in table 3.1 separately for the fourth-grade and eighth-grade samples. They are interpreted as follows, using the fourth-grade numbers to provide a detailed example.

3.1.1 Schools

The fourth-grade school sample consisted of 300 schools. Ten ineligible schools were identified on the basis that they served special student populations, or had closed or altered their grade makeup since the sampling frame was developed. This left 290 schools eligible to participate, of which 202 agreed to do so. The fourth-grade school participation rate before substitution was 70 percent (unweighted). The analogous weighted school participation rate was also 70 percent.

In addition to the 202 participating schools from the original sample, 55 substitute schools also participated for a total of 257 participating schools at the fourth grade in the United States. This gave a weighted (and unweighted) school participation rate after substitution of 89 percent.

Table 3-1. Number of U.S. schools, classrooms and students participating in TIMSS, and participation rates, by grade: 2007

	Numbers		Rates			
	Grade 4	Grade 8	Grade 4		Grade 8	
			Unweighted	Weighted	Unweighted	Weighted
Schools						
Sampled	300	300				
Excluded	10	13				
Eligible	290	287				
Participating	202	197	70	70	69	68
Substitutes	55	42				
Participating (all schools)	257	239	89	89	83	83
Classrooms in participating schools						
Total	1,108	3,125				
Excluded	99	350				
Eligible	1,009	2,775				
Sampled	521	539				
Participating	521	536	100	100	99	99
Students in participating schools						
Bridge study	2,454	2,346				
TIMSS 2007						
Sampled	9,000	8,447				
Excluded	543	272				
Withdrawn	140	202				
Eligible	8,317	7,973				
Absent	421	596				
Assessed	7,896	7,377	95	95	93	93

NOTE: NCES standards (Standard 1-3-8) indicate that participation rates should be calculated without including substitute schools since substitute schools do not have an independent probability of selection (National Center for Education Statistics 2002). However, the participation rates shown in this table are those reported by TIMSS and do include substitute schools in the calculations.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

3.1.2 Classrooms

In accord with the international requirements, schools agreeing to participate were asked to list their fourth grade mathematics classes as the basis for sampling at the classroom level. Schools appeared to be able to identify classes in this way without any problems. A total of 1,108 mathematics classrooms were identified as a result. At this time schools, were given the opportunity to identify special classes—classes containing all, or a majority of, students with intellectual and/or functional disabilities, or students who were non-native-language speakers. While these classes were regarded as eligible, the students as a group were treated as ”excluded” since, in the opinion of the school, their disabilities or language capabilities would render meaningless their performance on the assessment. Some 876 fourth-grade students in a total of 99 classrooms in 63 schools were excluded in this way. This left a pool of 1,009 eligible classrooms from which the sample was drawn. While the students in these excluded classrooms

did not figure in the participation rate calculations, they did count in the population coverage calculations and this is reflected in the higher exclusion rate for the U.S. In the international report the U.S. is annotated to reflect this fact.

Classrooms with fewer than 15 students were collapsed into pseudo-classrooms prior to sampling so that each eligible classroom in a school had at least 20 students. Two classrooms (pseudo-classrooms) were selected per school where possible. In schools where there was only one classroom, this classroom was selected with certainty. Some 521 classrooms were selected as a result of this process and all participated in TIMSS. Weighted and unweighted classroom participation rates were 100 percent.

Subsequently, schools were asked to list the students in each of the 521 sampled classrooms at the fourth grade, along with the teachers who taught mathematics and science to these students. At this time, schools were given the opportunity to identify particular students not suited to take the test because of functional and/or intellectual disabilities and/or because they were non-native language speakers (students with disabilities or non-native language speakers who had been mainstreamed; see definitions in section 3.3).

3.1.3 Students

A total of 11,454 fourth-grade students were listed as being in these classrooms. (In mixed-grade classrooms only students in the target population were considered.) Of these, 2,454 students were not part of TIMSS 2007 proper but rather had been designated to participate in the special Bridge Study designed to ensure the proper equating of the 2007 test scores with those from earlier administrations of TIMSS. For this purpose they completed a TIMSS 2003 assessment booklet rather than the TIMSS 2007 assessment. These students are not included in any of the participation rate calculations or tabulations and are not considered further in this report.

The remaining 9,000 students comprised the student sample for the TIMSS 2007 assessment proper. At the outset, 543 of these were excluded because of functional/intellectual disabilities or because they were non-native language speakers. Additionally, in the months between the listing of students and the time of the assessment 140 students were classified as ”withdrawn” as they were no longer in the school/classroom at the time of the assessment. As a consequence, 8,317 students were considered eligible to take the assessment. On the day of the assessment some 421 students were absent, leaving 7,896 students who completed a TIMSS 2007 assessment booklet. Participation rates are calculated on the number of eligible students (8,317). Since 7,896 of the 8,317 eligible students were assessed, the weighted (and unweighted) student participation rate was 95 percent.

3.1.4 Combined participation rates

The combined school, classroom, and student weighted participation rate standard of 75 percent used by TIMSS in situations in which it was necessary to recruit substitute schools was met in this instance. Both the weighted and unweighted product of the separate participation rates exceeded 84 percent. The application of International guidelines means, however, that U.S. statistics describing fourth-grade students are annotated in international reports to indicate that coverage of the defined student population was less than the IEA standard of 95 percent; and, that participation rates were met only after substitute schools were included.

3.2 Participation rates for all countries

For comparable fourth- and eighth-grade school, classroom and student participation rates in other nations, see exhibits 9.8 through 9.15 in Olson, Martin, and Mullis (2008).

3.3 Exclusions

The following discussion draws on the TIMSS 2007 School Sampling Manual (Foy, Joncas, and Zuhlke 2005). The national defined target population is described in chapter 2 of this report. All schools and students excluded from this population are referred to as the excluded population. Exclusions could occur at the school level, with entire schools being excluded, or within schools, with specific students or entire classrooms excluded. TIMSS 2007 did not provide accommodations for students with disabilities or students who were unable to read or speak the language of the test.

3.3.1 School exclusions

Countries could exclude schools that

- Were geographically inaccessible;
- Were of extremely small size;
- Offered a curriculum, or school structure, radically different from the mainstream educational system; or
- Provided instruction only to students in the excluded categories defined under "within-school exclusions," such as schools for the blind.

3.3.2 Within-school exclusions

Countries were asked to adapt the following international within-school exclusion rules to define excluded students:

- **Students with intellectual disabilities –** Students who, in the professional opinion of the school principal or other qualified staff members, were considered to be intellectually disabled or who had been tested psychologically as such. This included students who were emotionally or mentally unable to follow even the general instructions of the test. Students were not to be excluded solely because of poor academic performance or normal disciplinary problems.
- **Students with functional disabilities –** Students who were permanently physically disabled in such a way that they could not perform in the TIMSS testing situation. Functionally disabled students who were able to respond were included in the testing.
- **Non-native-language speakers –** Students who were unable to read or speak the language(s) of the test and were unable to overcome the language barrier of the test. Typically, a student who had received less than 1 year of instruction in the language(s) of the test was excluded.

3.3.3 Exclusions in the U.S. samples

As noted earlier, schools were given the opportunity to exclude any special classes among the total number of classes in the fourth or eighth grade. These classes were made up largely of students with functional or intellectual disabilities or students who were non-native-language speakers, as defined above. Classes identified in this way were excluded from the class sampling procedure. Subsequently, schools were given the opportunity to exclude students from the sampled classes—essentially, students

with functional or intellectual disabilities, or non-native-language speaking students in the United States who had been mainstreamed.

These procedures resulted in a (weighted) student exclusion rate of 9.2 percent in the fourth-grade and 7.9 percent in the eighth grade, based on the combination of whole-class and within-class exclusions. IEA standards define this degree of coverage of the target population (90.8 percent for fourth grade and 92.1 percent for eighth grade) as acceptable though falling below the desired range of 95 percent or better. The tabulations shown in the international reports show the United States annotated to indicate this fact.

3.4 Nonresponse bias analysis

The National Center for Education Statistics (NCES) standards for assessment surveys stipulate that a nonresponse bias analysis is required at any stage of data collection with a weighted unit response rate less than 85 percent at both grades. Since the U.S. TIMSS weighted school response rates are below 85 percent, NCES required an investigation into the potential magnitude of nonresponse bias at the school level in the U.S. sample, which is the focus of this section.

3.4.1 Methodology

To measure the potential nonresponse bias at the school level, the characteristics of participating schools were compared to those of the total eligible sample of schools. The alternative of comparing participants to non-participants, while resulting in the same tests of significance, makes it more difficult to judge the potential for bias. This analysis is similar to other NCES nonresponse bias studies on the 2007 National Assessment of Educational Progress (NAEP) (Krenzke and Lieber 2007) and the 2003 TIMSS (Ferraro and Van de Kerckhove 2006).

The analysis for each grade was conducted in three parts:

- First, the distribution of the responding original school sample was compared with that of the total eligible original school sample. The original sample is the sample before substitution. In each sample, schools were weighted by their school base weights that did not include a nonresponse adjustment factor. The base weight for each original school was the reciprocal of its selection probability.

- Second, the distribution of the responding sample, which includes the participating substitutes that were used as substitutes for schools from the original sample that did not participate, was compared to the total eligible final sample. The final sample is the sample after substitution. Again, school base weights were used for both the eligible sample and the participating schools. The base weight for each substitute school was set to the base weight of the original school that it replaced.

- Third, the same sets of schools were compared as in the second analysis but, this time, when analyzing the responding schools alone, school nonresponse adjustments were applied to the weights. The international weighting procedures created a nonresponse adjustment class[2] for each explicit stratum. For both U.S. TIMSS grade 4, or TIMSS-4, and U.S. TIMSS grade 8, or TIMSS-8, there was no explicit stratification and thus a single adjustment class.

[2] In general, nonresponse adjustment classes are formed based on characteristics related to response rates or to values of survey estimates where respondents and nonrespondents are similar within each class. The nonresponse adjustment is applied within each of these classes.

The first analysis indicates the potential for nonresponse bias that was introduced through school nonresponse. The second analysis suggests the remaining potential for nonresponse bias after the mitigating effects of substitution have been accounted for. The third analysis indicates the potential for bias after accounting for the mitigating effects of both substitution and nonresponse weight adjustments. However, the international weighting procedures created a nonresponse adjustment class for each explicit stratum. There was no explicit stratification in either grade and consequently a single adjustment class with a constant nonresponse adjustment factor was applied across all responding schools (by grade). Thus the adjustment had no effect on the characteristics of the weighted responding sample of schools, and the results in the third analysis are identical to those in the second analysis, so the results are not repeated. The second analysis, however, may provide an overly optimistic scenario, since even though substitution may correct somewhat for deficiencies in the few characteristics examined here, there is no guarantee that they are equally as effective for other characteristics, and in particular for student achievement.

To compare TIMSS participants and the total eligible sample, the sample of schools was matched to the sample frame to compare as many characteristics as possible that might provide information about the presence of nonresponse bias. Comparing frame characteristics for participants and the total eligible sample is not an ideal measure of nonresponse bias if the characteristics are unrelated or weakly related to more substantive items in the survey; however, this is often the only approach available.

Frame characteristics for public schools were taken from the 2003–04 Common Core of Data (CCD) and, for private schools, from the 2003–04 Private School Survey (PSS).

The following categorical variables were available for all schools:

- **School control –** Indicates whether the school is under public control (operated by publicly elected or appointed officials) or private control (operated by privately elected or appointed officials and derives its major source of funds from private sources);

- **Community type –** The location of a school relative to populous areas, i.e., central city, urban fringe/large town, rural/small town;

- **Region** (see section 3.5 for state listing); and

- **Poverty level**[3] **–** For public schools, a high poverty school is defined as one in which 50 percent or more of the students are eligible for participation in the National School Lunch Program (NSLP), and a low poverty school is defined as one in which less than 50 percent are eligible; all private schools are treated as low poverty schools.

[3] The sample frame did not contain a direct measure of poverty. No National School Lunch Program (NSLP) data were available for private schools.

The following continuous variables were available for all schools:

- Number of grade 4 or grade 8 students enrolled;
- Total number of students; and
- Percentage of students in five race/ethnicity categories (White, non-Hispanic; Black, non-Hispanic; Hispanic; Asian or Pacific Islander; and American Indian or Alaska Native).[4]

An additional continuous variable, the percentage of students eligible to participate in the NSLP, was available only for public schools.

For categorical variables, the distribution of frame characteristics for participants was compared with the distribution for all eligible schools. The hypothesis of independence between the characteristic and participation status was tested using a Rao-Scott modified Chi-square statistic at the 5 percent level (Rao and Thomas 2003). For continuous variables, summary means were calculated and the difference between means was tested using a *t* test. The statistical significance of differences between participants and the total eligible sample is identical to that which would result from comparing participants and non-participants, since all significance tests account for the fact that the participants are a subset of the full sample. The bias and relative bias are also given in each table. The bias is the difference between the respective estimates for the participants and the eligible sample. The relative bias is calculated as the bias divided by the estimate from the eligible sample. The relative bias is a measure of the size of the bias compared to the eligible sample estimate.

In addition to these tests, logistic regression models were used to provide a multivariate analysis in which the conditional independence of these school characteristics as predictors of participation was examined. It may be that only one or two variables are actually related to participation status. However, if these variables are also related to the other variables examined in the analyses, then other variables, which are not related to participation status, will appear as significant in simple bivariate tables. Dummy variables were created for each component of the categorical variables so that each component was included separately. The last component of each categorical variable is always the reference category and is not included in the model explicitly. The p-value of a dummy variable indicates whether there is a significant difference at the 5 percent level from the effect of the (omitted) reference category. It is not possible to include all the frame characteristics in a single model because the five race/ethnicity variables are linearly dependent (i.e., they sum up to 100 percent for every school). Therefore, two models were used. In the first model, four race/ethnicities (Black, non-Hispanic; Hispanic; Asian or Pacific Islander; and American Indian or Alaska Native) were included in the model with percentage White, non-Hispanic as the reference category. In addition, an *F*-test was used to determine whether the parameter estimates of these four characteristics were simultaneously equal to zero. In the second model, the summed percentage of the four race/ethnicities (Black, non-Hispanic; Hispanic; Asian or Pacific Islander; and American Indian or Alaska Native) replaced the four race/ethnicity variables with percentage White, non-Hispanic again as the reference category. All other frame characteristics were included in both models.

The logistic regression was performed using WesVar (http://www.westat.com/westat/statistical_software/WesVar/index.cfm) and replicate weights to properly account for the complex sample design. The paired jackknife replication method was used to create the replicate weights (Brick, Morganstein, and Valliant 2000).

[4] Black includes African American and Hispanic includes Latino. Racial categories exclude Hispanic origin.

3.4.2 Results for original respondent sample—Grade 4

This section presents the results of the nonresponse bias analysis, based exclusively on the original sample of 290 eligible U.S. schools for TIMSS-4. The distribution of the responding original school sample was compared with that of the total eligible original school sample using base weights in each case. All original schools in the sample that declined to participate in the survey were treated as non-participants regardless of whether they were replaced by a substitute school. The weighted and unweighted response rates were 70 percent, with 202 out of 290 eligible schools participating. See table 3.1 for details on the TIMSS-4 school participation rates.

Categorical Variables (TIMSS-4). The distribution of participating and eligible schools in the U.S. TIMSS-4 sample by the four characteristics is shown in table 3-2. Based on these analyses, the Chi-square statistic for community type and region are significant and suggests that there is evidence of relationships with participating in the assessment. In particular, schools in central cities were underrepresented among participating schools relative to eligible schools (28.2 vs. 32.1 percent, respectively), while schools in rural/small towns were overrepresented among participating schools (30.2 vs. 26.2 percent, respectively). Similarly, schools in the Central region were overrepresented among participating schools (27.2 vs. 22.1 percent, respectively), and schools in the West were underrepresented among participating schools (30.7 vs. 34.1 percent, respectively). There are no statistically significant relationships between participation status and any of the other characteristics shown in table 3-2.

Table 3-2. Percentage distribution of eligible and participating schools in the U.S. TIMSS fourth-grade original sample, by selected categorical variables: 2007

School characteristic	Sample schools: Eligible (percent) (N=290)	Sample schools: Participating (percent) (N=257)	Bias	Relative bias	Chi-square *p*-value
School control					0.482
Public	91.7	92.6	0.85	0.009	
Private	8.3	7.4	-0.85	-0.103	
Community type					0.041
Central city	32.1	28.2	-3.85	-0.120	
Urban fringe/large town	41.7	41.6	-0.14	-0.003	
Rural/small town	26.2	30.2	3.99	0.152	
Region					0.018
Northeast	20.0	18.8	-1.19	-0.059	
Southeast	23.8	23.3	-0.53	-0.022	
Central	22.1	27.2	5.16	0.234	
West	34.1	30.7	-3.45	-0.101	
Poverty level					0.464
High	37.6	36.1	-1.45	-0.038	
Low	62.4	63.9	1.45	0.023	

NOTE: For public schools, a high poverty school is defined as one in which 50 percent or more of the students are eligible for participation in the NSLP; all private schools are treated as low poverty schools. Region is the state-based region of the country (see section 3.5 for state listing.) Eligible schools contained at least one fourth-grade class. Participating schools agreed to have their students assessed. The bias is the difference between the respective estimates for the participants and the eligible sample. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Continuous Variables (TIMSS-4). Summary means for each continuous variable for participating and eligible schools are shown in tables 3-3 through 3-5. Three schools had a missing value for the total school enrollment, four schools had missing values for the race/ethnicity variables, and 31 out of the 266 public schools had a missing value for the free or reduced-price lunch variable; these schools were excluded from the analysis.

There were no statistically significant enrollment differences detected between participating and eligible schools (table 3-3). Participating schools had a higher mean percentage than the eligible sample of White, non-Hispanic students (62.6 vs. 59.9 percent, respectively; table 3-4) and American Indian or Alaska Native students (1.8 vs. 1.4 percent, respectively; table 3-4). There were no statistically significant differences detected between the participating and eligible public schools for the free or reduced-price lunch (table 3-5). However, this must be interpreted with caution because this variable was missing for 31 public schools.

Table 3-3. Mean enrollment of eligible and participating schools in the U.S. TIMSS fourth-grade original sample: 2007

	Sample schools				
Student enrollment	Eligible (mean)	Participating (mean)	Bias	Relative bias	*t* test *p*-value
Total school	566.0[1]	551.8[2]	-14.22	-0.025	0.188
Fourth grade	94.4[3]	91.2[4]	-3.19	-0.034	0.154

[1] N=287
[2] N=201
[3] N=290
[4] N=202

NOTE: Information on total school enrollment is missing for three of the 290 eligible schools in the sample. Eligible schools contained at least one fourth-grade class. Participating schools agreed to have their students assessed. The bias is the difference between the respective estimates for the participants and the eligible sample. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.

SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 3-4. Mean percentage of students in eligible and participating schools in the U.S. TIMSS fourth-grade original sample, by race/ethnicity: 2007

Race/ethnicity	Sample schools: Eligible (percent) (N=286)	Sample schools: Participating (percent) (N=198)	Bias	Relative bias	*t* test *p*-value
White, non-Hispanic	59.9	62.6	2.68	0.045	0.041
Black, non-Hispanic	15.7	14.7	-1.01	-0.064	0.280
Hispanic	19.0	17.3	-1.67	-0.088	0.110
Asian or Pacific Islander	4.0	3.6	-0.43	-0.105	0.249
American Indian or Alaska Native	1.4	1.8	0.42	0.301	0.009

NOTE: Information on race/ethnicity is missing for four non-participating schools of the 290 eligible schools in the sample. Black includes African American, and Hispanic includes Latino. Racial categories exclude Hispanic origin. Eligible schools contained at least one fourth-grade class. Participating schools agreed to have their students assessed. The bias is the difference between the respective estimates for the participants and the eligible sample. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 3-5. Mean percentage of students eligible for free or reduced-price lunch, in eligible and participating public schools in the U.S. TIMSS fourth-grade original sample: 2007

Students	Sample schools: Eligible (percent) (N = 235)	Sample schools: Participating (percent) (N = 167)	Bias	Relative bias	*t* test *p*-value
Percentage of students eligible for free or reduced-price lunch	45.8	45.0	-0.80	-0.017!	0.475!

! Interpret data with caution.
NOTE: Information on percentage of students eligible for free or reduced-price lunch is missing for 31 of the 266 public schools in the sample. Eligible schools contained at least one fourth-grade class. Participating schools agreed to have their students assessed. The bias is the difference between the respective estimates for the participants and the eligible sample. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Logistic Regression Model (TIMSS-4). To examine the joint relationship of various characteristics to school nonresponse, the analysis utilized a logistic regression model with participation status as the binary dependent variable and frame characteristics as predictor variables. Public and private schools were modeled together using the variables available for all schools. Seven schools were excluded from the analysis due to missing information for race/ethnicity and total school enrollment.

Standard errors and tests of hypotheses for the full model parameter estimates are presented in tables 3-6a (with four race/ethnicity variables) and b (with summed race/ethnicity percentage). Central region and the percentage of American Indian or Alaska Native are the two significant predictors of school participation in table 3-6a. The positive parameter estimates indicate that, relative to schools in the West region, schools in the Central region were somewhat overrepresented among the participating schools and the percentage of American Indian or Alaska Native students in participating schools was lower than in all eligible schools. The *F*-test statistic to determine whether the race/ethnicity characteristics were simultaneously equal to 0 was 1.48 with a p-value of 0.2178, which indicates no significant relationship detected with participation. The *F*-test indicates that since the percentage of American Indian or Alaska

Native students is very small in the vast majority of schools it had little effect on all the other races/ethnicities.

Central region was again a significant predictor of school participation in table 3-6b. Additionally, fourth grade enrollment is also significant. The negative parameter estimate for fourth-grade enrollment indicates that participating schools tended to be smaller than non-participating schools, i.e., the larger the fourth grade enrollment, the less likely a school was to participate. This model also shows that the summed race/ethnicity percentage is not significantly related to participation.

Table 3-6a. Logistic regression model parameter estimates (with four race/ethnicity variables) using the U.S. TIMSS fourth-grade original sample: 2007

Parameter	Parameter estimate	Standard error	*t* test for H_0: parameter = 0	*p*-value
Intercept	0.445	0.4512	0.9869	0.3269
Central city	-0.395	0.3962	-0.9961	0.3224
Rural/small town	0.778	0.4865	1.5994	0.1139
Private school	-0.242	0.5318	-0.4560	0.6497
High poverty	0.010	0.3886	0.0270	0.9785
Northeast	0.386	0.4388	0.8804	0.3815
Southeast	0.248	0.4518	0.5486	0.5849
Central	1.484	0.4185	3.5458	0.0007
Total school enrollment	0.001	0.0010	1.2421	0.2181
Fourth grade enrollment	-0.009	0.0044	-1.9887	0.0504
Black, non-Hispanic	-0.003	0.0078	-0.4112	0.6821
Hispanic	0.002	0.0072	0.2480	0.8048
Asian or Pacific Islander	0.002	0.0276	0.0848	0.9326
American Indian or Alaska Native	0.209	0.0953	2.1900	0.0316

NOTE: Seven of the 290 eligible schools in the sample were excluded due to missing information for race/ethnicity and total school enrollment. For public schools, a high poverty school is defined as one in which 50 percent or more of the students are eligible for participation in the National School Lunch Program; all private schools are treated as low poverty schools. Region is the state-based region of the country (see section 3.5 for state listing). Black includes African American, and Hispanic includes Latino. Racial categories exclude Hispanic origin. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 3-6b. Logistic regression model parameter estimates (with summed race/ethnicity percentage) using the U.S. TIMSS fourth-grade original sample: 2007

Parameter	Parameter estimate	Standard error	*t* test for H_0: parameter = 0	*p*-value
Intercept	0.826	0.4317	1.9122	0.0597
Central city	-0.395	0.3803	-1.0378	0.3027
Rural/small town	0.735	0.4672	1.5725	0.1201
Private school	-0.325	0.5405	-0.6015	0.5494
High poverty	0.078	0.3636	0.2142	0.8309
Northeast	0.112	0.3886	0.2873	0.7746
Southeast	-0.067	0.3636	-0.1852	0.8536
Central	1.269	0.3667	3.4591	0.0009
Total school enrollment	0.001	0.0009	1.1651	0.2477
Fourth grade enrollment	-0.008	0.0042	-1.9955	0.0496
Summed race/ethnicity percentage	-0.001	0.0060	-0.2243	0.8231

NOTE: Seven of the 290 eligible schools in the sample was excluded due to missing information for race/ethnicity and total school enrollment. For public schools, a high poverty school is defined as one in which 50 percent or more of the students are eligible for participation in the NSLP; all private schools are treated as low poverty schools. Region is the state-based region of the country (see section 3.5 for state listing.) Summed race/ethnicity percentage includes Black, non-Hispanic; Hispanic; Asian or Pacific Islander; and American Indian or Alaska Native. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

3.4.3 Results for respondent sample with substitutes (final sample)

This section presents the nonresponse bias analysis based on the final sample of 290 eligible schools for TIMSS-4. The distribution of the responding sample, including participating substitutes, was compared to the total eligible original sample. School base weights were used for both the eligible sample and the participating schools. Only eligible original schools that refused and were not successfully replaced by a substitute were treated as non-participants. All other eligible original sample schools were treated as participating. Through the use of substitutes, the weighted and unweighted school response rates for TIMSS-4 were 89 percent, with 257 out of 290 schools participating.

Categorical Variables (TIMSS-4). The distribution of participating and eligible schools by the four characteristics is shown in table 3-7. Only community type was found to be statistically significant among the categorical variables. In particular, schools in central cities were underrepresented among participating schools relative to eligible schools (29.6 vs. 32.1 percent, respectively), while schools in rural/small towns were overrepresented among participating schools relative to eligible schools (28.4 vs. 26.2 percent, respectively). These differences were both smaller than that shown in table 3-1, in which only the original sample was considered. Thus while there is no evidence that the use of substitute schools eliminated the potential for bias, as indicated by this variable, it certainly has also not added to it. There were no statistically significant relationships detected between participation status and the other characteristics shown in table 3-7.

