Appendix A
Supporting Documentation

TIMSS Advanced 2008 Advanced Mathematics and Physics Assessments

The *TIMSS Advanced 2008 Assessment Frameworks*[1] define the advanced mathematics and physics to be addressed in the TIMSS Advanced 2008 assessment and provide an outline of the assessment design. As described in that document, the frameworks for advanced mathematics and physics were built around content and cognitive domains within each subject. Algebra (35%), calculus (35%), and geometry (30%) made up the advanced mathematics content domains and their targeted percentages, and knowing (35%), applying (35%), and reasoning (30%) were the cognitive domains and target percentages. The physics content domains were mechanics (30%), electricity and magnetism (30%), heat and temperature (20%), and atomic and nuclear physics (20%). The cognitive domains in physics were the same as in mathematics, but the target percentages slightly different—knowing (30%), applying (40%), and reasoning (30%). Exhibit A.1 presents these content and cognitive domains together with the number of items and score points in each domain and the distribution of score points across domains.

Exhibit A.1   **Distribution of Advanced Mathematics and Physics Items by Content Domain and Cognitive Domain**

TIMSSAdvanced 2008

SOURCE: IEA TIMSS Advanced 2008 ©

Advanced Mathematics

Content Domain	Number of Multiple-choice Items	Number of Constructed-response Items	Total Number of Items	Total Number of Score Points[1]	Percentage of Score Points
Algebra	17	9	26	30	37
Calculus	13	12	25	29	35
Geometry	16	5	21	23	28
Total	46	26	72	82	100

Cognitive Domain	Number of Multiple-choice Items	Number of Constructed-response Items	Total Number of Items	Total Number of Score Points[1]	Percentage of Score Points
Knowing	21	7	28	30	37
Applying	14	13	27	31	38
Reasoning	11	6	17	21	26
Total	46	26	72	82	100

Physics

Content Domain	Number of Multiple-choice Items	Number of Constructed-response Items	Total Number of Items	Total Number of Score Points[1]	Percentage of Score Points
Mechanics	11	9	20	24	29
Electricity and Magnetism	13	8	21	24	29
Heat and Temperature	7	8	15	20	24
Atomic and Nuclear Physics	11	4	15	16	19
Total	42	29	71	84	100

Cognitive Domain	Number of Multiple-choice Items	Number of Constructed-response Items	Total Number of Items	Total Number of Score Points[1]	Percentage of Score Points
Knowing	12	6	18	18	21
Applying	25	11	36	41	49
Reasoning	5	12	17	25	30
Grand Total	42	29	71	84	100

[1]   In scoring the tests, correct answers to most items were worth one point. However, responses to some constructed-response items were evaluated for partial credit with a fully correct answer awarded two points. Thus, the number of score points exceeds the number of items in the test.

TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

The advanced mathematics assessment had a total of 72 items: 26 items in algebra, 25 items in calculus, and 21 items in geometry. Each item also was categorized according to its cognitive domain, with 28 items in the knowing domain, 27 in the applying domain, and 17 in the reasoning domain. A little more than one third of the items (26) were in constructed-response format and the rest (46) were multiple-choice items. The constructed-response items required students to generate and write their own answers. Some items required short answers while others demanded a more elaborate response. In scoring the assessment, correct answers to most questions (including all those in multiple-choice format) were worth 1 point. However, responses to questions seeking more elaborate responses were evaluated for partial credit, with a fully-correct answer being awarded 2 points. Thus, the total number of score points available for analyses (82) somewhat exceeds the number of items in the assessment. The percentages of score points for the content and cognitive domains were nearly identical to the target percentages designated in the advanced mathematics assessment framework.

In the physics assessment, there were 71 items in total: 20 mechanics items, 21 electricity and magnetism items, 15 heat and temperature items, and 15 atomic and nuclear physics items. Of these, 18 were classified as measuring knowing, 36 as measuring applying, and 17 as measuring reasoning skills. Compared to the target percentages in the physics framework, there was a relatively greater percentage of applying items and lesser percentage of knowing and reasoning items in the assessment. Two fifths of the items (28) were constructed response and the remainder (42) multiple choice.

Sample Implementation and Participation Rates

The TIMSS Advanced 2008 assessment was administered to scientifically-selected random samples of students from the target populations for advanced mathematics and physics in each country. These target populations were defined as follows: *students in the final year of secondary schooling who have taken courses in advanced mathematics*, and *students in the final year of secondary schooling who have taken courses in physics*. The tracks or programs that define these target populations are presented in detail in Chapter 1 for advanced mathematics and Chapter 7 for physics.

