# Appendix A

## Overview of Procedures and Methods

## The TIMSS Assessment

TIMSS in 1995-96 tested students in primary school (third and fourth grades – Population 1) and middle school (seventh and eighth grades – Population 2) in mathematics and science, and final-year high-school students (Population 3) in mathematics and science literacy, advanced mathematics, and physics. The data used in this study were from the upper grade of Population 2, which was eighth grade in most countries. Six content areas were covered by the mathematics tests taken by the eighth-grade students. These areas, and the percentage of test items devoted to each, include fractions and number sense (34%); measurement (12%); proportionality (7%); data representation, analysis, and probability (14%); geometry (15%); and algebra (18%). The eighth-grade science test consisted of just five content areas: earth science (16%); life science (30%); physics (30%); chemistry (14%); and environmental issues and the nature of science (10%). There were 151 mathematics items and 135 science items in the eighth-grade TIMSS assessment.

To maximize the content coverage of the TIMSS tests, yet minimize the burden on individual students, TIMSS used a multiple matrix sampling design whereby each student responded to just a subset of the total item pool.[1] By combining student responses across the item pool using sophisticated scaling techniques, TIMSS was able to derive estimates of average mathematics and science achievement for the entire population of eighth-grade students in each country.

In each subject, approximately one-quarter of the items were in the free-response format, requiring students to generate and write their own answers. Designed to take up about one-third of students' response time, some of these questions asked for short answers while others required extended responses in which students needed to show their work. The remaining questions were in multiple-choice format. In scoring the tests, correct answers to most questions were worth one point. Consistent with the approach of allotting longer response times for constructed-response questions than for multiple-choice questions, responses to some of these questions (particularly those requiring extended responses) could earn partial credit, with a fully correct answer being awarded two or three points.

### Target Population

The target population (internationally desired population in IEA parlance) for Population 2 was the two adjacent grades that contained the largest proportion of 13-year-old students at the time of testing. These were the seventh and eighth grade in most countries. In a few situations where TIMSS testing could not be done for the

[1] The TIMSS test design is fully described in Adams, R.J., and Gonzalez, E.J. (1996); "TIMSS Test Design," in M.O. Martin and D.L. Kelly (eds.), *Third International Mathematics and Science Study Technical Report, Volume I*, Chestnut Hill, MA: Boston College.

entire internationally desired population, countries were permitted to define a national desired population that excluded part of the internationally desired population. The results for such countries were annotated in international reports. Because coverage fell below 65% for Latvia, the Latvian results have been labeled "Latvia (LSS)," for Latvian-Speaking Schools, throughout the report.

### School and Student Sampling

Within countries, TIMSS used a two-stage sample design, where the first stage involved selecting 150 public and private schools within each country. Within each school, each country was required to use a random sampling procedure to select one mathematics class at the eighth grade and one at the seventh grade (or the corresponding upper and lower grades in that country). All of the students in those two classes were to participate in the TIMSS testing. This approach was designed to yield, for each population, a representative sample of at least 7,500 students per country, with approximately half students at each grade. Countries were, however, permitted to extend the basic sampling design to meet domestic concerns, provided they complied with TIMSS standards for population coverage and sampling precision. For example, four countries, Australia, Cyprus, Sweden and the United States, sampled two intact mathematics classes in each sampled school. Korea sampled students within the sampled mathematics classes, and England used within-school sampling.

### Indicating Compliance with Sampling Guidelines[2]

In Exhibit A.1, countries are grouped by how they met the TIMSS sampling requirements. Countries that achieved acceptable participation rates – 85% of both the schools and students, or a combined rate (the product of school and student participation) of 75%, with or without replacement schools – and that complied with the TIMSS guidelines for grade selection and classroom sampling are shown in the first panel of the exhibit. Countries that met the guidelines only after including replacement schools are annotated in international reports.

Countries not reaching at least 50% school participation without the use of replacement schools, or that failed to reach the participation standard even with the inclusion of replacement schools, are shown in the second panel of the figure.