Table 3-7. Percentage distribution of eligible and participating schools in the U.S. TIMSS fourth-grade final sample, by selected categorical variables: 2007

School characteristic	Sample schools: Eligible (percent) (N=290)	Sample schools: Participating (percent) (N=257)	Bias	Relative bias	Chi-square *p*-value
School control					0.871
Public	91.7	91.8	0.11	0.001	
Private	8.3	8.2	-0.11	-0.013	
Community type					0.023
Central city	32.1	29.6	-2.50	-0.078	
Urban fringe/large town	41.7	42.0	0.30	0.007	
Rural/small town	26.2	28.4	2.20	0.084	
Region					0.231
Northeast	20.0	19.1	-0.93	-0.047	
Southeast	23.8	24.5	0.72	0.030	
Central	22.1	23.3	1.28	0.058	
West	34.1	33.1	-1.06	-0.031	
Poverty level					0.697
High	36.6	36.2	-0.37	-0.010	
Low	63.4	63.8	0.37	0.006	

NOTE: For public schools, a high poverty school is defined as one in which 50 percent or more of the students are eligible for participation in the National School Lunch Program; all private schools are treated as low poverty schools. Region is the state-based region of the country (see section 3.5 for state listing). Eligible schools contained at least one fourth-grade class. Participating schools agreed to have their students assessed. The bias is the difference between the respective estimates for the participants and the eligible sample. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Continuous Variables (TIMSS-4). Summary means for each continuous variable for participating and eligible schools are shown in tables 3-8 through 3-10. Two schools had a missing value for the total number of students, four schools had a missing value for one or more of the race/ethnicity variables, and 33 out of the 266 public schools had a missing value for the free or reduced-price lunch variable; these schools were excluded from the analysis.

There were no statistically significant enrollment differences detected between participating and eligible schools (table 3-8). Participating schools had a higher mean percentage of White, non-Hispanic and American Indian or Alaska Native students than the eligible sample (60.9 vs. 59.2 percent and 1.6 vs.1.4, respectively, table 3-9). The difference in the percentage of students who are White, non-Hispanic was smaller than that shown table 3-4, in which only the original sample was considered. The difference in the percentage of students who are American Indian or Alaska Native was considerably smaller than that shown in table 3-4, in which only the original sample was considered. Thus, as in the case with community type, while there is no evidence that the use of substitute schools substantially reduced the potential for bias, as indicated by these variables, it has also not added to it. There were no statistically significant differences detected in the mean percentage of students of the other races and ethnicities (Black, non-Hispanic, Hispanic, and Asian or Pacific Islander) between the participating and eligible schools.

There was no statistically significant difference detected between the participating and eligible public schools for free or reduced-price lunch (table 3-10). However, this must be interpreted with caution because the variable is missing for 33 schools.

Table 3-8. Mean enrollment of eligible and participating schools in the U.S. TIMSS fourth-grade final sample: 2007

	Sample schools				
Student enrollment	Eligible (mean)	Participating (mean)	Bias	Relative bias	*t* test *p*-value
Total school	572.8[1]	575.1[2]	2.36	0.004	0.678
Fourth grade	94.4[3]	94.3[4]	-0.06	-0.001	0.961

[1] N=288
[2] N=255
[3] N=290
[4] N=257

NOTE: Information on total school enrollment is missing for two of the 290 eligible schools in the sample. Eligible schools contained at least one fourth-grade class. Participating schools agreed to have their students assessed. The bias is the difference between the respective estimates for the participants and the eligible sample. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 3-9. Mean percentage of students in eligible and participating schools in the U.S. TIMSS fourth-grade final sample, by race/ethnicity: 2007

	Sample schools				
Race/ethnicity	Eligible (percent) (N=286)	Participating (percent) (N=253)	Bias	Relative bias	*t* test *p*-value
White, non-Hispanic	59.2	60.9	1.73	0.029	0.019
Black, non-Hispanic	15.8	14.8	-0.98	-0.062	0.057
Hispanic	19.0	18.6	-0.44	-0.023	0.399
Asian or Pacific Islander	4.6	4.1	-0.44	-0.096	0.170
American Indian or Alaska Native	1.4	1.6	0.13	0.091	0.018

NOTE: Information on race/ethnicity is missing for four non-participating school of the 290 eligible schools in the sample. Black includes African American, and Hispanic includes Latino. Racial categories exclude Hispanic origin. Eligible schools contained at least one fourth-grade class. Participating schools agreed to have their students assessed. The bias is the difference between the respective estimates for the participants and the eligible sample. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 3-10. Mean percentage of students eligible for free or reduced-price lunch, in eligible and participating public schools in the U.S. TIMSS fourth-grade final sample: 2007

Students	Sample schools: Eligible (percent) (N = 233)	Sample schools: Participating (percent) (N = 209)	Bias	Relative bias	*t* test *p*-value
Percentage of students eligible for free or reduced-price lunch	46.0	46.0	0.00	0.000	0.968

NOTE: Information on percentage of students eligible for free or reduced-price lunch is missing for 33 of the 266 public schools in the sample. Eligible schools contained at least one fourth-grade class. Participating schools agreed to have their students assessed. The bias is the difference between the respective estimates for the participants and the eligible sample. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Logistic Regression Model (TIMSS-4). To examine the joint relationship of various characteristics to school nonresponse, the analysis utilized a logistic regression model with participation status as the binary dependent variable and frame characteristics as predictor variables. Six schools were excluded from the analysis due to missing information for race/ethnicity and total school enrollment.

Standard errors and tests of hypotheses for the full model parameter estimates are presented in tables 3-11a (with four race/ethnicity variables) and b (with summed race/ethnicity percentage). None of the parameter estimates are significant which indicates that there were no significant relationships detected with participation status. The *F*-test statistic to determine whether the race/ethnicity characteristics were simultaneously equal to 0 was 1.55 with a p-value of 0.1999, which indicates no significant relationship detected with participation.

Table 3-11a. Logistic regression model parameter estimates (with four race/ethnicity variables) in the U.S. TIMSS fourth-grade final sample: 2007

Parameter	Parameter estimate	Standard error	*t* test for H_0: parameter = 0	*p*-value
Intercept	1.497	0.6229	2.4041	0.0187
Central city	-0.520	0.4995	-1.0417	0.3009
Rural/small town	1.084	0.7475	1.4498	0.1513
Private school	-0.024	0.9305	-0.0257	0.9795
High poverty	0.131	0.5213	0.2517	0.8020
Northeast	-0.044	0.6264	-0.0703	0.9441
Southeast	0.409	0.7130	0.5735	0.5680
Central	1.034	0.7156	1.4455	0.1525
Total school enrollment	0.003	0.0019	1.3902	0.1686
Fourth grade enrollment	-0.011	0.0094	-1.1421	0.2571
Black, non-Hispanic	-0.011	0.0108	-1.0344	0.3043
Hispanic	-0.002	0.0100	-0.2004	0.8417
Asian or Pacific Islander	-0.020	0.0174	-1.1427	0.2568
American Indian or Alaska Native	0.272	0.1717	1.5839	0.1174

NOTE: Six participating schools of the 290 eligible schools in the sample were excluded due to missing information for race/ethnicity and total school enrollment. For public schools, a high poverty school is defined as one in which 50 percent or more of the students are eligible for participation in the National School Lunch Program; all private schools are treated as low poverty schools. Region is the state-based region of the country (see section 3.5 for state listing). Black includes African American, and Hispanic includes Latino. Racial categories exclude Hispanic origin. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 3-11b. Logistic regression model parameter estimates (with summed race/ethnicity percentage) in the U.S. TIMSS fourth-grade final sample: 2007

Parameter	Parameter estimate	Standard error	*t* test for H_0: parameter = 0	*p*-value
Intercept	1.830	0.5863	3.1210	0.0026
Central city	-0.485	0.4643	-1.0435	0.3001
Rural/small town	1.112	0.7121	1.5613	0.1227
Private school	-0.120	0.8855	-0.1357	0.8924
High poverty	0.262	0.4281	0.6123	0.5422
Northeast	-0.261	0.4753	-0.5501	0.5839
Southeast	0.071	0.4507	0.1580	0.8749
Central	0.755	0.6237	1.2111	0.2297
Total school enrollment	0.003	0.0017	1.5088	0.1356
Fourth grade enrollment	-0.010	0.0087	-1.1971	0.2350
Summed race/ethnicity percentage	-0.010	0.0074	-1.2954	0.1992

NOTE: Six participating schools of the 290 eligible schools in the sample were excluded due to missing information for race/ethnicity and total school enrollment. For public schools, a high poverty school is defined as one in which 50 percent or more of the students are eligible for participation in the National School Lunch Program; all private schools are treated as low poverty schools. Region is the state-based region of the country (see section 3.5 for state listing). Summed race/ethnicity percentage includes Black, non-Hispanic; Hispanic; Asian or Pacific Islander; and American Indian or Alaska Native. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

3.4.4 Summary—Grade 4

For the original sample of schools in TIMSS-4 in the United States, three variables were found to be related to participation—community type, region, and racial/ethnic composition. Central city schools were underrepresented among participating schools and rural/small-town schools were overrepresented. Similarly, schools in the Central region were overrepresented, and schools in the West underrepresented. And, in regard to racial/ ethnic composition, both the percentage of White, non-Hispanic and the percentage of American Indian or Alaska Native students were higher in participating schools than in the eligible sample. The regression analysis showed that Central region and the percentage of American Indian or Alaska Native students were significant predictors of participation when the individual race/ethnicities were included in the model. Additionally, Central region and fourth grade enrollment were significant predictors when the summed race/ethnicity percentage was included.

The bivariate results for the final sample of TIMSS-4 indicated that two of the three variables were still found to be related to participation: community type, and racial/ethnic composition. As in the earlier analysis, central city schools were underrepresented among participating schools and rural/small-town schools were overrepresented. Similarly, both the percentage of White, non-Hispanic and the percentage of American Indian or Alaska Native students were higher in participating schools than in the eligible sample. In each instance the differences were substantially reduced over those seen in connection with the original sample. These same differences could not be demonstrated in the multivariate regression analysis which failed to show any variables as significant predictors of participation.

These results suggest that there is some potential for nonresponse bias in the U.S. TIMSS-4 original sample based on the characteristics studied. It also suggests that the use of substitute schools somewhat reduced the potential for bias. The application of school nonresponse adjustments had no effect on bias, as just a single nonresponse adjustment factor was used for all schools.

3.4.5 Results for original respondent sample–grade 8

The following nonresponse bias analysis is based exclusively on the original sample of 287 eligible schools for TIMSS-8. The distribution of the responding original school sample was compared with that of the total eligible original school sample using base weights in each case. All original schools in the sample that declined to participate in the survey were treated as non-participants regardless whether they were replaced by a substitute school. The weighted and unweighted response rates were both 69 percent with 197 out of 287 original schools participating in TIMSS. See table 3.1 for details on the TIMSS-8 school participation rates.

Categorical Variables (TIMSS-8). The distribution of participating and eligible schools in the U.S. TIMSS-8 sample by the four characteristics is shown in table 3-12. There were no statistically significant relationships detected between participation status and the characteristics shown in table 3-12.

Table 3-12. Percentage distribution of eligible and participating schools in the U.S. TIMSS eighth-grade original sample, by selected categorical variables: 2007

School characteristic	Sample schools: Eligible (percent) (N=287)	Sample schools: Participating (percent) (N=197)	Bias	Relative bias	Chi-square *p*-value
School control					0.661
Public	91.8	91.4	-0.46	-0.005	
Private	8.2	8.6	0.46	0.056	
Community type					0.061
Central city	30.0	34.0	3.98	0.133	
Urban fringe/large town	42.9	39.1	-3.86	-0.090	
Rural/small town	27.0	26.9	-0.12	-0.004	
Region					0.347
Northeast	20.6	18.8	-1.82	-0.088	
Southeast	23.5	23.4	-0.18	-0.008	
Central	23.0	25.9	2.84	0.123	
West	32.8	32.0	-0.84	-0.026	
Poverty level					0.153
High	29.7	32.0	2.30	0.078	
Low	70.3	68.0	-2.30	-0.033	

NOTE: Detail may not sum to totals because of rounding. For public schools, a high poverty school is defined as one in which 50 percent or more of the students are eligible for participation in the National School Lunch Program; all private schools are treated as low poverty schools. Region is the state-based region of the country (see section 3.5 for state listing). Eligible schools contained at least one eighth-grade class. Participating schools agreed to have their students assessed. The bias is the difference between the respective estimates for the participants and the eligible sample. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Continuous Variables (TIMSS-8). Summary means for each continuous variable for participating and eligible schools are shown in tables 3-13 through 3-15. One school had a missing value for the total number of students, four schools had a missing value for one or more of the race/ethnicity variables, 34 out of the 263 public schools had a missing value for the free or reduced-price lunch variable; these schools were excluded from the analysis.

No statistically significant differences were detected between participating and eligible schools for the variables shown in tables 3-13 through 3-15. However, this must be interpreted with caution for the free or reduced-price lunch variable because it is missing for 34 schools.

Table 3-13. Mean enrollment of eligible and participating schools using the U.S. TIMSS eighth-grade original sample: 2007

Student enrollment	Sample schools: Eligible (mean)	Sample schools: Participating (mean)	Bias	Relative bias	*t* test *p*-value
Total school	813.2[1]	805.9[2]	-7.27	-0.009	0.685
Eighth grade	276.5[3]	276.3[4]	-0.14	-0.001	0.984

[1] N=286
[2] N=196
[3] N=287
[4] N=197

NOTE: Information on total school enrollment is missing for one participating school of the 287 eligible schools in the sample. Eligible schools contained at least one eighth-grade class. Participating schools agreed to have their students assessed. The bias is the difference between the respective estimates for the participants and the eligible sample. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 3-14. Mean percentage of students in eligible and participating schools in the U.S. TIMSS eighth-grade original sample, by race/ethnicity: 2007

Race/ethnicity	Sample schools: Eligible (percent) (N=283)	Sample schools: Participating (percent) (N=194)	Bias	Relative Bias	*t* test *p*-value
White, non-Hispanic	62.0	62.2	0.14	0.002	0.908
Black, non-Hispanic	16.3	15.7	-0.64	-0.039	0.465
Hispanic	17.1	17.4	0.29	0.017	0.741
Asian or Pacific Islander	3.7	3.7	0.05	0.014	0.835
American Indian or Alaska Native	0.9	1.0	0.16	0.190	0.098

NOTE: Information on race/ethnicity is missing for four participating schools of the 287 eligible schools in the sample. Black includes African American, and Hispanic includes Latino. Racial categories exclude Hispanic origin. Eligible schools contained at least one eighth-grade class. Participating schools agreed to have their students assessed. The bias is the difference between the respective estimates for the participants and the eligible sample. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 3-15. Mean percentage of students eligible for free or reduced-price lunch, in eligible and participating public schools using the U.S. TIMSS eighth-grade original sample: 2007

Students	Sample schools: Eligible (percent) (N = 229)	Sample schools: Participating (percent) (N = 157)	Bias	Relative bias	*t* test *p*-value
Percentage of students eligible for free or reduced-price lunch	42.3	43.9	1.60	0.038!	0.133!

! Interpret data with caution.
NOTE: Information on percentage of students eligible for free or reduced lunch program is missing for 34 of the 263 public schools in the sample. Eligible schools contained at least one eighth-grade class. Participating schools agreed to have their students assessed. The bias is the difference between the respective estimates for the participants and the eligible sample. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Logistic Regression Model (TIMSS-8). To examine the joint relationship of various characteristics to school nonresponse, the analysis utilized a logistic regression model with participation status as the binary dependent variable and frame characteristics as predictor variables. Public and private schools were modeled together using the variables available for all schools. Five schools were excluded from the analysis due to missing information for race/ethnicity and total school enrollment.

Standard errors and tests of hypotheses for the full model parameter estimates are presented in tables 3-16a (with four race/ethnicity variables) and b (with summed race/ethnicity percentage). Being located in a central city and Central region are the two significant predictors of school participation in table 3-16a. The positive parameter estimates indicate that, relative to urban fringe/large town schools, central city schools were overrepresented among the participating schools. And, relative to schools in the West region, schools in the Central region were overrepresented among the participating schools. The *F*-test statistic to determine whether the race/ethnicity characteristics were simultaneously equal to 0 was 1.45 with a p-value of 0.2254, which indicates no significant relationship detected with participation.

Being located in a central city was also a significant predictor of school participation in table 3-16b. The positive parameter estimates indicate that schools in central cities were more likely than schools in rural areas to participate.

Table 3-16a. Logistic regression model parameter estimates (with four race/ethnicity variables) using the U.S. TIMSS eighth-grade original sample of schools: 2007

Parameter	Parameter estimate	Standard error	*t* test for H_0: parameter = 0	*p*-value
Intercept	0.111	0.5318	0.2096	0.8346
Central city	0.778	0.3047	2.5517	0.0128
Rural/small town	0.114	0.3846	0.2959	0.7681
Private school	0.344	0.5244	0.6552	0.5143
High poverty	0.709	0.4221	1.6797	0.0972
Northeast	0.258	0.3934	0.6554	0.5142
Southeast	0.531	0.4860	1.0933	0.2778
Central	0.897	0.4231	2.1193	0.0374
Total school enrollment	#	0.0006	-0.6666	0.5071
Fourth grade enrollment	0.001	0.0013	0.9655	0.3374
Black, non-Hispanic	-0.014	0.0069	-1.9536	0.0545
Hispanic	-0.004	0.0073	-0.4893	0.6261
Asian or Pacific Islander	0.006	0.0110	0.5596	0.5774
American Indian or Alaska Native	0.075	0.0455	1.6591	0.1013

Rounds to zero.

NOTE: Five of the 287 eligible schools in the sample were excluded due to missing information for race/ethnicity and total school enrollment. For public schools, a high poverty school is defined as one in which 50 percent or more of the students are eligible for participation in the NSLP; all private schools are treated as low poverty schools. Region is the state-based region of the country (see section 3.5 for state listing). Black includes African American, and Hispanic includes Latino. Racial categories exclude Hispanic origin. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.

SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 3-16b. Logistic regression model parameter estimates (with summed race/ethnicity percentage) using the U.S. TIMSS eighth-grade original sample of schools: 2007

Parameter	Parameter estimate	Standard error	*t* test for H_0: parameter = 0	*p*-value
Intercept	0.387	0.4948	0.7831	0.4361
Central city	0.835	0.3106	2.6878	0.0089
Rural/small town	0.194	0.3803	0.5097	0.6118
Private school	0.408	0.5161	0.7897	0.4322
High poverty	0.654	0.3895	1.6788	0.0973
Northeast	-0.023	0.3523	-0.0664	0.9472
Southeast	0.124	0.4058	0.3067	0.7599
Central	0.539	0.3547	1.5193	0.1329
Total school enrollment	#	0.0006	-0.6683	0.5060
Fourth grade enrollment	0.001	0.0013	1.0170	0.3124
Summed race/ethnicity percentage	-0.008	0.0057	-1.4135	0.1616

Rounds to zero.

NOTE: Five of the 287 eligible schools in the sample were excluded due to missing information for race/ethnicity and total school enrollment. For public schools, a high poverty school is defined as one in which 50 percent or more of the students are eligible for participation in the National School Lunch Program; all private schools are treated as low poverty schools. Region is the state-based region of the country (see section 3.5 for state listing). Summed race/ethnicity percentage includes Black, non-Hispanic; Hispanic; Asian or Pacific Islander; and American Indian or Alaska Native. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.

SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

3.4.6 Results for respondent sample with substitutes (final sample)

The following nonresponse bias analysis is based on the final sample of 287 schools for TIMSS-8. The distribution of the responding sample, including participating substitutes was compared to the total eligible original sample. School base weights were used for both the eligible sample and the participating schools. Only eligible original schools that refused and were not successfully replaced by a substitute were treated as non-participants. All other eligible original sample schools were treated as participating. Through the use of substitutes, the weighted and unweighted school response rates in TIMSS-8 were 83 percent, with 239 out of 287 schools participating.

Categorical Variables (TIMSS-8). The distribution of participating and eligible schools by the four characteristics is shown in table 3-17. The Chi-square statistic for community type is significant and suggests that there is evidence of a relationship with participation. In particular, schools in central cities were overrepresented among participating schools relative to eligible schools (33.5 vs. 30.0 percent, respectively), while schools in rural/small towns were underrepresented among participating schools relative to eligible schools (40.2 vs. 42.9 percent, respectively). Even though community type did not show a significant difference when only the original sample was considered (table 3-12), the relative bias is now smaller than in the original sample for two of the three categories. This is due to the smaller standard errors on the participating school estimate and thus being able to detect smaller differences with the additional substitute participating schools. Thus while there is no evidence that the use of substitute schools reduced the potential for bias, as indicated by this variable, it has also not substantially added to it. There were no statistically significant relationships detected between participation status and the other characteristics shown in table 3-17.

Table 3-17. Percentage distribution of eligible and participating schools in the U.S. TIMSS eighth-grade final sample, by selected categorical variables: 2007

School characteristic	Sample schools: Eligible (percent) (N=287)	Sample schools: Participating (percent) (N=239)	Bias	Relative bias	Chi-square *p*-value
School control					0.314
Public	91.8	91.2	-0.62	-0.007	
Private	8.2	8.8	0.62	0.076	
Community type					0.022
Central city	30.0	33.5	3.45	0.115	
Urban fringe/large town	42.9	40.2	-2.78	-0.065	
Rural/small town	27.0	26.4	-0.67	-0.025	
Region					0.789
Northeast	20.6	19.7	-0.94	-0.045	
Southeast	23.5	24.3	0.73	0.031	
Central	23.0	23.4	0.39	0.017	
West	32.8	32.6	-0.19	-0.006	
Poverty level					0.392
High	30.4	31.4	1.00	0.033	
Low	69.6	68.6	-1.00	-0.014	

NOTE: Detail may not sum to totals because of rounding. For public schools, a high poverty school is defined as one in which 50 percent or more of the students are eligible for participation in the National School Lunch Program; all private schools are treated as low poverty schools. Region is the state-based region of the country (see section 3.5 for state listing). Eligible schools contained at least one eighth-grade class. Participating schools agreed to have their students assessed. The bias is the difference between the respective estimates for the participants and the eligible sample. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Continuous Variables (TIMSS-8). Summary means for each continuous variable for participating and eligible schools are shown in tables 3-18 through 3-20. One school had a missing value for the total number of students, four schools had a missing value for one or more of the race/ethnicity variables, and 36 out of the 263 public schools had a missing value for the free or reduced-price lunch variable; these schools were excluded from the analysis.

There were no statistically significant enrollment differences detected between participating and eligible schools (table 3-18). Participating schools had a statistically significantly higher mean percentage of American Indian or Alaska Native students than the eligible sample (1.5 vs. 1.3 percent, respectively; table 3-19). The difference in the percentage of students who are American Indian or Alaska Native was similar to that shown in table 3-14, in which only the original sample was considered. Thus while there is no evidence that the use of substitute schools reduced the potential for bias, as indicated by this variable, it has also not substantially added to it as the relative bias was reduced (0.190 vs. 0.137). There were no statistically significant differences detected for the other race or ethnicity categories (White, non-Hispanic; Black, non-Hispanic; Hispanic; and Asian or Pacific Islander).

There was also no statistically significant difference detected in the mean percentage of students eligible for free or reduced-price lunch between participating and eligible public schools (table 3-20). However, this must be interpreted with caution because the variable is missing for 36 out of the 263 public schools.

Table 3-18. Mean enrollment of eligible and participating schools in the U.S. TIMSS eighth-grade final sample: 2007

	Sample schools				
Student enrollment	Eligible (mean)	Participating (mean)	Bias	Relative bias	*t* test *p*-value
Total school	816.5[1]	814.4[2]	-2.06	-0.003	0.859
Eighth grade	277.2[3]	274.6[4]	-2.62	-0.009	0.561

[1] N=286
[2] N=238
[3] N=287
[4] N=239

NOTE: Information on total school enrollment is missing for one participating school of the 287 eligible schools in the sample. Eligible schools contained at least one eighth-grade class. Participating schools agreed to have their students assessed. The bias is the difference between the respective estimates for the participants and the eligible sample. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.

SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 3-19. Mean percentage of students in eligible and participating schools in the U.S. TIMSS eighth-grade final sample, by race/ethnicity: 2007

Race/ethnicity	Sample schools: Eligible (percent) (N=283)	Sample schools: Participating (percent) (N=235)	Bias	Relative Bias	*t* test *p*-value
White, non-Hispanic	61.5	61.7	0.23	0.004	0.816
Black, non-Hispanic	15.7	15.1	-0.58	-0.037	0.433
Hispanic	17.8	18.0	0.20	0.011	0.745
Asian or Pacific Islander	3.7	3.7	-0.04	-0.010	0.816
American Indian or Alaska Native	1.3	1.5	0.18	0.137	0.012

NOTE: Information on race/ethnicity is missing for four participating schools of the 287 eligible schools in the sample. Black includes African American, and Hispanic includes Latino. Racial categories exclude Hispanic origin. Eligible schools contained at least one eighth-grade class. Participating schools agreed to have their students assessed. The bias is the difference between the respective estimates for the participants and the eligible sample. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 3-20. Mean percentage of students eligible for free or reduced-price lunch, in eligible and participating public schools in the U.S. TIMSS eighth-grade final sample: 2007

Students	Sample schools: Eligible (percent) (N = 227)	Sample schools: Participating (percent) (N = 189)	Bias	Relative bias	*t* test *p*-value
Percentage of students eligible for free or reduced-price lunch	43.1	43.7	0.60	0.014	0.502

NOTE: Information on percentage of students eligible for free or reduced lunch program is missing for 36 of the 263 public schools in the sample. Eligible schools contained at least one eighth-grade class. Participating schools agreed to have their students assessed. The bias is the difference between the respective estimates for the participants and the eligible sample. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Logistic Regression Model (TIMSS-8). To examine the joint relationship of various characteristics to school nonresponse, the analysis utilized a logistic regression model with participation status as the binary dependent variable and frame characteristics as predictor variables. Five schools were excluded from the analysis due to missing information for race/ethnicity and total school enrollment.

Standard errors and tests of hypotheses for the full model parameter estimates are presented in tables 3-21a (with four race/ethnicity variables) and b (with summed race/ethnicity percentage). Being located in a central city and the percentage of Black, non-Hispanic are the two significant predictors of school participation in table 3-21a. The positive parameter estimates indicate that, relative to urban fringe/large town schools, central city schools were overrepresented among the participating schools. The negative parameter estimate indicates that the percentage of Black, non-Hispanic students in participating schools was lower than in all eligible schools. The *F*-test statistic to determine whether the race/ethnicity characteristics were simultaneously equal to 0 was 1.77 with a p-value of 0.1441, which indicates no significant relationship detected with participation.

