U.S. Department of Education
Institute of Education Sciences
NCES 2007-044

Trends in International Mathematics and Science Study (TIMSS) 2003 Nonresponse Bias Analysis

Technical Report

David Ferraro
Wendy Van de Kerckhove
Westat

Patrick Gonzales
Project Officer
National Center for Education Statistics

U.S. Department of Education
Margaret Spellings
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Mark Schneider
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 nations. 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

November 2006

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.

This publication is only available online. To download, view, and print the report as a PDF file, go to the NCES World Wide Web Electronic Catalog address shown above.

Suggested Citation

Ferraro, D., and Van de Kerckhove, W. (2006). *Trends in International Mathematics and Science Study (TIMSS) 2003 Nonresponse Bias Analysis* (NCES 2007-044). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [date] from http://nces.ed.gov/pubsearch.

Content Contact

Patrick Gonzales
415-920-9229
E-mail: patrick.gonzales@ed.gov

Executive Summary

The Trends in International Mathematics and Science Study (TIMSS) is a large international comparative study that was developed by the International Association for the Evaluation of Educational Achievement (IEA) to measure trends in the mathematics and science achievement of fourth- and eighth-grade students. The United States and other nations participate in this program on a 4-year cycle. The TIMSS assessments are closely linked to the curricula of the participating countries, providing an indication of the degree to which students have learned concepts in mathematics and science they are likely to encounter in school.

This report is concerned with the extent of potential bias introduced into the U.S. study through nonresponse on the part of schools. Data from the third cycle of TIMSS, conducted in April-June, 2003, are the basis for the current analyses. Specifically, the National Center for Education Statistics (NCES), which funded the U.S. study, requires that a nonresponse bias analysis be completed for any stage of data collection which has a unit response rate less than 85 percent. In the 2003 U.S. TIMSS, the weighted response rates (prior to the use of replacement schools) for grade 4 and grade 8 schools were 70 and 71 percent, respectively, substantially below the NCES established standard.

Effects of School Nonresponse

The analysis compares selected characteristics likely to reflect bias in participation from participating and non-participating schools. Frame characteristics for public schools come from the 2000–01 Common Core of Data (CCD); for private schools, the characteristics were obtained from the 2000–01 Private School Survey (PSS). The selected variables include school control, community type; geographic region; poverty level; number of grade-eligible students enrolled; total number of students; and percentage of students by race/ethnicity categories.[1] The percentage of students eligible for the free or reduced-price lunch under the National School Lunch Program, available only for public schools, also was included.

Two forms of analysis were undertaken: a test of the independence of each school characteristic and participation status, and logistic regression in which the conditional independence of these school characteristics as predictors of participation was examined. Significance tests used with analyses in the first of these categories were based on a Rao-Scott modified Chi-square statistic in the case of categorical variables and a *t* test on the differences between means for continuous variables. The 95 percent confidence level was used for all tests of significance. All analyses were performed using WesVar

[1] The definitions of the variables are found in section 2, Methodology, p. 4.

(Westat 2002) with replicate weights to properly account for the complex sample design. Two sets of school weights were used: the school base weights that did not include a nonresponse adjustment factor, and nonresponse adjusted school weights.

The statistical analyses used two types of samples: the original and the final sample. The original sample treats all the schools that were replaced as non-participants; thus replacement schools are not included in this analysis. The final sample analysis includes the replacement schools in the participating schools group, and treats the nonresponding schools that were not replaced as non-participants. The final sample analysis was repeated using the nonresponse adjusted school weights.

For the original sample of schools in grade 4 TIMSS (TIMSS-4), there was no statistically significant relationship detected between participation and any school characteristic in either the bivariate or logistic regression analyses. The bivariate results for the final sample of schools in TIMSS-4 indicate that participating schools had statistically significantly higher percentages of Black, non-Hispanic students than the eligible sample of schools (table 3-8). The bivariate analysis using the National School Lunch Program status also reveals that participating schools had a statistically significantly lower proportion of students eligible for this program than the eligible sample of schools (table 3-9). However, the regression analysis results do not detect any variables as significant predictors of participation in TIMSS-4 (table 3-10). For the final sample of schools in TIMSS-4 with school nonresponse adjustments applied to the weights, there was no statistically significant relationship detected between participation and any available school characteristic in both the bivariate and regression analysis.

For the original sample of schools in grade 8 TIMSS (TIMSS-8), the bivariate analysis indicates statistically significant differences in three variables related to participation in: community type (table 4-1); the total number of students enrolled in participating schools was lower than in all schools (table 4-2); and the percentage of Hispanic students in participating schools was lower than the eligible sample (table 4-3). The regression analysis shows only that the total number of students enrolled in participating schools to be statistically significantly lower than in all schools in TIMSS-8 (table 4-5). For the final sample, the bivariate analysis shows two of the three variables remained statistically significant: community type (table 4-6) and the total number of students enrolled was statistically significantly lower in participating schools than in the eligible sample (table 4-7). The regression analysis indicated that rural/small town schools were more likely than urban fringe/large town schools to participate (4-10). For the final sample of schools in TIMSS-8 with school nonresponse adjustments applied to the weights, the results are identical to the previous analysis.

The investigation into nonresponse bias at the school level for U.S. TIMSS 2003 samples for grades 4 and 8 shows that there was no statistically significant relationship detected between participation status and the majority of school characteristics that are available for analysis. At grade 4, the analyses suggest that there is little potential for nonresponse bias in the TIMSS-4 original sample based on the characteristics studied. It also suggests that while there is certainly no evidence that the use of replacement schools reduced the potential for bias it has also not substantially added to it. At grade 8, however, the analyses suggest that there is some potential for nonresponse bias in the TIMSS-8 original sample based on the characteristics studied. While there is no evidence that the use of replacement schools reduced the potential for bias, the analyses indicate that it has also not substantially added to it.

Acknowledgments

Many people contributed to making this report possible, and the authors wish to thank all those who have assisted with various aspects of the report.

We wish to thank Keith Rust of Westat for his comments and technical assistance and William Wall of Westat for programming support.

We also wish to thank the technical reviewers, Trevor Williams and Dan Levine of Westat; Jack Buckley, Patrick Gonzales, Steve Kaufman, Ralph Lee, Val Plisko, Elois Scott, Bruce Taylor, and Marilyn Seastrom of the National Center for Education Statistics; and Juan Carlos Guzmán, formerly of the Education Statistics Services Institute.

Contents

Contents

List of Tables

1. Introduction

The Trends in International Mathematics and Science Study (TIMSS) is a large international comparative study of the knowledge, skills, and competencies of fourth- and eighth-grade students in the domains of mathematics and science. The study was carried out in some 45 countries, including the United States. The student population sampled in TIMSS is defined as all students in each country who were enrolled in the upper of the two adjacent grades that contained the greatest proportions of 9- and 13-year-olds respectively at the time of testing. In most participating nations, including the United States, this corresponds to all fourth-graders and eighth-graders.

The U.S. TIMSS study, supported by the National Center for Education Statistics (NCES) and the National Science Foundation, utilized a two-stage stratified cluster sampling design. The first stage made use of a systematic probability-proportionate-to-size technique to select schools. Though efforts were made to secure the participation of all schools selected in the first stage, 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 (immediately preceding and following it) were designated as replacement schools. If an original school refused to participate, the first replacement was then contacted. If that school also refused to participate, the second school was then contacted.

The second stage of sampling consisted of selecting classrooms within sampled schools. At the classroom level, TIMSS sampled intact mathematics classes that were available to students in the target grades. The U.S. fourth-grade sample consisted of 300 eligible schools[2] containing at least one fourth-grade class. Low-income schools—defined as those in which at least 50 percent of students were eligible for participation in the federal National School Lunch Program (NSLP)—were oversampled. The U.S. eighth-grade sample consisted of 296 eligible[3] schools containing at least one eighth-grade class. There was no oversampling of any type of schools at grade eight. Where feasible, two classrooms were selected per school in the United States. In schools containing only one class, this class was selected. Data were collected in April, May, and June 2003. Detailed information on technical aspects of the study is included in appendix A.