Because the accuracy of the TIMSS results depends on the quality of the national samples, TIMSS Advanced worked with participating countries on all phases of sampling to ensure efficient sampling design and implementation. National coordinators were trained in how to select the school and student samples, and how to use the *WinW3S* sampling software provided by the IEA Data Processing and Research Center. Staff from Statistics Canada reviewed the national sampling plans, sampling data, sampling frames, and sample selections. The sampling documentation was used by the TIMSS & PIRLS International Study Center (in consultation with Statistics Canada and the sampling referee) to evaluate the quality of the samples.

Exhibit A.2 shows that the TIMSS Advanced countries were very successful in developing comprehensive national sampling schemes that covered 100 percent of their intended advanced mathematics and physics target populations. No country found it necessary to restrict coverage by, for example, omitting specific regions or language groups. The countries also were successful in ensuring that the percentage of students excluded from the target populations was below the 5% limit. In fact, in no country was the percentage of excluded students greater

Exhibit A.2: **Coverage of TIMSS Advanced 2008 Target Populations for Advanced Mathematics and Physics**

Advanced Mathematics

Country	Coverage	School-level Exclusions	Within-sample Exclusions	Overall Exclusions
Armenia	100%	0.0%	0.0%	0.0%
Iran, Islamic Rep. of	100%	0.0%	0.0%	0.0%
Italy	100%	0.0%	0.5%	0.5%
Lebanon	100%	1.3%	0.0%	1.3%
Netherlands	100%	2.5%	0.0%	2.5%
Norway	100%	0.9%	0.1%	1.0%
Philippines	100%	0.0%	0.0%	0.0%
Russian Federation	100%	0.0%	0.0%	0.0%
Slovenia	100%	0.0%	1.3%	1.3%
Sweden	100%	1.5%	0.2%	1.7%

Physics

Country	Coverage	School-level Exclusions	Within-sample Exclusions	Overall Exclusions
Armenia	100%	0.0%	0.0%	0.0%
Iran, Islamic Rep. of	100%	0.0%	0.0%	0.0%
Italy	100%	0.0%	0.9%	0.9%
Lebanon	100%	1.3%	0.0%	1.3%
Netherlands	100%	2.5%	0.2%	2.7%
Norway	100%	0.4%	0.0%	0.5%
Russian Federation	100%	0.0%	0.0%	0.0%
Slovenia	100%	0.0%	0.5%	0.5%
Sweden	100%	2.1%	0.1%	2.3%

TIMSSAdvanced 2008

TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

than 3 percent, and Armenia, Iran, the Philippines, and the Russian Federation had no excluded students at all. Usually when students are excluded from testing in large-scale assessments, it is because they are in schools that would be very difficult or resource intensive to test (e.g., schools that were very small or located in remote rural areas) or because they do not have sufficient knowledge of the language of the test or have a disability. However, in order to be part of the advanced mathematics or physics target populations in TIMSS Advanced, students have had to demonstrate a strong track record of achievement in these subjects so there may be relatively fewer students with language learning limitations or disabilities than might be encountered at lower grades.

The differences in how countries organize their education systems to provide advanced courses in mathematics and physics are reflected in marked differences across countries in the proportion of the age cohort that attend such courses in the final year of secondary education. In some countries, only a very select group of students were considered eligible for the study, while in others a much larger group was included. To measure differences in coverage of the national age cohorts, coverage indices were calculated for both the advanced mathematics and physics populations. The TIMSS Advanced Mathematics Coverage Index (TAMCI) and the TIMSS Advanced Physics Coverage Index (TAPCI) identify the percentage of the school-leaving age cohort represented by the advanced mathematics and physics samples, respectively.

The TIMSS Advanced coverage indices for advanced mathematics and physics are defined as follows:

$$\text{TAMCI} = \frac{\text{Estimated total number of students in advanced mathematics target population in 2008}}{\text{Total national population in the corresponding age cohort in 2008}} \times 100\%$$

$$\text{TAPCI} = \frac{\text{Estimated total number of students in physics target population in 2008}}{\text{Total national population in the corresponding age cohort in 2008}} \times 100\%$$

TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

For each index, the *numerator* is the total number of students eligible for TIMSS Advanced, estimated from the weighted sample data. The *denominator* is size of the population age cohort corresponding to the average age of the students in the target populations. Exhibit A.3[2] presents these data for the advanced mathematics and physics populations for each country and the resulting coverage indices. Data on the size of the age cohort were provided by National Research Coordinators from official national statistics (except for Armenia, where data were obtained from the U.S. Census Bureau's International Database). For Armenia, Lebanon, and Slovenia, population data were not available for the specific age cohort corresponding to the TIMSS Advanced sample, but had to be estimated from data spanning several years. In the case of Armenia and Slovenia, the available population figure for the age group 15–19 was divided by 5 to derive an estimate of the single year age cohort: 18-year-olds for Armenia and 19-year-olds for Slovenia. In Lebanon, the population figure for the 18–20 age group was divided by 3 to get an estimate of the size of the 18-year age cohort.