Being located in a central city was also a significant predictor of school participation in table 3-21b. Additionally, the summed race/ethnicity percentage was also significant. The negative parameter estimate indicates that the summed race/ethnicity percentage in participating schools was lower than in all eligible schools.

Table 3-21a. Logistic regression model parameter estimates (with four race/ethnicity variables) in the U.S. TIMSS eighth-grade final sample: 2007

Parameter	Parameter estimate	Standard error	*t* test for H_0: parameter = 0	*p*-value
Intercept	1.173	0.5651	2.0764	0.0413
Central city	1.562	0.5496	2.8417	0.0058
Rural/small town	-0.137	0.4625	-0.2963	0.7678
Private school	0.436	0.8992	0.4852	0.6289
High poverty	0.921	0.5617	1.6388	0.1054
Northeast	0.152	0.4907	0.3100	0.7574
Southeast	0.649	0.6173	1.0511	0.2966
Central	0.418	0.5248	0.7973	0.4278
Total school enrollment	#	0.0007	0.4770	0.6347
Fourth grade enrollment	-0.001	0.0014	-0.4734	0.6373
Black, non-Hispanic	-0.023	0.0092	-2.5090	0.0143
Hispanic	-0.011	0.0106	-1.0241	0.3091
Asian or Pacific Islander	-0.006	0.0138	-0.4610	0.6461
American Indian or Alaska Native	0.130	0.1155	1.1281	0.2629

Rounds to zero.

NOTE: Five participating schools of the 287 eligible schools in the sample were excluded due to missing information for race/ethnicity and total school enrollment. For public schools, a high poverty school is defined as one in which 50 percent or more of the students are eligible for participation in the National School Lunch Program; all private schools are treated as low poverty schools. Region is the state-based region of the country (see section 3.5 for state listing). Black includes African American, and Hispanic includes Latino. Racial categories exclude Hispanic origin. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.

SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Table 3-21b. Logistic regression model parameter estimates (with summed race/ethnicity percentage) in the U.S. TIMSS eighth-grade final sample: 2007

Parameter	Parameter estimate	Standard error	*t* test for H_0: parameter = 0	*p*-value
Intercept	-0.060	0.7331	-0.0819	0.9349
Central city	1.583	0.5345	2.9608	0.0041
Rural/small town	0.007	0.4397	0.0159	0.9873
Private school	0.500	0.8385	0.5966	0.5526
High poverty	0.873	0.4923	1.7730	0.0803
Northeast	-0.205	0.4557	-0.4507	0.6535
Southeast	0.165	0.4958	0.3326	0.7404
Central	0.006	0.4642	0.0128	0.9898
Total school enrollment	#	0.0007	0.4185	0.6768
Fourth grade enrollment	-0.001	0.0014	-0.4436	0.6586
Summed race/ethnicity percentage	-0.016	0.0077	-2.0598	0.0429

Rounds to zero.
NOTE: Five participating schools of the 287 eligible schools in the sample were excluded due to missing information for race/ethnicity and total school enrollment. For public schools, a high poverty school is defined as one in which 50 percent or more of the students are eligible for participation in the National School Lunch Program; all private schools are treated as low poverty schools. Region is the state-based region of the country (see section 3.5 for state listing). Summed race/ethnicity percentage includes Black, non-Hispanic; Hispanic; Asian or Pacific Islander; and American Indian or Alaska Native. Schools were weighted by their school base weights that did not include a nonresponse adjustment factor.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

3.4.7 Summary—Grade 8

For the original sample of schools in TIMSS-8 in the United States, no characteristic was found to be significantly related to school participation in the bivariate analysis. However, the multivariate regression analysis showed that, relative to urban fringe/large town schools, central city schools were overrepresented among the participating schools. And, relative to schools in the West region, schools in the Central region were similarly overrepresented.

The bivariate results for the final sample indicated that two variables were related to participation: community type, and the percentage of American Indian or Alaska Native. Central city schools were overrepresented among participating schools, and schools in urban fringe/large town were underrepresented. And, in regard to racial/ethnic composition, the percentage of American Indian or Alaska Native students in participating schools was higher than in all eligible schools. The multivariate regression analysis indicated that, relative to urban fringe/large town schools, central city schools were overrepresented among the participating schools, and the percentage of Black, non-Hispanic students in participating schools was lower than in all eligible schools when the individual race/ethnicities were included in the model. Additionally, the summed race/ethnicity percentage in participating schools was lower than in all eligible schools when it was included in the model.

The application of school nonresponse adjustments had no effect on bias, as just a single nonresponse adjustment factor was used for all schools.

3.4.8 Conclusions

The investigation into nonresponse bias at the school level for U.S. TIMSS 2007 samples for grades 4 and 8 has shown that there was no statistically significant relationship detected between participation status and the majority of school characteristics that are available for analysis.

For the original sample of schools in TIMSS-4 in the United States, four variables were found to be statistically significantly related to participation in the bivariate analysis: community type (table 3-2); region (table 3-2); the percentage of White, non-Hispanic students (table 3-4); and the percentage of American Indian or Alaska Native students (table 3-4). Although each of these findings indicates some potential for nonresponse bias, when all of these factors were considered simultaneously in a regression analysis, Central region and the percentage of American Indian or Alaska Native students were significant predictors of participation (table 3-6a). The second model showed that Central region and fourth grade enrollment were significant predictors of participation (table 3-6b, with summed race/ethnicity percentage).

For the final sample of schools in TIMSS-4, three of the four variables remained statistically significant in the bivariate analysis: community type (table 3-7); the percentage of White, non-Hispanic students (table 3-9); and the percentage of American Indian or Alaska Native students (table 3-9). In each instance the differences were substantially reduced over those seen in connection with the original sample. These same differences found in the final sample could not be demonstrated in the regression analysis, which failed to show any variables as significant predictors of participation (tables 3-11a and b). For the final sample of schools in TIMSS-4 with school nonresponse adjustments applied to the weights, the results were identical to those just described.

These results suggest that there is some potential for nonresponse bias in the U.S. TIMSS-4 original sample based on the characteristics studied. It also suggests that the use of substitute schools somewhat reduced the potential for bias. The application of school nonresponse adjustments had no effect on bias, as just a single nonresponse adjustment factor was used for all schools.

For the original sample of schools in TIMSS-8 in the United States, no characteristic was found to be significantly related to school participation in the bivariate analysis. However, the multivariate regression analysis showed that, relative to urban fringe/large town schools, central city schools were overrepresented among the participating schools. And, relative to schools in the West region, schools in the Central region were similarly overrepresented (tables 3-16a and 3-16b).

For the final U.S. TIMSS-8 sample of schools, two variables were found to be statistically significantly related to participation in the bivariate analysis: community type (table 3-17); and the percentage of American Indian or Alaska Native students (table 3-19). The logistic regression showed that being located in a central city and the percentage of Black, non-Hispanic students were statistically significant (table 3-21a). Additionally, the summed race/ethnicity percentage was significant (table 3-21b). For the final sample of schools in TIMSS-8 with school nonresponse adjustments applied to the weights, the results were identical to those just described.

These results suggest that there is some potential for nonresponse bias in the TIMSS-8 original sample based on the characteristics studied. It also suggests that, while there is no evidence that the use of substitute schools reduced the potential for bias, it has not added to it substantially. The application of school nonresponse adjustments had no effect on bias, as just a single nonresponse adjustment factor was used for all schools.

3.5 Technical Notes

3.5.1 Description of Variables

Frame characteristics for public schools were taken from the 2003–04 Common Core of Data (CCD) and, for private schools, from the 2003–04 Private School Survey (PSS).

Race/Ethnicity. Students' race/ethnicity was obtained through student responses to a two-part question. Students were asked first whether they were Hispanic or Latino, and then asked whether they were members of the following racial groups: American Indian/Alaska Native; Asian; Black or African American; Native Hawaiian or other Pacific Islander; or White. Multiple responses to the race classification question were allowed.

Community Type. Community type was derived from the locale variable based on how the school is situated in a particular location relative to populous areas, based on the school's address. Central city consists of large (a principal city of a Metropolitan Core Based Statistical Area (CBSA), with the city having a population greater than or equal to 250,000) and mid-size (a principal city of a Metropolitan CBSA, with the city having a population less than 250,000) cities. Urban fringe/large town consists of urban fringe of a large city (any incorporated place, Census-designated place, or non-place territory within a Metropolitan CBSA of a Large City and defined as urban by the Census Bureau), urban fringe of a mid-size city (any incorporated place, Census-designated place, or non-place territory within a CBSA of a Mid-Size City and defined as urban by the Census Bureau) or large town (an incorporated place or Census-designated place with a population greater than or equal to 25,000 and located outside a Metropolitan CBSA or inside a Micropolitan CBSA). Rural/small town consists of small town (an incorporated place or Census Designated Place (CDP) with population less than 25,000 and greater than or equal to 2,500 and located outside a CBSA or inside a Micropolitan CBSA, rural, outside CBSA (any incorporated place, Census-designated place, or non-place territory not within a CBSA and defined as rural by the Census Bureau) or rural, inside CBSA (any incorporated place, Census-designated place, or non-place territory within a Metropolitan CBSA and defined as rural by the Census Bureau).

Region. Region is a "state-based" region of the country developed for the NAEP. This grouping of states differs from the region defined by the Census Bureau. Northeast consists of Connecticut, Delaware, District of Columbia, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont. Central consists of Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin. West consists of Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oklahoma, Oregon, Texas, Utah, Washington, and Wyoming. Southeast consists of Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia.

Poverty Level in Public Schools (Percentage of Students Eligible for Free or Reduced Price Lunch). The poverty level in public schools was obtained from principal responses to the school questionnaire. The question asked what percentage of students at the school was eligible to receive free to reduced-price lunch through the National School Lunch Program around the first of October, 2004. For the analyses included in this report, responses were grouped into high poverty—schools in which 50 percent or more of students were eligible—and low poverty—less than 50 percent of students were eligible.

3.5.2 Statistical Procedures

Weighting. Before the data are analyzed, responses from the groups of students assessed are assigned sampling weights to ensure that their representation in TIMSS 2007 results matches their actual percentage of the school population in the grade assessed.

Responses from the groups of students were assigned sampling weights to adjust for over- or under-representation during the sampling of a particular group. The use of sampling weights is necessary for the computation of sound, nationally representative estimates. The weight assigned to a student's responses is the inverse of the probability that the student would be selected for the sample. When responses are weighted, none are discarded, and each contributes to the results for the total number of students represented by the individual student assessed. Weighting also adjusts for various situations (such as school and student nonresponse) because data cannot be assumed to be randomly missing. The internationally defined weighting specifications for TIMSS require that each assessed student's sampling weight should be the product of (1) the inverse of the school's probability of selection, (2) an adjustment for school-level nonresponse, (3) the inverse of the classroom's probability of selection, and (4) an adjustment for student-level nonresponse.

In the analyses in this report, sometimes the appropriate weight (base weight) includes only the components of the reciprocals of the respective selection probabilities. This is the case when estimates are made based on the entire sample. In other cases nonresponse adjustments, as computed by the TIMSS International Study Center, are also applied. In each case the text and tables make clear which of these weighting procedures has been applied. Whereas for substantive analyses using the TIMSS data, one would normally apply the nonresponse adjustments when analyzing the data from the respondents in the sample, this is not always when the case when carrying out analyses of potential nonresponse bias analyses.

Sampling Errors. Sampling errors occur when the discrepancy between a population characteristic and the sample estimate arises because not all members of the reference population are sampled for the survey. The size of the sample relative to the population and the variability of the population characteristics both influence the magnitude of sampling error. The particular sample of students in fourth and eighth grade from the 2006-07 school year was just one of many possible samples that could have been selected. Therefore, estimates produced from the TIMSS sample may differ from estimates that would have been produced had another student sample been drawn. This type of variability is called sampling error because it arises from using a sample of students in fourth or eighth grade, rather than all students in the grade in that year.

The standard error is a measure of the variability due to sampling when estimating a statistic, and is often included in reports containing estimates from survey data. The approach used for calculating sampling variances in TIMSS was the Jackknife Repeated Replication (JRR). In this report we do not show estimates of standard errors for each estimate. Rather the effects of sampling error are reflected in the test statistics (for *t* tests and chi-square tests, and *t* test used in logistic regression analyses) that are presented for each analysis. These are described below.

The first step to compute the variance with replication is to calculate the estimate of interest from the full sample as well as each subsample or replicate. The variation between the replicate estimates and the full-sample estimate is then used to estimate the variance for the full sample. Suppose that $\hat{\theta}$ is the full-sample estimate of some population parameter θ. The variance estimator, $v\left(\hat{\theta}\right)$, takes the form

$$v\left(\hat{\theta}\right) = \sum_{g=1}^{G}\left(\hat{\theta}_{(g)} - \hat{\theta}\right)^2$$

where

$\hat{\theta}_{(g)}$ is the estimate of θ based on the observations included in the *g*-th replicate, and

G is the total number of replicates formed (G=75 for U.S. TIMSS).

The standard error is then

$$se\left(\hat{\theta}\right) = \sqrt{v\left(\hat{\theta}\right)}$$

The JRR algorithm used in TIMSS 2007 assumes that there are *G* replicates, each containing two sampled schools selected independently. The element $\hat{\theta}_{(g)}$ denotes the estimate using the *g*-th jackknife replicate. This is computed using all cases except those in the *g*-th replicate of the sample. For those in the *g*-th replicate, the replicate weights for all cases associated with one of the randomly selected units of the pair are multiplied by zero, and the replicate weights for the elements associated with the other unit in the replicate are doubled. The computation of the JRR variance for any estimate requires the computation of the statistic 76 times for any given country: once to obtain the estimate for the full sample, and 75 times to obtain the estimate for each of the jackknife replicates ($\hat{\theta}_{(g)}$).

Tests of Significance. Comparisons made in the text of this report have been tested for statistical significance. For example, when comparing results obtained from the full sample for a given grade, with those obtained only from the responding sample units, tests of statistical significance were used to establish whether or not the observed differences are statistically significant. The estimation of the standard errors that are required in order to undertake the tests of significance is complicated by the complex sample and assessment designs which both generate error variance. Together they mandate a set of statistically complex procedures in order to estimate the correct standard errors. As a consequence, the estimated standard errors contain a sampling variance component estimated by Jackknife Repeated Replication (JRR). Details on the procedures used can be found in the WesVar 4.3 User's Guide (Westat 2007).

Two kinds of statistical tests are included in the report: *t* tests and chi-square tests. In addition, logistic regression analyses were conducted.

***t* Tests.** *t* tests were used for testing for the hypothesis that no difference exists between the means of continuous variables for two groups (namely, the full sample and the responding sample). Suppose that $\bar{x}_A$ and $\bar{x}_B$ are the means for two groups that are being compared and $se\left(\bar{x}_A - \bar{x}_B\right)$ is the standard error of the difference between the means which accounts for the complex survey design. Then the *t* test is defined as

$$t = \frac{\left|\bar{x}_A - \bar{x}_B\right|}{se\left(\bar{x}_A - \bar{x}_B\right)}$$

This statistic is then compared to the critical values of the appropriate Student t-distribution, to determine whether the difference is statistically significant. The appropriate number of degrees of freedom for the distribution is given by the number of primary sampling units in the design (in this case the number of schools), minus the number of sampling strata.

Note that this procedure took account of the fact that the two samples in question were not independent samples, but in fact the responding sample was a subsample of the full sample. This effect was accounted for in calculating the standard error of the difference. Note also that, in those cases where both samples were weighted just using base weights the test is exactly equivalent to testing that the mean of the respondents was equal to the mean of the nonrespondents.

Consider for example the data in table 3-3. The first row shows that the weighted mean total school enrollment for the full eligible sample of grade 4 schools is 566.0. For the subsample of schools that

participated, the corresponding mean is 551.8, resulting in a difference of 14.22. The standard error of this estimated difference, calculated so as to reflect the dependency between these two samples, and the complex sample design, is 10.7 (not shown in table). This gives rise to a *t*-statistic of -1.33 (not shown in table), and using 75 degrees of freedom (the appropriate figure for the TIMSS design), the resulting significance (or *p*-value) is 0.188. This last figure appears in the table.

t tests were also used in the logistic regression for testing for the hypothesis for whether each estimated parameter estimate is significantly different from 0. Then the *t* test is defined as

$$t = \frac{b_k}{\sqrt{v(b_k)}}$$

where b_k is a parameter estimate and $v(b_k)$ is the replication variance estimate for that parameter. This statistic is then compared to the critical values of the appropriate Student *t*-distribution, as described above, to determine whether the difference is statistically significant. The appropriate number of degrees of freedom for the distribution is again given by the number of primary sampling units in the design (in this case the number of schools), minus the number of sampling strata.

Chi-Square Tests. Chi-square tests are used for testing whether two distributions of a given categorical variable are different, conducted in a way that reflects the impact of the complex sample design on sampling variance. In this instance one distribution is for the full sample, and one for the responding sample. Suppose that the categorical variable in question has *c* levels, cross-tabulated producing weighted proportions *p*. The usual Pearson chi-square statistic is calculated as

$$X^2 = n\sum_{i=1}^{2}\sum_{j=1}^{c}\left(p_{ij} - p_{i\cdot}p_{\cdot j}\right)^2 \Big/ p_{i\cdot}p_{\cdot j}$$

where *j* denotes the categories of the categorical variable, and *i* indexes the samples (full sample and respondents), and *n* indicates the overall sample size. This statistic is not suitable for use directly in a statistical test with these data, for two reasons. First, the fact that the respondents are a subset of the full sample violates the standard assumptions for a chi-square test of this kind. Second, this statistic does not account for the complex sample design used to collect the data.

Thus the Pearson Chi-square statistic is modified appropriately to account for the impact of these two features. The resulting test statistic is referred to as the Rao-Scott Adjusted chi-square statistic. It is sometimes also referred to as the Satterthwaite-adjusted chi-square statistic. The number of degrees of freedom for the chi-square test, normally given as (*c* -1), where *c* is the number of categories of the categorical variable for each distribution, is also modified on account of the complex design. The modified test statistic is then compared to the chi-square distribution with the appropriate number of degrees of freedom, to determine whether the difference in the two distributions is statistically significant. For a detailed description of the technique, see Rao and Scott (1984) or Rao and Thomas (2007).

The first step in the calculation of the Satterthwaite-adjusted chi-square statistic is to form the following vector:

$$\mathrm{Y} = \sqrt{n}\begin{pmatrix} p_{11} - p_{1\cdot}p_{\cdot 1} \\ p_{12} - p_{1\cdot}p_{\cdot 2} \\ \\ p_{rc} - p_{r\cdot}p_{\cdot c} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \\ y_{rc} \end{pmatrix}$$

An *rc* x 1 vector made up of the products of the marginal proportions is defined as

$$\mathrm{p} = \begin{pmatrix} p_{1\cdot} p_{\cdot 1} \\ p_{1\cdot} p_{\cdot 2} \\ \\ p_{r\cdot} p_{\cdot c} \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ \\ p_{rc} \end{pmatrix}$$

For each replicate, an *rc* x *rc* matrix is calculated whose *ij*-th element is made up of

$$\left(y_{ig} - y_i\right)\left(y_{jg} - y_j\right),$$

where y_{ig} and y_{jg} are the *i*-th and *j*-th elements of **Y** calculated for the *g*-th replicate and y_i and y_j are the corresponding full-sample values. The *ij*-th element of the estimated covariance matrix for Y, B=cov(Y), is calculated using the following formula:

$$B_{ij} = \sum_{g=1}^{G} \left(y_{ig} - y_i\right)\left(y_{jg} - y_j\right),$$

where *c* is the constant appropriate to the replication method (c=0.05 for U.S. TIMSS). The Satterthwaite's approximation to degrees of freedom for the chi-square statistic to be calculated is

$$\nu = \frac{\left(\sum_{i=1}^{rc} \frac{B_{ii}}{p_i}\right)^2}{\sum_{i=1}^{rc}\sum_{j=1}^{rc} \frac{B_{ij}^2}{p_i p_j}}.$$

Since ν will generally not be an integer, interpolation in standard chi-square tables is required.

Finally, the adjusted chi-square statistic is defined as

$$RS3 = \frac{X^2}{\sum_{i=1}^{rc} \frac{B_{ii}}{p_i}}.$$

Logistic Regression Models. Let p_i denote the probability that the *i*-th sampled school will participate. Under the logistic regression model, the log odds of response propensity (expressed in terms of the logarithm of $p_i/(1-p_i)$), is assumed to have the following linear form:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi}$$

where X_{1i}, X_{2i} ..., X_{pi} are *p* auxiliary variables associated with the *i*-th sampled school, and β_0, β_1, ..., β_p are coefficients to be estimated. Asymptotic assumptions are used to develop statistical tests to determine which, if any, of the coefficients are significantly different from zero. In the analyses in this report the standard procedures for carrying out logistic regression analyses have been modified both to incorporate the sampling weights in the estimation of the coefficients, and to reflect the effect of the complex sample design on the variance-covariance matrix of the coefficients.

The Newton-Raphson algorithm is used to iteratively solve for parameter solutions in the logistic regression. Let $q(\beta) = \partial L_n(\beta)/\partial\beta$ be the vector of first partial derivatives of the sample log-likelihood with respect to β. Let $H(\beta)$ be the matrix of second partial derivatives (or Hessian) of the sample log-likelihood having entries $\partial^2 L/\partial\beta_a\partial\beta_b$, where β_a and β_b are two separate components of β. Denote by q^t and H^t the values of $q(\beta)$ and $H(\beta)$ evaluated at b^t, the value of the estimate b at step *t*.

The general approach is to approximate the sample log-likelihood at the desired estimate, $L_n(b)$, at step *t* in the iterative process near the point b^t by a second-order Taylor series expansion:

$$L_n^t(b) \cong L_n(b^t) + q^{t\prime}(b - b^t) + \tfrac{1}{2}(b - b^t)' H^t (b - b^t).$$

Solving $\partial L^t/\partial b = q^t + H^t(b\text{-}b^t) = 0$ for b yields the iteration equations

$$b^{t+1} = b^t - \left[H^t\right]^{-1} q^t,$$

assuming H^t has an inverse. Given an initial value for $t = 0$, the set of iteration equations is solved for b^1, b^1 is used to solve for b^2, and so on, until the convergence criterion is satisfied. The $se(\hat{\beta})$ is calculated using JRR and repeating the procedure for each replicate.

Page intentionally left blank.

4. Survey Operations

This chapter describes data collection and related activities undertaken in connection with TIMSS 2007 in the United States. These activities included: recruitment of schools; sampling of students within schools; development of the instrumentation used; field operations undertaken in order to administer the assessment; post-assessment activities associated with scoring and data entry; and, several activities associated with the preparation of the data to meet international standards.

4.1 Recruiting districts and schools

The protocol followed in seeking the participation of schools was: first, state education authorities were notified of the intention to approach schools within their jurisdiction; second, permission was sought at the district level to make contact with the sampled schools; and, third, the schools themselves were contacted. Participation may be refused at any of these levels. Three matters took on particular importance in this context. First, it was necessary to establish the value of participation. Second, it was important to establish the timing of the TIMSS assessment window *vis a vis* mandatory federal, state and local assessments. Third, concerns about the extent of burden on the schools needed to be addressed.

4.1.1 Timing of recruitment activities

There was an important time element to this process as well. The school calendar fills quickly each year such that assessment dates needed to be established early in the school year in which the assessment was to take place or, better still, toward the conclusion of the previous year. In total, the recruitment phase for TIMSS 2007 extended from April 2006 through May 2007. States were contacted toward the end of April 2006. Contact with districts began in May of 2006 and continued to April 2007. Schools were contacted as soon as district permission was received. School recruitment activities continued from late in May 2006 through April 2007, with the exception of the summer months when the schools were not in session.

4.1.2 The impact of the sampling design on recruitment activities

The sampling design played an important role in the design of recruitment activities. At both grade levels 300 schools were sampled, along with 300 first substitutes and 300 second substitutes, for a total of 900 schools at each grade level. The 300 originally sampled schools were approached first. Operationally this meant, first, seeking permission from the districts in which these schools were located and then, with permission in hand, approaching the schools themselves. If a sampled school refused to participate, the district of the first substitute school was approached and the district-school permission procedure began anew. And, if the first substitute school refused as well, then the district of the second substitute school was approached and, then, the second substitute school itself was contacted.

There was considerable variation in the amount of time required to gain the permission (or to have it denied) at either the district or school level. Many districts required proposals in their own unique format, and the participation question, with or without a tailored proposal, could be subject to an extended review process which may include school board approval. Similarly, the decision process within schools could be drawn out by the practice of holding school-wide discussions with teachers and, sometimes, parents.

Arriving at a balance between the potential for an extended recruitment period, and the need to get on the school calendar as early as possible, was accomplished in three ways. First, in order to provide for a maximum recruiting period, the permission process with respect to sampled schools was begun in the spring of 2006, approximately 12 months prior to the actual assessment in 2007. Second, refusals by

sampled schools were evaluated using information from field staff to decide whether further attempts should be made to enlist the cooperation of the school, or whether the first substitute district and school should be approached. Third, during the schools' summer vacation "permission in principle" was sought from districts to approach, if necessary, the substitute school(s) in that district when the new school year began.

4.1.3 Contacting states

The Chief State School Officer and State Test Director in each of the 50 states and the District of Columbia were each contacted beginning in March 2006. Each person received a TIMSS package which included: cover letter from the NCES commissioner, a list of the districts containing the sampled fourth- and eighth-grade schools in the state, a brochure describing the study, a TIMSS publications sheet, and a TIMSS resource kit on CD-ROM. Three states (Delaware, Hawaii, and Vermont) were contacted with a somewhat different letter since the districts in these states contained only substitute schools. A copy of the letter sent to states is provided as exhibit B.1 in appendix B. Several items of TIMSS information materials (exhibits C.1 through C.4 in appendix C), and a list of sampled schools, were included with the letter.

4.1.4 Contacting districts

Shortly after notification to the states, the first of several contacts was initiated with the superintendent and the test director of each district (or diocese) containing sampled schools. A copy of the letter sent to districts is provided as exhibit B.2 in appendix B. Several items of TIMSS information materials (exhibits C.1 through C.4 in appendix C), and a list of sampled schools, were included with the letter.