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. For grade 4 (hereafter referred to as TIMSS-4), the weighted response rate for

[2] Of the 310 original schools selected for the sample, there were 10 ineligible or closed schools at grade 4.

[3] Of the 301 original schools selected for the sample, there were 5 ineligible or closed schools at grade 8.

eligible schools was 70 percent, with 212 out of 300 schools participating. For grade 8 (hereafter referred to as TIMSS-8), the unweighted response rate for eligible schools was 71 percent, with 211 out of 296 schools participating. See table A-1 for details on the U.S. TIMSS school participation rates. Since the U.S. TIMSS weighted school response rates are below 85 percent, NCES requires an investigation into the potential magnitude of nonresponse bias at the school level in the U.S. sample, which is the focus of this report. The methodology used to investigate nonresponse bias analysis in the TIMSS-4 and TIMSS-8 U.S. samples is provided in chapter 2, and the results are provided in chapters 3 and 4. Conclusions follow in chapter 5.

2. 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 2005 National Assessment of Educational Progress (NAEP) (Kali and Rust 2005) and the 2001 Program for International Reading Literacy Study (Piesse and Rust 2003).

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 replacements 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 replacement school was equal 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 for each explicit stratum. For TIMSS-4, the explicit strata were defined simply by school control. For TIMSS-8, there was no explicit stratification and thus a single adjustment class.

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. Both the second and third analyses, however, may provide an overly optimistic scenario, since even though substitution and nonresponse adjustments 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 2000–01 Common Core of Data (CCD) and, for private schools, from the 2000–01 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;
- National Assessment of Educational Progress (NAEP) region (see appendix for state listing); and
- poverty level[4]—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.

The following continuous variables were available for all schools:

- number of grade eligible (4 or 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).

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 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

[4] The sample frame did not contain a direct measure of poverty.

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. All the frame characteristics were included in the model.

The logistic regression was performed using WesVar (Westat 2002) 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).

3. Results—TIMSS Grade 4

Original Respondent Sample

This section presents the results of the nonresponse bias analysis, based exclusively on the original sample of 300 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 whether they were substituted by a replacement school. The unweighted response rate was 71 percent, with 212 out of 300 eligible schools participating. The weighted response rate was 70 percent. See table A-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 3-1. 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. The *p*-values for the Chi-square tests are presented in the table.

Based on these analyses, no significant differences were detected between participation status and any of the characteristics shown in table 3-1.

Table 3-1. Percentage distribution of eligible and participating schools in the U.S. TIMSS fourth-grade original sample, by selected categorical variables: 2003

School characteristic	Sample schools: Eligible (percent)	Sample schools: Participating (percent)	Bias	Relative bias	Chi-square *p*-value
School control					0.713
Public	90.4	90.0	-0.40	-0.004	
Private	9.6	10.0	0.40	0.041	
Community type					0.482
Central city	31.5	31.7	0.17	0.005	
Urban fringe/large town	40.5	38.6	-1.85	-0.046	
Rural/small town	28.0	29.7	1.68	0.060	

See notes at end of table.

Table 3-1. Percentage distribution of eligible and participating schools in the U.S. TIMSS fourth-grade original sample, by selected categorical variables: 2003—Continued

School characteristic	Sample schools: Eligible (percent)	Sample schools: Participating (percent)	Bias	Relative bias	Chi-square *p*-value
NAEP region					0.312
Northeast	20.8	18.0	-2.81	-0.135	
Southeast	23.5	24.0	0.52	0.022	
Central	22.8	24.4	1.59	0.070	
West	33.0	33.6	0.70	0.021	
Poverty level					0.668
High	33.4	34.1	0.70	0.021	
Low	66.6	65.9	-0.70	-0.010	

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 NSLP; all private schools are treated as low poverty schools. NAEP region is the state-based region of the country (see appendix for state listing.) Eligible schools contained at least one fourth-grade class. Participating schools agreed to have their students assessed. 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), 2003.

Continuous Variables (TIMSS-4)

Summary means for each continuous variable for participating and eligible schools are shown in tables 3-2 through 3-4. The difference between means of participating and eligible schools was tested using a *t* test. The *p*-values for the *t* tests are presented in the tables. One school had a missing value for the total number of students, one school had a missing value for one or more of the race/ethnicity variables, and 26 out of the 273 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 as shown in tables 3-2 through 3-4. However, this must be interpreted with caution for the free or reduced-price lunch variable because it is missing for 26 schools.

Table 3-2. Mean enrollment of eligible and participating schools in the U.S. TIMSS fourth-grade original sample: 2003

Student enrollment	Sample schools: Eligible (mean)	Sample schools: Participating (mean)	Bias	Relative bias	*t* test *p*-value
Total school	583.9	576.4	-7.50	-0.013	0.483
Fourth grade	99.5	99.3	-0.26	-0.003	0.912

NOTE: Information on total school enrollment is missing for one non-participating school of the 300 eligible schools in the sample. Eligible schools contained at least one fourth-grade class. Participating schools agreed to have their students assessed. 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), 2003.

Table 3-3. Mean percentage of students in eligible and participating schools in the U.S. TIMSS fourth-grade original sample, by race/ethnicity: 2003

Race/ethnicity	Sample schools: Eligible (percent)	Sample schools: Participating (percent)	Bias	Relative bias	*t* test *p*-value
White, non-Hispanic	58.4	56.8	-1.64	-0.028	0.155
Black, non-Hispanic	16.4	17.6	1.20	0.073	0.138
Hispanic	17.2	17.5	0.25	0.014	0.762
Asian or Pacific Islander	4.0	4.0	-0.08	-0.020	0.800
American Indian or Alaska Native	1.2	1.5	0.27	0.221	0.146

NOTE: Information on race/ethnicity is missing for one non-participating school of the 300 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 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), 2003.

Table 3-4. 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: 2003

	Sample schools				
Students	Eligible (percent) (N = 247)	Participating (percent) (N = 172)	Bias	Relative bias	t test p-value
Percentage of students eligible for free or reduced-price lunch	43.9	45.2	1.35	0.031	0.243

NOTE: Information on percentage of students eligible for free or reduced-price lunch is missing for 26 of the 273 public schools in the sample. Eligible schools contained at least one fourth-grade class. Participating schools agreed to have their students assessed. 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), 2003.

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. One school was 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 table 3-5. None of the parameter estimates are significant, which indicates that there were no significant relationships detected with participation status.

Table 3-5. Logistic regression model parameter estimates using the U.S. TIMSS fourth-grade original sample: 2003

Parameter	Parameter estimate	Standard error	*t* test for H_0: parameter = 0	*p*-value
Intercept	0.907	1.1536	0.7858	0.4344
Central city	-0.138	0.3232	-0.4275	0.6702
Rural/small town	0.441	0.3310	1.3331	0.1864
Private school	0.655	0.5835	1.1219	0.2654
High poverty	-0.233	0.3222	-0.7239	0.4713
Northeast	-0.446	0.4010	-1.1112	0.2699
Southeast	-0.035	0.4757	-0.0727	0.9423
Central	0.337	0.4422	0.7631	0.4477
Total school enrollment	-0.001	0.0007	-0.8958	0.3732
Fourth grade enrollment	0.003	0.0036	0.7905	0.4317
White, non-Hispanic	-0.004	0.0104	-0.3916	0.6964
Black, non-Hispanic	0.011	0.0111	0.9661	0.3370
Hispanic	0.006	0.0129	0.4436	0.6586
Asian or Pacific Islander	0.004	0.0188	0.2266	0.8213
American Indian or Alaska Native	0.033	0.0248	1.3421	0.1835

NOTE: One non-participating school of the 300 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. NAEP region is the state-based region of the country (see appendix 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), 2003.