2 Details of the sampling compliance can be found in Beaton, A. E., Martin, M . O., Mullis, I. V. S., Gonzalez, E. J., Smith, T. A. and Kelly, D. L. (1996a); *Science Achievement in the Middle Years: IEA's Third International Mathematics and Science Study,* Chestnut Hill, MA: Boston College, and Beaton, A. E., Mullis, I. V. S., Martin, M . O., Gonzalez, E. J., Kelly, D. L. and Smith, T. A. (1996b); *Mathematics Achievement in the Middle Years: IEA's Third International Mathematics and Science Study,* Chestnut Hill, MA: Boston College.

# Exhibit A.1 Countries Grouped for Reporting According to Their Compliance with Guidelines for Sample Implementation and Participation Rates

| | | |
|---|---|---|
| **Countries satisfying guidelines for sample participation rates, grade selection and sampling procedures** | † Belgium (Fl) | [1] Latvia |
| | Canada | [1] Lithuania |
| | Cyprus | New Zealand |
| | Czech Republic | Norway |
| | †2 England | Portugal |
| | France | Russian Federation |
| | Hong Kong | Singapore |
| | Hungary | Slovak Republic |
| | Iceland | Spain |
| | Iran, Islamic Republic | Sweden |
| | Ireland | [1] Switzerland |
| | Japan | † United States |
| | Korea | |
| **Countries not satisfying guidelines for sample participation** | Australia | |
| | Austria | |
| | Belgium (Fr) | |
| | Netherlands | |
| | Scotland | |
| **Countries not meeting age/grade specifications (high percentage of older students)** | Colombia | |
| | †1 Germany | |
| | Romania | |
| | Slovenia | |

† Met guidelines for sample participation rates only after replacement schools were included.

1 National Desired Population does not cover all of the International Desired Population. Because coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.

2 National Defined Population covers less than 90 percent of National Desired Population.

## Data Collection Procedures

Each participating country was responsible for carrying out all aspects of the data collection, using standardized procedures developed for the study. International quality control monitors interviewed the National Research Coordinator (NRC) in each country about data collection plans and procedures. They also selected about ten schools to visit, where they observed testing sessions and interviewed school coordinators.[3] The results indicate that, in general, NRCs were well prepared for data collection and that the TIMSS tests were administered in compliance with international specifications and guidelines.

## Scoring the Free-Response Items

Because about one-third of the written test time was devoted to free-response items, TIMSS developed procedures for reliably evaluating student responses within and across countries. The scoring used a system of two-digit codes with rubrics specific to each item.

To gather and document empirical information about the within-country agreement among scorers, TIMSS had systematic subsamples of some 10% of the students' responses coded independently by two scorers. The percentage of exact agreement between scorers was computed for each free-response item. A very high percentage of exact agreement at the score level was observed for the free-response items on all TIMSS tests.[4]

## Data Processing

To ensure the availability of comparable, high-quality data for analysis, TIMSS undertook a set of rigorous quality control steps to create the international database.[5] TIMSS prepared manuals and software for countries to use in recording their data on computer files so that the information would be in a standard international format before being forwarded to the IEA Data Processing Center in Hamburg. Upon arrival at the Center, the data from each country underwent an exhaustive cleaning process designed to identify, document, and correct deviations from the international instruments, file structures,

3 The results of the interviews and observations by the quality control monitors are presented in Martin, M.O., Hoyle, C.D., and Gregory, K.D. (1996), "Monitoring the TIMSS Data Collection" and "Observing the TIMSS Test Administration," both in M.O. Martin and I.V.S. Mullis (eds.), *Third International Mathematics and Science Study: Quality Assurance in Data Collection*, Chestnut Hill, MA: Boston College.

4 Summaries of the scoring reliability data for each test are included in the appendices of the international reports (see Appendix A in Beaton, A. E., Martin, M . O., Mullis, I. V. S., Gonzalez, E. J., Smith, T. A. and Kelly, D. L. (1996a); *Science Achievement in the Middle Years: IEA's Third International Mathematics and Science Study*, Chestnut Hill, MA: Boston College, and in Beaton, A. E., Mullis, I. V. S., Martin, M . O., Gonzalez, E. J., Kelly, D. L. and Smith, T. A. (1996b); *Mathematics Achievement in the Middle Years: IEA's Third International Mathematics and Science Study*, Chestnut Hill, MA: Boston College.)