Districts of first substitute schools were contacted after the originally sampled schools were judged firm refusals; districts of second substitute schools were contacted after first substitute schools were judged firm refusals.

During the summer of 2006 another form of district contact was initiated. Given the need to complete the recruitment of schools as early in the 2006-07 school year as possible, the districts of substitute schools linked to sampled schools whose participation was not confirmed were asked to provide "provisional permission" to approach these substitute schools early in the new school year. Provisional permission meant asking for permission to approach a substitute school if necessary – if the sampled school (or first substitute school) refused. Some 500 districts were contacted in this way early in the summer of 2006. A package similar to that sent to other districts was mailed along with a special letter asking for provisional permission early in the summer. A copy of this letter is provided as exhibit B.3 in appendix B.

Follow-up contact. In each case field staff made follow-up calls to the district superintendent after a few days to discuss the study, answer questions and attempt to gain explicit permission to approach schools. Districts varied considerably in the nature of the procedures to be followed before granting permission. For example, 61 school districts required a formal application process involving the completion of district-specific applications and school board approval.

Special security requirements. Each TIMSS field staff member had a current FBI clearance and fingerprints on file. In most instances this satisfied school security requirements. However, a number of districts required that additional procedures be met before allowing staff to enter schools. Many of these districts had an additional "volunteer form" that needed to be completed and presented at the school. Other districts and schools had more stringent requirements. For example, New York City required staff to obtain additional, local fingerprinting at a local office in New York City. Additionally, for each school, Westat was also required to complete a "signature form" with signatures from both the principal and the district superintendent. Schools in Massachusetts required an additional background check based on the

Criminal Offender Record Information Act (CORI). However, this requirement was waived for TIMSS 2007 since field staff carried a letter from the TIMSS project director stating that all staff entering schools had current clearance and fingerprints on file with the FBI. In Florida, the Jessica Lundsford Act required that staff entering Broward County schools undergo additional fingerprinting required for the issuance of a specific security badge. Hillsborough County and Miami-Dade also required additional fingerprinting and a local security check. Miami-Dade County required the completion of an Affidavit of Good Moral Character. In many cases, field staff working in these districts had these clearances on file from their recent NAEP work.

4.1.5 Contacting schools

Once approval to contact each school was received from the district, schools were contacted with an initial school information packet. Private schools and some parochial schools not linked with a diocese were contacted directly. Each school information package was sent on a flow basis governed by receipt of district approval. A copy of the covering letter included in the school package is shown as exhibit B.4 in appendix B. Copies of items of TIMSS information materials (exhibits C.1 through C.4 in appendix C) were included with the letter.

Follow-up contact. After a few days, each school was contacted by the member of field staff assigned to the school with the view to discussing the school's participation in TIMSS. In-person visits were made in a small number of schools where efforts to secure participation proved difficult.

Nomination of School Coordinator. At this time, school principals were asked to identify an individual within the school that would act as the TIMSS School Coordinator. Principals, deputy principals, teachers and guidance counselors took on this role. School Coordinator responsibilities included:

- Working with Westat to establish an assessment date;
- Preparing a list of all mathematics classrooms in the specified grade;
- Identifying excluded classrooms;
- Identifying students and teachers in the sampled classrooms;
- Identifying excluded students in the sampled classrooms;
- Liaising with staff, students, and parents as necessary;
- Arranging for space and for the release of students from classes on assessment day;
- Ensuring completion and return of the School and Teacher Questionnaires; and,
- Holding secure, until 2008, the confidential files that linked student names with IDs and, then, destroying them.

4.1.6 Informational materials and gifts

Since the initial contact with states, districts and schools was by mail, particular attention was paid to the development of informative materials designed to establish the value of participation and assure all concerned that the burden on schools would be minimal. These materials included the following:

- A TIMSS brochure describing TIMSS, its history and its importance for the United States (exhibit C.1 in appendix C);
- A TIMSS publications sheet, showing a selection of publications arising out of TIMSS (exhibit C.2 in appendix C);
- A "key information and summary of activities" brochure designed to indicate the specifics of school participation in TIMSS (exhibit C.3 in appendix C); and,
- A brochure describing "incentives" to be provided to schools and students; in particular, the TIMSS resource kit designed to provide teachers with curriculum-relevant mathematics and science information from TIMSS linked to an archive of actual assessment items (exhibit C.4 in appendix C).

Schools were provided with copies of the TIMSS resource kit at this time and, subsequent to the assessment, an all-in-one printer. The school coordinator was given a TIMSS satchel, and students each received a clock-compass carabiner.

Some additional documents were used to maintain contact with schools throughout the year and to inform school coordinators of coming TIMSS activities; see, for example, the sheet describing School Coordinator responsibilities (exhibit C.5 in appendix C). Schools were sent a holiday card toward the end of the 2006 year and, at various points, school coordinators were sent informational materials to prepare them for upcoming TIMSS tasks. Drafts of parent approval letters, forms and fact sheets were supplied to those schools indicating that parent approval was required. Details on the parent approval materials is provided in exhibits B.5 through B.7 in appendix B and discussed further below.

4.1.7 Field staff

Twenty-seven Field Supervisors experienced with recruiting districts and schools on similar studies (NAEP, PISA, TIMSS, and PIRLS), along with four Field Managers, were assigned the task of recruiting the districts and schools. On average, each Field Supervisor had responsibility for 22 schools.

Field Manager responsibilities. The four Field Managers engaged in the recruiting phase of the study had the following responsibilities:

- Coordinating recruitment activities of the Field Supervisors within their assignment;
- Holding weekly one-on-one telephone meetings with their Field Supervisors to monitor progress on gaining-cooperation activities and to troubleshoot recruitment strategies;
- Monitoring and maintaining the Field Management System, ensuring that the disposition codes provided by Field Supervisors gave an accurate portrayal of participation status;

- Attending weekly conference calls involving Westat home office staff and other Field Managers; and,

- Acting as trouble-shooters to handle special-issues that arose during the recruiting process.

Field Supervisor responsibilities. The primary responsibility of the 27 Field Supervisors was person-to-person interaction with the districts and schools along with some more administrative tasks, as follows:

- Making telephone contact with the districts of sampled schools within their assignment to obtain permission to recruit schools;

- Making telephone contact to sampled schools within their assignment to obtain permission to conduct TIMSS;

- Making in-person refusal conversion attempts when necessary and requested by a Field Manager;

- Completing an Electronic Record of Calls (EROC) of each contact within the Field Management System (FMS);

- Maintaining the most current disposition codes within the FMS; and,

- Meeting weekly with Field Manager by conference call to discuss progress.

4.1.8 Training field staff

A recruitment training workshop for field staff was held at Westat on April 22-23, 2006. A Field Supervisor Manual was assembled and sent to field staff 5 days prior to training. Four hours of home study were provided for field staff to familiarize themselves with the material before arriving at training.

Day one of the training began with a presentation on gaining an understanding of TIMSS and its goals. The history of TIMSS, the participating countries, an overview of the school, classroom and student sampling, the development of the instruments and examples of published reports and data were presented. All of the materials that would be sent to district contacts in the recruitment package were provided and explained. A subsequent discussion ensured that all field staff engaged in the recruitment of districts and schools understood the mission of TIMSS, the materials used in the United States and the importance of satisfactory participation rates.

A presentation explained the responsibilities of a TIMSS Recruitment Supervisor. Preliminary activities, such as planning, organizing, coordinating and scheduling were presented as well as an examination of the administrative tasks that were expected. At the conclusion of day one, the field staff were given the assignment of familiarizing themselves with the *TIMSS Frequently Asked Questions* and to develop a "TIMSS story" of talking points that would be valuable in a recruiting situation.

Day two focused on how to gain cooperation from schools. Training included open discussions about interpersonal skills important for gaining the cooperation of districts and schools. These included: introducing the assessment; answering questions; dealing with schools' concerns; overcoming objections; and, avoiding refusals. Participants were invited to share strategies and techniques that they had found to be successful in past recruitment experiences. Common recruitment strategies and techniques were demonstrated and discussed. A presentation about refusal avoidance was undertaken with the view to assisting participants in recognizing potential refusals and gearing the conversation toward a successful

outcome. Further, approaches that might be used with schools for the purposes of refusal conversion were introduced and practiced.

To reinforce these presentations and discussions, field staff participated in several role-playing exercises. The role plays were cumulative, building from constructing a simple story, or discussion of what TIMSS was about, to interactive one-on-one conversations where various types of objections had to be recognized and countered. The goal was to overcome any objections convincingly.

The final activity was a hands-on tutorial of the TIMSS Field Management System (FMS). Each member of the field staff had a laptop workstation connected to the FMS. A live demonstration was given showing how to navigate the functions of the system during the several stages of recruitment. The demonstration also explained iFMS functions, such as setting of disposition codes and code meanings. A novel function of the TIMSS FMS was an Electronic Record of Calls (EROC). This functioned similarly to a traditional paper Record of Calls, but allowed call information to be collected in a central data base at the home office. The EROC function provided a screen with a combination of discrete and open ended entry fields.

Following the demonstration, the participants performed two exercises. In the first exercise, they simply entered information obtained from a fictional call to an initial school into the FMS to generate a final disposition code. This was used to familiarize them with using the system and entering information correctly. In the second exercise, the participants entered information to a point and then switched work stations with another participant, picking up the case as if it was transferred to them. The exercise was designed to illustrate the required level of detail that should be used in recording call information.

4.1.9 Monitoring recruiting progress

Progress in the recruitment of districts and schools was monitored on a daily basis through the FMS. Field staff were required to update the FMS with each contact that was made with districts or schools. This involved the completion of an electronic record of calls (EROC) detailing their conversation with district or school staff, along with the updating of the disposition code relating to the district/school in question. The disposition code indicated whether the district or school as pending, refusing, or cooperating. Using these status codes, the Westat home office was able to track the progress of recruitment and daily reports were created using the FMS data. These daily reports enabled the operations of recruitment (and eventually assessment) to be closely monitored. The reports were used to determine the status of recruitment efforts and the performance of each Field Supervisor. Matters arising from this monitoring were taken up by the Field Managers with the particular supervisors.

4.1.10 Difficulties in gaining cooperation

The principal reasons given by both districts and schools for refusing to participate were, in approximate order of priority:

- Conflict with mandatory federal, state or local assessments whose outcomes had direct implications for districts, schools, teachers and students;

- The related matter of the burden that additional testing placed on students at the cost of instructional time; and

- The limited return on the school's investment of time since they would not receive much usable information on the school, and none on particular students.

Some additional problems arose as a result of the year-long recruitment period, though they were fairly minor: schools forgot that they had agreed to participate; others had second thoughts; and, in some cases the district or school administration changed and the new administration saw no need to honor the original commitment. Field staff attempted to address these issues as they arose.

4.2 Sampling students within schools

Students in participating schools were sampled in a two-stage process. In the first stage, schools were asked to provide lists of fourth- or eighth-grade mathematics classrooms with the number of students in each class indicated. An equal probability sample of two classrooms (or pseudo-classrooms) was identified from the classrooms listed for each school. In the second stage all students in sampled classrooms (pseudo-classrooms) were selected for assessment.

The student sampling process was performed using WinW3S software developed by the IEA Data Processing Center (DPC) and made available to each country (IEA Data Processing Center 2006). The WinW3S system provides for forms generation, data entry, class sampling, student sampling, student-teacher linkages, the random assignment of assessment booklets to students, the production of various survey tracking forms, and the printing of labels for test instruments and questionnaires.

Westat home office staff attended a training session in the use of WinW3S, and conducted a 2-day training for the data entry staff to introduce them to the software and familiarize them with the listing forms and the procedures for entering data. This training provided practice exercises in all aspects of the work. The workflow was organized around four independent workstations. Four mutually exclusive projects, one for each data processor, were created in WinW3S and stored in a project directory on a secure network. Two data processors were assigned to each grade with each person receiving half of the cooperating schools for their grade. Most of the data entry occurred over the period January to mid-May, 2007 as schools returned completed Class Listing and Student-Teacher Listing forms.

4.2.1 Obtaining class lists from schools

A Class Listing Form (CLF) was sent to the School Coordinator. The CLF was used to create a list of the eligible classes, some attributes of each class, and the names of the teacher(s) teaching each class. A copy of a Class Listing Form containing fictitious information is reproduced in exhibit 4.1 below.

The information relating to school ID and name, the name of the School Coordinator, and the grade level in question was filled before the form was dispatched. Schools were asked to complete the remaining information for each eligible mathematics class in the school: the name of the class; the grade-level, and whether it is a mixed grade; whether the students are grouped by ability; the number of students in the class; whether the class is a “special” class that should be excluded from the assessment; and, the name(s) of the mathematics teacher(s) for the class.

These forms were emailed to schools in the form of Excel files which schools could complete. Westat developed a secure site designed to allow schools to provide the completed form electronically via a secure upload and provided help-desk support for schools. This procedure was trial-tested on some 40 schools. It became clear within a week that the schools were not responding as anticipated. On this basis the decision was made to revert to hardcopy forms dispatched via Fedex to all schools. Schools were given the choice of sending back the completed form by either faxing it to a secure fax site, or sending it back via Fedex. Responses began to flow back immediately. Schools not responding within a few days were followed up with emails and telephone calls to the School Coordinator. (Notably, a few schools requested the form as a spreadsheet that could be filled in on-line and sent back via email. These schools were sent the original Excel form.)

Processing the Class Listing Forms. Receipt of the CLF was tracked using a ledger that included the return status of each form and detailed any anomalies. The status of schools was reviewed weekly. As CLFs were returned, they were processed as follows:

- The forms were reviewed for clarity and completeness;
- Any incomplete forms generated calls to the schools in question to provide the information or clarification necessary;
- Completed forms were passed to data entry staff where the information was entered into WinW3S, and a CLF in WinW3S system format was generated; and,
- The WinW3S CLF was compared with the original form as a check on the accuracy of data entry.

4.2.2 Sampling classes

WinW3S generated an equal probability sample of two classes (or pseudoclasses) where possible, based on the information in the CLF. If only one grade-appropriate class was available in a school, this class was selected with certainty. Subsequently, a Class Sampling Form (CSF) was generated which identified the selected classes and assigned class IDs. An example of a fictitious CSF form is reproduced below in exhibit 4.2.

Exhibit 4.1. Example of Class Listing Form

CLASS LISTING FORM

SCHOOL ID	SCHOOL NAME	SCHOOL COORDINATOR	GRADE
9950	SMALLTOWN SCHOOL	CINDY JENKINS	8

PAGE _____ OF _____

PLEASE LIST ALL MATHEMATICS CLASSES THAT HAVE ANY 8th-GRADE STUDENTS IN THEM

	CLASS NAME (e.g., 8A, Jefferson, red8, etc)	GRADE (Or, if mixed-grade, enter 'M')	IS THIS CLASS IS GROUPED BY ABILITY? (0= not grouped 1= low ability 2= average ability 3= high ability)	NUMBER OF 8TH-GRADE STUDENTS IN CLASS (Or, if mixed grade, how many 8th-grade students?)	IF THIS IS A SPECIAL CLASS, ENTER (0 = not a special class 1 = functional disabilities 2 = intellectual disabilities 3 = LEP)	MATHEMATICS TEACHER'S NAME Please do not include teacher aides (first)	(last)	2ND MATHEMATICS TEACHER'S NAME Please do not include teacher aides (first)	(last)
1	8A	8	0	24	0	IRIS	CHAVEZ		
2	JEFFERSON	8	0	12	0	JIM	SMITH		
3	RED8	8	0	26	0	MARY	ROBBINS	JAMES	MARCH
4	BLUE8	8	0	14	0	BRUCE	EVANS		
5									
6									
7									
8									
9									
10									
11									
12									

PLEASE MAKE A COPY OF THIS FORM IF YOU NEED TO LIST ADDITIONAL CLASSES

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Exhibit 4.2. Example of Class Sampling Form

TIMSS 2007 - Class Sampling Form

School Name: Smalltown School

TIMSS Participant: UNITED STATES

Page 1 of 1

[a] School ID	[b] Grade	[a] School ID	[b] Grade	[c] Average Class Size	[d] Random Number	[e] Selections	[f] Sampling Interval	[g] Random Start
9950	8	9950	8	21	0.5436	2	2	2.1045

(1) Class Name	(2) Class ID	(3) Class ID	(4) Stream (sorted)	(5) Number of Students	(6) Excluded	(7) Pseudo-class	(8) Cum. Count	(9) Line Number	(10) Selected
8A	995001	995001	1	21				1	Yes
Jefferson	995002	995002	1	23				2	
RED8	995003	995003	1	26				3	
BLUE8	995004	995004	1	14				4	Yes

Use additional sheets if necessary

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

4.2.3 Identifying students and their teachers

With the classes identified, a Student-Teacher Linkage Form (STLF) was generated for each of the sampled classes. These forms were created in Excel using information from the CLF. An example of an STLF containing fictitious information is provided in exhibit 4.3 below.

The upper part of the form containing the school ID, school name, class ID, School Coordinator name and grade was completed using information from the Class Listing Form. The remainder of the form was to be completed by the school, beginning with a listing of the names of the students in the named class. Schools were then asked to work across the form line-by-line to add in the information relating to each student and that which related to the mathematics and science teacher(s) of each student.

As a function of the size of the form it was printed on 11x17 inch paper. Copies of the forms sufficient to describe each of the sampled classes, along with instructions and examples, were sent to each school coordinator. A prepaid FedEx return envelope was provided for the return of the completed forms.

All things considered the response by schools to completing this somewhat formidable task was encouraging, though it required more follow up than with the simpler Class Listing Form. The return of the STLF forms was tracked and reviewed weekly with the view to initiating reminder calls from field staff to the schools.

4.2.4 Student Tracking Forms (STF)

Information from the STLF was entered into the WinW3S system which generated a Student Tracking Form (STF) for each class. The student tracking form was designed to provide Field Supervisors with student IDs, student identifiers in the form of date of birth and sex, the booklet assignment to each student, and the means to record the completion of the assessment and associated questionnaire. The student names or IDs shown in the first column of the form were removed following the assessment and retained by the school as a means of ensuring confidentiality. An example of a STF containing fictitious information is provided in exhibit 4.4 below.

4.2.6 Teacher Tracking Form (TTF)

A Teacher Tracking Form was also generated from information provided on the STLF. This form was designed to provide Field Supervisors with a means to record the participation of teachers and to ensure that each teacher received the correct Teacher Questionnaire. Teacher names provided in the TTF were removed following the assessment and retained by the school as a means of ensuring confidentiality. An example of a TTF containing fictitious information is provided in exhibit 4.5 below.

Exhibit 4.3. Example of Student Teacher Linkage Form

TIMSS2007 STUDENT-TEACHER LINKAGE FORM PAGE 1 of 2

SCHOOL ID	SCHOOL NAME	CLASS ID	CLASS NAME	SCHOOL COORDINATOR	GRADE
9950	SMALLTOWN SCHOOL	995001	8A	CINDY SULLIVAN	8

PLEASE TELL US ABOUT THE 8TH GRADE STUDENTS IN THE 8A MATHEMATICS CLASS

SCROLL DOWN TO LIST MORE STUDENTS

	STUDENT NAME		DATE OF BIRTH		GENDER	DOES THIS STUDENT TAKE MATH?	DOES THIS STUDENT TAKE SCIENCE?	THIS STUDENT SHOULD NOT BE ASSESSED BECAUSE THEY HAVE:	WHICH STUDENTS ARE TAUGHT MATHEMATICS BY IRIS CHAVEZ	WHICH STUDENTS ARE TAUGHT MATHEMATICS BY <2ND TEACHER'S NAME>
	(first)	(last)	month (mm)	year (yyyy)	F = female, M = male	Y = yes N = No	Y = yes N = No	Enter 1 = functional disability 2= intellectual disabiity 3= LEP 0 = not disabled	Place an X against the name of each student taught by this teacher	Place an X against the name of each student taught by this teacher
1	STUART	HUGHES	7	1993	M	Y	Y	0	X	
2	FRANK	BILLINGS	2	1993	M	Y	Y	0	X	
3	ROBIN	JAVIER	5	1992	F	Y	N	1	X	X
4	LESLEY	MILES	9	1992	F	Y	Y	0	X	
5	CHARLES	HUANG	4	1993	M	Y	Y	0	X	
6	MARY	SILVERS	3	1990	F	Y	Y	0	X	
7										
8										
9										
10										

See note at end of table.

Exhibit 4.3. Example of Student Teacher Linkage Form—Continued

TIMSS2007 STUDENT-TEACHER LINKAGE FORM **PAGE 2 of 2**

SCHOOL ID	SCHOOL NAME	CLASS ID	CLASS NAME	SCHOOL COORDINATOR	GRADE
9950	SMALLTOWN SCHOOL	995001	8A	CINDY SULLIVAN	8

PLEASE TELL US ABOUT EACH OF THE SCIENCE COURSES TAKEN BY THE 8TH GRADE STUDENTS IN THE 8A MATHEMATICS CLASS

SCROLL RIGHT TO SEE MORE COURSES →

	Name of Science Class/Course #1 (e.g., science 8A, biology, animal kingdom, etc)	Name of Science Class/Course #2 (e.g., science 8A, biology, animal kingdom, etc)	Name of Science Class/Course #3 (e.g., science 8A, biology, animal kingdom, etc)	Name of Science Class/Course #4 (e.g., science 8A, biology, animal kingdom, etc)	Name of Science Class/Course #5 (e.g., science 8A, biology, animal kingdom, etc)	Name of Science Class/Course #6 (e.g., science 8A, biology, animal kingdom, etc)
1. Name of class/course (Please enter name)	SCIENCE 8A	IB SCIENCE	INTRO TO SCIENCE			
Teacher Name (first)	MARIA	GORDON				
Teacher Name (last)	ALETTO	REMMERS				

2. Name of teacher (Please enter name)

3. Please tell us which students take this science class/course.

SCROLL DOWN TO SEE MORE STUDENTS ↓

	STUDENT NAME (first)	STUDENT NAME (last)	Place an X against the name of each student taking this course with this teacher	Place an X against the name of each student taking this course with this teacher	Place an X against the name of each student taking this course with this teacher	Place an X against the name of each student taking this course with this teacher	Place an X against the name of each student taking this course with this teacher	Place an X against the name of each student taking this course with this teacher
1	STUART	HUGHES	X					
2	FRANK	BILLINGS		X				
3	ROBIN	JAVIER	X					
4	LESLEY	MILES		X				
5	CHARLES	HUANG						
6	MARY	SILVERS		X				
7								
8								
9								
10								
11								
12								
13								

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Exhibit 4.4. Example of Student Tracking Form

TIMSS 2007 - Student Tracking Form

School Name: SMALLTOWN SCHOOL

TIMSS Participant: USA

Page 1 of 1

[a] School ID	[b] Class Name	[a] School ID	[b] Class Name	[c] Class ID	[d] Grade	[e] Random Number	[f] First Booklet
9950	8A	9950	8A	995001	8		4

(1) Student Name or Number	(2) Student ID	(3) Check Sum	(4) Exclu-sion Status	(5) Date Of Birth (MM/YY)	(6) Sex	(7) Book-let	(8) Participation Status Session ACH	(8) Session BG	(9) Make-Up ACH	(9) Make-Up BG
	99500101	628		7/93	1	4				
	99500102	331		2/93	1	B1 15				
	99500103	305		5/92	2	5				
	99500104	538		9/92	2	6R				
	99500105	34		4/93	1	7				
	99500106	793		3/90	2	B2 16				

Use additional sheets if necessary

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

Exhibit 4.5. Example of Teacher Tracking Form

TIMSS 2007 - Teacher Tracking Form

School Name:

SMALLTOWN
SCHOOL

TIMSS Participant:

[a] School ID	[b] Grade	[a] School ID	[b] Grade

[1] Class (Course) Name	[2] Teacher Name	[3] Teacher ID	[4] Link No	[5] Check Sum	[6] Selected Class ID	[7] Class (Course) Name	[8] Subject Code	[9] Eligible Students	[10] Questionnaire	[11] Return Status
8A	IRIS CHAVEZ	995001	01	896	995001	8A	1	21	M	
SCIENCE 8A	MARIO ALETTO	995002	02	498	995001	SCIENCE 8A	6	2	S	
IB SCIENCE	GORDON REMMERS	995003	03	735	995001	IB SCIENCE	6	3	S	

Use additional sheets if necessary

SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2007.

4.3 Instrumentation

All TIMSS instrumentation was developed by the IEA as a cooperative effort involving representatives from every country participating in the study. At each grade, the primary instrumentation for TIMSS 2007 consisted of: a combined mathematics and science assessment; a student questionnaire; a school questionnaire; and, teacher questionnaires to be completed by the teachers teaching mathematics and/or science to the students in the sampled classrooms.

The assessment instrument took the form of 14 booklets containing mathematics and science items in rotated item blocks. While some of these items have to be held in confidence for use in coming TIMSS assessments, others have been released in order to provide actual examples of assessment items. TIMSS 2007 items released by the IEA are available on the TIMSS and PIRLS International Study Center website: http://timssandpirls.bc.edu/TIMSS2007/items.html.

The student questionnaire was grade-specific and bound into the assessment booklet with the assessment items for the particular grade. The school questionnaire was designed to be completed by the school principal. It differed a little between fourth-grade and eighth-grade schools in that the eighth-grade questionnaire differentiates between mathematics and science teachers with regard to the evaluation of teaching practice. The fourth-grade questionnaire made no such distinction. In all other respects, the two questionnaires were identical. In the case of the teacher questionnaires, a single fourth-grade questionnaire was designed to be completed by the mathematics and/or science teachers of the students in the sampled classrooms. By contrast, separate questionnaires were provided for the mathematics teachers and science teachers of the students in the sampled eighth-grade classrooms. The international versions of the questionnaires are available on the TIMSS and PIRLS International Study Center website http://timssandpirls.bc.edu/TIMSS2007/context.html. The U.S. versions of all seven U.S. questionnaires are provided in appendix E and are available on the NCES website http://nces.ed.gov/timss.

4.3.1 The TIMSS Mathematics and Science Assessment

The following discussion provides a summary of the rationale for and development of the TIMSS 2007 assessment. Complete detail is provided in the *TIMSS 2007 Technical Report* (Olson, Martin, and Mullis 2008).