The TIMSS Advanced coverage indices show that the population of students taking advanced course in mathematics or physics (i.e., enrolled in the tracks or programs targeted by TIMSS Advanced 2008) represented a low percentage of the students in the corresponding age cohort. Highest values on the TIMSS Advanced Mathematics Coverage Index were found in Slovenia (40.5%), Italy (19.7%), Sweden (12.8%), and Norway (10.9%), each of which had more than 10 percent of the age cohort eligible for TIMSS Advanced. Countries with lower values included Iran (6.5%), Lebanon (5.9%), Armenia (4.3%), the Netherlands (3.5%), the Russian Federation (1.4%), and the Philippines (0.7%). Only Sweden (11.0%) had more than 10 percent of the age cohort enrolled in advanced physics courses. For the other

---

2   Exhibit A.3 is derived from Exhibits 1.2 in Chapter 1 and 7.2 in Chapter 7.

Exhibit A.3 **Size of the TIMSS Advanced 2008 Target Populations for Advanced Mathematics and Physics, Age Cohorts, and Coverage Indices**

TIMSSAdvanced2008

SOURCE: IEA TIMSS Advanced 2008 ©

Advanced Mathematics

Country	Estimated Size of the Population of Students in the Final Year of Secondary School Taking the Advanced Mathematics Track or Program Targeted by TIMSS Advanced (Derived from TIMSS Advanced Student Sample)	Age Cohort Corresponding to the Final Year of Secondary School	Size of the Age Cohort Corresponding to the TIMSS Advanced Population Based on National Census Figures[a]	TIMSS Advanced Mathematics Coverage Index – the Percentage of the Entire Corresponding Age Cohort Covered by the TIMSS Advanced Target Population
Armenia	2,684	18	62,758	4.3%
Iran, Islamic Rep. of	111,298	18	1,705,000	6.5%
Italy	119,162	19	605,507	19.7%
Lebanon	4,702	18	79,784	5.9%
Netherlands	7,091	18	205,200	3.5%
Norway	6,668	19	61,093	10.9%
Philippines	14,007	16	1,900,656	0.7%
Russian Federation	29,672	17	2,073,041	1.4%
Slovenia	8,836	19	21,815	40.5%
Sweden	16,116	19	125,923	12.8%

Physics

Country	Estimated Size of the Population of Students in the Final Year of Secondary School Taking the Physics Track or Program Targeted by TIMSS Advanced (Derived from TIMSS Advanced Student Sample)	Age Cohort Corresponding to the Final Year of Secondary School	Size of the Age Cohort Corresponding to the TIMSS Advanced Population Based on National Census Figures[a]	TIMSS Physics Coverage Index – the Percentage of the Entire Corresponding Age Cohort Covered by the TIMSS Advanced Target Population
Armenia	2,684	18	62,758	4.3%
Iran, Islamic Rep. of	111,908	18	1,705,000	6.6%
Italy	23,176	19	605,507	3.8%
Lebanon	4,724	18	79,784	5.9%
Netherlands	6,889	18	205,200	3.4%
Norway	4,181	19	61,093	6.8%
Russian Federation	52,934	17	2,073,041	2.6%
Slovenia	1,635	19	21,815	7.5%
Sweden	13,873	19	125,923	11.0%

[a] Armenia: Estimate derived by dividing the population of 15–19-year olds by 5 for the single year estimate for the year 2008. Data taken from the U.S. Census Bureau's International Database (www.census.gov/). Iran, Islamic Rep. of: Total population of 18-year olds in Iran in 2008. Data taken from the Statistical Center of Iran (SCI) (http://www.sci.org.ir/portal/faces/public/sci_en). Italy: Total population of 19-year olds in Italy for the year 2008. Data taken from the Italian Bureau of Statistics (ISTAT) (http://demo.istat.it/pop2008/index.html). Lebanon: Estimate derived by dividing the population of 18–20-year olds by 3 for the single year estimate. Data taken from the Central Bureau for Statistics in the Ministry of Interior. Netherlands: Estimate based on data taken from the Central Bureau of Statistics in the Netherlands (www.cbs.nl). Norway: Total population of 19-year olds in Norway on 1 January 2008. Data taken from the Norwegian National Bureau of Statistics (SSB) (http://www.ssb.no/english/). Philippines: Population of 16-year olds for 2008 projected from the 2000 census. Data taken from the National Statistics Office, Philippines (NSO) (http://www.census.gov.ph/). Russian Federation: Total population of 17-year olds in 2008. Data taken from the Federal State Statistics Service (http://www.gks.ru/wps/portal/english). Slovenia: Estimate was derived by dividing the population of 15–19-year olds by 5 for the single year estimate for the year 2008. Data taken from the Statistical Office of the Republic of Slovenia (www.stat.si). Sweden: Total population of 19-year olds in Sweden for the year 2008. Data taken from Statistics Sweden (SCB) (http://www.scb.se/default____2154.aspx). Data provided by National Research Coordinators.

TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

countries, the values of the TIMSS Advanced Physics Coverage Index were as follows: Slovenia (7.5%), Norway (6.8%), Iran (6.6%), Lebanon (5.9%), Armenia (4.3%), Italy (3.8%), the Netherlands (3.4%), and the Russian Federation (2.6%).

The basic sampling design used in TIMSS Advanced 2008 was a two-stage stratified cluster design, with schools sampled at the first stage and one or more intact classes from a list of eligible classes in the school at the second stage.[3] In countries with large school populations (Iran, Italy, and the Russian Federation), schools were selected with probability proportional to size. In Lebanon, the Netherlands, Norway, the Philippines, and Sweden, which had smaller school populations, schools were sampled with equal probabilities, and in Armenia and Slovenia, all schools were included in the sample. In all countries, classes within sampled schools were selected using a systematic random sampling method.

Although TIMSS Advanced aimed for a uniform sampling approach for all countries, the implementation was influenced by the relationship between the advanced mathematics and physics populations and how classrooms were organized in each country. In Armenia, Iran, and Lebanon, with completely overlapping populations (i.e., the populations of advanced mathematics and physics students were identical), there was a single school and class sample, with half the students in the sampled classes randomly assigned a mathematics booklet and the other half a physics booklet. In the Netherlands, Norway, and Sweden, where students could belong to the advanced mathematics population, the physics population, or both, separate school samples were selected, with only the advanced mathematics classes listed for sampling in one sample and only physics classes listed in the other. The Philippines, which assessed students in advanced mathematics only, could be considered a special case of this

3   See LaRoche, S., Zuehlke, O., and Joncas, M. (2009). TIMSS Advanced 2008 sampling. In A. Arora, P. Foy, M.O. Martin, & I.V.S. Mullis (Eds.), *TIMSS Advanced 2008 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

approach. Italy, the Russian Federation, and Slovenia each had specific issues that required more complex adaptations to the basic sampling approach. These are described in detail in the *TIMSS Advanced 2008 Technical Report.*

Most countries sampled at least 120 schools and at least one intact classroom from each school for each population. This approach was designed to yield a representative sample of at least 2,000 students for each population in each country. Armenia and Slovenia had fewer than 120 eligible schools, and so all were included in the sample.

Exhibits A.4 and A.5 present achieved sample sizes for schools and students, respectively. Exhibit A.6 shows the participation rates for schools, classes, students, and overall—both with and without the use of replacement schools. With the exception of the physics sample in Slovenia, all countries achieved the minimum acceptable participation rates—85 percent of both the schools and students, or a combined rate (the product of school and student participation) of 75 percent—although the Netherlands did so only after including replacement schools. The results for the Netherlands in both subjects and for Slovenia in physics have been annotated in the achievement exhibits contained in this report (see Chapters 2, 3, 8, and 9).

Because an important goal for the TIMSS Advanced 2008 countries that also participated in 1995—Italy, the Russian Federation, Slovenia, and Sweden in advanced mathematics and Norway, the Russian Federation, Slovenia, and Sweden in physics—was to measure changes in students' achievement since 1995, it was important to track any changes in population composition and coverage since they might be related to student achievement. Exhibit A.7 presents, for each of these countries, five attributes of the advanced mathematics and physics populations sampled in 2008 and 1995: number of years of formal schooling, average student age at time of testing, percentage

Exhibit A.4: **School Sample Sizes – Advanced Mathematics and Physics**

Advanced Mathematics

Country	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample That Participated	Number of Replacement Schools That Participated	Total Number of Schools That Participated
Armenia	38	38	38	0	38
Iran, Islamic Rep. of	120	120	119	0	119
Italy	100	92	88	3	91
Lebanon	240	240	203	9	212
Netherlands	135	133	102	10	112
Norway	120	120	107	0	107
Philippines	121	120	118	0	118
Russian Federation	143	143	143	0	143
Slovenia	87	82	79	0	79
Sweden	127	126	111	5	116

Physics

Country	Number of Schools in Original Sample	Number of Eligible Schools in Original Sample	Number of Schools in Original Sample That Participated	Number of Replacement Schools That Participated	Total Number of Schools That Participated
Armenia	38	38	38	0	38
Iran, Islamic Rep. of	120	120	119	0	119
Italy	112	91	91	0	91
Lebanon	240	240	201	9	210
Netherlands	135	133	98	18	116
Norway	120	120	101	0	101
Russian Federation	149	149	149	0	149
Slovenia	66	64	54	0	54
Sweden	127	125	119	2	121