5 These steps are detailed in Jungclaus, H., and Bruneforth, M. (1996), "Data Consistency Checking Across Countries," in M.O. Martin and D.L. Kelly (eds.), *Third International Mathematics and Science Study Technical Report, Volume I*, Chestnut Hill, MA: Boston College.

and coding schemes. The process also ensured consistency of information within national data sets and appropriate linking among the many student, teacher, and school data files. Throughout the data-cleaning process, the data were checked and double-checked by the IEA Data Processing Center, the International Study Center, and the national centers. The national centers were in constant contact with the DPC and had multiple opportunities to review their data.

## IRT Scaling and Data Analysis

The mathematics and science achievement results were summarized using an item response theory (IRT) scaling method based on the Rasch one-parameter model.[6] This method produces a test score by averaging the responses to the items each student took in a way that takes into account the difficulty of each item. The method used in TIMSS includes refinements that enable reliable scores to be produced even though individual students responded to only subsets of the total item pool. Analyses of the response patterns of students from participating countries indicated that, although the items in each TIMSS test address a wide range of mathematics or science content, the performance of the students across the items was consistent enough that it could usefully be summarized in a single score per test. The IRT method was preferred for developing comparable estimates of performance for all students, since students answered different test items depending upon which test booklet they received. The IRT analysis provides a common scale on which performance can be compared across countries.

## Estimating Sampling Error

Because the statistics presented in this report are national estimates based on samples of schools and students rather than the values that could be calculated if every school and student in a country answered every question, it is important to have measures of the degree of uncertainty of the estimates. The jackknife procedure was used to estimate the standard error associated with each statistic presented in chapter 1.[7] The use of confidence intervals based on the standard errors allows inferences to be made about the population means and proportions in a manner that reflects the uncertainty associated with the sample estimates. An estimated sample statistic plus or minus two standard errors represents a 95% confidence interval for the corresponding population result.

6 The TIMSS scaling model is fully documented in Adams, R.J., Wu, M.L., and Macaskill, G. (1997), "Scaling Methodology and Procedures for the Mathematics and Science Scales," in M.O. Martin and D.L. Kelly (eds.), *Third International Mathematics and Science Study Technical Report, Volume II*, Chestnut Hill, MA: Boston College.

7 The jackknife repeated replication technique for estimating sampling errors is documented in Gonzalez, E.J., and Foy, P. (1997), "Estimation of Sampling Variability, Design Effects, and Effective Sample Sizes," in M.O. Martin and D.L. Kelly (eds.), *Third International Mathematics and Science Study Technical Report, Volume II*, Chestnut Hill, MA: Boston College.

## Comparing High- and Low-Performing Schools

The purpose of the analyses reported in Chapter 1 of this report was to contrast schools with high and low average student performance in mathematics and science in terms of student, teacher, classroom, and school characteristics, with a view to identifying characteristics associated with high performing schools.

To identify the schools for the high- and low-performing groups, schools were first ranked by average achievement. Schools in the top third were assigned to the high-achieving group, and those in the bottom third to the low-achieving group. The one-third of schools in the middle of the distribution were not used in these analyses (they were, however, used in the analyses reported in Chapter 2). This procedure was followed separately for mathematics achievement and science achievement. Since the TIMSS sampling procedure was based on intact mathematics classes, in most cases the average achievement for a school was based on the students from a single class. Differences between schools therefore also reflect differences between classes, and probably overestimate the actual difference between schools. In countries where two classes were sampled, each class was treated separately for the purpose of these analyses. For the science analyses, the school mean was computed across all students in the school, regardless of the class to which they belonged.

The variables examined in the contrast between the high- and low-achieving schools were drawn from the student, teacher, and school questionnaires that were administered as part of the TIMSS assessment. TIMSS researchers reviewed the questionnaires in the light of the effective-schools literature to identify variables that were likely to characterize effective schools. These variables were correlated with student achievement in science and mathematics in an extensive exploratory analysis. Variables that were significantly related to achievement were retained for the contrast study. Where possible, individual variables were combined to form an index that was more global and more stable than the original variables.