Assessment frameworks. For TIMSS 2007, the test development effort began with a revision of the frameworks used to guide the construction of the TIMSS 2003 assessment (Mullis et al. 2005). The frameworks were updated to reflect changes in the curriculum and instruction of participating countries. Extensive input from experts in mathematics and science education, assessment, and curriculum, and representatives from national educational centers around the world contributed to the final shape of the frameworks. Maintaining the ability to measure change over time was an important factor in revising the frameworks.

Content and cognitive domains. The TIMSS mathematics and science assessments were designed along two dimensions: the topics or content that students are expected to learn, and the cognitive skills students are expected to have developed. In TIMSS 2007 the cognitive domains are *knowing*, *applying*, and *reasoning* and are common to each grade and subject area. The content areas differ between grades and subject areas. The mathematics topical or content domains covered at grade four are *number, geometric shapes and measures*, and *data display*. At grade eight, the mathematics content domains are *number*, *algebra*, *geometry*, and *data and chance*. The science content domains covered at grade four are *life science*, *physical science*, and *Earth science*. At grade eight, the science content domains are *biology*, *chemistry*, *physics*, and *Earth science*.

Item development. Approximately one-half of the assessment items used in TIMSS 2003 were held in confidence and included in the 2007 assessment. To replace assessment items that had been released

following the 2003 assessment, countries submitted items for review by subject-matter specialists, and additional items were written by the IEA Science and Mathematics Review Committee in consultation with item-writing specialists. This expert consultation was designed to ensure that the content, as explicated in the frameworks, was covered adequately. Items were reviewed by the Science and Mathematics Item Review Committee and field-tested in most of the participating countries. Results from the field test were used to evaluate item difficulty, how well items discriminated between high- and low-performing students, the effectiveness of distracters in multiple-choice items, scoring suitability and reliability for constructed-response items, and evidence of bias toward or against individual countries or in favor of boys or girls. As a result of this review, 196 new fourth-grade items were selected for inclusion in the international assessment. In total, 353 mathematics and science items were included in the fourth-grade TIMSS assessment booklets. At the eighth grade, the review of the item statistics from the field test led to the inclusion 240 new eighth-grade items in the assessment. In total, 429 mathematics and science items were included in the eighth-grade TIMSS assessment booklets. More detail on the distribution of new and trend items can be found in Olson, Martin, and Mullis (2008).

Assessment booklets. To keep the testing burden to a minimum while ensuring broad subject-matter coverage, TIMSS used a rotated block design that included both mathematics and science items. The rotated item design meant that, while no students responded to all of the items, each student encountered both mathematics and science items during the assessment.

The 2007 fourth-grade assessment consisted of 14 booklets, each requiring approximately 72 minutes of response time. To ensure that TIMSS 2007 maintained the trend established by previous TIMSS assessments, and to provide for a correction through equating if necessary, four additional “bridge” booklets were required but only for countries that participated in TIMSS 2003, of which the United States was one. These bridge study booklets were identical to booklets used in 2003. Performance on the bridge booklets did not contribute to the overall score for TIMSS 2007 but the data were used in the trend scaling. This aspect of scaling placed the 2007 results on the same scale as previous TIMSS assessments and so allowed for comparisons across the years.

For the United States and other countries that participated in the 2003 assessment, the additional four booklets meant a total of 18 booklets in all. The 18 booklets were rotated among students, with each participating student completing 1 booklet only. In the TIMSS 2007 booklets the mathematics and science items were each assembled separately into 14 blocks, or clusters, of items. Each block contained either mathematics items or science items only. The secure trend items used in prior assessments were included in 3 blocks, with the other 11 blocks containing new items. Each of the 14 TIMSS 2007 booklets contained 4 blocks in total. The 4 additional bridge study booklets from TIMSS 2003 contained 6 blocks of items each. All 18 booklets were completed within the timeframe indicated in section 4.4.6 below.

The 2007 eighth-grade assessment followed the same pattern and consisted of 18 booklets, each requiring approximately 90 minutes of response time. The 18 booklets were rotated among students, with each participating student completing 1 booklet only. The mathematics and science items for TIMSS 2007 were assembled into 14 blocks of items. Each block contained either mathematics items or science items only. The secure trend items used in prior assessments were included in 3 blocks, with the other 11 blocks containing new items. Each of the 14 TIMSS 2007 booklets contained 4 blocks in total. The 4 additional bridge study booklets from TIMSS 2003 contained 6 blocks of items each. Performance on the bridge booklets did not contribute to the overall score for TIMSS 2007 but the data were used in the trend scaling. This aspect of scaling placed the 2007 results on the same scale as previous TIMSS assessments and so allowed for comparisons across the years. All 18 booklets were completed within the timeframe indicated in section 4.4.6 below.

4.3.2 Calculator usage

Calculator use was not permitted during the TIMSS fourth-grade assessment. However, the TIMSS policy on calculator use at the eighth grade was to give students the best opportunity to operate in settings that mirrored their classroom experiences. As a consequence, calculators were permitted, but not required, for the eighth-grade assessment materials. In the United States this meant that students assigned one of the 14 TIMSS 2007 booklets were allowed, but not required, to use calculators. However, students assigned one of the trend (or "bridge") booklets from the 2003 assessment were required to follow the 2003 rules in this respect. These students could use a calculator only for the second section of the booklet.

4.3.3 Background questionnaires

As in prior administrations, TIMSS 2007 included self-administered questionnaires for principals, teachers, and students. To create the questionnaires for 2007, the 2003 versions were reviewed extensively by the national research coordinators from the participating countries as well as a Questionnaire Item Review Committee (QIRC). Based on this review, the QIRC eliminated or revised some questions, and added several new ones. Like the assessment items, all questionnaire items were field-tested and the results reviewed. As a consequence, some of the questionnaire items needed to be revised prior to their inclusion in the final questionnaires. The questionnaires requested information to help provide a context for the performance scores, focusing on such topics as students' attitudes and beliefs about learning, their habits and homework, and their lives both in and outside of school; teachers' attitudes and beliefs about teaching and learning, teaching assignments, class size and organization, instructional practices, and participation in professional development activities; and principals' viewpoints on policy and budget responsibilities, curriculum and instruction issues and student behavior, as well as descriptions of the organization of schools and courses. Detailed results from the student, teacher, and school questionnaires are available in the two international reports: the *TIMSS 2007 International Mathematics Report* (Mullis, Martin, and Foy 2008) and *TIMSS 2007 International Science Report* (Martin, Mullis, and Foy 2008).

4.3.4 U.S. adaptations to the assessment items and questionnaires

Source versions of all instruments (assessment booklets, questionnaires, and manuals) were prepared by the IEA in English and translated by the participating countries into the primary language or languages of instruction in each country. In addition, it was sometimes necessary to adapt the instruments for better fit to language usage even in countries that use English as the primary language of instruction. Other adaptations to fit national education characteristics were sometimes required as well. All adaptations were reviewed and approved by the IEA to ensure they did not change the substance or intent of the question or answer choices.

U.S. adaptations to the assessment items. As in previous cycles of TIMSS, the U.S. adaptations to the international instruments were minimal and designed to make the assessment more readable to U.S. students without changing the essence of the assessment item. For example, at times names of individuals were changed to more familiar forms (for example, "Ahmed" to "Andrew"), nouns with British origins were changed to their U.S. equivalent (for example, "cinema" to "movie theater"), Imperial English spellings were changed to American English (for example, "organisation" to "organization", "programme" to "program") , and some changes to the text of instructions were also made to better mirror the administration procedures in the U.S.

U.S. adaptations to the school, teacher and student background questionnaires. Adaptations made to the school, teacher and student questionnaires were of five main kinds:

- Changes to general instructions made in the interests of enhancing clarity;
- Changes designed to make question text more readable to U.S. students, similar to those made to the assessment items as described above;
- Changes to response alternatives where the international response set did not adequately reflect the U.S. context;
- Additional questionnaire items included to address particular issues of national interest; and
- International items that were omitted from the U.S. questionnaires because they violated the federal Pupil Privacy Rights Act (for example, questions on bullying and violence in the school).

A detailed list of changes made to the questionnaires is provided in appendix F. Both the original text from the international version of the questionnaire and the changed text from the U.S. version are shown. Text that has been changed in the U.S. version is underlined in that version. Both international and U.S. questionnaire item numbers, or other location indicators, are provided in each instance. Where appropriate, a crosswalk between the U.S. and international versions of the set of response categories of items is provided in the "Comments" column.

4.3.5 Translation and verification of instruments

Each country prepared translations of the instruments according to translation guidelines established by the IEA. Since the international versions of the instrumentation are produced in English, the United States did not need to engage in the full-fledged translation required of many nations. However, the adaptations made to the U.S. instruments required verification by the IEA to ensure their suitability for the current cycle of TIMSS and, if trend items, their continuity with previous cycles. Further details on the translation process can be found in Olson, Martin, and Mullis (2008).

4.3.6 Production of assessment booklets and questionnaires

On receiving IEA approval of the adaptations, Pearson applied the adaptations to the international questionnaires and item blocks and assembled the final assessment booklets. Quality control procedures for this process included a review of each adaptation made to the questionnaires and item blocks as well as a full review of the assembled instruments in blue-line layout.

In mid-December 2006, electronic files were sent to the IEA DPC for verification of the national changes and to the TIMSS and PIRLS International Study Center for layout verifications. The student assessment booklets and student questionnaires were printed in scannable form. The teacher and school questionnaires were printed in non-scannable form.

Scannable document preparation and printing. The student test assessment and student questionnaire were printed as scannable documents combined into 14 single booklets. Document production was divided into two phases. The preparation phase included the mockup and design of forms, typesetting, composition and editing for text accuracy and processability. The production phase included final platemaking, printing, binding and any finishing procedures (for example, counting, wrapping, etc.) that were required prior to the packaging and distribution processes.

After a form was created, it was thoroughly inspected for grammar, spelling and punctuation to ensure that it matched the approved electronic files. Subsequently, copies of these proofs were subject to a technical review of scan-ability, oval placement, and spine code assignment. Immediately after printing, sample documents were selected from predetermined locations throughout the print run for testing. Before shipping, a sample from each carton of multi-page documents was inspected to ensure that the pages of the booklets were in the correct sequence. After binding, all documents were boxed to assure that material quality was maintained during transit.

Preparation and printing of non-scannable documents. The teacher and school questionnaires were produced as non-scannable documents. In a first stage proofs of each document were reviewed against the original electronic files. Once accuracy was certified, printing was initiated. During this process staff checked a ten percent sample of the printed form against the approved document to ensure that accuracy was maintained throughout the printing process.

4.4 Field operations

The activities discussed under this heading refer to those associated with the administration of the assessment in participating schools. In the United States the administration of the assessment was carried out by professional staff trained according to the international guidelines. School personnel were asked only to assist with listings of students, the identification of school space for the assessment, and the specification of parental consent procedures, if any, needed for sampled students.

Field operations centered on three main tasks; recruiting and training of field staff; scheduling the assessments; and, administration of the assessments within the schools.

4.4.1 Recruiting and training field staff

Many of the field supervisors working on the recruitment of districts and schools were retained for the assessment phase of the study. However, it was necessary to recruit additional staff at the field supervisor level from the pool of experienced Westat field staff. All had previous experience with other educational assessments in schools and all had FBI clearance based on fingerprint and background checks. All Field Supervisors also signed a statement of nondisclosure indicating that they would maintain confidentiality of all survey materials and of the data collected. The Field Supervisors recruited local Assessment Administrators to assist in the administration of the assessment since in most schools two separate classrooms were assessed simultaneously. Assessment Administrators were also experienced with the administration of educational assessments in schools, and held the same kind of FBI clearance. They too signed confidentiality statements.

The total complement of field staff consisted of four Field Managers, 25 Field Supervisors and 94 Assessment Administrators. The Field Managers reported directly to the Westat home office and met weekly to discuss progress and any problematic issues arising in the field. Field Supervisors reported to a Field Manager who coordinated and monitored their work and, in turn, the Field Supervisors coordinated and supervised the work of the Assessment Administrators.

4.4.2 Responsibilities of Field Managers, Field Supervisors and Assessment Administrators

Field Managers had responsibility for:

- Providing input to the TIMSS data collection training;
- Tracking the Field Supervisors' receipt of assessment booklets and other materials;
- Coordinating data collection activities undertaken by their assigned Field Supervisors;
- Holding weekly one-on-one telephone meetings with their Field Supervisors to monitor progress and to trouble-shoot any problems arising;
- Ensuring that their Field Supervisors followed TIMSS procedures and guidelines; and
- Reporting progress and problems in weekly conference calls with Westat home office staff and other Field Managers.

Field Supervisors had responsibility for:

- Attending TIMSS data collection training;
- Receiving and securing assessment materials;
- Training Assessment Administrators;
- Preparing and assigning assessment materials for students;
- Collecting the completed school and teacher questionnaires;
- Conducting the assessment according to TIMSS specified procedures;
- Completing the Test Administration Form, Student Tracking Form, Teacher Tracking Form, and Student Response Rate Form;
- Determining if a follow-up session was needed;
- Securing, packing, and shipping all assessment materials to Pearson at the conclusion of the assessment;
- Recording the status of the assessment in the Field Management System; and
- Reporting progress to their Field Manager on a regular basis.

Assessment Administrators had responsibility for:

- Attending the TIMSS training conducted by their Field Supervisor;
- Administering the assessment according to TIMSS specified procedures;
- Completing the Test Administration Form, Student Tracking Form, Teacher Tracking Form, and Student Response Rate Form; and,
- Consulting regularly with their Field Supervisor.

4.4.3 Training

A 2-day, in-person training for Field Supervisors was held on March 23 and 24. The attendees received a Field Supervisor manual 5 days prior to the training session and were given 4 paid "study hours" to become familiar with the information prior to training. The agenda for this training session is provided as exhibit D.1 in appendix D.

The first day of training focused on the responsibilities of a TIMSS Field Supervisor, TIMSS data collection materials from the perspective of item security and student privacy, the Pre-Assessment Call checklist, and assessment day activities. The second day included discussion about the procedures to be followed once the assessment was completed, and the appropriate methods for packaging and shipping the assessment materials to Pearson. Subsequently, the focus shifted to the administrative duties of the Field Supervisors and instruction on the use of the Field Management System.

Field staff were assigned laptop computers to take with them for the duration of TIMSS data collection. Printers for use with the laptops were issued and sent to each Field Supervisor after training. Field Supervisors were also provided with an official TIMSS 2007 photo-ID badge to wear while representing TIMSS in the schools.

Training for Assessment Administrators (AA) was conducted as a separate exercise by their respective Field Supervisor. Approximately two hours was allocated for this training. Training materials consisted of an Assessment Administrator Manual, session scripts, and a sub-set of material from the Field Supervisor training.

4.4.4 Scheduling assessments

The 496 schools taking part in TIMSS 2007 were dispersed across the country though with concentrations in the more populous states and cities. Scheduling assessment dates for these schools required the optimization of school preferences for a particular date with the assessment date preferences of nearby schools and the location of field staff in an effort to keep travel and related expenses to a minimum. The basic approach adopted involved the geographic mapping of schools and their preferred assessment dates, along with the location of field staff. This formed the foundation for discussions with schools, and the assignment of schools to field staff. In essence, geographical clusters of schools were assigned assessment dates clustered in time insofar as this was possible.

Obtaining assessment dates from schools. At the time of recruitment, schools were asked to provide basic contact information. Westat field staff confirmed the school name and school address and identified a School Coordinator. Schools were also asked to indicate preferred assessment dates or date ranges as well as black-out dates associated with state testing, holidays, breaks, field trips, and the like. This information was recorded on School Information Worksheets.

Mapping schools. In a second step a geographic map of participating schools in the fourth- and eighth-grade samples was developed based on the addresses of the schools. The map allowed the ready identification of the more obvious clusters of schools in large metropolitan areas, and at the same time allowed the identification of less concentrated clusters spread across broader geographical areas, as well as isolated schools with no obvious relationship to any cluster. With the view to defining optimal work areas for assessment administrators, geographic clusters of schools were subsequently defined judgmentally in an iterative process which took into account the location of field staff and their case load. Some 40 work areas of varying size were identified in this way.

Assigning tentative assessment dates. The provisional assessment dates provided by schools were then mapped to the schools in each cluster and represented as a spreadsheet showing both preferred dates and black-out dates in each defined work area. Within clusters, preferred assessment dates were balanced against the location of schools in relation to one another. With some minor modifications to work areas, schools in relatively close proximity were assigned to provisional assessment weeks. This process identified smaller clusters of schools that could be assessed within the one week without undue travel and related expenses. For the most part, the intent was to schedule assessments on Tuesday through Thursday leaving Monday and Friday as travel days.

Negotiating final assessment dates. Once tentative dates were established, field staff contacted schools in an effort to negotiate a final assessment date. For the most part, schools accepted the dates proposed. However, when this was not possible, field staff used the particular work area spreadsheet and a work area map to negotiate an assessment date suitable to the school and consistent with a minimum of staff travel. In each case the final assessment date and time was confirmed by email with the School Coordinator.

4.4.5 Assignment of schools to Field Supervisors

For the most part, field staff were assigned a work area based on their location and availability with most work areas relatively close to the Field Supervisor's home address. The balancing of these several demands resulted in some variation in the case load of Field Supervisors. During the course of the assessments, which ran from March 2007 through to June 2007, some reassignments of schools and work areas became necessary.

Trouble-shooting. Four of the Field Supervisors were recruited as trouble-shooters on a regional basis. In this way provision was made to cover last minute changes in assessment dates, staff illnesses and the like. Each of these staff had assignments of their own as well, essentially schools that did not easily fit into an identified work area or were unable to schedule an assessment on dates similar to those of other schools in the work area.

4.4.6 The assessment and related activities

Field staff engaged in a number of activities in advance of the actual assessment. These included:

- Working with the School Coordinator to gain the permission of parents and students, if this was required by the school;
- Making arrangements with the school for the assessment sessions; and,
- Obtaining the materials to be used in the assessment.

Recruiting parents and students. During recruitment and scheduling contacts with schools, field staff asked about district and school requirements for notifying parents about their child's participation in TIMSS. School requirements fell into three main categories:

1. **Notification –** The school would simply send parents a notification of the child's participation in TIMSS along with some informational material;

2. **Passive consent –** The school was required to ask parents for permission for the child to participate but permission would be assumed unless there was a formal objection; and,

3. **Active consent –** The school was required to ask parents for permission for the child to participate and the child could not participate until the parents provided formal approval.

A majority of schools (190) opted for parent notification. Some 138 schools used passive consent and a further 32 used active consent. In the remaining schools parent notification/permission was not required.

To assist schools in this task the School Coordinator was provided with three draft letters to parents, one for each of the three forms of parent permission. The intent was that these letters could be edited as appropriate and sent out on school stationery. Consent forms to accompany the passive and active consent letters were provided as well along with an information sheet describing TIMSS. English and Spanish versions of each of these documents were made available to the schools. Copies of these materials as they apply to fourth-grade students are provided as exhibits B.5 through B.7 in appendix B. The eighth-grade materials were essentially the same.

Organizing the assessment session at the school. Approximately 2 weeks prior to each school's assessment date, the Field Supervisor called the assigned School Coordinator. Using a Pre-Assessment Call Checklist, Supervisors were instructed to verify previously obtained information on items such as the school's address, Principal's name, assessment date, session location, requirements for entering the school, parking arrangements and the like. The information obtained was maintained within the School Folder and updated as a basic reference. The Pre-Assessment Call Checklist is provided as exhibit D.2 in appendix D.

On assessment day each Field Supervisor, accompanied by an Assessment Administrator, arrived at the school with all of the materials needed for the assessment. One "session box" of materials was provided for each of the sampled mathematics classes. Each session box contained the estimated number of student assessment booklets required, plus three unassigned booklets to accommodate any changes in class enrollments. Upon arrival, the Field Supervisor met with the School Coordinator to make any updates to the Student Tracking Form that would impact the preparation of student materials (for example, the addition of new students, the withdrawal of listed students from the school or class, or a change in exclusion status of a sampled student).

Field Supervisors arrived at the school about an hour before the scheduled assessment to prepare booklets for distribution and take care of other arrangements for the assessment. Following the prescribed international procedures, the supervisor did not open the booklet bundles until 45 minutes before the assessment. At that time, the booklets were assigned to students in the random order established by the IEA sampling software, labels were placed on the booklets, and the manipulatives associated with the assessment inserted in the appropriate booklets. TIMSS pencils were provided to all students. As required by international rules, simple-function calculators were provided at the eighth-grade level only, and then only at the discretion of the school. Students kept the TIMSS pencils as gifts, but manipulatives and calculators distributed during the assessments were collected with the booklets after the assessment.

Administering the assessment. Assessments were administered by reading verbatim from a standardized script according to the instructions set forth in the international TIMSS Test Administrator

Manual. A copy of the fourth-grade script is provided as exhibit D.3 in appendix D. The eighth-grade session script is essentially the same. The script began with a brief introduction to the study, following which the assessment booklets were distributed. Immediately following this distribution, the students were instructed to begin Part 1 of the assessment. After 36/45 minutes (grade 4/grade 8), a short break was provided. After the break, students were instructed to begin Part 2 of the assessment. After the allotted 36/45 minutes for this part of the assessment, students were instructed to stop work. A longer break was provided at this time, following which the student questionnaire was administered. The questionnaire part of the assessment was not time-limited, but was usually completed in about 30 minutes.

Post-assessment activities. Following the assessment, supervisors instructed students to seal the booklets with a security seal and remove an identifying label from the cover of the booklet. Students then handed their booklets to the supervisor, received their gift, and were dismissed. The Field Supervisor then recorded participation codes for each session, following which the booklets along with the school and teacher questionnaires were packed into the shipping box. In the next-to-final step, the Field Supervisors made copies of the Student Tracking Form and the Session Report Form, placing the original documents in the TIMSS Storage Envelope to be kept at the school. In order to maintain the security of student names the Field Supervisor removed and destroyed the column of student names from the copies of the STF and placed the de-identified STF and Session Report Form in with the assessment booklets and questionnaires. In the final step, the session materials were sealed and shipped to Pearson via Fedex for processing.

4.5 Receipt control, scoring, coding, and data entry

Field staff sent the completed assessments and questionnaires along with any related materials directly to Pearson following the completion of the assessment session at a school. Pearson then recorded the receipt of materials, scored the open-ended responses in the assessment, coded the multiple-choice assessment items and the questionnaire responses, and created data files from this information.

4.5.1 Receipt control

Approximately 11,500 grade four assessment booklets, 11,000 grade eight assessment booklets, 900 grade four teacher questionnaires, 1,070 grade eight teacher questionnaires, and 465 school questionnaires were received by Pearson.

Receipt control system specification. Specific procedures were developed to account for materials distributed and returned. Two systems were used to monitor the receipt and processing of materials for this study, the Process Control System (PCS) and the Work Flow Management (WFM) system. These systems enabled project staff to: determine the status of any selected school; verify that materials from a completed school had been received; identify discrepancies in student or school information; and, obtain information on the status of data processing activities for a particular batch of materials.

The Process Control System was modified for use in monitoring the status of all sampled schools and test materials for each study. The PCS contained a list of all participating schools and the completion status of the schools for the TIMSS assessment. Shipments were returned packaged in their original boxes. A bar coded label, applied prior to shipping the booklets out for the assessment, contained the school ID and was scanned upon arrival. A receipt report was run each day and forwarded to the Westat home office on a weekly basis.

When shipments were received staff checked each shipment to verify that the contents of the box matched the school indicated on the label. Each shipment was then checked for completeness and accuracy based on procedures outlined in the Survey Operations Manual. Any discrepancies were recorded and project staff alerted to determine the cause. Once a shipment was opened and verified as

being complete, the documents were organized into work units and batched. The computerized Work Flow Management System (WFM) was used to track the documents through every processing step, thus enabling project staff to easily locate materials for a particular school.

Booklet accountability. Prior to the distribution of materials, all assessment instruments were read into a file by units of a specific predetermined number of documents called "bundles". This file was used to control distribution by assigning specific bundles to field staff or a particular school. This assignment was recorded in the Materials Distribution System.

When shipments were received, a manual count was made to ensure that all booklets from the original bundle had been returned. The booklets were then submitted for scanning and key entry. The unused booklets were batched and the booklet identification bar code read into a file by the bar code scanner. This file and the processed documents file were compared to the original bundle security file created in the packaging phase before distribution. A list of unmatched booklet identification numbers was printed in a report used to confirm any non-receipt of individual booklets.

After the batches of documents had successfully passed the coding and editing process, they were sent to the warehouse for storage. The storage locations of all documents were recorded on the inventory control system which permits the rapid retrieval of any document, should it be necessary. Unused materials were sent to temporary storage until the study was completed and the data files accepted by the Westat home office, at which time the extra inventory was destroyed.

4.5.2 Scoring the assessment items

The TIMSS assessment items included both multiple-choice and constructed-response items. A scoring rubric developed internationally following the field test of the assessment items was available to guide the scoring of each constructed-response item. In the United States, the scoring of the open-ended student responses according to these rubrics was the responsibility of Pearson.

Training. Two subject-specific Scoring Directors participated in the TIMSS training sessions sponsored by IEA. Materials from these sessions along with additional materials constructed specifically for this purpose were used to train the 12 experienced leaders and supervisors recruited for this work. All were hired based on their experience with similar mathematics and science scoring projects. Training activities for the 109 scorers followed the same routine with supervisors leading each small team reading the item prompt; reading the rubric or scoring guide aloud; reading aloud each of the Anchor papers and explaining the reasoning behind the score; allowing the scorers time to complete the practice papers; review each of the practice papers, and open individual scoring on the electronic Performance Evaluation Network (ePEN). Items were trained on and scored one at a time with the exception of the linked items which were trained together.

Scoring. Each of 12 teams worked on different items. Scorers were able to view only the one active item that they were assigned by their supervisors and did not have the ability to assign scores to other items from any other team. The exception was with the linked items from the "themed" blocks. In this situation a question or questions required reference to a previous answer. For instance, question number four in the block may have required data or the answer from questions one and two in order to answer the question. In this situation, the ePEN set up was designed so that the scorer had access to screens with all of the pertinent responses, again only as assigned by the supervisor. The answers could not be submitted to ePEN until all questions from the set were scored in order to make sure that the scorer had all of the information necessary to score all related items.