Respondent Sample With Replacements (Final Sample)

This section presents the nonresponse bias analysis based on the final sample of 300 schools for TIMSS-4. The distribution of the responding sample, including participating replacements 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 replacements, the unweighted school response rate for TIMSS-4 was 83 percent, with 248 out of 300 schools participating. The weighted response rate was 82 percent.

Categorical Variables (TIMSS-4)

The distribution of participating and eligible schools by the four characteristics is shown in table 3-6. 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. The *p*-values for the Chi-square tests are presented in the table. There were no statistically significant relationships detected between participation status and any of the characteristics.

Table 3-6. Percentage distribution of eligible and participating schools in the U.S. TIMSS fourth-grade final sample, by selected categorical variables: 2003

	Sample schools				
School characteristic	Eligible (percent)	Participating (percent)	Bias	Relative bias	Chi-square *p*-value
School control					0.252
Public	90.4	91.4	1.04	0.0115	
Private	9.6	8.6	-1.04	-0.1081	
Community type					0.340
Central city	31.5	32.5	0.96	0.0304	
Urban fringe/large town	40.5	38.6	-1.88	-0.0465	
Rural/small town	28.0	29.0	0.93	0.0330	
NAEP region					0.854
Northeast	20.8	19.9	-0.88	-0.0426	
Southeast	23.5	23.3	-0.20	-0.0086	
Central	22.8	23.4	0.58	0.0252	
West	33.0	33.5	0.51	0.0156	
Poverty level					0.133
High	33.4	35.1	1.65	0.0493	
Low	66.6	64.9	-1.65	-0.0248	

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 NSLP; all private schools are treated as low poverty schools. NAEP region is the state-based region of the country (see appendix for state listing.) Eligible schools contained at least one fourth-grade class. Participating schools agreed to have their students assessed. 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), 2003.

Continuous Variables (TIMSS-4)

Summary means for each continuous variable for participating and eligible schools are shown in tables 3-7 through 3-9. The difference between means of participating and eligible schools was tested using a *t* test. The *p*-values for the *t* tests are presented in the tables. One school had a missing value for the total number of students, one school had a missing value for one or more of the race/ethnicity variables, 28 out of the 273 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 3-7). Participating schools had a statistically significantly higher mean percentage of Black, non-Hispanic students than the eligible sample (18.0 vs. 16.6 percent, respectively; table 3-8). The difference in the percentage of students who are Black was small in absolute bias but substantial in relative bias, and was similar to that shown table 3-3, in which only the original sample was considered. Thus while there is certainly no evidence that the use of replacement schools reduced the potential for bias, as indicated by this variable, it has also not substantially added to it as the change in relative bias is less than one percent (0.073 vs. 0.081). There were no statistically significant differences detected in the mean percentage of students of the other races and ethnicities (White, Hispanic, Asian or Pacific Islander; and American Indian or Alaska Native) between the participating and eligible schools (table 3-8).

Participating schools had a statistically significantly higher mean percentage of students eligible for free or reduced-price lunch than for the eligible sample of public schools (45.8 vs. 44.0 percent, respectively; table 3-9). However, this must be interpreted with caution because the variable is missing for 28 schools. The difference in the percentage of students eligible was small in absolute bias but substantial in relative bias, and was similar to that shown Table 3-4, in which only the original sample was considered. Thus, and in the case with the percentage of students who are Black, while there is certainly no evidence that the use of replacement schools reduced the potential for bias, as indicated by the free or reduced-price lunch variable, it has also not substantially added to it as the change in relative bias is one percent (0.031 vs. 0.041).

Table 3-7. Mean enrollment of eligible and participating schools in the U.S. TIMSS fourth-grade final sample: 2003

	Sample schools				
Student enrollment	Eligible (mean)	Participating (mean)	Bias	Relative bias	*t* test *p*-value
Total school	579.5	569.7	-9.71	-0.017	0.214
Fourth grade	99.3	98.6	-0.71	-0.007	0.699

NOTE: Information on total school enrollment is missing for one non-participating school of the 300 eligible schools in the sample. Eligible schools contained at least one fourth-grade class. Participating schools agreed to have their students assessed. 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), 2003.

Table 3-8. Mean percentage of students in eligible and participating schools in the U.S. TIMSS fourth-grade final sample, by race/ethnicity: 2003

	Sample schools				
Race/ethnicity	Eligible (percent)	Participating (percent)	Bias	Relative bias	*t* test *p*-value
White, non-Hispanic	58.0	56.5	-1.46	-0.025	0.077
Black, non-Hispanic	16.6	18.0	1.35	0.081	0.021
Hispanic	17.0	17.5	0.46	0.027	0.441
Asian or Pacific Islander	4.4	4.3	-0.14	-0.031	0.562
American Indian or Alaska Native	1.3	1.5	0.14	0.108	0.192

NOTE: Information on race/ethnicity is missing for one non-participating school of the 300 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 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), 2003.

Table 3-9. 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: 2003

Students	Sample schools: Eligible (percent) (N = 245)	Sample schools: Participating (percent) (N = 203)	Bias	Relative bias	t test p-value
Percentage of students eligible for free or reduced-price lunch	44.0	45.8	1.81	0.041	0.019

NOTE: Information on percentage of students eligible for free or reduced-price lunch is missing for 28 of the 273 public schools in the sample. Eligible schools contained at least one fourth-grade class. Participating schools agreed to have their students assessed. 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), 2003.

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. One school was 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 table 3-10. None of the parameter estimates are significant which indicates that there were no significant relationships detected with participation status.

Table 3-10. Logistic regression model parameter estimates in the U.S. TIMSS fourth-grade final sample: 2003

Parameter	Parameter estimate	Standard error	*t* test for H_0: parameter = 0	*p*-value
Intercept	1.353	1.2544	1.0782	0.2843
Central city	0.101	0.4525	0.2232	0.8240
Rural/small town	0.474	0.4178	1.1339	0.2604
Private school	-0.345	0.5374	-0.6428	0.5223
High poverty	-0.124	0.3822	-0.3251	0.7460
Northeast	-0.251	0.5095	-0.4934	0.6231
Southeast	-0.169	0.6148	-0.2756	0.7836
Central	0.174	0.5435	0.3197	0.7501
Total school enrollment	-0.001	0.0008	-1.5526	0.1246
Fourth grade enrollment	0.001	0.0041	0.1949	0.8460
White, non-Hispanic	0.003	0.0113	0.2684	0.7891
Black, non-Hispanic	0.025	0.0150	1.6846	0.0961
Hispanic	0.015	0.0144	1.0663	0.2896
Asian or Pacific Islander	0.024	0.0228	1.0469	0.2984
American Indian or Alaska Native	0.003	0.0202	0.1502	0.8810

NOTE: One non-participating school of the 300 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. NAEP region is the state-based region of the country (see appendix 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), 2003.

Final Sample, With Nonresponse Adjustments Applied

This section presents the nonresponse bias analysis based on the final sample of 300 schools for TIMSS-4. This is the same sets of schools that were compared as in the previous analysis, but this time when analyzing the responding schools alone, school nonresponse adjustments were applied to the weights.

Categorical Variables (TIMSS-4)

The distribution of participating and eligible schools by the four characteristics is shown in table 3-11. 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. The *p*-values for the Chi-square tests are presented in the table.

There were no statistically significant relationships detected between participation status and any of the characteristics, shown in table 3-11.