Each variable and index was dichotomized at a point that seemed to maximally discriminate between schools in the high-achieving group and those in the low-achieving group. A t-test was applied to the data from each country to determine whether there was a significant difference between the two groups in the frequency of occurrence of the dichotomized variable. Variables and indices that showed significant differences in most of the participating countries, or that showed particularly big differences in a few countries, were included in this report. For example, students were asked how many books they had in their homes, and could respond “none or very few (0-100 books),” “about one shelf (11-25 books),” “about one bookcase (26-

100 books)," "about two bookcases (101-200 books)," or "three or more bookcases (more than 200 books)." The dichotomous version of this variable was "having at least 100 books". The analysis for Chapter 1 then contrasted the percentage of students in the low-achieving schools having at least 100 books with the percentage in high-achieving schools, showing the difference between them and presenting it graphically. The presentation also included the jack-knife standard errors of the percentages and of the difference between the percentages. Since each exhibit in Chapter 1 contains a statistical test for all of the countries, a Bonferroni correction for multiple a priori comparisons (the number of countries minus one) was applied to the results of the t-tests for each of the exhibits reported in this chapter.

## Hierarchical Analyses

The hierarchical nature of the TIMSS data, where students are nested within schools, readily lends itself to analysis with hierarchical linear models (HLM). The analyses reported in Chapter 2 were conducted by fitting a series of two-level models (school and student) and summarizing the results across countries.

For the analyses presented in this report, three types of two-level HLMs were constructed. The **between-schools** model was used to examine how student achievement differed between schools across countries. The **home background** model was used to examine how much of the difference between schools in average student achievement could be attributed to differences in the home background of the students. The **exploratory** models examined how home and school factors related to differences between schools in science and mathematics achievement after controlling for the home background of the students.

### The Hierarchical Linear Models

**Between-Schools Model**
This HLM is similar to a one-way analysis of variance in that variation in the dependent variable (student achievement) is partitioned into two components: between schools and within the school. This model also was used for analyzing between-school differences in student home background (Exhibit 2.3)

Within School

$Y_{ij} = \beta_{oj} + e_{ik}$

The score $Yij$ of student $i$ in school $j$ is expressed in terms of the school mean $\beta_{oj}$ for school $j$ plus a deviation for student $i$.

$Y_{ij}$ is the achievement score (or home background index) for student $i$ in school $j$,

$\beta_{oj}$ represents the mean of each school,

$e_{ik}$ is a random error assumed normally distributed with a variance that is constant across individuals and schools.

School Level

$\beta_{oj} = \gamma_{oo} + U_{oj}$

where:

$\beta_{oj}$ is the school mean for school $j$,

$\gamma_{oo}$ is the grand mean (mean of $\beta_{oj}$'s),

$Uoj$ is a normally distributed random error with variance $\tau^2$. This variance is constant across schools and is independent of the first-level error term.

**Home Background Model**
This model further decomposes the between-school variance in achievement into that which is due to differences in average home background and that which is not. The intercept term represents the school/classroom mean adjusted for students' home background.

Within School

$Y_{ij} = \beta_{oj} + \beta_{1j}\text{HBI}_{ij} + \text{e}_{ik}$

Here, the relationship between home background and student achievement in each school is represented by a linear regression equation for that school,

where:

$\beta_{oj}$ is the intercept of the regression line,

$\beta_{1j}$ is the slope of the regression line relating student home background to achievement,

$HBI_{ij}$ is the home background index for student i in school j.

School Level

$$\beta_{oj} = \gamma_{oo} + \gamma_{1j}W_{1j} + U_{oj}$$

Here, differences between schools are modeled in terms of differences between the school mean (represented by the intercept in the linear equation) and the overall (grand) mean.

$\gamma_{1j}$ is the school-level regression coefficient,

$W_{1j}$ is the school mean (the intercept) of the student home background index.

**Exploratory Models**

The exploratory analyses make use of a generalized form of the home background model to study the relationship between a range of home and school variables and average school achievement while controlling for average student home background. In all, seven variations on the model were used in these analyses. The between-schools model was used as a baseline for evaluating the utility of the other, more complex, models.

- Model 1: Classroom Characteristics. In modeling science achievement, five classroom characteristics variables were used as school level predictors: daily doing homework in three subjects, amount of science homework, liking science, belief in the efficacy of science, and frequency of experiments. The corresponding predictors for the mathematics model were daily doing homework in three subjects, amount of mathematics homework, checking of homework in class, liking mathematics, classroom environment, and mathematics class size.