Scoring quality was monitored continuously. Using the ePEN system allowed inter-rater reliability reports to be run almost as soon as scoring began. Another monitoring method used was back-reading of already scored responses. This allowed the scoring supervisor to look at responses by category. The scoring

supervisor had the option to pull up responses either by scorer or by score point agreements or splits. This helped the supervisor establish whether there was a need to correct a specific scorer and/or clarify a score point category to the whole team. The scoring supervisor could also print scoring responses in order to create calibration sets which are used as a review tool. Scoring supervisors also checked completion statistics.

Cross-country scoring reliability study. Responses to TIMSS items from Southern Hemisphere countries were sent to the Northern Hemisphere countries for scoring. After the scoring of the U.S. responses for an item with their team, two scorers then completed scoring of the international responses that were pre-loaded on desktop computers. After training with and successfully scoring a portion of the 2007 responses, two scorers were asked to score all the responses for an item. While the individuals making up the pairs varied, no scorer worked on items that they had not been trained on.

Trend scoring. The trend data for fourth and eighth grades were provided to the content leaders prior to training and scoring activities. The responses were loaded onto two desktop computers. After training and successfully scoring a portion of the 2007 responses, two scorers were asked to score the responses from 2003 for each individual item as their team completed working on the current year's responses. The two scorers were assigned to score either the first or second scorer pool of responses until completed at which time the next scorer would begin.

4.5.3 Scanning the student questionnaire responses

Responses to the student questionnaires items were scanned using optical-scanning equipment which also captured images of the constructed-response items and intelligent character recognition (ICR) fields. The data values captured from booklets and questionnaires were coded as numeric data. Unmarked fields were coded as blanks and the editing staff was alerted to missing or uncoded critical data. The images of constructed-response items were saved as a digitized computer file. In addition to capturing the student responses, the bar code identification numbers used to maintain process control were decoded and transcribed to the TIMSS computerized data file. The ICR engine was used to read various hand and machine printing on the assessment booklets and student questionnaires. Image clips of the fields on the booklet covers were displayed to on-line editing staff for verification. Image clips were also taken for the open-ended items to be used for scoring.

4.5.4 Key-entry of teacher and school questionnaire responses

The teacher questionnaire data were entered into a data file using standard key entry methods. The data entry screens were set up based on the international codebook. A second person keyed the same data into a verification file. The original data and the verified data were then programmatically compared and discrepancies were corrected on the spot. The key-entry quality rate was 99.95 percent. The school questionnaire data were keyed directly into the WinDem software provided by the IEA International Data Processing Center (IEA International Data Processing Center 2006b).

4.5.5 Data validation

Student booklets. Each data set produced by the scanning system was validated for type and range of response. The data-entry and resolution system used was able to simultaneously process a variety of materials from all age groups, subject areas, and assessment booklets as the materials were submitted to the system from scannable media.

The data records in the scan file were organized in the same order in which the paper materials were processed by the scanner. As the program processed each record within a batch from the scan file, it

wrote the edited and reformatted data records to the pre-edit file and recorded all errors on the edit file. The program generated an on-line edit file of the data problems and resolution guidelines. Image clips requiring edits were routed to on-line editing stations for the imaged scanned documents.

All data values that were out of range were read “as is” but were flagged as suspect. All data fields that were read as asterisks (*) were recorded on the on-line edit file. Since the asterisk code indicated a double-response, these items were identified for possible resolution by editing staff. Each field was validated for range response and any values outside of the specified range. Corrections were made immediately. The system employed an edit/verify system which meant two different people viewed the same suspect data and operated on it separately. The verifier made sure the two responses (one from either the entry operator or the ICR engine) were the same before the system accepted the item as being correct. If it could not be determined, it was escalated to a supervisor.

When the edit process produced an error-free file, the booklet ID number was posted to the TIMSS tracking file by grade and school. This permitted staff to monitor the TIMSS processing effort by accurately measuring the number of documents processed. The posting of booklet IDs also ensured that a booklet ID was not processed more than once.

Teacher questionnaires. If a teacher questionnaire contained suspect fields, the cover information was recorded on the edit log along with a description of the suspect data. When the entire document was processed, the completed string of data was written to the data file. When the program encountered the end of a file, it closed the dataset and generated an edit listing.

In a second phase of data editing, experienced editing staff used a predetermined set of specifications to review the field errors and record necessary corrections to the student data file. The computerized edit list used in phase one was used to perform this function. The editing staff reviewed the computer-generated edit log and the area of the source document that was noted as being suspect or as containing possible errors. The composition of the field was shown in the edit box. The editing staff checked this piece of information against the TIMSS source document.

The corrected edit log was then forwarded to the key entry staff for processing. When all corrections were entered and verified for a batch, an extract program pulled the corrected records into a mainframe dataset. At this point, the mainframe edit program was initiated. The edit criteria were again applied to all records. If there were further errors, a new edit listing was printed and the cycle was repeated.

School questionnaires. The school questionnaires were manually entered in the WinDem software. If data was recorded outside of the possible ranges, the system immediately alerted the user and would not continue until resolution was made.

4.5.6 File creation and consistency checks

In a final step the data from the assessment score files were merged with the student scanned data. At this time, final output files were produced for each file type. The final files were checked to ensure the data was in the correct format. In earlier editing functions, data was checked for completeness and compliance with codebook specifications. In addition, a check was performed to verify correct linking and matching of student, teacher, and school data files. Student and teacher files were loaded in the WinDem software such that all data from the assessments and questionnaires was available in the format required by IEA.

4.6 Data preparation

The data collected for TIMSS 2007 were entered into data files according to a common international format, as specified in the WinDEM data entry software. The software facilitated the checking and correction of data by providing various data consistency checks

The data files in this format were sent to the IEA Data Processing Center (DPC) where they were subjected to an extensive series of data cleaning and consistency checks. The overriding concern of these checks was to ensure that all information in the database conformed to the internationally defined data structure, national adaptations to questionnaires were reflected appropriately in the codebooks and documentation, and all variables used for international comparisons were comparable across countries.

4.6.1 International data cleaning procedures

The DPC was responsible for checking the data files from each country, applying standard cleaning rules to verify the accuracy and consistency of the data, and documenting electronically any deviations from the international file structure. Queries arising during this process were addressed to national centers and this process repeated as many times as necessary to ensure the data were consistent and comparable within and between countries.

Following this cleaning step countries were provided national univariate and reliability statistics along with data almanacs containing international univariate statistics and item statistics. This allowed countries to examine their data relative to that of other participating nations. Detailed information on the entire data entry and cleaning process can be found in Olson, Martin, and Mullis (2008).

4.6.2 Data confidentiality safeguards

The Education Sciences Reform Act of 2002 explicitly requires that NCES protect the confidentiality of all those responding to NCES-sponsored surveys so that no individual respondent can be identified. More specifically, NCES Standard 4-2, *Maintaining Confidentiality* (NCES 2002), provides guidelines for limiting the risk of data disclosure for data released by NCES. Data disclosure occurs when an individual respondent has been identified through the use of the survey item responses and other external data sources. The following discussion describes the procedures used to reduce the risk of data disclosure for TIMSS 2007, in accordance with the guidelines specified in NCES Standard 4-2.

All students, teachers, and schools participating in the TIMSS do so with the assurance that their identities will not be disclosed. Confidentiality procedures in place included the following: all employees with access to the data signed affidavits of data confidentiality; questionnaires were sealed by students after completion; and, names of students, teachers and schools were removed by field staff from the assessment booklets, the questionnaires and all other related materials, and replaced with unique identification numbers. In addition to data collected directly from schools, teachers and students, additional information was used during the TIMSS sampling, data collection, and weighting processes and these variables too were considered as part of the review to determine disclosure risk levels.

The confidentiality analysis review described below used a three-step process to reduce disclosure risk:

1. Determining the disclosure risk arising from existing external data;
2. Deriving the race data; and
3. Swapping the data.

In this process additional assurance is provided that individual schools, teachers and students participating in TIMSS could be not identified through comparison with public data collections once the TIMSS data are released for public use. While no public data collections identify students or teachers by name, three publicly available data collections do identify schools by name. These are: the Common Core of Data (CCD), a detailed public school listing; the Private School Survey (PSS), a detailed private school listing; and, the QED data collections produced by Quality Education Data, Inc., a privately owned education research firm. The QED data contain a school-based file that provides demographic information for both public and private schools along with the names of the schools. Thus, there is some possibility that schools at least, and perhaps teachers and students as well, could be identified if comparisons of these data sets with the TIMSS data set allowed the identification of schools.

It might be possible to identify TIMSS schools by taking variables from the TIMSS school data and searching the publicly available data files (QED, CCD, and/or PSS files) for schools with a matching profile. However, since the variables in the TIMSS data files were obtained from responses to the school questionnaire, for the most part, exact profile matches are unlikely. Even then, one would not know for certain whether any of the matched schools were the actual TIMSS schools, or whether the match had simply arisen by chance.

Nevertheless, school matching analyses were undertaken using probabilistic matching algorithms approved by the IES Disclosure Review Board (DRB) for use in disclosure analyses. These algorithms identify schools with some potential for identification. In order to provide further protection, elements of the data from schools identified as “disclosure risks” in this way were perturbed using the procedures approved by the DRB. After perturbation, the data were subjected to another round of analyses to ensure that the potential for identification no longer existed.

An additional measure was taken to reduce further the risk of disclosure of an individual respondent. This measure is referred to as “data swapping”, a DRB requirement that reduces risk by modifying microdata. In data swapping, a probability sample of records is paired with other records on the file using selected characteristics, and then some identifying variables are swapped between pairs of records (see Kaufman et al. 2005). The sampling rate for TIMSS swapping was designed to protect the confidentiality of the data without affecting the usability of the dataset. All questionnaire data (school, teacher, and student) were involved in the swapping. This method is an effective way of keeping as much valuable data as possible while protecting respondent identity. Swapping preserves the univariate frequencies, means, and variances, although it may affect multivariate relationships a little. Pre- and post-swapping percentage distributions (unweighted and weighted) and correlations were reviewed to ensure data quality was maintained.

Confidentiality analyses of this kind were conducted before the U.S. data files were delivered to the DPC for cleaning, and prior to the IRT scaling and estimation of sampling weights.

4.6.3 International variable creation procedures

Once any problems arising during the data cleaning process were resolved, sampling weights produced by Statistics Canada and IRT-scaled student proficiency scores in mathematics and science generated by the international study center were added to the files for each country.

Student proficiencies. TIMSS 1995, 1999, 2003, and 2007 used Item Response Theory (IRT) methods to produce score scales that summarized the achievement results. With this method, the performance of a sample of students in a subject area or sub-area could be summarized on a single scale or a series of scales, even when different students had been administered different items.

IRT scaling provides estimates of item parameters (for example, item difficulty and item discrimination) that define the relationship between the item and the underlying variable measured by the test.

Parameters of the IRT model are estimated for each test item, with an overall scale being established as well as scales for each content area and cognitive domain specified in the assessment framework. As noted earlier, in addition to mathematics and science total scores, the TIMSS 2007 assessments had scales describing mathematics and science content areas, as well as cognitive domains.

In order to allow for the calculation of trends in achievement, comparisons of scores across the four TIMSS assessments conducted in 1995, 1999, 2003 and 2007 were necessary. To this end achievement scores from all four assessments were placed on the same scale. Information obtained from the bridge study was incorporated into this scaling to ensure strict comparability of scores across the four assessments. Details are provided in the *TIMSS 2007 Technical Report* (Olson, Martin, and Mullis 2008) and in appendix A to this report.

Plausible values. During the scaling phase, plausible values were used to characterize scale scores for students participating in the assessment. To keep student burden to a minimum while ensuring content coverage, TIMSS administered a limited number of assessment items to each student—too few to produce accurate scale scores for each student. To account for this, TIMSS generated five possible scale scores for each student, each representing a random selection from the distribution of scale scores of students with similar backgrounds who answered the assessment items the same way.

This plausible-values methodology was used to represent what the true performance of an individual might have been, had it been observed. This is done by using a small number of random draws from an empirically derived distribution of score values based on the student's observed responses to assessment items and on background variables. Each random draw from the distribution is considered a representative value from the distribution of potential scale scores for all students in the sample who have similar characteristics and identical patterns of item responses. The draws from the distribution are different from one another to quantify the degree of precision (the width of the spread) in the underlying distribution of possible scale scores that could have caused the observed performances. The TIMSS plausible values function like point estimates of scale scores for many purposes, but they are unlike true point estimates in several respects. They differ from one another for any particular student, and the amount of difference quantifies the spread in the underlying distribution of possible scale scores for that student.

This approach to the estimation of scale scores ensures that the estimates of the average performance of student populations and the estimates of variability in those estimates are more accurate than those determined through traditional procedures, which estimate a single score for each student. An accessible treatment of the derivation and use of plausible values can be found in Beaton and Gonzalez (1995). A more technical treatment can be found in the *TIMSS 2007 Technical Report* (Olson, Martin, and Mullis 2008).

International benchmarks. International benchmarks for achievement were developed in an attempt to provide a concrete interpretation of what the scores on the TIMSS mathematics and science achievement scales mean. TIMSS used scale anchoring to summarize and describe student achievement at four points on the mathematics and science scales—Advanced International Benchmark (625), High International Benchmark (550), Intermediate International Benchmark (475), and Low International Benchmark (400). Scale anchoring involves selecting benchmarks (scale points) on the TIMSS achievement scales to be described in terms of student performance and then identifying items that students scoring at the anchor points can answer correctly. Subsequently, these items are grouped by content area within benchmarks and reviewed by mathematics and science experts. These experts focus on the content of each item and describe the kind of mathematics or science knowledge demonstrated by students answering the item correctly. The experts then provide a summary description of performance at each anchor point leading to a content-referenced interpretation of the achievement results. Detailed information on the creation of the benchmarks is provided in the international TIMSS reports (Mullis, Martin, and Foy 2008; Martin, Mullis, and Foy 2008) and in appendix A to this report.

Sampling weights. Because of the complex sampling design used in TIMSS, students were assigned sampling weights. In general the sampling weight assigned to a student is the inverse of the probability that the student would be selected for the sample. When responses are weighted and each contributes to the results for the total number of students represented by the individual student assessed. Weighting also adjusts for school and student nonresponse. The internationally defined weighting specifications for TIMSS require that each assessed student's sampling weight should be the product of: the inverse of the school's probability of selection; an adjustment for school-level nonresponse; the inverse of the classroom's probability of selection; and an adjustment for student-level nonresponse. Sampling weights should be used in all TIMSS analyses. A detailed description of this process is provided in Olson, Martin, and Mullis (2008).

5. The TIMSS 2007 Data for the United States

The TIMSS 2007 international database contains student achievement data as well as student, teacher, school, and curricular background data for 59 countries and 8 benchmarking participants. In total, data are available on 433,785 students, 46,770 teachers, and 14,753 school principals. This database provides comparable information across countries on student achievement in mathematics and science and; contextual information about students, teachers, classrooms, schools, and education systems.

5.1 U.S. international and national data files

The TIMSS 2007 data for the United States are available through three avenues:

1. *U.S. international data file* which is part of the TIMSS international database and is directly comparable to that of other nations. As such it allows comparisons of the United States with any of the other 58 countries in virtually all respects. This file is available from the TIMSS and PIRLS International Study Center as a SAS export file or an SPSS ".sav" file through the following link: http://timssandpirls.bc.edu/TIMSS2007; and

2. *U.S. national public-use data file* which includes the U.S.-specific adaptations not part of the U.S. international datafile. These adaptations affect only a few variables and take the form of: some elaboration of international response alternatives in a few items; the omission of the item asking about violence in schools; the addition of measures of "race/ethnicity" and "language spoken at home" to the student questionnaire; and the addition of three questions to the school questionnaire. These additional questions refer to: the percentage of students in the school eligible for free- or reduced-price lunch; the percentage of students in the school who have limited-English proficiency; and a specification of the type of school (see question 8 in the school questionnaires, appendix E, which list the 11 categories of schools defined). The U.S. national public-use data file is available from the National Center for Education Statistics by downloading from their website http://nces.ed.gov/timss/datafiles.asp, or by requesting a TIMSS 2007 CD-ROM which also contains an electronic codebook and other related files. This CD-ROM is described below.

3. *U.S. national restricted-use data file* is only available from the National Center for Education Statistics as a CD-ROM. This CD-ROM contains all of the information on the public-use CD-ROM described above plus supplemental link files which link TIMSS school ID numbers to the ID numbers of schools as these appear in the publicly available Common Core of Data or the Private School Universe Survey. The Common Core of Data (Hoffman 2009) is an NCES program that annually collects fiscal and non-fiscal data about all public schools, public school districts and state education agencies in the United States. The data include information that describes schools and school districts, including name, address, and phone number; descriptive information about students and staff, including demographics; and fiscal data, including revenues and current expenditures. The Private School Universe Survey (Broughman, Swaim and Keaton 2009) provides similar data for private schools. Information collected includes: religious orientation; level of school; size of school; length of school year, length of school day; total enrollment (K-12); number of high school graduates, whether a school is single-sexed or coeducational and enrollment by sex; number of teachers employed; program emphasis; existence and type of kindergarten program. The supplemental link files allow users to merge school-level data from the CDD and PSS with TIMSS school data and, subsequently with other TIMSS teacher and student data. Since school names are revealed in this process the data require a restricted-use license (for information on how to obtain a license, see http://nces.ed.gov/pubsearch/licenses.asp).

Since the U.S. national files differ little from their international counterparts, the bulk of the variables describing U.S. students, teachers and schools are as described in the *TIMSS 2007 User Guide for the International Database* (Foy and Olson 2009), the most comprehensive and detailed reference for the TIMSS 2007 data. This publication should be seen as the primary reference for the TIMSS 2007 data and is available for downloading from http://timssandpirls.bc.edu/TIMSS2007.

5.2 TIMSS international data files

There are five basic types of data files available for each country in the international dataset:

1. Achievement files containing item response data and scale scores for the TIMSS assessment;

2. Background files with information from students, their mathematics teachers, their science teachers, and the principals of their schools;

3. Student-teacher linkage files that contain the information needed to link student data to that of their teachers;

4. Constructed-response scoring reliability files providing data on the reliability of scoring for this type of item; and

5. Curriculum data files which contain the responses of countries to the curriculum questionnaires.

The following discussion focuses on the first three categories of files since these are the ones most likely to be used in data analyses by most users. For the remaining two categories of files the reader is referred to the *TIMSS 2007 User Guide for the International Database* (Foy and Olson 2009).

5.2.1 Data file naming convention

The filenames of the data files consist of an eight character string followed by a three character file extension using the following conventions. These conventions are illustrated by reference to the SPSS file containing the fourth grade school background data for the United States (ACGUSAM4.sav).

1. The first character of the file name indicates grade-level (ACGUSAM4):
    a. A — fourth grade; and
    b. B — eighth grade.

2. The second character indicates the source or level of the information in the file (ACGUSAM4):
    a. C — school-level file;
    b. T — teacher-level file; and
    c. S — student-level file.

3. The third character indicates the type of data in the file (ACGUSAM4):
    a. A — student achievement items;
    b. G — general background questionnaire items (school and student questionnaires), and the general teacher questionnaire for the fourth grade;
    c. M — mathematics teacher background questionnaire;
    d. S — science teacher background questionnaire; and
    e. T — student-teacher linkage files.

4. Characters four through six identify the country using a three-character alphanumeric country code based on the ISO coding scheme (ACGUSAM4). See Exhibit 4.1 in Foy and Olson (2009) for a list of the ISO codes for each participating country.

5. The seventh and eighth characters indicate the study cycle (ACGUSAM4):
    a. M1 — TIMSS 1995 files;
    b. M2 — TIMSS 1999 files;
    c. M3 — TIMSS 2003 files; and
    d. M4 — TIMSS 2007 files.

6. The three-character file extensions used for the data files are as follows:
    a. .EXP — SAS export data; and
    b. .SAV — SPSS data.

Exhibit 5.1 shows the formats of the TIMSS 2007 data file names according to this scheme. For example: ASGUSAM4.SAV is an SPSS file that contains the U.S. TIMSS 2007 fourth-grade student background data. BTSUSAM4.EXP is a SAS export file that contains the U.S. TIMSS 2007 eighth-grade science teacher background data. For each file type, a separate data file is provided for each participating country. Country data files are distinguished by the ISO code in positions 4 through 6.

Exhibit 5.1. Data file names

File Names	Descriptions
ACG●●●M4	Fourth-grade school background data files
ASA●●●M4	Fourth-grade student achievement data files
ASG●●●M4	Fourth-grade student background data files
AST●●●M4	Fourth-grade student-teacher linkage files
ATG●●●M4	Fourth-grade teacher background data files
BCG●●●M4	Eighth-grade school background data files
BSA●●●M4	Eighth-grade student achievement data files
BSG●●●M4	Eighth-grade student background data files
BST●●●M4	Eighth-grade student-teacher linkage files
BTM●●●M4	Eighth-grade mathematics teacher background data files
BTS●●●M4	Eighth-grade science teacher background data files

●●● indicates country code as part of filename.
SOURCE: Foy, P. & Olson, J.F. (2009). Exhibit 4.2. p.85.

5.3 TIMSS international achievement data files and variable names

The data files containing the IRT-scaled achievement scores for overall mathematics and science, the several mathematics and science content domains, and the three mathematics and science cognitive domains are identified by the first three characters in the file name. A set of five plausible values characterizes each of these achievement scores.

Files beginning with ASA are fourth grade achievement score files, and those beginning with BSA are eighth grade achievement score files. For analytic convenience, these same achievement scores are also provided as an addition to the student background data files.

The achievement score variable names are based on an 8-character string defined as follows. In exhibit 5.2 these conventions are illustrated by reference to the first plausible value for each of the total, content domain and cognitive domain achievement scales at each grade level.

1. First character of the variable name:
    a. A — fourth-grade score; and
    b. B — eighth-grade score.

2. Second character of the variable name:
    a. S — indicates that this a score variable.

3. Third character of the variable name:
    a. M — mathematics score; and
    b. S — science score.

4. Fourth through sixth characters of the variable name:
    a. three-character code identifying the achievement scale, as indicated in Exhibit 5.2 on the following page.

5. Seventh and eighth characters of the variable name:
    a. 01 — first plausible value;
    b. 02 — second plausible value;
    c. 03 — third plausible value;
    d. 04 — fourth plausible value; and
    e. 05 — fifth plausible value.

For example, ASMMAT01 is the first plausible value for the fourth grade mathematics total score. Again, ASSSCI03 is the third plausible value for the total science score, and ASSEAR02 is the second plausible value for the content domain “Earth science.”

Exhibit 5.2 shows the nomenclature for identifying the various total, content domain and cognitive domain achievement scale score variables, each of which is represented by five plausible values.

5.3.1 TIMSS benchmark achievement variables

The achievement files also contain a set of variables indicating which international benchmark the students reached. For the overall mathematics and science scales at both grades there are five plausible values for each of the four benchmark levels defined (high, advanced, intermediate, low). The international benchmark variables follow the achievement score variable naming convention but substitute the letters “IBM” in the fourth through sixth positions of the variable name. Thus, ASMIBM01, ASMIBM02, ASMIBM03, ASMIBM04, and ASMIBM05 are the five benchmark variables associated with the fourth-grade overall mathematics score. Similarly, BSMIBM01 through BSMIBM05 are the benchmark variables associated with the eighth-grade overall mathematics score and BSSIBM01 through BSSIBM05 are the benchmark variables associated with the eighth-grade overall science score. Details are provided in Foy and Olson (2009, pp. 92-93).

Exhibit 5.2. Three-digit codes identifying total, content domain, and cognitive domain achievement variable names

Achievement score	Score identifier	Examples of variable names	
		Grade 4	Grade 8
Mathematics overall score	MAT	ASMMAT01	BSMMAT01
Science overall score	SCI	ASSSCI01	BSSSCI01
Mathematics content domains			
a. Number	NUM	ASMNUM01	BSMNUM01
b. Geometric shapes/geometry	GEO	ASMGEO01	BSMGEO01
c. Data display/data and chance	DAT	ASMDAT01	BSMDAT01
d. Algebra	ALG		BSMALG01
Science content domains			
a. Life science	LIF	ASSLIF01	
b. Physical science/physics	PHY	ASSPHY01	BSSPHY01
c. Earth science	EAR	ASSEAR01	BSSEAR01
d. Chemistry	CHE		BSSCHE01
e. Biology	BIO		BSSBIO01
Mathematics cognitive domains			
a. Knowing	KNO	ASMKNO01	BSMKNO01
b. Applying	APP	ASMAPP01	BSMAPP01
c. Reasoning	REA	ASMREA01	BSMREA01
Science cognitive domains			
a. Knowing	KNO	ASSKNO01	BSSKNO01
b. Applying	APP	ASSAPP01	BSSAPP01
c. Reasoning	REA	ASSREA01	BSSREA01

SOURCE: Foy, P. and Olson, J.F. (2009). *TIMSS 2007 User Guide for the International Database*. Chestnut Hill, MA: Boston College.

5.4 TIMSS international background questionnaire data files

Student, teacher and school files contain the responses to the questions contained in the respective background questionnaires administered in TIMSS 2007, along with a fourth file used to link the student and teacher background data appropriately when student and teacher files are merged.

5.4.1 TIMSS student background data files (ASG/BSG)

The student background data files contain students' responses to questions in the student questionnaire along with students' mathematics and science achievement scores (as plausible values). At the fourth grade, there was a single version of the student questionnaire. However, internationally there were two versions of the student questionnaire at the eighth grade: a version for educational systems where science is taught as an integrated subject (general science version) as it is in the United States; and another version for educational systems where the sciences (biology or life science, earth science, physics, chemistry, and environmental science) are taught separately. For eighth-grade students who were administered the general science version, as is the case for the United States, questions that appeared only in the separate science version were coded as "not administered." For students in those

countries assigned the separate science version, questions asked only in the general science version were coded as “not administered.”

The student background data files also contain a number of identification variables, tracking variables, sampling and weighting variables, and derived variables that were used to produce some of the exhibits in the international reports.

5.4.2 TIMSS teacher background data files (ATG/BTM/BTS)

The mathematics and science teachers of the students sampled for TIMSS 2007 were administered at least one questionnaire for each TIMSS class taught. The teacher background data files contain one record for each of the classes taught. Teachers who taught more than one class were expected to complete only one set of general background questions (part A) irrespective of the number of classes taught, and a separate part B (class-specific questions) for each class they taught.