Table 3-11. Percentage distribution of eligible and participating schools in the U.S. TIMSS fourth-grade final sample after nonresponse adjustment, by selected categorical variables: 2003

School characteristic	Sample schools: Eligible (percent)	Sample schools: Participating (percent)	Bias	Relative bias	Chi-square *p*-value
School control					0.4065
Public	90.4	91.2	0.78	0.009	
Private	9.6	8.8	-0.78	-0.081	
Community type					0.450
Central city	31.5	32.1	0.60	0.019	
Urban fringe/large town	40.5	38.9	-1.60	-0.040	
Rural/small town	28.0	29.0	1.00	0.036	
NAEP region					0.830
Northeast	20.8	20.1	-0.73	-0.035	
Southeast	23.5	23.1	-0.42	-0.018	
Central	22.8	23.7	0.93	0.041	
West	33.0	33.2	0.22	0.007	
Poverty level					0.751
High	33.4	33.1	-0.34	-0.010	
Low	66.6	66.9	0.34	0.005	

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 NSLP; all private schools are treated as low poverty schools. NAEP region is the state-based region of the country (see appendix for state listing.) Eligible schools contained at least one fourth-grade class. Participating schools agreed to have their students assessed. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school nonresponse adjusted weights.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Continuous Variables (TIMSS-4)

Summary means for each continuous variable for participating and eligible schools are shown in tables 3-12 through 3-14. The difference between means of participating and eligible schools was tested using a *t* test. The *p*-values for the *t* tests are presented in the tables. One school had a missing value for the total number of students, one school had a missing value for one or more of the race/ethnicity variables, 28 out of the 273 public schools had a missing value for the free or reduced-price lunch variable; these schools were excluded from the analysis.

Based on the analyses shown in tables 3-12 through 3-14, there were no statistically significant differences detected between participating and eligible schools. However, this must be interpreted with caution for the free or and reduced-price lunch variable because it is missing for 28 schools.

Table 3-12. Mean enrollment of eligible and participating schools in the U.S. TIMSS fourth-grade final sample after nonresponse adjustment: 2003

Student enrollment	Sample schools: Eligible (mean)	Sample schools: Participating (mean)	Bias	Relative bias	*t* test *p*-value
Total school	579.5	567.1	-12.32	-0.021	0.114
Fourth grade	99.3	98.2	-1.13	-0.011	0.544

NOTE: Information on total school enrollment is missing for one non-participating school of the 300 eligible schools in the sample. Eligible schools contained at least one fourth-grade class. Participating schools agreed to have their students assessed. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school nonresponse adjusted weights.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Table 3-13. Mean percentage of students in eligible and participating schools in the U.S. TIMSS fourth-grade final sample after nonresponse adjustment, by race/ethnicity: 2003

Race/ethnicity	Sample schools: Eligible (percent)	Sample schools: Participating (percent)	Bias	Relative bias	*t* test *p*-value
White, non-Hispanic	58.0	57.2	-0.77	-0.013	0.348
Black, non-Hispanic	16.6	17.6	1.01	0.061	0.078
Hispanic	17.0	17.1	0.07	0.004	0.910
Asian or Pacific Islander	4.4	4.2	-0.16	-0.035	0.507
nerican Indian or Alaska Native	1.3	1.4	0.13	0.095	0.271

NOTE: Information on race/ethnicity is missing for one non-participating school of the 300 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 relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school nonresponse adjusted weights.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Table 3-14. 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 after nonresponse adjustment: 2003

Students	Sample schools: Eligible (percent) (N = 245)	Sample schools: Participating (percent) (N = 203)	Bias	Relative bias	t test p-value
Percentage of students eligible for free or reduced-price lunch	44.0	44.7	0.70	0.016	0.353

NOTE: Information on percentage of students eligible for free or reduced-price lunch is missing for 28 of the 273 public schools in the sample. Eligible schools contained at least one fourth-grade class. Participating schools agreed to have their students assessed. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school nonresponse adjusted weights.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Logistic Regression Model (TIMSS-4)

The analysis of the final sample of original and replacement schools utilized a logistic regression model, with participation status as the binary dependent variable and frame characteristics and nonresponse adjustment class as predictor variables. The international weighting procedures created a nonresponse adjustment class for each explicit stratum. For grade 4, the explicit strata were defined by school control, resulting in two cells: public and private schools. Since school control was included in the analysis described in the previous section, these results are identical to those in table 3-10. An F-test was used to determine whether the frame characteristics, in aggregate, were significantly related to response, once the nonresponse adjustment classes were taken into account. The F-test statistic is 0.79 with a p-value of 0.673, which indicates no significant relationship detected with participation.

Summary—Grade 4

For the original sample of schools in TIMSS-4 in the United States, there was no statistically significant relationship detected between participation and any available school characteristic in both the bivariate and logistic regression analyses.

For the final sample of schools in TIMSS-4, the bivariate analysis indicated a significant effect of percentage of Black, non-Hispanic students and free or reduced lunch program eligibility (participating schools were more likely to have a larger percentage of Black students and students eligible for free or reduced-price lunch). However, the regression analysis did not show any variables to be significant predictors of participation.

For the final sample of schools in TIMSS-4 with school nonresponse adjustments applied to the weights, there was no statistically significant relationship detected between participation and any available school characteristic in both the bivariate and regression analyses.

4. Results–TIMSS Grade 8

Original Respondent Sample

The following nonresponse bias analysis is based exclusively on the original sample of 296 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 substituted by a replacement school. The weighted and unweighted response rates were both 71 percent with 211 out of 296 original schools participating in TIMSS. See table A-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 4-1. 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. The *p*-values for the Chi-square tests are presented in the table.

Based on these analyses, the Chi-square statistic for community type is significant and suggests that there is evidence of a relationship with participating in the assessment. There were no statistically significant relationships detected between participation status and the other characteristics shown in table 4-1.

Table 4-1. Percentage distribution of eligible and participating schools in the U.S. TIMSS eighth-grade original sample, by selected categorical variables: 2003

School characteristic	Sample schools: Eligible (percent)	Sample schools: Participating (percent)	Bias	Relative bias	Chi-square *p*-value
School control					0.195
Public	90.8	92.3	1.50	0.017	
Private	9.2	7.7	-1.50	-0.163	
Community type					0.005
Central city	29.2	26.7	-2.47	-0.085	
Urban fringe/large town	42.2	39.3	-2.93	-0.070	
Rural/small town	28.6	34.0	5.40	0.189	
NAEP region					0.138
Northeast	20.4	20.6	0.18	0.009	
Southeast	24.1	25.8	1.67	0.069	
Central	22.8	24.8	2.08	0.091	
West	32.7	28.7	-3.93	-0.120	
Poverty level					0.602
High	22.4	21.5	-0.93	-0.041	
Low	77.6	78.5	0.93	0.012	

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 NSLP; all private schools are treated as low poverty schools. NAEP region is the state-based region of the country (see appendix for state listing). Eligible schools contained at least one eighth-grade class. Participating schools agreed to have their students assessed. 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), 2003.

Continuous Variables (TIMSS-8)

Summary means for each continuous variable for participating and eligible schools are shown in tables 4-2 through 4-4. The difference between means of participating and eligible schools was tested using a *t* test. The *p*-values for the *t* tests are presented in the tables. One school had a missing value for the total number of students, one school had a missing value for one or more of the race/ethnicity variables, 40 out of the 268 public schools had a missing value for the free or reduced-price lunch variable; these schools were excluded from the analysis.

Participating schools in U.S. TIMSS-8 had a statistically significantly lower total enrollment than the eligible sample (table 4-2). On the other hand, there were no statistically differences detected in the mean

eighth-grade enrollment between participating and eligible schools. Participating schools had a statistically significantly lower mean percentage of Hispanic students than the eligible sample (table 4-3). There were no statistically significant differences detected in the mean percentage of students of the other races and ethnicities (White, non-Hispanic; Black, non-Hispanic; Asian or Pacific Islander; and American Indian or Alaska Native) between the participating and eligible schools.

There was no statistically significant difference detected in the mean percentage of students eligible for the free or reduced-price lunch between participating and the eligible sample of public schools (table 4-4). However, this must be interpreted with caution because the variable is missing for 40 schools.

Table 4-2. Mean enrollment of eligible and participating schools using the U.S. TIMSS eighth-grade original sample: 2003

	Sample schools				
Student enrollment	Eligible (mean)	Participating (mean)	Bias	Relative bias	*t* test *p*-value
Total school	783.3	741.3	-41.98	-0.054	0.007
Eighth grade	256.7	246.4	-10.25	-0.040	0.096

NOTE: Information on total school enrollment is missing for one participating school of the 296 eligible schools in the sample. Eligible schools contained at least one eighth-grade class. Participating schools agreed to have their students assessed. 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), 2003.