- Model 2: Model 1 with Teacher Characteristics. Model 2 combines the classroom characteristics of Model 1 with a set of teacher characteristics. For science, teaching experience and the ability to teach a general science course were the new predictors. Mathematics used teaching experience only.

- Model 3: Model 2 with School Climate. The third exploratory model was constructed by adding two school climate variables to Model 2 for both science and mathematics. The two new predictors were administrative violations and serious misbehavior.

- Model 4: Model 3 with School Location and Size. Two additional predictors, urban location and class size, were added to Model 3 for both science and mathematics.
- Model 5: Model 4 with Home-School Interaction. The fifth model was constructed by including variables for aspirations for future education, self press, and maternal press with the Model 4 variables.
- Model 6: Model 5 with Home Background Index. The most complex model was constructed by combining the school average on the home background index with the predictors in Model 5.
- Model 7: Home Background Index Only. The final model, designed to show the explanatory power of home background at the school level, was constructed using the school mean on the home background index as the sole predictor for both science and mathematics.

Since each model was unique, seven separate analyses were conducted for each of the fourteen countries included in the science analyses and the sixteen countries included in the mathematics analyses. The common structure of models 1 through 7 was the following.

<u>Within School</u>

$Y_{ij} = \beta_{oj} + \beta_{1j} HBI_{ij} + e_{ik}.$

This is identical to the home background model.

<u>School Level</u>

$\beta_{oj} = \gamma_{oo} + \gamma_{1j} W_{1j} + \gamma_{2j} W_{2j} \ldots. U_{oj}$

$\beta_{1j} = \gamma_{1o}$

where:

$\gamma_{1j}, \gamma_{2j}..$ are the school-level regression coefficients,

$W_{1j}$, $W_{2j}$...are the school means of the predictor variables.

## Data for Hierarchical Analyses

The procedure when conducting the hierarchical analyses involved fitting a linear regression model within each school or classroom to adjust for students' home background. A requirement of the HLM program used for these analyses was that there be no missing data at the second level of the model. This necessitated the exclusion of a number of variables from the final hierarchical analyses and effectively reduced final sample sizes. To ensure the stability of the estimates, a minimum sample size of at least ten students per school/classroom was used. Therefore, schools in the science analyses and classes in the mathematics analyses with fewer than ten students remaining after other cases had been removed were deleted from the exploratory analysis sample.

In producing measures of student achievement in science and mathematics for use in secondary analysis, TIMSS made use of imputed score or "plausible value" technology.[8] Student achievement scores were represented by random draws from achievement-score distributions the parameters of which were estimated from student responses to achievement items and from student background data. To capture the uncertainty due to the imputation process, each student has five imputed scores for science and five for mathematics. The version of the HLM program used for the analyses in this report combined the results from all five imputed scores to give the most appropriate results.

### Sampling Weights in Hierarchical Analyses

Given the complexities of the sampling design employed by TIMSS, appropriate sampling weights were applied to obtain unbiased results. The weighting for each country reflected the probability of selection for each student in each school and had to account for non-participation.[9] For the Chapter 1 analysis, a single weighing variable was used. Since the hierarchical analyses reported in Chapter 2 consist of two levels, a student level and a school level, appropriate weights had to be applied for each level. The school sampling weight was the inverse of the probability of selection for the school, adjusted for non-participating schools in the sample. The weight applied at the student level for the science and mathematics analyses consisted of the inverse of the probability of selection of students within each selected school, and also was adjusted to account for non-participating students.

## Derived Variables for Comparison of High- and Low-Achieving Schools

In the upper versus lower one-third analyses, and the hierarchical analyses, variables were derived from student, teacher, and school questionnaire data. These derived variables and the procedures used to construct them are described in the following sections.

8 Adams, R.J., Wu, M.L., and Macaskill, G. (1997), "Scaling Methodology and Procedures for the Mathematics and Science Scales," in M.O. Martin and D.L. Kelly (eds.), *Third International Mathematics and Science Study Technical Report, Volume II*, Chestnut Hill, MA: Boston College.

9 Detailed information about applying sampling weights to TIMSS data can be found in Gonzalez, E.J., and Smith, T. A. (1997)., "Users' Guide to the TIMSS International Database: Primary and Middle School Assessment," Chestnut Hill, MA: Boston College.