Exhibit A.5:   **Student Sample Sizes – Advanced Mathematics and Physics**

TIMSSAdvanced**2008**

SOURCE: IEA TIMSS Advanced 2008 ©

Advanced Mathematics

Country	Within School Student Participation (Weighted Percentage)	Number of Sampled Students in Participating Schools	Number of Students Withdrawn from Class/ School	Number of Students Excluded	Number of Students Eligible	Number of Students Absent	Number of Students Assessed
Armenia	95%	899	0	0	899	41	858
Iran, Islamic Rep. of	97%	2,556	55	0	2,501	76	2,425
Italy	96%	2,269	15	8	2,246	103	2,143
Lebanon	95%	1,767	36	0	1,731	116	1,615
Netherlands	92%	1,876	200	0	1,676	139	1,537
Norway	89%	2,206	17	2	2,187	255	1,932
Philippines	96%	4,253	3	0	4,250	159	4,091
Russian Federation	98%	3,269	11	0	3,258	73	3,185
Slovenia	85%	2,577	3	22	2,552	396	2,156
Sweden	89%	2,645	26	1	2,618	315	2,303

Physics

Country	Within School Student Participation (Weighted Percentage)	Number of Sampled Students in Participating Schools	Number of Students Withdrawn from Class/ School	Number of Students Excluded	Number of Students Eligible	Number of Students Absent	Number of Students Assessed
Armenia	97%	926	0	0	926	32	894
Iran, Islamic Rep. of	97%	2,556	43	0	2,513	79	2,434
Italy	97%	1,968	18	15	1,935	74	1,861
Lebanon	94%	1,755	35	0	1,720	120	1,600
Netherlands	90%	1,911	203	3	1,705	194	1,511
Norway	86%	1,935	17	1	1,917	275	1,642
Russian Federation	97%	3,269	9	0	3,260	94	3,166
Slovenia	82%	1,404	0	6	1,398	278	1,120
Sweden	92%	2,537	29	4	2,504	213	2,291

TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

Exhibit A.6: **Participation Rates (Weighted) – Advanced Mathematics and Physics**

Advanced Mathematics

Country	School Participation Before Replacement	School Participation After Replacement	Class Participation	Student Participation	Overall Participation Before Replacement	Overall Participation After Replacement
Armenia	100%	100%	100%	95%	95%	95%
Iran, Islamic Rep. of	99%	99%	100%	97%	96%	96%
Italy	97%	99%	100%	96%	93%	95%
Lebanon	86%	89%	99%	95%	81%	83%
Netherlands	77%	84%	100%	92%	71%	77%
Norway	94%	94%	100%	89%	83%	83%
Philippines	98%	98%	100%	96%	95%	95%
Russian Federation	100%	100%	100%	98%	98%	98%
Slovenia	96%	96%	100%	85%	81%	81%
Sweden	90%	94%	100%	89%	80%	84%

Physics

Country	School Participation Before Replacement	School Participation After Replacement	Class Participation	Student Participation	Overall Participation Before Replacement	Overall Participation After Replacement
Armenia	100%	100%	100%	97%	97%	97%
Iran, Islamic Rep. of	99%	99%	100%	97%	96%	96%
Italy	100%	100%	100%	97%	97%	97%
Lebanon	85%	88%	99%	94%	80%	82%
Netherlands	73%	87%	100%	90%	65%	78%
Norway	85%	85%	100%	86%	73%	73%
Russian Federation	100%	100%	100%	97%	97%	97%
Slovenia	83%	83%	98%	82%	67%	67%
Sweden	97%	97%	100%	92%	89%	89%

TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

Exhibit A.7:   **Trends in Characteristics of TIMSS Advanced Student Populations**   TIMSSAdvanced2008

SOURCE: IEA TIMSS Advanced 2008 ©

Advanced Mathematics

Country	Years of Formal Schooling*		Average Age at Time of Testing		Exclusion Rates		Mathematics Coverage Index		Overall Participation Rate (After Replacement)	
	2008	1995	2008	1995	2008	1995**	2008	1995	2008	1995
Italy	13	13	19.0	19.1	0.5%	3.8%	19.7%	20.2% ***	94.8%	67.5%
Russian Federation	10/11	11	17.0	16.9	0.0%	2.0%	1.4%	2.0%	97.6%	95.9%
Slovenia	12	12	18.8	18.9	1.3%	6.0%	40.5%	75.4%	81.4%	42.4%
Sweden	12	12	18.8	18.9	1.7%	0.2%	12.8%	16.2%	83.6%	88.6%