Table 4-3. Mean percentage of students in eligible and participating schools in the U.S. TIMSS eighth-grade original sample, by race/ethnicity: 2003

	Sample schools				
Race/ethnicity	Eligible (percent)	Participating (percent)	Bias	Relative bias	*t* test *p*-value
White, non-Hispanic	65.0	66.6	1.55	0.024	0.171
Black, non-Hispanic	14.1	14.1	0.02	0.001	0.980
Hispanic	13.7	12.1	-1.59	-0.117	0.044
Asian or Pacific Islander	4.3	3.8	-0.47	-0.111	0.158

See notes at end of table.

Table 4-3. Mean percentage of students in eligible and participating schools in the U.S. TIMSS eighth-grade original sample, by race/ethnicity: 2003—Continued

Race/ethnicity	Sample schools: Eligible (percent)	Sample schools: Participating (percent)	Bias	Relative bias	*t* test *p*-value
American Indian or Alaska Native	1.2	1.5	0.27	0.219	0.084

NOTE: Information on race/ethnicity is missing for one participating school of the 296 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 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), 2003.

Table 4-4. 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: 2003

Students	Sample schools: Eligible (percent) (N = 228)	Sample schools: Participating (percent) (N = 160)	Bias	Relative bias	*t* test *p*-value
Percentage of students eligible for free or reduced-price lunch	35.9	35.7	-0.20	-0.006	0.863

NOTE: Information on percentage of students eligible for free or reduced lunch program is missing for 40 of the 268 public schools in the sample. Eligible schools contained at least one eighth-grade class. Participating schools agreed to have their students assessed. 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), 2003.

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. One school was 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 table 4-5. Total school enrollment is the only significant predictor of school participation of the characteristics examined. The negative parameter estimate indicates that participating schools tended to be smaller than non-participating schools, i.e., the larger the total student enrollment, the less likely a school was to participate.

Table 4-5. Logistic regression model parameter estimates using the U.S. TIMSS eighth-grade original sample of schools: 2003

Parameter	Parameter estimate	Standard error	*t* test for H_0: parameter = 0	*p*-value
Intercept	1.170	11.2538	0.1039	0.9175
Central city	0.085	0.3203	0.2656	0.7913
Rural/small town	0.819	0.4137	1.9805	0.0513
Private school	-0.790	0.5787	-1.3652	0.1763
High poverty	-0.513	0.4792	-1.0706	0.2878
Northeast	0.692	0.4187	1.6540	0.1023
Southeast	0.800	0.4656	1.7172	0.0901
Central	0.747	0.4624	1.6149	0.1105
Total school enrollment	-0.002	0.0007	-2.2687	0.0262
Eighth grade enrollment	0.002	0.0020	1.1450	0.2558
White, non-Hispanic	-0.004	0.1133	-0.0323	0.9743
Black, non-Hispanic	0.001	0.1136	0.0070	0.9945
Hispanic	0.002	0.1134	0.0148	0.9882
Asian or Pacific Islander	0.076	0.1341	0.5634	0.5748
American Indian or Alaska Native	0.001	0.1138	0.0090	0.9929

NOTE: One non-participating school of the 296 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. NAEP region is the state-based region of the country (see appendix 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), 2003.

Respondent Sample With Replacements (Final Sample)

The following nonresponse bias analysis is based on the final sample of 296 schools for TIMSS-8. The distribution of the responding sample, including participating replacements (N = 232) was compared to the total eligible original sample (N = 296). 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 replacements, the weighted and unweighted school response rate in TIMSS-8 was 78 percent, with 232 out of 296 schools participating.

Categorical Variables (TIMSS-8)

The distribution of participating and eligible schools by the four characteristics is shown in table 4-6. 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. The p-values for the Chi-square tests are presented in the table.

The Chi-square statistic for community type is significant and suggests that there is evidence of a relationship with participating in TIMSS-8 (table 4-6). The results are similar to those of the previous section, indicating that the use of replacements was not effective in eliminating this discrepancy. There were no statistically significant relationships detected between participation status and the other characteristics.

Table 4-6. Percentage distribution of eligible and participating schools in the U.S. TIMSS eighth-grade final sample, by selected categorical variables: 2003

School characteristic	Sample schools: Eligible (percent)	Sample schools: Participating (percent)	Bias	Relative bias	Chi-square *p*-value
School control					0.354
Public	90.8	91.7	0.91	0.010	
Private	9.2	8.3	-0.91	-0.098	
Community type					0.005
Central city	29.2	27.3	-1.87	-0.064	
Urban fringe/large town	42.2	39.6	-2.61	-0.062	
Rural/small town	28.6	33.1	4.47	0.156	
NAEP region					0.088
Northeast	20.4	21.8	1.34	0.066	
Southeast	24.1	23.9	-0.25	-0.010	
Central	22.8	24.7	1.98	0.087	
West	32.7	29.6	-3.08	-0.094	
Poverty level					0.571
High	22.4	21.7	-0.72	-0.032	
Low	77.6	78.3	0.72	0.009	

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 NSLP; all private schools are treated as low poverty schools. NAEP region is the state-based region of the country (see appendix for state listing.) Eligible schools contained at least one eighth-grade class. Participating schools agreed to have their students assessed. 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), 2003.

Continuous Variables (TIMSS-8)

Summary means for each continuous variable for participating and eligible schools are shown in tables 4-7 through 4-9. The difference between means of participating and eligible schools was tested using a *t* test. The *p*-values for the *t* tests are presented in the tables. One school had a missing value for the total number of students, one school had a missing value for one or more of the race/ethnicity variables, 38 out of the 268 public schools had a missing value for the free or reduced-price lunch variable; these schools were excluded from the analysis.

Participating schools had a statistically significantly lower total enrollment than the eligible sample (table 4-7). This result was similar to that shown in table 4.2, suggesting that the use of replacements was

not effective in eliminating this discrepancy. However, there were no statistically significant differences detected in the mean eighth-grade enrollment between participating and eligible schools.

There were no statistically significant differences detected in mean race/ethnicity percentages between participating and eligible schools (table 4-8). 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 4-9). However, this must be interpreted with caution because the variable is missing for 38 schools.

Table 4-7. Mean enrollment of eligible and participating schools in the U.S. TIMSS eighth-grade final sample: 2003

	Sample schools				
Student enrollment	Eligible (mean)	Participating (mean)	Bias	Relative bias	*t* test *p*-value
Total school	777.8	745.2	-32.61	-0.042	0.004
Eighth grade	256.6	248.0	-8.60	-0.034	0.074

NOTE: Information on total school enrollment is missing for one participating school of the 296 eligible schools in the sample. Eligible schools contained at least one eighth-grade class. Participating schools agreed to have their students assessed. 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), 2003.

Table 4-8. Mean percentage of students in eligible and participating schools in the U.S. TIMSS eighth-grade final sample, by race/ethnicity: 2003

	Sample schools				
Race/ethnicity	Eligible (percent)	Participating (percent)	Bias	Relative bias	*t* test *p*-value
White, non-Hispanic	64.7	65.9	1.14	0.018	0.222
Black, non-Hispanic	15.0	15.1	0.09	0.006	0.895
Hispanic	12.9	11.8	-1.10	-0.085	0.062
Asian or Pacific Islander	4.3	3.9	-0.35	-0.082	0.209
American Indian or Alaska Native	1.2	1.4	0.17	0.139	0.111

NOTE: Information on race/ethnicity is missing for one participating school of the 296 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 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), 2003.

Table 4-9. 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: 2003

Students	Sample schools: Eligible (percent) (N = 230)	Sample schools: Participating (percent) (N = 178)	Bias	Relative bias	t test p-value
Percentage of students eligible for free or reduced-price lunch	35.8	35.2	-0.62	-0.017	0.506

NOTE: Information on percentage of students eligible for free or reduced lunch program is missing for 38 of the 268 public schools in the sample. Eligible schools contained at least one eighth-grade class. Participating schools agreed to have their students assessed. 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), 2003.