**At least 100 Books in Family Home (Science, Mathematics)**
Derived from the number of books in the family home, this variable was coded to 1 if students reported having 100 or more books in the home and 0 if they reported having fewer than 100 books.

**Having a Study Desk, Dictionary, and Computer in Family Home (Science, Mathematics)**
This variable was derived from three items on the student questionnaire asking each student whether he or she had a study desk, a dictionary, and a computer in the home. This variable was coded 1 if all three items were present and 0 if the student reported having fewer than all three items.

**Number of Possessions in the Family Home (Science, Mathematics)**
The index of possessions in the home is based on the ratio of the student's reported number of possessions to the total number of possible possessions in each country. Students who reported having over half of the possible number of home possessions in that country were coded as 1. Students with less than half of the total number of possible possessions were coded as 0.

**At Least One Parent Reported to Have Finished University (Science, Mathematics)**
As part of the student questionnaire, students were asked to report the highest level of education attained by each parent. In the TIMSS database, a single variable was derived to capture the educational status of both parents. For this analysis the variable was recoded to 1 if at least one parent had finished university and to 0 if neither parent had done so.

**Student Works One or More Hours at Home (Science, Mathematics)**
On the student questionnaire, students were asked how many hours they worked at home. If a student reported doing jobs at home for one or more hours each day, he or she was given a code of 1. Less was coded as 0.

**Student Thinks that it is Important to do Well in Mathematics, Science, and Language (Science, Mathematics)**
This indicator was constructed out of three variables from the student questionnaire that asked the student how important it was to do well in mathematics, science, and the language of the test. The new variable representing students' press was coded as 1 in cases where students agreed that it was important to do well in all three areas, and 0 if they did not.

**Mother Thinks that it is Important to do Well in Mathematics, Science, and Language (Science, Mathematics)**
This variable was constructed using three variables asking the student to report whether his or her mother thinks it important to do well in science, mathematics, and the language of the test. The new variable representing mother's press was coded as a 1 in cases where students agreed that their mothers thought it important to do well in all three areas and 0 if they did not.

**Student Plans to Attend University (Science, Mathematics)**
Students reported how much education they anticipate receiving. For the purposes of this report, educational aspirations were coded as 1 if the student expected to attain at least some university education and 0 otherwise.

**Student Daily Works on Homework in Mathematics, Science, and Other Subjects (Science, Mathematics)**
The student questionnaire asked students to report how much homework they do daily in mathematics, science, and other subjects. For this report, the three variables were combined into an index that was coded as 1 only if students reported doing at least some homework in all three subjects daily. If a student had missing data on any of the variables, the case was coded as missing.

**School Located in Urban Area (Science, Mathematics)**
In the school questionnaire, each principal was asked to identify the type of community in which the school was located. For these analyses, cases where principals reported that their schools were located either in or on the outskirts of a major town or city were coded as 1. If the school was reported to be in a geographically isolated area, a village, or a rural (farm) area it was coded as 0.

**School Enrollment Greater than the Country Mean (Science, Mathematics)**
This variable was coded as 1 when the principal reported school enrollment to be in excess of the computed average for that country and as 0 if it was less than that average.

**Average Class Size Greater than the Country Mean (Science, Mathematics)**
To construct this variable, the principal's report of average class size was compared with the average class size for that country. If the average class size reported was greater than that of the country involved, then the new variable was coded as 1. If less, it was coded to 0.

**Student Administrative Violations (Science, Mathematics)**
The administrative violations index was created by taking the mean of principals' reports of student tardiness at school, unjustifiable absenteeism, skipping class, and violation of dress code. The occurrence of each item was rated on a four-point scale from "rarely" to "daily." If no more than one of these variables was found to be missing and the mean was greater than 1.5, then the student misbehavior variable was coded as 1. If the mean was less than or equal to 1.5 it was coded as 0. Cases for which more than one of the component variables was missing were coded as missing.

**Serious Student Misbehavior (Science, Mathematics)**
The creation of the serious student misbehavior variable involved taking the mean of principals' reports of serious problem behavior among students. Such behavior included disrupting the work of other students, cheating, profanity, vandalism, and intimidation. The occurrence of each behavior was rated on a four-point scale from "rarely" to "daily." If the mean value was greater than 1.5, the new variable was coded as 1. If the mean was less than or equal to 1.5, it was coded as 0. At least five of the component items had to be present or the variable would be coded as missing.