Physics

Country	Years of Formal Schooling*		Average Age at Time of Testing		Exclusion Rates		Physics Coverage Index		Overall Participation Rate (After Replacement)	
	2008	1995	2008	1995	2008	1995**	2008	1995	2008	1995
Norway	12	12	18.8	19.0	0.5%	3.8%	6.8%	8.4%	73.0%	83.0%
Russian Federation	10/11	11	17.1	16.9	0.0%	2.0%	2.6%	1.5%	97.3%	95.1%
Slovenia	12	12	18.7	18.8	0.5%	6.0%	7.5%	38.6%	67.1%	43.0%
Sweden	12	12	18.8	18.9	2.3%	0.2%	11.0%	16.3%	89.3%	88.6%

*   Represents years of schooling counting from the first year of primary or basic education (first year of ISCED Level 1).

**   In 1995 exclusion rates for Advanced Mathematics and Physics were computed based on exclusion rates among all students in the final year of schooling. In the case of the Russian Federation, the figure presented in the 1995 International Report (43.0%) greatly overestimates the level of exclusions in the advanced mathematics population. The figure presented above (2.0%) includes two regions, North Ossetia and Chechen Republic, as well as non-Russian speaking students.

***   The 1995 mathematics coverage index for Italy was recomputed for this report and is different from the figure reported in 1995.

TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

of students excluded from the assessment, the advanced mathematics or physics coverage index, and overall sampling participation rate (after replacement).

The participating countries were very similar in 2008 and 1995 for both the advanced mathematics and physics populations with regard to years of formal schooling, average age at time of testing, and exclusion rates. The greatest changes involved Slovenia, which had an TIMSS Advanced Mathematics Coverage Index of 40.5 percent in 2008 compared to 75.4 percent in 1995, and a TIMSS Advanced Physics Coverage Index of 7.5 percent in 2008 compared to 38.6 percent in 1995. Sweden also had a lower coverage index in 2008 than in 1995, although the difference was not so great (12.8% in 2008 compared to 16.2% in 1995 for mathematics, and 11.0% in 2008 compared to 16.3% in 1995 for physics). Slovenia had higher student sampling participation in 2008 than in 1995 (81.4% vs. 42.4% for advanced mathematics, and 67.1% vs. 43.0% for physics). Slovenia did not meet the TIMSS standards for sampling participation in 1995.

Translation and Layout Verification

Participants were given detailed guidelines for translating the TIMSS Advanced 2008 instruments developed in English into their target language(s) and adapting them to be appropriate for their cultural contexts. They also were urged to work with an experienced translator who would be well-suited to the task of working with the TIMSS materials. Because the goal was to create a set of instruments comparable to the originals in terms of difficulty and accessibility, the instruments were subjected to a stringent international translation verification process. Each participant was asked to submit the following materials for verification prior to both the field test and main data collection: items and directions; questionnaires for students, teachers,

TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

and schools; manuals; and scoring guides for constructed-response items, where necessary. Verifiers documented their suggestions, and the National Research Coordinators were responsible for reviewing the suggestions and revising the instruments. The verified instruments were used to generate the booklets and questionnaires in their final form and these were submitted to the TIMSS & PIRLS International Study Center for international layout verification. Participants who tested in English also were required to go through the verification steps. Although they had not translated the instruments, the materials were reviewed for national adaptations and comparable layout. Further information is provided in the *TIMSS Advanced 2008 Technical Report.*

Survey Operations for Data Collection

Designing the survey operations for data collection was a collaborative effort between the TIMSS & PIRLS International Study Center, the IEA Secretariat, the IEA Data Processing and Research Center, and Statistics Canada. Data collection involved contacting schools and sampling classes, preparing materials for data collection, administering the assessment, conducting quality control, scoring the assessment, and creating the data files. Detailed information is provided in the *TIMSS Advanced 2008 Technical Report*. However, in brief, guidelines for each of these activities were described in an international set of materials, software, and manuals provided to each National Research Coordinator; for example, manuals for the school coordinator, the test administrators, and the national quality control observers. The school coordinator was responsible for coordinating the testing, including arranging for test administrators, receiving the testing materials, and returning the completed materials to the national center. Within the schools, the assessment was conducted by the test administrator for each class which involved distributing materials to the appropriate

students, following the script for the administration, and timing the sessions accurately. During the test administrations, 10 percent of the schools were visited by an International Quality Control Monitor hired by the IEA Secretariat and trained to verify the quality of the materials and adherence to the test administration procedures in each country. Additionally, countries were asked to conduct their own quality control procedures in another 10 percent of sampled schools, based on the international quality control program.

