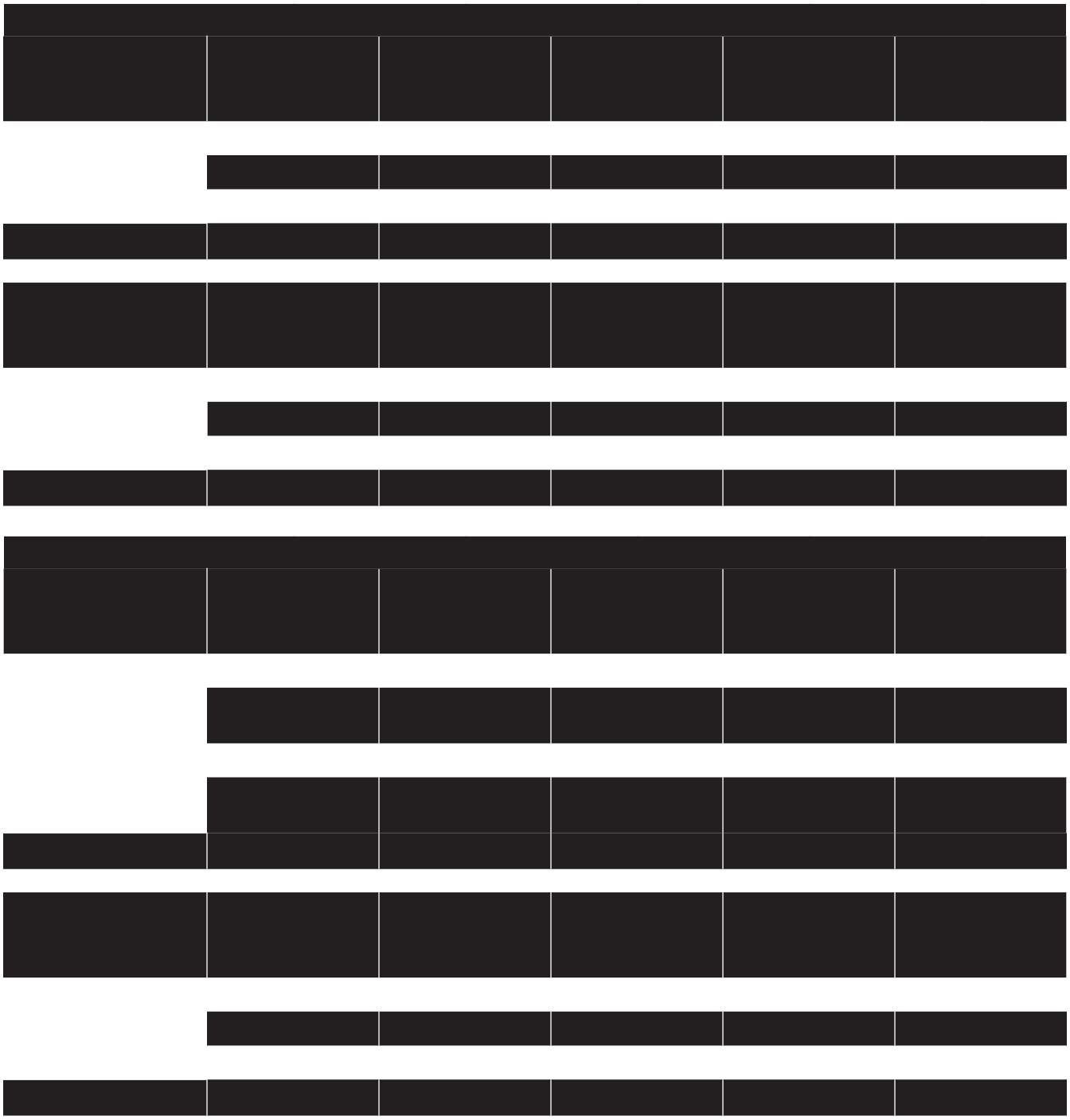
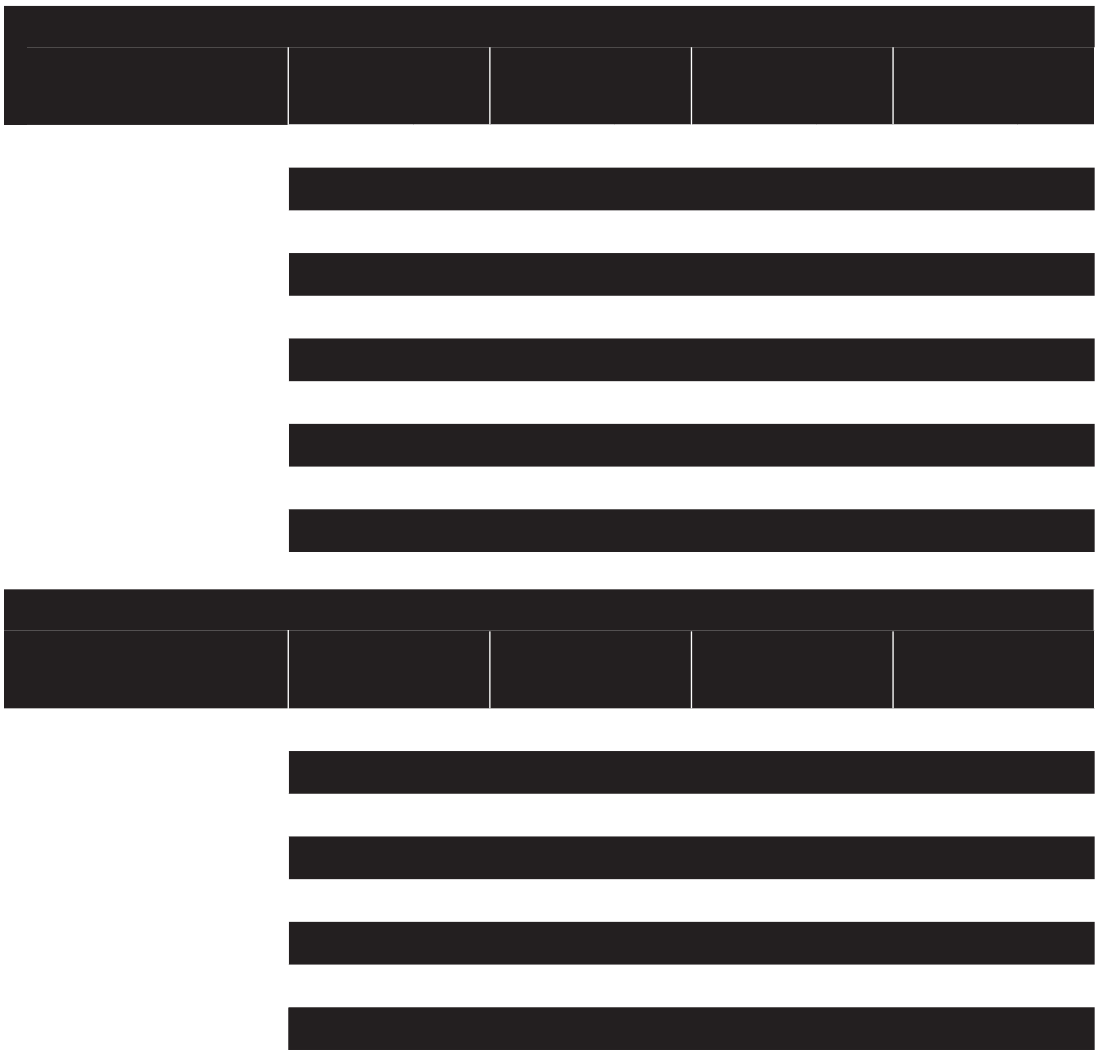
assessment and provide an outline of the assessment design. As