**Positive Attitude towards Science (Science only)**
The index of attitude towards science was only based on three statements: I like science; I enjoy learning science; and science is boring. In countries where science subjects are taught separately, students were asked about earth science, life science, physics, and chemistry individually. Each of the statements was rated by students on a five-point scale. If a student reported liking, and enjoying, any of the subject areas and found at least one area interesting, his or her attitude was considered to be positive and coded as 1.

**Positive Attitude towards Mathematics (Mathematics only)**
The item used to measure student attitude towards only mathematics was based upon 5 items from the student questionnaire: How much do you like mathematics; I enjoy learning mathematics (reversed); Mathematics is boring; Mathematics is important to everyone's life (reversed); and I would like a job that involves using mathematics (reversed). Where the mean of these items was 2.5 or higher, a student's attitude towards mathematics was considered to be positive, and the new variable was coded 1. If the mean was less then 2.5, it was coded to 0.

**Belief in the Efficacy of Science (Science only)**
The index was based on responses to questions about the following environmental problems: air pollution; water pollution; destruction of forests; endangered species; damage to the ozone layer; problems

from nuclear power plants. The students who reported believing that science application can help “somewhat” or “a great deal” in addressing all six problems were given a code of 1. Students that did not believe that science could help “somewhat” or “a great deal” to address all six problems were given a 0 on the efficacy of science variable.

**Doing Experiments or Practical Investigations in Class (Science only)**
This index was based on student reports of doing experiments in the following five areas: science (integrated) lessons; biology lessons; chemistry lessons; earth science lessons; physics lessons. Students who report they “almost always” or “pretty often” do experiments in these areas were coded as 1 on this variable.

## Derived Variables for Hierarchical Analyses

A number of the variables in the upper versus lower third analyses also were used in the HLM analyses.

**The Home Background Index**
The Home Background Index (HBI) was constructed by standardizing each component variable and then taking the mean of all non-missing variables. The component variables were: number of people in the family home, number of natural parents in the family home, books in the home, percentage of possessions from the international option list of items, study desk in home, computer in home, highest level of education of father, and highest level of education of mother.

**Homework in Mathematics, Science, and Other Subjects (Science, Mathematics)**
For the hierarchical analyses, this variable was constructed in three stages. First, three variables were made indicating whether or not students’ questionnaire replies reported doing any homework in math, science, and other subjects on a daily basis. Next, the three variables were summed for each student, creating one general homework variable that could range from 0 to 3. A 0 indicated that a student reported not doing homework in the three subject areas on a daily basis. A 3 indicated that a student did homework in all three areas on a daily basis. Finally, the school average was computed for the science analyses and the classroom average was computed for mathematics analyses.

**Amount of Science Homework (Science only)**
For the hierarchical analyses, the amount of time doing science homework was computed as the school average of the amount of time students reported spending doing science homework on a daily basis. The response options provided to students were no time; less than 1 hour; 1-2 hours; 3-5 hours; and more than 5 hours.

**Efficacy of Science (Science only)**
The hierarchical analysis version of this variable was formed by summing each student's response to the following environmental problems: air pollution, water pollution, destruction of forests, endangered species, damage to the ozone layer, and problems from nuclear power plants, and then calculating the school mean.

**Attitude to Science (Science only)**
Attitude to science was based on three statements: I like science, I enjoy learning science, and science is boring (reversed). In countries where science subjects are taught separately, students were asked about earth science, life science, physics, and chemistry individually. Each of the statements was rated by students on a five-point scale ranging from "strongly disagree" to "strongly agree." The variable used in the hierarchical analyses was the school average of this composite student variable.

**Experiments (Science only)**
The index of students doing experiments used in the hierarchical analyses was based on student reports of doing experiments in the following 5 areas: science (integrated) lessons; biology lessons; chemistry lessons; earth science lessons; physics lessons. Each of the items was based upon a 4-point scale ranging from "never" to "almost always" that assessed how often students did experiments or practical investigations in the given area. The maximum value attained by a student in any of these areas was taken as the value of the experiments variable for that student. The school average of this variable was then computed for the hierarchical analyses.