Scoring the Constructed-response Items

Because more than half of the score points on the assessment came from constructed-response items, TIMSS Advanced 2008 had to develop procedures for reliably evaluating student responses within and across countries. To ensure reliable scoring procedures based on the TIMSS Advanced scoring rubrics, the TIMSS & PIRLS International Study Center prepared detailed guides containing the rubrics and explanations of how to implement them, together with example student responses for the various rubric categories. These guides, along with training packets containing extensive examples of student responses for practice in applying the rubrics, were used as a basis for intensive training in scoring the constructed-response items. The training sessions were designed to help representatives of national centers, who would then be responsible for training personnel in their own countries to apply the scoring rubrics reliably.

To gather and document information about the within-country agreement among scorers, TIMSS Advanced arranged to have systematic sub-samples of at least 200 students' responses to each item scored independently by two scorers. Scoring reliability within countries was high. The percentage of agreement for score points, on average, across countries, was 98 percent for advanced mathematics

and 97 percent for physics. Country-by-country results are provided in the *TIMSS Advanced 2008 Technical Report*.

While the double scoring of a sample of the student test booklets provided a measure of the consistency with which the constructed-response questions were scored within each country, TIMSS Advanced also took steps to monitor the consistency with which the scoring rubrics were applied across countries. TIMSS Advanced assembled a sample of 100 student responses in English to each of 9 constructed-response items in advanced mathematics and in physics. The set of 900 student responses in each subject was then sent to each TIMSS Advanced participant that had scorers proficient in English, and all responses were scored independently by two of these scorers. Seven countries participated in this exercise for each subject—Armenia, Iran, Italy, Norway, Slovenia, and Sweden participated for both subjects, and were joined by the Philippines for advanced mathematics and by the Russian Federation for physics. With 2 scorers from each of the 7 countries, each student response to an item was scored independently by 14 scorers. Comparing each assigned score with all others gives 91 comparisons for each student response (the number of different pairs of scores that can be made from 14 scores is $^{14}C_2 = (14 \times 13) \div 2$). Since there were 100 responses to each item, this gives 9,100 comparisons for each item, and further multiplying by 9 items gives 81,900 comparisons in total for each of advanced mathematics and physics. Agreement across countries was defined in terms of the percentage of these comparisons that were in exact agreement, and it was high: 94 percent for advanced mathematics and 88 percent for physics. Details may be found in the *TIMSS Advanced 2008 Technical Report*.

TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

Test Reliability

As an indication of the reliability of the measurement of student achievement, TIMSS calculated a test reliability coefficient for each country. This coefficient is the median KR-20 reliability across the four test booklets for each subject. Reliabilities ranged across countries from 0.70 to 0.90 for advanced mathematics and from 0.68 to 0.88 for physics. Across all countries, the median reliability coefficient was 0.80 for advanced mathematics and 0.82 for physics. More information may be found in the *TIMSS Advanced 2008 Technical Report*.

Scaling the Achievement Data

The primary approach to reporting the TIMSS Advanced 2008 achievement data was based on item response theory (IRT) scaling methods.[4] Student achievement in advanced mathematics and physics was summarized using 2- and 3-parameter IRT models for dichotomously-scored items (right or wrong), and generalized partial credit models for constructed-response items with two available score points. The IRT scaling method produces a score by averaging the responses of each student to the items that he or she took in a way that takes into account the difficulty and discriminating power of each item. The methodology used in TIMSS Advanced included refinements enabling reliable scores to be produced even though individual students responded to just one assessment booklet. With four advanced mathematics booklets and four physics booklets, each booklet contained about three sevenths of the TIMSS Advanced achievement items in one subject. Thus, TIMSS Advanced has two separate achievement scales: one for advanced mathematics and one for physics.

---

4   For a detailed description of the TIMSS Advanced 2008 scaling, see Foy, P., Galia, J., & Li, I. (2009). Scaling the data from the TIMSS Advanced 2008 mathematics and physics assessments. In A. Arora, P. Foy, M.O. Martin, & I.V.S. Mullis (Eds.), *TIMSS Advanced 2008 Technical Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

To improve the estimation of summary statistics for student subpopulations, the TIMSS Advanced scaling made use of conditioning and plausible-value technology, whereby five separate estimates of each student's score were generated on each scale, based on the student's responses to the items in that student's booklet and on the student's background characteristics. The five score estimates are known as "plausible values," and the variability between them encapsulates the uncertainty inherent in the score estimation process. The IRT analysis provides a common scale on which performance can be compared across countries. In addition to providing a basis for estimating mean achievement, scale scores permit estimates of how students within countries vary and provide information on percentiles of performance.