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. One school was 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 table 4-10. Being located in a rural/small town is the only significant predictor of school participation of the characteristics examined. The positive parameter estimate indicates that rural/small town schools were more likely than urban fringe/large town schools to participate in TIMSS-8.

Table 4-10. Logistic regression model parameter estimates in the U.S. TIMSS eighth-grade final sample: 2003

Parameter	Parameter estimate	Standard error	*t* test for H_0: parameter = 0	*p*-value
Intercept	1.379	4.3976	0.3135	0.7548
Central city	0.089	0.3398	0.2631	0.7932
Rural/small town	1.197	0.5259	2.2769	0.0256
Private school	-0.570	0.5967	-0.9549	0.3427
High poverty	-0.551	0.4707	-1.1699	0.2458
Northeast	0.959	0.4914	1.9510	0.0548
Southeast	0.383	0.4523	0.8477	0.3993
Central	0.821	0.4774	1.7200	0.0896
Total school enrollment	-0.001	0.0008	-1.8496	0.0683
Eighth grade enrollment	0.002	0.0022	1.0308	0.3059
White, non-Hispanic	-0.004	0.0420	-0.0934	0.9258
Black, non-Hispanic	0.004	0.0428	0.0840	0.9333
Hispanic	0.002	0.0437	0.0566	0.9550
Asian or Pacific Islander	0.055	0.0974	0.5683	0.5715
American Indian or Alaska Native	0.001	0.0493	0.0274	0.9782

NOTE: One non-participating school of the 296 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. NAEP region is the state-based region of the country (see appendix 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), 2003.

Final Sample, With Nonresponse Adjustments Applied

This section presents the nonresponse bias analysis based on the final sample of 296 schools for TIMSS-8. This is the same sets of schools that were compared as in the previous analysis; but this time when analyzing the responding schools alone, school nonresponse adjustments were applied to the weights.

Note that the results for this section are identical to those of the previous section. This is because the school nonresponse adjustment for TIMSS-8, following the international weighting procedures, was constant across all schools, and thus had no effect on the characteristics of the weighted responding sample of schools.

Categorical Variables (TIMSS-8)

The distribution of participating and eligible schools by the four characteristics is shown in table 4-11. 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. The *p*-values for the Chi-square tests are presented in the table.

The Chi-square statistic for community type is significant and suggests that there is evidence of a relationship with participating in TIMSS-8. There were no statistically significant relationships detected between participation status and the other characteristics.

Table 4-11. Percentage distribution of eligible and participating schools in the U.S. TIMSS eighth-grade final sample after nonresponse adjustment, by selected categorical variables: 2003

School characteristic	Sample schools: Eligible (percent)	Sample schools: Participating (percent)	Bias	Relative bias	Chi-square *p*-value
School control					0.353
Public	90.76	91.67	0.91	0.010	
Private	9.2	8.3	-0.91	-0.098	
Community type					0.007
Central city	29.2	27.3	-1.87	-0.064	
Urban fringe/large town	42.2	39.6	-2.61	-0.062	
Rural/small town	28.6	33.1	4.47	0.156	
NAEP region					0.088
Northeast	20.4	21.8	1.34	0.066	
Southeast	24.1	23.9	-0.25	-0.010	
Central	22.8	24.7	1.98	0.087	
West	32.7	29.6	-3.08	-0.094	
Poverty level					0.571
High	22.4	21.7	-0.72	-0.032	
Low	77.6	78.3	0.72	0.009	

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 NSLP; all private schools are treated as low poverty schools. NAEP region is the state-based region of the country (see appendix for state listing.) Eligible schools contained at least one eighth-grade class. Participating schools agreed to have their students assessed. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school nonresponse adjusted weights.
SOURCE: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Continuous Variables (TIMSS-8)

Summary means for each continuous variable for participating and eligible schools are shown in tables 4-12 through 4-14. The difference between means of participating and eligible schools was tested using a *t* test. The *p*-values for the *t* tests are presented in the tables. One school had a missing value for the total number of students, one school had a missing value for one or more of the race/ethnicity variables, 38 out of the 268 public schools had a missing value for the free or reduced-price lunch variable; these schools were excluded from the analysis.

Participating schools had a statistically significantly lower total enrollment than the eligible sample (table 4-12). However, there were no statistically significant differences detected in the mean eighth-grade enrollment between participating and eligible schools.

There were no statistically significant differences detected in mean race/ethnicity percentages between participating and eligible schools (table 4-13).

There was no statistically significant difference detected in the mean percentage of students eligible for the free or reduced lunch program between participating and eligible public schools (table 4-14). However, this must be interpreted with caution because the variable is missing for 38 schools.

Table 4-12. Mean enrollment of eligible and participating schools in the U.S. TIMSS eighth-grade final sample after nonresponse adjustment: 2003

	Sample schools				
Student enrollment	Eligible (mean)	Participating (mean)	Bias	Relative bias	*t* test
Total school	777.8	745.2	-32.61	-0.042	0.004
Eighth grade	256.6	248.0	-8.60	-0.034	0.074

NOTE: Information on total school enrollment is missing for one participating school of the 296 eligible schools in the sample. Eligible schools contained at least one eighth-grade class. Participating schools agreed to have their students assessed. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school nonresponse adjusted weights.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Table 4-13. Mean percentage of students in eligible and participating schools in the U.S. TIMSS eighth-grade final sample after nonresponse adjustment, by race/ethnicity: 2003

	Sample schools				
Race/ethnicity	Eligible (percent)	Participating (percent)	Bias	Relative bias	*t* test *p*-value
White, non-Hispanic	64.7	65.9	1.14	0.018	0.222
Black, non-Hispanic	15.0	15.1	0.09	0.006	0.895
Hispanic	12.9	11.8	-1.10	-0.085	0.062
Asian or Pacific Islander	4.3	3.9	-0.35	-0.082	0.209
American Indian or Alaska Native	1.2	1.4	0.17	0.139	0.111

NOTE: Information on race/ethnicity is missing for one participating school of the 296 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 relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school nonresponse adjusted weights.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Table 4-14. 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 after nonresponse adjustment: 2003

Students	Sample schools: Eligible (percent) (N = 230)	Sample schools: Participating (percent) (N = 178)	Bias	Relative bias	t test p-value
Percentage of students eligible for free or reduced lunch	35.8	35.2	-0.62	-0.017	0.506

NOTE: Information on percentage of students eligible for free or reduced lunch program is missing for 38 of the 268 public schools in the sample. Eligible schools contained at least one eighth-grade class. Participating schools agreed to have their students assessed. The relative bias is calculated as the bias divided by the estimate from the eligible sample. Schools were weighted by their school nonresponse adjusted weights.
SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Summary—Grade 8

For the original sample of schools in TIMSS-8 in the United States, community type, the total number of students enrolled and the percentage of Hispanic students were found to be significantly related to school participation in the bivariate analysis. Only the total number of students enrolled was found to be significant in the logistic regression and indicated that the larger the total student enrollment, the less likely a school was to participate, however.

For the final sample of schools in TIMSS-8 (the participating original and replacement schools), community type and the total number of students enrolled remained significant in the bivariate analysis. However, the logistic regression showed that being located in a rural/small town was statistically significant, indicating that the larger the total student enrollment, the less likely a school was to participate.

The application of school nonresponse adjustments in TIMSS-8 had no effect on bias, as just a single nonresponse adjustment factor was used for all schools.

5. Conclusions

The investigation into nonresponse bias at the school level for U.S. TIMSS 2003 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, there was no statistically significant relationship detected between participation and any school characteristic in either the bivariate or logistic regression analyses. The bivariate results for the final sample of schools in TIMSS-4 indicated that participating schools had statistically significantly higher percentages of Black, non-Hispanic students than the eligible sample of schools (table 3-8). The bivariate analysis using the free or reduced-price lunch program also revealed that participating schools had a statistically significantly lower proportion of students eligible for this program than the eligible sample of schools (table 3-9). However, the regression analysis results do not show any variables as significant predictors of participation in TIMSS-4 (table 3-10). For the final sample of schools in TIMSS-4 with school nonresponse adjustments applied to the weights, there was no statistically significant relationship detected between participation and any available school characteristic in the bivariate analysis.