**Checking Homework in Class (Mathematics only)**
This variable represented how often the teacher checked mathematics homework in class. It was constructed by taking the classroom average of students' reports of the amount of time spent checking mathematics homework. The initial item appears in the student questionnaire as a 4-point scale with responses ranging from "never" to "almost always."

**Amount of Mathematics Homework (Mathematics only)**
The amount of mathematics homework variable consisted of the classroom mean of students' reports of the amount of time they spend doing mathematics homework. The student variable was measured on a 5-point scale with options ranging from "no time" to "more than 5 hours."

**Attitude to Mathematics (Mathematics only)**
Attitude to mathematics was derived from 5 items from the student questionnaire: How much do you like mathematics, I enjoy learning mathematics (reversed), mathematics is boring, mathematics is important to everyone's life (reversed), and I would like a job that

involves using mathematics (reversed). Each of the statements was rated by students on a 5-point scale ranging from "strongly disagree" to "strongly agree." The variable used in the hierarchical analyses was the school average of this composite student variable.

**Classroom Behavior (Mathematics only)**
The classroom behavior index used in the mathematics hierarchical analyses was constructed from student agreement to three statements: students often neglect their work (reversed), students are orderly and quiet during lessons (reversed), and students do exactly as the teacher says. Each of the statements was rated by students on a 4-point scale ranging from "strongly disagree" to "strongly agree." The variable used in the hierarchical analyses was the school average of this composite student variable.

**Mathematics Class Size (Mathematics only)**
In the mathematics analyses the size of the classroom was determined by adding the number of boys to the number of girls reported in each mathematics class.

**Teaching Experience in Science (Science, Mathematics)**
The number of years the teacher has been teaching is used in the mathematics hierarchical analyses as a proxy for teaching experience. In science, the school mean of the variable was used, as the TIMSS sampling design allowed for more than one science teacher to be represented in each intact mathematics classroom at the grade tested.

**Readiness to Teach General Science (Science only)**
This index was created from a series of items from the teacher questionnaire that asked teachers to report their readiness to teach earth features, energy, light, human tissues and organs, metabolism, reproduction, genetics, measurement, and data organization. Readiness to teach each subject area was rated on a 3-point scale ranging from "not well prepared" to "confident teaching this topic." The mean of each of these items was computed. School means were then calculated.

**Student Administrative Violations (Science, Mathematics)**
The variable representing administrative violations was computed as the mean of principals' reports of students arriving late at school, unjustifiable absenteeism, students skipping class, and violation of dress code. The occurrence of each item was rated on a 4-point scale from "rarely" to "daily."

**Serious Student Misbehavior (Science, Mathematics)**
For the hierarchical analyses this variable was computed as the mean of principals' reports of serious problem behavior among students. Such behavior included disrupting the work of other students, cheating, profanity, vandalism, and intimidation. The occurrence of each item was rated on a four-point scale from "rarely" to "daily."

**School Location (Science, Mathematics)**
The school location variable used in the hierarchical analyses was the principal's report of the type of community in which the school was located. The response options were a geographically isolated area; village or rural (farm) area; on the outskirts of a town/city; close to the center of a town/city. Higher numbers indicated generally greater urbanization.

**Average Class Size (Science, Mathematics)**
Average class size was as reported by principals on the school questionnaire. It does not refer specifically either to science or mathematics classes in the school.

**Aspirations for Future Education (Science, Mathematics)**
In the student questionnaire, each student was asked to identify the level of education that he or she expected to receive, with "finished university" being the highest option available. The school mean was used to represent this variable in the hierarchical analyses.

**Mother's Press (Science, Mathematics)**
The mother's press variable in the hierarchical analyses was derived by first taking for each student the mean of the three variables asking the student to report whether his or her mother thinks it is important to do well in science, mathematics, and the language of the test. Responses were on a 4-point scale ranging from "strongly agree" to "strongly disagree." The school average was used in the hierarchical analyses.

**Self Press (Science, Mathematics)**
The self press variable in the hierarchical analyses was derived by first taking for each student the mean of the three variables asking the student to report whether his or her mother thinks it is important to do well in science, mathematics, and the language of the test. Responses were on a 4-point scale ranging from "strongly agree" to "strongly disagree." The school average was used in the hierarchical analyses.

**School Average Home Background Index (Science, Mathematics)**
The school average on the home background index was used as a school-level variable in some of the hierarchical analyses.