An overall advanced mathematics achievement scale and an overall physics achievement scale were produced. In order to measure trends in advanced mathematics achievement and physics achievement between the 1995 and 2008 assessments, the TIMSS Advanced mathematics and physics achievement scales were designed to provide reliable measures on a common scale spanning 1995 and 2008. However, because achievement scaling in TIMSS Advanced 1995 was originally conducted using a 1-parameter model, the 1995 assessment was rescaled using the 2- and 3-parameter model approach.[5] The metric of the scales was established with the re-scaled 1995 assessment data. Treating all countries participating in TIMSS Advanced 1995 equally, the TIMSS Advanced scale average for each subject across those countries was set at 500, and the standard deviation was set at 100. The average and standard deviation of the scale scores are arbitrary and do not affect scale interpretation. Since the countries varied in size, each country was weighted to contribute equally to the mean and standard deviation of the scale. To preserve the metric of the original 1995 scale for use with the 2008 data, the 2008 assessment was scaled using

5   The rescaling of the TIMSS Advanced 1995 data is described in the scaling chapter by Foy, Galia, & Li in the *TIMSS Advanced 2008 Technical Report*. The rescaled 1995 data have been used in all trend analyses.

TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

students from all countries that participated in 1995 and all countries that participated in 2008. All advanced mathematics and physics items from 1995 and 2008 were included in this scaling, including about one third of the items that were used in both assessments and that formed the foundation for linking the 1995 and 2008 assessment data.

Scale Anchoring Analysis

For the scale anchoring analysis described in Chapters 3 and 9, the students' achievement results from all the participating countries were pooled so that the benchmark descriptions refer to all students achieving at that benchmark level. Thus, in determining performance in relation to the benchmarks, it does not matter what country a student is from, only how he or she performed on the test. Considering students' advanced mathematics and physics achievement scores, criteria were applied to identify the sets of items that students reaching each international benchmark were likely to answer correctly and that those at the next lower benchmark were unlikely to answer correctly.

For example, a multiple-choice item anchored at the Advanced International Benchmark if at least 65 percent of students scoring at 625 answered the item correctly and fewer than 50 percent of students scoring at the High International Benchmark (550) answered correctly. Similarly, a multiple-choice item anchored at the High International Benchmark if at least 65 percent of students scoring at 550 answered the item correctly and fewer than 50 percent of students scoring at the Intermediate International Benchmark (475) answered it correctly. A multiple-choice item anchored at the Intermediate International Benchmark if at least 65 percent of students scoring at 475 answered correctly. Since constructed-response questions virtually eliminate guessing, the criterion for the constructed-response items was simply

TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College

50 percent at the particular benchmark. Also, the analysis was conducted based on the percentage of students receiving full credit.

The sets of items identified by the scale anchoring analysis represented the accomplishments of students reaching each successively higher benchmark, and were used by the committee of experts[6] that worked with staff of the TIMSS & PIRLS International Study Center to develop the benchmark descriptions. For each benchmark, the work of the committee involved developing a short description for each anchor item that characterized the content knowledge and skills demonstrated by students answering it successfully. These item-by-item descriptions were then summarized by the committee members to provide the more general statements of achievement at each of the benchmarks. The item-by-item descriptions and further details about the analysis can be found in the *TIMSS Advanced 2008 Technical Report*.

The descriptions of achievement at the benchmarks are based solely on student performance on the TIMSS Advanced 2008 items and do not purport to be comprehensive. There are undoubtedly other curriculum elements on which students at the various benchmarks would have been successful if they had been included in the assessment. Also, some students scoring below a benchmark may indeed know or understand some of the concepts that characterize a high level. Finally, describing mathematics or physics concepts or familiarity with procedures was more straightforward than describing the cognitive behavior necessary to answer the item correctly. An item may require only simple recall for a student familiar with the item's content, but necessitate problem-solving strategies from a student unfamiliar with the material. The descriptions are based on what the committee believed to be the way the great majority of advanced mathematics or physics students could be expected to respond to the item.

---

6   In addition to Robert A. Garden, the TIMSS Advanced Mathematics Coordinator, and Svein Lie, the TIMSS Physics Coordinator, committee members included Carl Angell, Wolfgang Dietrich, Liv Sissel Gronmo, Torgier Onstad, and David F. Robitaille.

Estimating Standard Errors

Because the statistics presented in this report are estimates of national performance based on samples of students—rather than on the values that could be calculated if every student in every country had answered every question—it is important to have measures for the degree of uncertainty of the estimates. The jackknife procedure was used to estimate the standard error associated with each statistic presented in this report.[7] As well as sampling error, the jackknife standard errors also include an error component due to variation among the 5 plausible values generated for each student. The use of confidence intervals (based on the standard errors) provides a way to make inferences 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 2 standard errors represents a 95 percent confidence interval for the corresponding population result.

---

7   Procedures for computing jackknifed standard errors are presented in the scaling chapter by Foy, Galia, & Li in the *TIMSS Advanced 2008 Technical Report*.

TIMSS & PIRLS
International Study Center
Lynch School of Education, Boston College