Exhibit A.1:



e 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. e 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. us, the total



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.

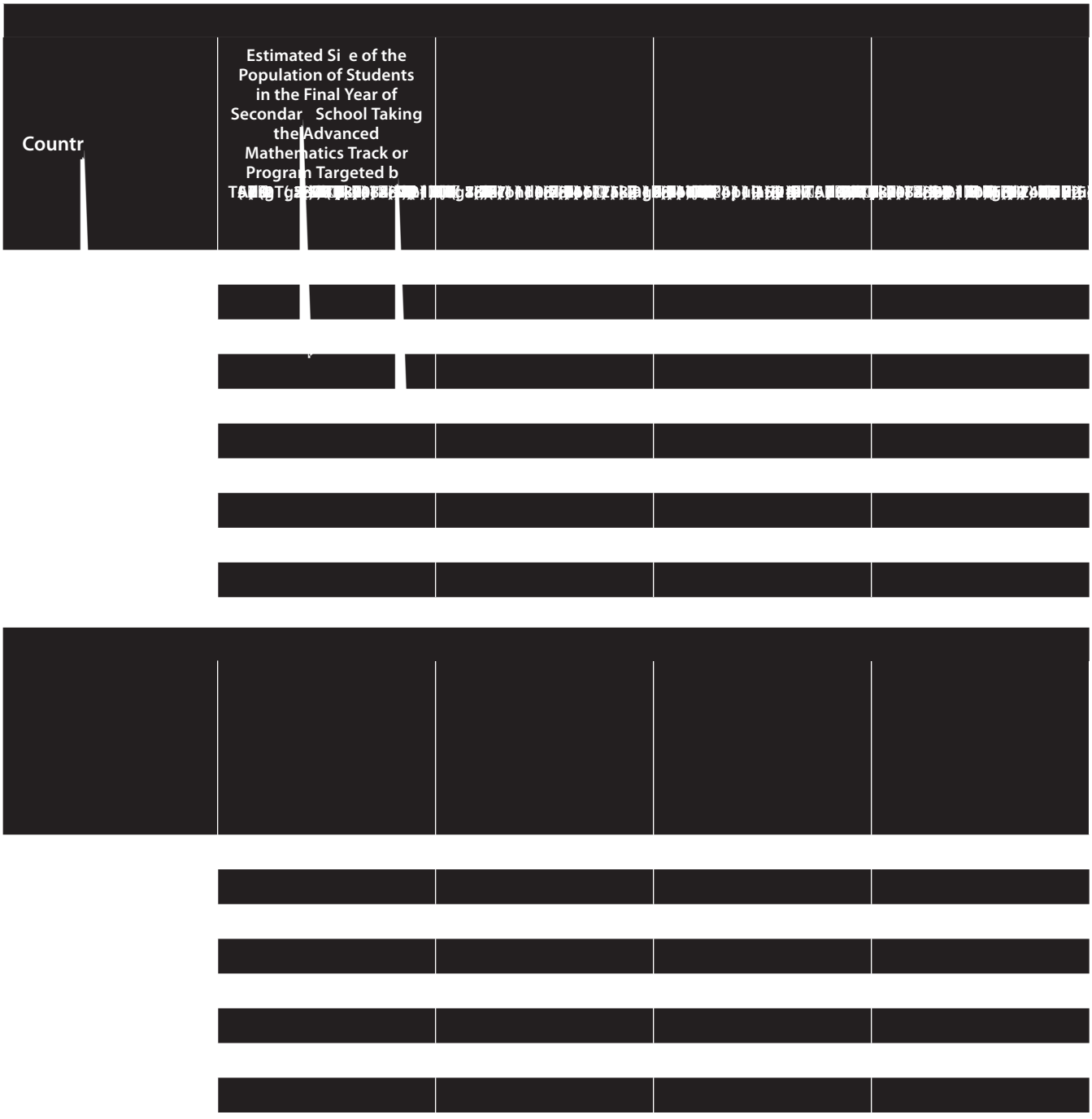
The differences in how countries organize their education systems to provide advanced courses in mathematics and physics are reflected in marked differences across countries in the proportion of the age cohort that attend such courses in the final year of secondary education. In some countries, only 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:

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

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

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



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

Most countries sampled at least 120 schools and at least one intact classroom from each school for each population. is 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. and A. present achieved sample sizes for schools and students, respectively. Exhibit A. 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 a er including replacement schools. e 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 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 teden

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

Countr	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					
Italy	100	92	88	3	91
Lebanon					

Exhibit A.6:

Countr	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						
Italy	97%	99%	100%	96%	93%	95%
Lebanon						
Netherlands	77%	84%	100%	92%	71%	77%
Norway						
Philippines	98%	98%	100%	96%	95%	95%
Russian Federation						
Slovenia	96%	96%	100%	85%	81%	81%
Sweden						

Countr	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						
Italy	100%	100%	100%	97%	97%	97%
Lebanon						
Netherlands	73%	87%	100%	90%	65%	78%
Norway						
Russian Federation	100%	100%	100%	97%	97%	97%
Slovenia						
Sweden	97%	97%	100%	92%	89%	89%

Exhibit A.7:

Countr	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										
Slovenia	12	12	18.8	18.9	1.3%	6.0%	40.5%	75.4%	81.4%	42.4%
Sweden										

Countr	Years of Formal Schooling*		Average Age at Time of Testing		Exclusion Rates		Ph sics 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										
Slovenia	12	12	18.7	18.8	0.5%	6.0%	7.5%	38.6%	67.1%	43.0%
Sweden										

* 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 di□erent from the figure reported in 1995.

and schools; manuals; and scoring guides for constructed-response items, where necessary. Veri ers documented their suggestions, and the National Research Coordinators were responsible for reviewing the suggestions and revising the instruments. e veri ed instruments were used to generate the booklets and questionnaires in their nal form and these were submitted to the TIMSS & PIRLS International Study Center for international layout veri cation. Participants who tested in English also were required to go through the veri cation 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 Technical Repo*

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 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

and 97 percent for physics. Country-by-country results are provided in the *TIMSS Advanced , , , 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 pro cient 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, e ed (r u


Test Reliabilit

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 sta 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. ese item-by-item descriptions were then summarized by the committee members to provide the more general statements of achievement at each of the benchmarks.

e item-by-item descriptions and further details about the analysis can be found in the *TIMSS Advanced Technical Report.*

The descriptions of achievement at the benchmarks are based solely on student performance on the TIMSS Advanced items and do not purport to be comprehensive. ere 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. e 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. e 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 re ects the uncertainty associated with the sample estimates. An estimated sample statistic plus or minus 2 standard errors represents a 95 percent con dence 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*.