Separate teacher questionnaires were administered to eighth-grade mathematics teachers and science teachers. The responses of teachers to the mathematics questionnaire are found in the BTM files and the responses of teachers to the science questionnaire are in the BTS files. Variable names for questions repeated in both questionnaires are the same. Responses to the single questionnaire administered to fourth-grade teachers are found in the ATG files.

In the teacher files each teacher has a unique identification number (IDTEACH) and a link number (IDLINK) that is specific to the class taught by the teacher and to which the information in the data record corresponds. The IDTEACH and IDLINK combination uniquely identifies, within a country, a teacher teaching a specific class. For example, students linked to teachers identified by the same IDTEACH but different IDLINK are taught by the same teacher but in different classes. It is important to note that the teachers in question do not constitute a representative sample of teachers in a country, but rather are the teachers who taught a representative sample of students. Therefore, the teacher data should be analyzed primarily only in conjunction with the student-teacher linkage data files and weighted with student sampling weights.

5.4.3 TIMSS school background data files (ACG/BCG)

The school background data file contains the responses of school principals to questions about school policy, resources, and environment asked in the school questionnaire. That file also contains a series of identification variables, link variables, and sampling variables. The school data files can be merged with the student data files by using the country and school identification variables. Details of the merging procedure using the SPSS-linked IEA IDB Analyzer, or using SAS programs, are described in Foy and Olson (2009).

5.4.4 TIMSS student-teacher linkage data files (AST/BST)

The TIMSS 2007 student-teacher linkage data files contain information required to link the student and teacher data files. These files contain one entry per student-teacher linkage combination in the data. For instance, if three teachers are linked to a student, there are three entries in the file corresponding to that student. The sole purpose of the student-teacher linkage data files is to link teacher-level data with student-level data to perform appropriate student-level analyses where teacher characteristics are disaggregated over students.

5.5 International variable naming convention for background questionnaires

The background variable naming convention is based on a 7- or 8-character string defined as follows. These conventions are illustrated by reference to an item in the fourth-grade school questionnaire. This item asks principals to report the population size of the community in which the school is located.

First character of the variable name (AC4GCOMU):

a. A — fourth-grade; and
b. B — eighth-grade.

Second character of the variable name (AC4GCOMU):

a. C — school principal;
b. T — teacher;
c. S — student; and
d. U — curriculum.

Third character of the variable name (AC4GCOMU):

a. 4 — a background variable (other than a derived variable or a curriculum questionnaire variable);
b. B — all background variables in the curriculum questionnaire data files; and
c. D — all derived variables.

Fourth character of the variable name (AC4GCOMU):

a. G — general question (not subject specific);
b. M — question related to mathematics;
c. S — question related to science;
d. B — question related to biology or life science;
e. C — question related to chemistry;
f. E — question related to earth science; and
g. P — question related to physics of physical science.

Fifth through eighth characters of the variable name (AC4GCOMU):

a. used to assign a unique content-related label to each question.

5.5.1 Summary indices and derived variables

The TIMSS questionnaires often devote several questions to a single construct. In these cases, responses to the individual items were combined to create a derived variable. A TIMSS index is a special type of derived variable that assigns students to one of three levels—high, medium, or low—on the basis of their responses to the component variables. The high category of an index represents the responses that are expected to characterize aspects of a positive learning environment, and the low category those responses that are least supportive of learning. These variables are described in detail in Supplement 3 to the *TIMSS 2007 User Guide for the International Database* (Foy and Olson 2009).

5.6 International sampling and weighting variables

Several sampling and weighting variables are included in the TIMSS data files. They are listed and described below in conjunction with a discussion of how and when these weights are used. Since TIMSS adopted a complex sampling design, sampling weights need to be used in order to generate accurate population estimates. The sampling weights account for the sample design, any stratification or disproportional sampling of subgroups, and include also adjustments for non-response (see Joncas 2008).

As noted, the sample of students is not a simple random sample and, as a consequence, students in the sample do not have an equal probability of selection. Sampling weights adjust for this unequal probability and, in so doing, provide for statistical estimates reflective of the student population from which the sample was drawn. Sampling weights also include adjustments for school and student nonresponse. All TIMSS analyses require the application of sampling weights. Provisions for weighting data are standard in many statistical analysis software packages.

The several sampling weights included in the TIMSS 2007 data files are described in exhibit 5.3. (Note that teacher background data files do not have any sampling weight variables since the analysis of teacher variable requires the merging of the teacher data with the student data and the use of student sampling weight variables.)

Exhibit 5.3. Sampling weight variables

Variable Names	Descriptions
TOTWGT	Total student weight – sums to the national population
SENWGT	Student senate weight – sums to 500 in each country
HOUWGT	Student house weight – sums to the student sample size in each country
SCHWGT	Weight for analyses in which the school is the unit of analysis
TCHWGT	Overall teacher weight
MATWGT	Mathematics teacher weight
SCIWGT	Science teacher weight

SOURCE: Foy, P. and Olson, J.F. (2009). *TIMSS 2007 User Guide for the International Database*. Chestnut Hill, MA: Boston College.

The characteristics of these sampling weight variables are as follows:

1. TOTWGT sums to the student population size in each country and is appropriate for within-country analyses and cross-country analyses where the analyses are conducted country-by-country and compared.

2. SENWGT is a transformation of TOTWGT that results in a weighted student sample size of 500 in each country. This weight may be appropriate for cross-country analyses that require each country to have the same number of students rather than proportionately more students from larger countries and fewer from smaller countries which is the case if TOTWGT is used.

3. HOUWGT, another transformation of TOTWGT, ensures that the weighted sample corresponds to the actual sample size in each country. This can be important since TOTWGT inflates sample sizes to approximate the population size and software systems that use the actual sample size to compute significance tests will give misleading results under these conditions.

    - TOTWGT, SENWGT and HOUWGT are designed for use in student-level analyses from all student-level files.

    - SCHWGT is designed for use in school-level analyses where the schools are the units of analysis.

    - TCHWGT, MATWGT, and SCIWGT are specifically designed for analyses that link teacher background data to student data.

- TCHWGT is used for analyses using all teachers.

- MATWGT and SCIWGT are used for analyses of mathematics and science teachers, respectively.

5.7 International structure and design variables in TIMSS 2007 data files

The TIMSS 2007 data files also contain unique numerical identification variables for each respondent along with sample design information.

5.7.1 Identification variables

In all TIMSS data files, several identification variables are included to label countries, students, teachers, or schools. These variables also are used to link cases between the different datafile types. The identification variables have the prefix "ID" and are described below.

1. IDCNTRY: a five-digit country identification code based on the ISO 3166 classification.

2. IDPOP: identifies the target grade; "1" for the fourth grade and "2" for the eighth grade.

3. IDGRADE: identifies the target grade of the participating students; "4" and "8" for most countries.

4. IDSCHOOL: a four-digit identification code that uniquely identifies the participating schools within each country but are not unique across countries.

5. IDCLASS: a six-digit identification code that uniquely identifies the sampled classrooms within a country.

6. IDSTUD: an eight-digit identification code that uniquely identifies each sampled student in a country.

7. IDBOOK: identifies the specific assessment booklet that was administered to each student.

8. IDSTRATE & IDSTRATI: identification variables generated by the school sampling process. IDSTRATE identifies the explicit strata and IDSTRATI the implicit strata from which the participating schools were sampled.

9. IDTEACH: a six-digit identification code that uniquely identifies a teacher within a school.

10. IDLINK: uniquely identifies the class for which a teacher answered a questionnaire.

5.7.2 Tracking variables

Information about students, teachers, and schools provided by the survey tracking forms described earlier is stored in the tracking variables. These variables have the prefix "IT." All tracking variables are included

in the student background data files. ITLANG is included in the student achievement and student background data files.

1. ITSEX: Gender of each student as stated in the Student Tracking Form.

2. ITBIRTHM and ITBIRTHY: month and year of birth of each student as stated in the Student Tracking Forms.

3. ITDATEM and ITDATEY: month and year of testing for each student.

4. ITLANG: language of testing for each student.

5.8 International TIMSS 2007 codebook files

All information related to the structure of the TIMSS 2007 data files, as well as the source, format, descriptive labels, and response option codes for all variables, is contained in codebook files. Each data file type in the database is accompanied by a codebook file, with the exception of the curriculum data files. These files are available from the TIMSS and PIRLS International Study Center website: http://timssandpirls.bc.edu/TIMSS2007/idb_ug.html .

5.9 The U.S. national data files

As noted earlier, the U.S. instrumentation differs from the international instrumentation in five ways.

1. Minor language/expression adaptations were made to some of the instructions.

2. Minor language adaptations were made to the wording of some assessment items.

3. For a few questionnaire items, response alternatives were changed but in a way that allowed a crosswalk to the international response alternatives.

4. One international question was not asked in the United States because of its sensitivity.

5. Several U.S.-specific questions without international counterparts were added to the student and schools questionnaires.

In other ways the U.S. instrumentation is exactly the same as the international instrumentation. This will become apparent in comparisons between U.S. and international questionnaires. The international versions of questionnaires are available at http://timssandpirls.bc.edu/TIMSS2007/context.html. The U.S. versions of the questionnaires can be found at http://nces.ed.gov/timss/questionnaire.asp. The variable names in the U.S. files are identical to those in the international files with the exception of the U.S.-specific variables added.

5.9.1 Items with U.S. adaptations to response alternatives

There were a number of relatively minor changes to the wording of item stems and response alternatives in the questionnaires. Most of these do not require comment as they are identical in format between the international and U.S. versions of the questionnaires. In some cases, however, the adaptations resulted in item response formats not immediately comparable between the international and national versions of the questionnaires. As indicated in appendix F, there are instances in which the international and U.S. versions of variables have different sets of response codes; “highest level of formal education” is one

example. What this means is that, for these items, the data in international and U.S. versions of the data files will not be identical. Using the same example, “highest level of formal education” will have six response categories in the U.S. international file and seven categories in the U.S. national file. However, as indicated in appendix F, crosswalks between international and U.S. versions of these questions allow for the conversion of the U.S. response codes to the international format.

5.9.2 U.S.-specific variables

U.S.-specific items were added to the student, teacher, and school questionnaires. Three questions were added to the student questionnaires: a two-part question designed to address students’ race/ethnicity; a question that asked for language other than English spoken at home; and a two-part question that asked students to indicate whether they had repeated a grade in elementary and/or middle/junior high school (grade 8 student questionnaire only). One question was added to the grade 8 teacher questionnaires: the question asked teachers to identify the title of the mathematics or science course being taught to the students being assessed. Three questions were added to the school questionnaire: the percentage of students in the school eligible for free- or reduced-price lunch; the percentage of students in the school who are limited-English proficient; and a specification of the type of school.

Race/ethnicity. Students’ race/ethnicity was obtained through student responses to a two-part question in the student questionnaire. Students were asked first to indicate whether or not they were “Hispanic or Latino”. In the second part of the question they were asked whether they were members of the following racial groups: “American Indian or Alaska Native”; “Asian”; “Black or African American”; “Native Hawaiian or other Pacific Islander”; or “White”. Multiple responses to the second part of the question were allowed. A composite variable was constructed in which results are shown separately for: “Hispanics of any race”; “Blacks, not Hispanic or Latino”; “Whites, not Hispanic or Latino”; “Asians, not Hispanic or Latino”; and, “Two or more races, not Hispanic or Latino”. Only a small number of students who indicated that they were not Hispanic or Latino also indicated that they were respectively, “American Indian or Alaska Native” or “Native Hawaiian or other Pacific Islander”.

Language other than English spoken at home. This item extended the international question about how often students spoke English at home to ask those who indicated that they did not always speak English about whether they spoke Spanish or another language.

Repeating a grade. Students were asked whether they had ever repeated a grade in either “elementary school” and/or “middle or junior high school”. The response alternatives were “yes” or “no” in each case.

Mathematics course taught. Eighth-grade mathematics teachers were asked about the nature of the mathematics course they taught to the TIMSS students. The response alternatives provided were as follows:

a. Basic or general eighth-grade math (not algebra or pre-algebra);

b. Introduction to algebra or pre-algebra;

c. Algebra I (one-year course);

d. Algebra I (first year of a two-year Algebra I course);

e. Algebra I (second year of a two-year Algebra I course);

f. Geometry;

g. Algebra II;

h. Integrated or sequential math; and

i. Other math class.

Science course taught. Eighth-grade science teachers were asked about the nature of the science course they taught to the TIMSS students. The response alternatives provided were as follows:

a. General science (several content areas of science taught separately);

b. Integrated science (several content areas of science taught combined and taught throughout the year);

c. Life science (e.g., biology, ecosystems, human health);

d. Physical science (e.g., physics or chemistry); and

e. Earth science (e.g., geology, earth and the solar system, fossils).

Poverty level in public schools (percentage of students eligible for free or reduced-price lunch). The measure of poverty level in public schools was obtained from principals' responses to the school questionnaire. The question asked the principal to report, as of approximately the first of October 2006, the percentage of students at the school eligible to receive free or reduced-price lunch through the National School Lunch Program. Responses were grouped into five categories: less than 10 percent; 10 to 24.9 percent; 25 to 49.9 percent; 50 to 74.9 percent; and 75 percent or more. Missing data on this variable were replaced with measures taken from the Common Core of Data. The effect of this replacement on the confidentiality of the data was examined as part of the confidentiality analyses described earlier. Note that this variable is available only for public schools.

Limited-English proficient students. Principals were asked to report the percentage of such students and were provided with the following response categories: 0 percent; 1 to 5 percent; 6 to 10 percent; 11 to 25 percent; 26 to 50 percent; 51 to 75 percent; 76 to 90 percent; and over 90 percent.

Type of school. Principals were asked to identify their schools using one of the following response categories: regular public school; regular public school with magnet program; magnet school or school with special program; special education; or, alternative curriculum.

5.9.3 Missing data

Data derived from the student, school and teacher questionnaires and from the student assessments contain missing data in varying amounts. Four sources of missing data are identified:

1. Not administered. The respondent was not administered the actual item. He or she had no chance to read and answer the question.
2. Omitted. The respondent had a chance to answer the question but did not do so. This code also was used for responses that were not interpretable.
3. Logically not applicable. The respondent answered a preceding filter question in a way that made the following dependent questions not applicable to him or her.
4. Not reached (only used in the achievement files). This code indicates those items not reached by the students due to a lack of time.

Imputation. No imputation for missing values was undertaken. However, missing data on the measure of school poverty (proportion of students eligible for free- or reduced-price lunch) reported by schools was replaced as described above.

System missing code. SAS and SPSS code that can be used to transform the several missing data codes to a single system missing code is provided below in exhibits 5.4 and 5.5.

Exhibit 5.4. SAS Code to Convert Missing or Not Administered codes to system missing (.)

```
Array Miss1 (*) Var1-VarN;                    /* Variables for which codebook format for missing = 9 */
  Do i=1 to dim(Miss1);
    If Miss1(i) = 9 then Miss1(i) = . ;
    Else if Miss1(i) = 8 then Miss1(i) = .A ;
  End;

  Array Miss2 (*) Var1-VarN;                  /* Variables for which codebook format for missing = 99 */
  Do i=1 to dim(Miss2);
    If Miss2(i) = 99 then Miss2(i) = . ;
    Else if Miss2(i) = 98 then Miss2(i) = .A ;
  End;

  Array Miss3 (*) Var1-VarN;                  /* Variables for which codebook format for missing = 999 */
  Do i=1 to dim(Miss3);
    If Miss3(i) = 999 then Miss3(i) = . ;
    Else if Miss3(i) = 998 then Miss3(i) = .A ;
  End;

  Array Miss4 (*) Var1-VarN;                  /* Variables for which codebook format for missing = 9999 */
  Do i=1 to dim(Miss4);
    If Miss4(i) = 9999 then Miss4(i) = . ;
    Else if Miss4(i) = 9998 then Miss4(i) = .A ;
  End;

  Array Miss5 (*) Var1-VarN;                  /* Variables for which codebook format for missing = 99999 */
  Do i=1 to dim(Miss5);
    If Miss5(i) = 99999 then Miss5(i) = . ;
    Else if Miss5(i) = 99998 then Miss5(i) = .A ;
  End;

  Array Miss6 (*) Var1-VarN;                  /* Variables for which codebook format for missing = 999999*/
  Do i=1 to dim(Miss6);
    If Miss6(i) = 999999 then Miss6(i) = . ;
    Else if Miss6(i) = 999998 then Miss6(i) = .A ;
  End;

  Array Miss7 (*) Var1-VarN;                  /* Variables for which codebook format for missing = 9999999 */
  Do i=1 to dim(Miss7);
    If Miss7(i) = 9999999 then Miss7(i) = . ;
    Else if Miss7(i) = 9999998 then Miss7(i) = .A ;
  End;

  Array Miss8 (*) Var1-VarN;                  /* Variables for which codebook format for missing = 99999999 */
  Do i=1 to dim(Miss8);
    If Miss8(i) = 99999999 then Miss8(i) = . ;
    Else if Miss8(i) = 99999998 then Miss8(i) = .A ;
  End;
```

Exhibit 5.5 SPSS Code to Convert Missing or Not Administered codes to system missing (sysmis)

```
*Converting to Sysmis-values for variables for which codebook format for Missing = 9.
RECODE  Var1 Var2 VarN  (9 = Sysmis).
EXECUTE.
*Converting to User-defined missing-values for variables for which codebook format for Not Administered
= 8.
MISSING VALUES Var1 Var2 VarN (8).
*Converting to Sysmis-values for variables for which codebook format for Missing = 99.
RECODE  Var1 Var2 VarN  (99 = Sysmis).
EXECUTE.
*Converting to User-defined missing-values for variables for which codebook format for Not Administered
= 98.
MISSING VALUES Var1 Var2 VarN (98).
*Converting to Sysmis-values for variables for which codebook format for Missing = 999.
RECODE  Var1 Var2 VarN  (999 = Sysmis).
EXECUTE.
*Converting to User-defined missing-values for variables for which codebook format for Not Administered
= 998.
MISSING VALUES Var1 Var2 VarN (998).
*Converting to Sysmis-values for variables for which codebook format for Missing = 9999.
RECODE  Var1 Var2 VarN  (9999 = Sysmis).
EXECUTE.
*Converting to User-defined missing-values for variables for which codebook format for Not Administered
= 9998.
MISSING VALUES Var1 Var2 VarN (9998).
*Converting to Sysmis-values for variables for which codebook format for Missing = 99999.
RECODE  Var1 Var2 VarN  (99999 = Sysmis).
EXECUTE.
*Converting to User-defined missing-values for variables for which codebook format for Not Administered
= 99998.
MISSING VALUES Var1 Var2 VarN (99998).
*Converting to Sysmis-values for variables for which codebook format for Missing = 999999.
RECODE  Var1 Var2 VarN  (999999 = Sysmis).
EXECUTE.
*Converting to User-defined missing-values for variables for which codebook format for Not Administered
= 999998.
MISSING VALUES Var1 Var2 VarN (999998).
*Converting to Sysmis-values for variables for which codebook format for Missing = 9999999.
RECODE  Var1 Var2 VarN  (9999999 = Sysmis).
EXECUTE.
*Converting to User-defined missing-values for variables for which codebook format for Not Administered
= 9999998.
MISSING VALUES Var1 Var2 VarN (9999998).
*Converting to Sysmis-values for variables for which codebook format for Missing = 99999999.
RECODE  Var1 Var2 VarN  (99999999 = Sysmis).
EXECUTE.
*Converting to User-defined missing-values for variables for which codebook format for Not Administered
= 99999998.
MISSING VALUES Var1 Var2 VarN (99999998).
```

5.10 TIMSS 2007 data files available from the National Center for Education Statistics

The data files containing the U.S. national data are in ASCII format and are named as indicated in exhibit 5.6. SAS and SPSS code for reading these data files is available and these SAS and SPSS files have the same file name as the data files but can be identified by the .SPS or .SAS extension. Supporting documentation in the form of user guides and codebooks is also available.

Exhibit 5.6. U.S. national data file names

Contents	File name
Grade 4 Public-use data files	
Student background	G4_STUDENT07.TXT
Student achievement	G4_ACHIEVE07.TXT
Teacher background	G4_TEACHER07.TXT
Student-teacher link	G4_STD_TCH_LINK07.TXT
School background	G4_SCHOOL07.TXT
Grade 4 Restricted-use data file	
Restricted-use ID link file	G4_Restricted_Use07.txt
Grade 8 Public-use data files	
Student background	G8_STUDENT07.TXT
Student achievement	G8_ACHIEVE07.TXT
Math teacher background	G8_MTEACHER07.TXT
Science teacher background	G8_STEACHER07.TXT
Student-teacher link	G8_STD_TCH_LINK07.TXT
School background	G8_SCHOOL07.TXT
Grade 8 Restricted-use data file	
Restricted-use ID link file	G8_Restricted_Use07.txt

The student background files at each grade level contain data on students' demographic information, home background and school experiences, achievement scores, the sampling weights, and sampling information.

The student achievement files contain the student response data for the individual achievement items in the TIMSS achievement test including the responses to both multiple-choice and open-ended questions.

The grade 4 teacher background file contains teacher background information from the survey, teacher views on mathematics and science education, the activities undertaken in mathematics and science education, and weights and sampling information.

The grade 8 mathematics teacher background file and the grade 8 science teacher background file contain teacher background information from the survey, teacher views on mathematics and science education, the activities undertaken in mathematics and science education, and weights and sampling information for the students' mathematics and science teachers respectively.

The student-teacher link files contain information to link the student and teacher files and correctly compute weighted teacher-level data using the student as the unit of analysis.

The school background files contain background information on the school provided by the school principal or designate, demographic information about the school, school programs related to mathematics and science, and weights and sampling information.

The restricted-use ID link files are supplemental files for each grade and are designed to provide the means by which TIMSS school ID numbers can be linked to the ID numbers of schools in the publicly available Common Core of Data or the Private School Universe Survey. This allows an appropriately certified user to merge school-level data from the CDD and PSS with TIMSS school data.

5.10.1 Downloading data files from the TIMSS website

The public-use data files containing the U.S. national data can be downloaded from the NCES website at: http://nces.ed.gov/timss/datafiles.asp. These national files are in ASCII format and are named as indicated in exhibit 5.6. SAS and SPSS code for reading these data files can also be downloaded from the NCES website as well. These SAS and SPSS files have the same file name as the data files but can be identified by the .SPS or .SAS extension. Supporting documentation can also be downloaded from this same website.

The two restricted-use data files noted in exhibit 5.6 cannot be downloaded from the website. They are only available on a licensed CD-ROM as described below.

5.10.2 TIMSS 2007 data on CD-ROM

The public-use data, control files and documentation are also available from NCES on a public-use CD-ROM. This CD-ROM also contains an electronic codebook (ECB) which is described in the following section. The ECB facilitates the review of the variables and also provides basic SAS and SPSS code which can be used to extract subsets of variables and create SAS/SPSS datasets. This CD-ROM can be ordered online at http://nces.ed.gov/timss.

The restricted-use dataset is also available from the National Center for Education Statistics on CD-ROM. This CD-ROM contains all of the information contained on the public-use CD-ROM plus a supplemental link-file for each grade (G4_Restricted_Use07.txt and G8_Restricted_Use07.txt) which links TIMSS school ID numbers to the ID numbers of schools as these appear in the publicly available Common Core of Data or the Private School Universe Survey. The supplemental link files allow a user to merge school-level data from the CDD and PSS with TIMSS school data and, subsequently with other TIMSS teacher and student data. Since school names are revealed in this process the data require a restricted-use license (for information on how to obtain a license, see http://nces.ed.gov/pubsearch/licenses.asp).

5.10.3 The Electronic Codebook for the TIMSS U.S. data

The Electronic Codebook (ECB) software tool enables analysts to review and extract United States TIMSS 2007 data. With this ECB software tool, an analyst can perform the following actions:

- Search the names and labels of variables in the TIMSS 2007 national data file catalog;
- Examine the unweighted response categories, frequencies, and percentages of responses for one or more catalog variables;

- Create a list of variables to be extracted from the catalog (taglist), save the list for later use, recall a previously saved taglist, print the list as a codebook, or taglist; and

- Automatically generate SAS and SPSS programs to extract selected variables from the whole data set or for a subset of defined cases.

The ECB software tool only works in a Windows-based environment (Windows 95 or higher). It will not run with other computer operating systems, such as Macintosh and Linux. The ECB software tool is included on both of the public- and restricted-use CD-ROMs described above. (The ECB is only available on CD-ROM and cannot be downloaded from this website.)

Note that the ECB is not essential in order to work with the TIMSS data. The raw data files and companion SAS and SPSS control files that can be downloaded from the NCES website, and which are included on the CD-ROMs, are sufficient for this purpose.

5.11 Merging TIMSS 2007 data files

In preparing TIMSS 2007 data for analysis it may be necessary to merge two (or more) of the data files named in exhibit 5.6. Not every analysis will require merging of file however. For example, analyses looking at the relationship between student background and achievement can be done using the student background file alone. However, analyses that wish to examine the relationships between school and/or teacher and/or student characteristics and student achievement will require that files be merged. Standard merging procedures as implemented in SPSS, SAS or Stata can be applied. Examples are provided below along with illustrative SAS and SPSS code. (These various merges are facilitated for SPSS users who choose to work with the IEA International Database Analyzer (IEA IDB Analyzer) described below.)

The merging procedures illustrated below follow the same pattern as previous TIMSS studies (see Martin 2005) and are illustrated with Grade 8 data. The user will need to replace file names in order to run the merges for Grade 4.

5.11.1 Merging student and school data

If the intent is to disaggregate school data across students the school-level data are merged to the student file using IDSCHOOL. The disaggregated data can be analyzed at the student level using the student-level weight TOTWGT. Exhibits 5.7 and 5.8 provide examples of how to merge the student and school data using SAS and SPSS. Additional examples are provided in chapters 4 and 5 in the *TIMSS 2003 User Guide* (Martin 2005; available at http://isc.bc.edu/timss2003i/userguide.html).