These results suggest that there is little potential for nonresponse bias in the TIMSS-4 original sample based on the characteristics studied. It also suggests that while there is certainly no evidence that the use of replacement schools reduced the potential for bias it has also not substantially added to it. The school nonresponse adjustment was at least partially effective in reducing the nonresponse bias as there were no significant results in tables 3-11 through 3-14.

For the original sample of schools in TIMSS-8, three variables were found to be statistically significant related to participation in the bivariate analysis: community type (table 4-1); the total number of students enrolled in participating schools is lower than in all schools (table 4-2); and the percentage of Hispanic students in participating schools was lower than the eligible sample (table 4-3). The regression analysis showed only the total number of students enrolled in participating schools was statistically significantly lower than in all schools in TIMSS-8 (table 4-5). For the final sample, two of the three variables remained statistically significant in the bivariate analysis: community type (table 4-6) and the total number of students enrolled are statistically significantly lower in participating schools than in the eligible sample (table 4-7). The regression analysis indicated that rural/small town schools were more likely than urban fringe/large town and central city schools to participate (4-10). For the final sample of schools in TIMSS-8 with school nonresponse adjustments applied to the weights, the results were identical. These results suggest that there is some potential for nonresponse bias in the TIMSS-8 original sample based on the

characteristics studied. Thus while there is certainly no evidence that the use of replacement schools reduced the potential for bias, it has also not substantially added to it. The school nonresponse adjustment had no effect on the characteristics of the weighted responding sample of schools.

References

Brick, M.J., Morganstein, D., and Valliant, R. (2000) *Analysis of Complex Sample Data Using Replication.* Retrieved January 21, 2005, from http://www.westat.com/wesvar/index.html.

Ferraro, D. and Rust, K. (2003). *U.S. TIMSS School Sample Final Report* (Prepared under contract to National Center for Education Statistics). Rockville, MD: Westat.

Gonzales, P., Guzmán, J.C., Partelow, L., Pahlke, E., Jocelyn, L., Kastberg, D., and Williams, T. (2004). *Highlights From the Trends in International Mathematics and Science Study (TIMSS) 2003* (NCES 2005-005). U.S. Department of Education. Washington, DC: Government Printing Office.

Kali, J., and Rust, K. (2005). *Analysis of School Nonresponse Bias in the 2005 NAEP –Grades 4 and 8.* (Prepared under contract to National Center for Education Statistics). Rockville, MD: Westat.

Martin, M.O., Mullis, I.V.S., and Chrostowski, S.J. (Eds.). (2004). *TIMSS 2003 Technical Report*. Chestnut Hill, MA: Boston College.

Piesse, A, and Rust, K. (2003). *U.S. 2001 PIRLS Nonresponse Bias Analysis* (NCES 2003-21). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Rao, J.N.K., and Scott, A.J. (1981). The Analysis of Categorical Data from Complex Sample Surveys: Chi-Squared Tests for Goodness of Fit and Independence in Two-Way Tables. *Journal of the American Statistical Association* 76: 221–230.

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.

Westat. (2002). *WesVar 4.2 User's Guide.* Rockville, MD: Westat, Inc.

This page left intentionally blank.

Appendix A—Technical Notes

Full details on the technical aspects of TIMSS 2003 can be found in Martin, Mullis, and Chrostowski (2004) and Gonzales et al. (2004). The sections below provide details on those aspects of the implementation of TIMSS that are relevant to the analyses included in this report.

Data Collection

The TIMSS 2003 data were collected by each country, following international guidelines and specifications. TIMSS required that countries select random, nationally representative samples of schools and students. TIMSS countries were asked to identify eligible students based on a common set of criteria, allowing for adaptation to country-specific situations. In IEA studies such as TIMSS, the target population for all countries is called the international desired population. For the fourth-grade assessment, the international desired population consisted of all students in the country who were enrolled in the upper of the two adjacent grades that contained the greatest proportion of 9-year-olds at the time of testing. In the United States and most other countries, this corresponded to fourth grade. For the eighth-grade assessment, the international desired population consisted of all students in the country who were enrolled in the upper of the two adjacent grades that contained the greatest proportion of 13-year-olds at the time of testing. In the United States and most other countries, this corresponded to eighth grade.

TIMSS used a two-stage stratified cluster sampling design. The first stage made use of a systematic probability-proportionate-to-size (PPS) technique to select schools. Although countries participating in TIMSS were strongly encouraged to secure the participation of schools selected in the first stage, it was anticipated that a 100 percent participation rate for schools would not be possible in all countries. Therefore, two replacement schools were identified for each originally sampled school, a priori. As each school was selected, the next school in the sampling frame was designated as a replacement school should the originally sampled school choose not to participate in the study. Should the originally sampled school and the replacement school choose not to participate, a second replacement school was chosen by going to the next school in the sampling frame.

The second stage consisted of selecting classrooms within sampled schools. At the classroom level, TIMSS sampled intact mathematics classes that were offered to students in the target grades. In most countries, one mathematics classroom per school was sampled, although some countries, such as the United States, chose to sample two mathematics classrooms per school.

Exclusions in the TIMSS Sample

All countries were required to define their national desired population to correspond as closely as possible to the definition of the international desired population. In some cases, countries needed to exclude schools and students in remote geographical locations or to exclude a segment of the education system. Any exclusions from the international desired population were clearly documented. Countries were expected to keep the excluded population to no more than 10 percent of the national desired population. Exclusions could take place at the school level, within schools, or both. Participants could exclude schools from the sampling frame for the following reasons:

- Locations were geographically remote;
- Size was extremely small;
- Curriculum or school structure was different from the mainstream education system; or
- Instruction provided was only to students in the categories defined as "within-school exclusions."

Within schools, exclusion decisions were limited to students who, because of some disability, were unable to take part in the TIMSS assessment. The general TIMSS rules for defining within-school exclusion included the following three groups:

- *Intellectually disabled students*. These students were considered, in the professional opinion of the school principal or other qualified staff members, to be intellectually disabled, or had been so diagnosed in psychological tests. This category included students who were emotionally or mentally unable to follow even the general instructions of the TIMSS test. It did not include students who merely exhibited poor academic performance or discipline problems.

- *Functionally disabled students.* These students were permanently physically disabled in such a way that they could not participate in the TIMSS assessment. Functionally disabled students who could perform were included in the testing.

- *Non-native-language speakers*. These students could not read or speak the language of the assessment and so could not overcome the language barrier of testing. Typically, a student who had received less than 1 year of instruction in the language of the assessment was excluded, but this definition was adapted in different countries.

Sampling, Data Collection, and Response Rates in the United States

The TIMSS 2003 school sample was drawn for the United States in November 2002. The sample design for this school sample was developed to follow international requirements as given in the TIMSS sampling manual. The U.S. sample for 2003 was a two-stage sampling process with the first stage a sample of schools, and the second stage a sample of students' classrooms from the target grade in sampled schools. Unlike TIMSS 1995 and 1999, the sample was not clustered at the geographic level for TIMSS 2003. This change was made in an effort to reduce the design effects and to spread the respondent burden across school districts as much as possible.

The sample design for TIMSS was a stratified systematic sample, with sampling probabilities proportional to measures of size. The U.S. TIMSS fourth grade sample had two explicit strata based on poverty. A high poverty school was defined as one in which 50 percent or more of the students were eligible for participation in the federal National School Lunch Program (NSLP); high poverty schools were oversampled (Ferraro and Rust 2003) This variable was not available for private schools, so they were all treated as low poverty schools. The target sample sizes were 120 high-poverty and 190 low-poverty schools.

Within the poverty strata, there are four categorical implicit stratification variables: type of school (public or private), region of the country[5] (Northeast, Southeast, Central, West), type of location relative to populous areas (eight levels), and minority status (above or below 15 percent). The last sort key within the implicit stratification was by grade enrollment in descending order.

The TIMSS eighth-grade sample had no explicit stratification. The frame was implicitly stratified (i.e., sorted for sampling) by four categorical stratification variables: type of school (public or private), region of the country, type of location relative to populous areas (eight levels), and minority status (above or below 15 percent). The last sort key within the implicit stratification was by grade enrollment in descending order.

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 (immediately preceding and following it) were designated as replacement schools. If an original school refused to participate, the first replacement was then contacted.

If that school also refused to participate, the second school was then contacted. There were several constraints on the assignment of substitutes. One sampled school was not allowed to substitute for another, and a given school could not be assigned to substitute for more than one sampled school. Furthermore, substitutes were required to be in the same 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. One was designated a first replacement and the other a second replacement. If an original school refused to participate, the first replacement was then contacted. If that school also refused to participate, the second school was then contacted.

The schools were selected with probability proportionate to the school's estimated enrollment of fourth- and eighth-grade students from the 2003 NAEP school frame with 2000-01 school data. The data for public schools were from the Common Core of Data (CCD), and the data for private schools were from the Private School Survey (PSS). Any school containing a fourth or an eighth grade as of the school year 2000-01 was included on the school sampling frame. Participating schools provided lists of fourth- or eighth-grade classrooms, and one or two intact mathematics classrooms were selected within each school in an equal probability sample. The overall sample design for the United States was intended to approximate a self-weighting sample of students as much as possible, with each fourth- or eighth-grade student having an equal probability of being selected.

The U.S. TIMSS fourth-grade school sample consisted of 310 schools, of which 300 were eligible schools and 212 agreed to participate. The school response rate before replacement was 70 percent (weighted; 71 percent unweighted). The weighted school response rate before replacement is given by the formula:

$$\text{weighted school response rate before replacement} = \frac{\sum_{i \in Y} W_i E_i}{\sum_{i \in (Y \cup N)} W_i E_i},$$

where Y denotes the set of participating original sample schools with age-eligible students, N denotes the set of eligible non-participating original sample schools, W_i denotes the base weight for school i, $W_i = 1/P_i$, where P_i denotes the school selection probability for school i, and E_i denotes the enrollment size of grade-eligible students, as indicated on the sampling frame. In addition to the 212 participating schools,

[5]See definition of regions later in appendix.

36 replacement schools also participated for a total of 248 participating schools at the fourth grade in the United States.

A total of 10,795 students were sampled for the fourth-grade assessment. Of these students, 49 were withdrawn from the school before the assessment was administrated. Of the eligible 10,746 sampled students, an additional 429 students were excluded using the criteria described above, for a weighted exclusion rate of 5 percent. Of the 10,317 remaining sample students, a total of 9,829 students participated in the assessment in the United States, since 488 students were absent. The student participation rate was 95 percent.

The U.S. TIMSS eighth-grade school sample consisted of 301 schools, of which 296 were eligible schools and 211 agreed to participate. The school response rate before replacement was 71 percent (weighted and unweighted). In addition to the 211 participating schools, 21 replacement schools also participated for a total of 232 participating schools at the eighth grade in the United States.

A total of 9,891 students were sampled for the assessment. Of these students, 90 were withdrawn from the school before the assessment was administrated. Of the eligible 9,801 sampled students, an additional 279 students were excluded using the criteria described above, for a weighted exclusion rate of 5 percent. Of the 9,522 remaining sample students, a total of 8,912 students participated in the assessment in the United States, since 610 students were absent. The student participation rate was 94 percent (weighted and unweighted).

Weighting

Before the data are analyzed, responses from the groups of students assessed are assigned sampling weights to ensure that their representation in TIMSS 2003 results matches their actual percentage of the school population in the grade assessed.

Sampling weights are used to adjust for over-representation or under-representation from particular student groups or types of schools. The use of sampling weights is necessary for the computation of statistically sound, nationally representative estimators. The base weight assigned to a school or student is the inverse of the probability that the school or 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 schools and

students that participated. Weighting also adjusts for various situations such as school and student nonresponse because data cannot be assumed to be randomly missing.

Response Rates

Table A-1 details the school participation rates for each of the two TIMSS target populations in the United States.

Table A-1. Selected characteristics of U.S. TIMSS grade 4 and 8 school samples: 2003

Grade	Schools in original sample	Eligible schools in sample	Schools in original sample that participated	Replacement schools	Total schools that participated	Weighted school participation rate before replacement (percent)	Weighted school participation rate after replacement (percent)
4	310	300	212	36	248	70	82
8	301	296	211	21	232	71	78

SOURCE: International Association for the Evaluation of Education Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS), 2003.

Description of Variables

Frame characteristics for public schools were taken from the 2000–01 Common Core of Data (CCD) and, for private schools, from the 2000–01 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 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 CSA), rural, outside CBSA (any incorporated place, Census-designated place, or non-place territory not within a CBSA or CSA 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).

NAEP Region

NAEP region is the 'state-based' region of the country. 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, 2002. 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.

Statistical Procedures

Weighting

Records from the sample schools and students were assigned sampling weights to adjust for over-representation or under-representation from a particular group. The use of sampling weights is necessary for the computation of statistically sound, nationally representative estimators. The weight assigned to a school's (or student's) data is the inverse of the probability that the school (or student) would be selected for the sample. When data are weighted each sample unit contributes to the results in proportion to the total number of schools or students represented by the individual unit. Weighting can also be used to adjust 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 reciprocal of the school's probability of selection, (2) an adjustment for school-level nonresponse, (3) the reciprocal 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 2002-03 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 2003 assumes that there are *G* replicates, each containing two sampled schools selected independently. The element $\hat{\theta}_{(g)}$ denotes the estimate using the *gth* jackknife replicate. This is computed using all cases except those in the *gth* replicate of the sample. For those in the *gth* 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 4.0 User's Guide (Westat 2000).

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 3-2. The first row shows that the weighted mean total school enrollment for the full eligible sample of grade 4 schools is 583.9. For the subsample of schools that participated the corresponding mean is 576.4, and difference of 7.5. The standard error of this estimated difference, calculated so as to reflect the dependency between these two samples, and the complex sample design, is 10.6. This gives rise to a *t*-statistic of -0.70, and using 75 degrees of freedom (the appropriate figure for the TIMSS design), the resulting significance (or *p*-value) is 0.483. 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

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

where b_k is a parameter estimate and $v\left(b_k\right)$ 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 (2003).

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:

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

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

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

Logistic Regression Models

A linear model for investigating the relationship between binary (dichotomous) outcomes and a set of explanatory variables is referred to as a *logistic regression model*. The data are assumed to follow a binomial distribution, with probabilities that depend on the independent variables. In this instance the binary outcome of interest is whether or not the sampled unit participated in TIMSS.

Let p_i denote the probability that the ith 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:

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

where X_{1i}, X_{2i} ..., X_{pi} are p auxiliary variables associated with the ith 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 $q(\boldsymbol{\beta}) = \partial L_n(\boldsymbol{\beta})/\partial \boldsymbol{\beta}$ be the vector of first partial derivatives of the sample log-likelihood with respect to $\boldsymbol{\beta}$. Let $\mathbf{H}(\boldsymbol{\beta})$ be the matrix of second partial derivatives (or Hessian) of the sample log-likelihood having entries $\partial^2 \mathrm{L} / \partial \beta_a \partial \beta_b$, where β_a and β_b are two separate components of $\boldsymbol{\beta}$. Denote by $\mathbf{q}^t$ and $\mathbf{H}^t$ the values of $q(\boldsymbol{\beta})$ and $\mathbf{H}(\boldsymbol{\beta})$ evaluated at $\mathbf{b}^t$, the value of the estimate $\mathbf{b}$ at step t.

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

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

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

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

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