Exhibit 5.7. Illustrative SAS code for merging U.S. TIMSS eighth-grade student and school data

```
libname bm4 “C:\TIMSS2007\TIMSS2007 Raw Data and Control Files” ;
data SCHOOL ;
set bm4.G8_SCHOOL07;
proc sort data= SCHOOL;
by IDSCHOOL ;
data STUDENT ;
set bm4.G8_STUDENT07;
proc sort data= STUDENT;
by IDSCHOOL;
data bm4.MERGE1 ;
merge STUDENT SCHOOL;
by IDSCHOOL;
run;
```

The example creates a temporary SAS dataset (SCHOOL) using the permanent school dataset bm4.G8_SCHOOL07. It then sorts the school data by school ID (IDSCHOOL). A similar procedure is used for the student file (STUDENT), which is also sorted by the school ID using the permanent student dataset bM4.G8_STUDENT07. The final dataset is a permanent dataset called bM4 .MERGE1 containing the merged file from SCHOOL and STUDENT using IDSCHOOL as the merge variable.

The SPSS example shown in exhibit 5.8 works in a similar way SPSS uses a file containing the school variables (G8_SCHOOL07.SAV) and sorts the cases by IDSCHOOL. The same procedure is used for the student dataset, G8_STUDENT07.SAV. The “match files” command merges the two files, and the final, merged output file is saved as MERGE1.SAV.

Exhibit 5.8. Illustrative SPSS code for merging U.S. TIMSS eighth-grade student and school data

```
get file = "C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\G8_SCHOOL07.SAV".
sort cases by IDSCHOOL .
save outfile = SCHOOL .
get file = "C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\G8_STUDENT07.SAV".
sort cases by IDSCHOOL.
save outfile = STUDENT .
match files
/ file= STUDENT
/ table= SCHOOL
/ by IDSCHOOL.
save outfile = "C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\MERGE1.SAV” .
```

5.11.2 Merging student and teacher data

In the United States, the student sample was based on intact classrooms. The teachers of the students selected in this way are not a sample of teachers and should be seen as the “teachers of the sampled students.” To maintain this linkage merges of teacher and student data must use the student-teacher link file (G8_STD_TCH_LINK07.TXT) which also contains the appropriate teacher sampling weights. Exhibits 5.9 and 5.10 provide illustrative code for merging the student and the mathematics teacher files in SAS and SPSS. The sample code provided illustrates the merge involving the eighth-grade mathematics teacher file with student file. The same logic applies to a merge involving the science teacher file (G8_STEACHER07.TXT). Additional examples of how to merge the student and teacher files using SAS and SPSS are provided in chapters 4 and 5 of the *TIMSS 2003 User Guide* (Martin 2005; available at http://isc.bc.edu/timss2003i/userguide.html).

Exhibit 5.9. Illustrative SAS code for merging U.S. TIMSS eighth-grade student and mathematics teacher data

```
data: 2007
Libname bm4 “C:\TIMSS2007\TIMSS2007 Raw Data and Control Files” ;
data TEACHER ;
set bM4.G8_MTEACHER07;
proc sort data= TEACHER;
by IDTEACH IDLINK ;
data STDTCH ;
set bM4.G8_STD_TCH_LNK07 ;
proc sort data= STDTCH;
by IDTEACH IDLINK;
data TEACHMRG;
merge TEACHER STDTCH;
by IDTEACH IDLINK;
if MATWGT > 0;
proc sort data = TEACHMRG;
by IDSTUD;
data STUDENT;
set bM4.G8_STUDENT07;
proc sort data = STUDENT;
by IDSTUD;
data bM4.MERGE2;
merge STUDENT TEACHMRG;
by IDSTUD;
run;
```

In the SAS example, the program creates a temporary SAS dataset (TEACHER) using the permanent mathematics teacher file, bm4.G8_MTEACHER07. It then sorts the teacher data by the teacher ID (IDTEACH) and the link ID (IDLINK). A similar procedure is used for the student-teacher link file (STDTCH), using the permanent file (bm4.G8_STD_TCH_LNK07) which is also sorted by the teacher ID and the link ID. The weight variable for mathematics teachers (MATWGT) is used as a selection variable because mathematics teachers have been selected. The result is a merged file called bm4.TEACHMRG with disaggregated teacher data. This file is merged with the student file (STUDENT). The final dataset is a permanent dataset called bM4.MERGE2 that contains the merged file from TEACHMRG and STUDENT using IDSTUD as the merge variable.

The SPSS student-teacher merge in Exhibit 5.10 uses a file containing the teacher variables (G8_MTEACHER07.SAV) and sorts the cases by and IDTEACH and IDLINK. The file is then saved as TEACHER. The same procedure is used for the student-teacher linkage dataset G8_STD_TCH_LNK07 .SAV. The “match files” command merges the two files by the ID variables IDTEACH and IDLINK, and the merged output file is saved as TEACHMRG. To include the student data, the student file is selected (G8_STUDENT07.SAV), sorted by IDSTUD and saved as STUDENT. This file is merged with TEACHMRG using IDSTUD to create the final file MERGE2.SAV containing both teacher and student variables.

Exhibit 5.10. Illustrative SPSS code for merging U.S. TIMSS eighth-grade student and mathematics teacher data

```
data: 2007
get file = "C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\G8_MTEACHER07.SAV".
sort cases by IDTEACH IDLINK .
save outfile = TEACHER .
get file = "C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\G8_STD_TCH_LNK07.SAV".
select if MATWGT > 0 .
sort cases by IDTEACH IDLINK.
save outfile= STDTCH.
match files
/ file=STDTCH
/ table=TEACHER
/ by IDTEACH IDLINK.
sort cases by IDSTUD.
save outfile = TEACHMRG.
get file = "C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\G8_STUDENT07.SAV".
sort cases by IDSTUD.
save outfile=STUDENT.
match files
/ file=TEACHMRG
/ table=STUDENT
/ by IDSTUD.
save outfile = “C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\MERGE2.SAV”.
```

5.11.3 Merging student, school, and teacher data

In merging student, teacher, and school data together to form a single dataset the procedures from sections 5.11.1 and 5.11.2 are combined. Exhibits 5.11 and 5.12 show illustrative SAS and SPSS code designed to achieve this three-way merge. As in the previous examples, the sample code is based on eighth-grade mathematics teacher file. A merge involving the eighth-grade science teacher file follows the same logic.

This example uses the same merging steps as with the previous school and teacher examples (MERGE1 and MERGE2), then merges the output files by the student id, IDSTUD, into a final file bM4 .MERGEALL containing linked student, school, and teacher data at the student level.

Exhibit 5.11. Illustrative SAS code for merging U.S. TIMSS eighth-grade school, mathematics teacher and student data

```
libname bm4 “C:\TIMSS2007\TIMSS2007 Raw Data and Control Files” ;
data SCHOOL ;
set bM4.G8_SCHOOL07;
proc sort data= SCHOOL;
by IDSCHOOL ;
data STUDENT ;
set bM4.G8_STUDENT07;
proc sort data= STUDENT;
by IDSCHOOL;
data MERGE1 ;
merge STUDENT SCHOOL;
by IDSCHOOL;
proc sort data=MERGE1;
by IDSTUD;
data TEACHER ;
set bm4.G8_MTEACHER07;
proc sort data= TEACHER;
by IDTEACH IDLINK ;
data STDTCH ;
set bM4.G8_STD_TCH_LNK07 ;
proc sort data= STDTCH;
by IDTEACH IDLINK;
data MERGE2;
merge STDTCH TEACHER;
by IDTEACH IDLINK;
if MATWGT > 0;
proc sort data = MERGE2;
by IDSTUD;
data bM4.MERGEALL;
merge MERGE1 MERGE2;
by IDSTUD;
run;
```

In the SPSS example shown in exhibit 5.12 the student and school data are first sorted by IDSCHOOL and then merged. The procedure followed for combining student and teacher data in exhibit 5.11 is used again. Then the saved student-school and student-teacher files are merged by IDSTUD, and a final dataset MERGEALL.SAV is saved.

Exhibit 5.12. Illustrative SPSS code for merging U.S. TIMSS eighth-grade school, mathematics teacher and student data

```
get file = “C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\G8_SCHOOL07.SAV".
sort cases by IDSCHOOL.
save outfile = SCHOOL.
get file = “C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\G8_STUDENT07.SAV".
sort cases by IDSCHOOL.
save outfile = STUDENT.
match files
/ file=STUDENT
/ table=SCHOOL
/ by IDSCHOOL.
save outfile = “C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\MERGE1.SAV" .
get file = “C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\G8_MTEACHER07.SAV".
sort cases by IDTEACH IDLINK .
save outfile = TEACHER .
get file = “C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\G8_STD_TCH_LNK07 .SAV".
select if MATWGT > 0 .
sort cases by IDTEACH IDLINK.
save outfile = STDTCH.
match files
/ file=STDTCH
/ table=TEACHER
/ by IDTEACH IDLINK.
save outfile = “C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\MERGE2.SAV".
Get file = “C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\MERGE1.SAV".
Sort cases by IDSTUD.
save outfile = MERGE1.
Get file = “C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\MERGE2.SAV".
Sort cases by IDSTUD.
save outfile = MERGE2.
match files
/ file=MERGE2
/ table=MERGE1
/ by IDSTUD.
save outfile = “C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\MERGEALL.SAV".
```

5.11.4 Merging TIMSS 2007 data with restricted-use data

Users who have been granted a license to use the restricted-use TIMSS 2007 data will receive a restricted-use CD-ROM which contains the public-use data and an additional link file that provides a way to merge TIMSS data with school data from the Common Core of Data (CCD) and the Private School Survey (PSS). The CCD_ID from the TIMSS file is used to merge with NCESSCH from the CCD file. The PSS_ID from the TIMSS file is used to merge with the PPIN from the PSS file. Illustrative SAS and SPSS code is provided in exhibits 5.13 and 5.14.

The code in question provides for a link between the TIMSS school data and the CCD/PSS data by school. Further merging to other TIMSS files (student, school, teacher) can be conducted using the IDSCHOOL for merging as has been shown in earlier examples.

Exhibit 5.13. Illustrative SAS code for merging U.S. TIMSS eighth-grade school data with CCD and PSS data

```
libname bm4 "C:\TIMSS2007\TIMSS2007 Raw Data and Control Files" ;
data SCHOOL ;
set bm4.Timss_2007_school_gr8_links;
proc sort data= SCHOOL;
by NCESSCH ;
data CCD ;
set bm4.CCD;
proc sort data= CCD;
by NCESSCH;
data MERGE1 ;
merge CCD(IN=IN1) SCHOOL (IN=IN2);
by NCESSCH;
IF  IN2;
run;

/* User can Merge in PSS data to the file containing CCD and TIMSS previously merged data */

Data SCHOOL2;
Set MERGE1;
proc sort data= SCHOOL2;
by PPIN ;
data PSS ;
set bm4.PSS;
proc sort data= PSS;
by PPIN;
data MERGE2 ;
merge PSS(IN=IN1) SCHOOL2 (IN=IN2);
by PPIN;
IF  IN2;
run;
```

SPSS code designed to provide this same link between the TIMSS school data and the CCD/PSS data by school is provided in exhibit 5.14. Further merging to other TIMSS files (student, school, teacher) can be conducted using the IDSCHOOL for merging as has been shown in earlier examples.

Exhibit 5.14. Illustrative SPSS code for merging U.S. TIMSS eighth-grade school data with CCD and PSS data

```
get file = "C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\ Timss_2007_school_gr8_links.SAV".
sort cases by NCESSCH.
save outfile = SCHOOL .
get file = "C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\ CCD.SAV".
sort cases by NCESSCH.
save outfile = CCD.
match files
/ file= CCD
/ table= SCHOOL
/ by NCESSCH.
save outfile = "C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\MERGE1.SAV" .

* Merge PSS to the Combined TIMSS/CCD school level file

get file = "C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\MERGE1.SAV ".
sort cases by PPIN.
save outfile = SCHOOL2 .
get file = "C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\PSS.SAV".
sort cases by PPIN.
save outfile = PSS.
match files
/ file= PSS
/ table= SCHOOL2
/ by PPIN.
save outfile = "C:\TIMSS2007\TIMSS2007 Raw Data and Control Files\MERGE2.SAV" .
```

5.12 Some notes on analyzing the TIMSS 2007 data

The design of TIMSS raises three special considerations for the analysis of the TIMSS data. First, the assessment design necessitates the use of *plausible values* rather than a single score for each of the various measures of mathematics and science achievement. Second, since the sampling design is not a simple random sample in which each student had an equal probability of selection, *sampling weights* need to be applied in order to generate unbiased estimates of population parameters. Third, the complex sampling design also means that the calculation of the *standard errors* of the various statistics generated requires special procedures.

5.12.1 Plausible values

The assessment design was based on Balanced Incomplete Block (BIB) spiraling of assessment items to increase content-area coverage without a concomitant increase in the assessment time demanded of students. Each student completed only a subset of the total pool of assessment items with the resulting data containing missing values for other items in the pool but not in the subset administered to the student. The trade-off for increased coverage through BIB spiraling is increased measurement error in the scores available for each student. This is accommodated through the estimation of five plausible values for each student rather than a single (unreliable) point estimate. Plausible values are random draws from the estimated distribution of a student's achievement. A detailed description of the TIMSS 2007 scaling can be found in Foy, Galia, and Li (2008).

What this means for those analyses of TIMSS data that include achievement measures is that the analyses need to be done five times and the results averaged. For example, if one was regressing mathematics achievement on a number of family and school attributes, it would be necessary to estimate this equation five times and then average each set of five parameter estimates. It would not be legitimate to take the mean of the five plausible values in the first instance and then regress this mean on a number of family and school attributes. The computing of standard errors is a more complicated process (see 5.11.2, below).

5.12.2 Estimating Sampling Variance

The complex sampling design used in TIMSS 2007 complicates the task of computing standard errors. Most standard analysis software systems such as SAS and SPSS provide estimates based on the assumption of a simple random sample. Given the TIMSS sampling design, such standard errors will underestimate the true standard errors.

There are, however, several options for estimating sampling errors that avoid the assumption of simple random sampling. TIMSS adopts the jackknife repeated replication technique (JRR) because it is computationally straightforward and provides approximately unbiased estimates of the sampling errors of means, totals, and percentages. The variables necessary for these JRR procedures are included as part of the TIMSS 2007 data files: JKZONE, the sampling zone (stratum) of the student's school; and, JKREP, the sampling replicate (primary sampling unit) of the student's school.

The SPSS-linked International Database (IDB) Analyzer software was designed specifically by IEA for analyzing TIMSS international data files. This software is freely available from the IEA website: http://www.iea.nl/iea_studies_datasets.html . Used in conjunction with SPSS, this software helps users analyze the TIMSS achievement data by conducting each analysis separately on each plausible value, averaging the resulting statistics, and applying the jackknife algorithm to provide appropriate standard errors for each statistic. It also simplifies management of the TIMSS database by providing a module for selecting subsets of countries and variables, and merging files for analysis. Most basic forms of analysis, including frequencies, crosstabulations, means, percentages, correlations, and regression analyses, are provided for with this software. The IDB Analyzer is described in detail in Foy and Olson (2009).

Special-use software is also available to use in estimating the standard errors of statistics generated from complex sampling designs. Among the packages available are: *AM*, available from the American Institutes for Research at http://am.air.org/about2.asp ; *Wesvar,* available from Westat at *http://www.westat.com/westat/statistical_software/wesvar; and, SUDAAN,* available from Research Triangle Institute at *http://www.rti.org/sudaan*. Some software packages provide for these capabilities as well. In addition, SAS macros suitable for this purpose are available as part of the *TIMSS 2007 User Guide for the International Database* (Foy and Olson 2009). See also, recent work by Stapleton (2008, 2009) which suggests procedures that can be used to generate appropriate standard errors for statistics generated by structural equation modeling techniques.

5.12.3 Using the IEA International Database Analyzer (IEA IDB Analyzer)

This software was developed by the IEA Data Processing and Research Center (IEA DPC) as a plug-in for SPSS and can only be used in conjunction with SPSS. It is not a stand-alone analysis system.

The IEA IDB Analyzer enables users to combine SPSS data files and conduct analyses using SPSS without actually writing programming code. The IEA IDB Analyzer generates SPSS syntax that takes into account information from the sampling design in the computation of statistics and their standard errors. In addition, the generated SPSS syntax makes appropriate use of plausible values for calculating estimates

of achievement scores and their standard errors, combining both sampling variance and imputation variance.

The IEA IDB Analyzer consists of two modules—a merge module and an analysis module. The *merge module* is used to create analysis datasets by combining data files of different types and from different countries, and selecting subsets of variables for analysis. The *analysis module* provides procedures for computing various statistics and their standard errors. All statistical procedures offered within the analysis module of the IEA IDB Analyzer make appropriate use of sampling weights and standard errors are computed using the jackknife repeated replication (JRR) method. Percentages, means, regressions, and correlations may be specified with or without achievement scores. When achievement scores are used, the analyses are performed five times—once for each plausible value—and the results are aggregated to produce accurate estimates of achievement and standard errors that incorporate both sampling and imputation errors.

The use of the IEA IDB Analyzer is described in detail with worked examples in chapter 2 of the *TIMSS 2007 User Guide for the International Database* (Foy and Olson 2009). Readers intending to use this software are urged to read this guide in detail.

5.12.4 SAS Programs and Macros

The *TIMSS 2007 User Guide for the International Database* also provides assistance for those investigators who wish to conduct their analyses using SAS. The *User Guide* DVD includes a number of SAS programs needed to process the SAS data files, compute survey results, and carry out example analyses. These are described in detail with worked examples in chapter 3 of the *User Guide***.** Readers intending to use SAS for their analyses are urged to read this chapter in detail.

The following SAS programs and macros are available:

- CONVERT.SAS: is used to convert SAS Export files into SAS data files.

- ASASCRM4.SAS, BSASCRM4.SAS: is used to convert the response codes on the achievement items to their corresponding score levels.

- JOIN.SAS: combines files of the same type from more than one country.

- JACKGEN.SAS: is a SAS macro used to compute weighted percentages of students within defined subgroups, along with their means on a specified continuous variable. This macro generates replicate weights and computes standard errors using the jackknife repeated replication (JRR) methodology. The analysis variable can be any continuous variable. (When computing mean achievement scores with plausible values, the macro JACKPV.SAS should be used.)

- JACKPV.SAS: is a SAS macro used to compute weighted percentages of students within defined subgroups, along with their mean achievement on a scale using the available plausible values. This macro generates replicate weights and computes standard errors using the jackknife repeated replication (JRR) and multiple imputation methodologies. This macro should be used when achievement plausible values are used in an analysis.

- JACKREG.SAS: is a SAS macro used to compute weighted regression coefficients and their standard errors within defined subgroups. This macro can be used with any analysis variable, but is not appropriate for analyzing achievement with plausible values.

- JACKREGP.SAS: is a SAS macro program used to compute weighted regression coefficients and their standard errors within defined subgroups when using achievement plausible values as the dependent variable.

5.12.5 Special considerations in using the teacher data

The teachers in the TIMSS 2007 international database are the teachers of nationally representative samples of students and are not representative samples of teachers in the participating countries. As a result, analyses with teacher data should be made with students as the units of analysis and reported in terms of students who are taught by teachers with a particular attribute.

When analyzing teacher data, it is first necessary to link the students to their respective teachers. The student-teacher linkage data files (AST/BST) were created for this purpose. Since student achievement scores (plausible values), jackknife replication information, and teacher weighting variables are found in the student-teacher linkage data files, it is only necessary to merge the teacher background data files with the student-teacher linkage data files. For analyses linking teacher variables to student background variables, it is also necessary to merge the student background data files with the teacher background data files after having been combined with the student-teacher linkage data files.

In general, to perform analyses using the teacher background data files, one should follow the steps outlined below:

1. Identify the variables of interest in the teacher background data files and note any specific national adaptations to the variables;

2. Retrieve the relevant variables from the teacher background data files, including analysis variables, classification variables, identification variables (IDCNTRY, IDTEACH, and IDLINK), and any other variables used in the selection of cases;

3. Retrieve the relevant variables from the student-teacher linkage data files, including plausible values of achievement, classification variables, identification variables (IDCNTRY, IDSTUD, IDTEACH, and IDLINK), sampling (JKZONE and JKREP) and weighting (MATWGT, SCIWGT, or TCHWGT) variables, and any other variables used in the selection of cases;

4. Merge the teacher background data files with the student-teacher linkage data files using the variables IDCNTRY, IDTEACH and IDLINK; and

5. If student background variables also are needed, merge the student background data files with the merged student-teacher data files from the previous step using the variables IDCNTRY and IDSTUD.

One further point: fourth-grade teachers were given a single questionnaire with mathematics and science sections. Teachers who taught only mathematics or only science to the TIMSS fourth-graders would complete only one of these sections. In frequency distributions of the variables in these sections, teachers who did not answer the questions for this reason are shown as “not administered.”

5.12.6 Special considerations in using the school data

In general, to perform analyses using the school background data files, you should do the following:

1. Identify the variables of interest in the school and student background data files and note any specific national adaptations to the variables;

2. Retrieve the relevant variables from the school background data files, including analysis variables, classification variables, identification variables (IDCNTRY and IDSCHOOL), and any other variables used in the selection of cases; and

3. Retrieve the relevant variables from the student background data files, including plausible values of achievement, classification variables, identification variables (IDCNTRY and IDSCHOOL), sampling (JKZONE and JKREP) and weighting (TOTWGT) variables, and any other variables used in the selection of cases.

Merge the school background data files with the student background data files using the variables IDCNTRY and IDSCHOOL.

5.13 Accessing data from other countries

Data on the other nations participating in TIMSS 2007 may be downloaded, along with documentation explaining the structure and content of the database, at http://timssandpirls.bc.edu/TIMSS2007/idb_ug.html. The international student, teacher, and school datasets are large, single datasets containing all countries. Subsets of countries may be created or data from other countries may be combined with the U.S. dataset using merge procedures similar to those described in chapters 2 and 3 of the *TIMSS 2007 User Guide for the International Database* (Foy and Olson 2009).

References

Beaton, A.E. and González, E. (1995) *The NAEP Primer*. Chestnut Hill, MA: Boston College.

Brick, M.J., Morganstein, D., and Valliant, R. (2000). *Analysis of Complex Sample Data Using Replication.* Retrieved June 20, 2008, from http://www.westat.com/westat/statistical_software/WesVar/index.cfm.

Broughman, S.P., Swaim, N.L., and Keaton, P.W. (2009). *Characteristics of Private Schools in the United States: Results From the 2007-08 Private School Universe Survey* (NCES 2009-313). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Ferraro, D., and Van de Kerckhove, W. (2006). *Trends in International Mathematics and Science Study (TIMSS) 2003 Nonresponse Bias Analysis* (NCES 2007-044). National Center for Education Statistics, Institute of Education Science, U.S. Department of Education. Washington, DC.

Foy, P. and Olson, J.F. (2009). *TIMSS 2007 User Guide for the International Database*. Chestnut Hill, MA: Boston College.

Foy, P., Joncas, M., and Zuhlke, O. (2005).*TIMSS 2007 School Sampling Manual.* Unpublished Manuscript. Chestnut Hill, MA: Boston College

Foy, P., Galia, J., and Li, I. (2008). Scaling the data from the TIMSS 2007 mathematics and science assessments. In J.F. Olson, M.O. Martin, and I.V.S. Mullis (Eds.), *TIMSS 2007 Technical Report* (pp. 225-280). Chestnut Hill, MA: Boston College.

Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., and Brenwald, S. (2008). *Highlights From TIMSS 2007: Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context* (NCES 2009–001). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Hoffman, L. (2009). *Numbers and Types of Public Elementary and Secondary Schools From the Common Core of Data: School Year 2006–07* (NCES 2009-304REV). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

IEA Data Processing Center. (2006a). *Within School Sampling Software (WinW3S).* Hamburg: IEA Data Processing Center.

IEA Data Processing Center. (2006b) *TIMSS 2007 Data Entry Manager Manual.* Hamburg: IEA Data Processing Center.

Joncas, M. (2008). TIMSS 2007 sample design. In J.F. Olson, M.O. Martin, and I.V.S. Mullis (Eds.), *TIMSS 2007 Technical Report* (pp. 77-92). Chestnut Hill, MA: Boston College.

Kaufman, S., Seastrom, M., and Roey, S. (2005). Do Disclosure Controls to Protect Confidentiality Degrade the Quality of the Data? *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 1218-1225.

Krenzke, T. and Lieber, V. (2007). *NAEP 2007 Nonresponse Bias Analysis Report* (Prepared under contract to National Center for Education Statistics). Rockville, MD: Westat.

Martin, M. O. (2005) *TIMSS 2003 User Guide for the International Database.* Chestnut Hill, MA: Boston College.

Martin, M.O., Mullis, I.V.S., and Foy, P. (2008). *TIMSS 2007 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., and Foy, P. (2008). *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades.* Chestnut Hill, MA: Boston College

Mullis, I.V.S., Martin, M.O., Ruddock, G.J., O'Sullivan, C.Y., Arora, A., and Erberber, E. (2005). *TIMSS 2007 Assessment Frameworks*. Chestnut Hill, MA: Boston College.

National Center for Education Statistics. (2002). *NCES Statistical Standards* (NCES 2003-601). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Olson, J.F., Martin, M.O., and Mullis, I.V.S. (2008). *TIMSS 2007 Technical Report.* Chestnut Hill, MA: Boston College.

Rao, J.N.K., and Scott, A.J. (1984). On Chi-squared Tests for Multiway Contingency Tables with Cell Proportions Estimated from Survey Data. *Annals of Statistics, 12*: 46-60.

Rao, J.N.K., and Thomas, D.R. (2003). Analysis of Categorical Response Data from Complex Surveys: An Appraisal and Update. In R.L. Chambers and C.J. Skinner (Eds.), *Analysis of Survey Data* (pp. 85-108). West Sussex, England: John Wiley and Sons.

Stapleton, L.M. (2006). An assessment of practical solutions for structural equation modeling with complex sample data. *Structural Equation Modeling,13*, 28-58.

Stapleton, L.M. (2008). Variance estimation using replication methods in structural equation modeling with complex sample data. *Structural Equation Modeling,15*, 183-210.

United Nations Educational, Scientific and Cultural Organization (UNESCO). (1999). *Classifying Educational Programmes Manual for ISCED-97 Implementation in OECD Countries* (1999 Edition). Paris: UNESCO.

Westat. (2007). *WesVar 4.3 User's Guide.* Rockville, MD: Westat, Inc.