Chapter 1

Overview of TIMSS 2003

Michael O. Martin and Ina V.S. Mullis

1.1 Introduction

Since pioneering cross-national studies of educational achievement with the First International Mathematics Study (FIMS) in 1964, the International Association for the Evaluation of Educational Achievement (IEA) has conducted almost 20 studies of student achievement in the curricular areas of mathematics, science, language, civics, and reading. The Third International Mathematics and Science Study (TIMSS) in 1994-1995 was the largest and most complex IEA study ever conducted, including both mathematics and science at third and fourth grades, seventh and eighth grades, and the final year of secondary school.

In 1999, TIMSS (now renamed the Trends in International Mathematics and Science Study) again assessed eighth-grade students in both mathematics and science to measure trends in student achievement since 1995. Also, 1999 represented four years since the first TIMSS, and the population of students originally assessed as fourth-graders had advanced to the eighth grade. Thus, TIMSS 1999 also provided information about whether the relative performance of these students had changed in the intervening years.

TIMSS 2003, the third data collection in the TIMSS cycle of studies, was

1.2 Participants in TIMSS 2003

Exhibit 1.1 lists all the countries that have participated in TIMSS in 1995, 1999, or 2003 at fourth or eighth grade. In all, 67 countries have participated in TIMSS at one time or another. Of the 49 countries that participated in TIMSS 2003, 48 participated at the eighth grade and 26 at the fourth grade. Yemen participated at the fourth but not the eighth grade. The exhibit shows that at the eighth grade 23 countries also participated in TIMSS 1995 and TIMSS 1999. For these participants, trend data across three points in time are available. Eleven countries participated in TIMSS 2003 and TIMSS 1999 only, while three countries participated in TIMSS 2003 and TIMSS 1995. These countries have trend data for two points in time. Of the 12 new countries participating in the study, 11 participated at eighth grade and 2 at the fourth grade. Of the 26 countries participating in TIMSS 2003 at the fourth grade, 16 also participated in 1995, providing data at two points in time.

Following the success of the TIMSS 1999 benchmarking initiative in the United States,[1] in which 13 states and 14 school districts or district consortia administered the TIMSS assessment and compared their students' achievement to student achievement world wide, TIMSS 2003 included an international benchmarking program, whereby regions of countries could participate in the study to compare to international standards. TIMSS 2003 included four benchmarking participants at the eighth grade: the Basque Country of Spain, the U.S. state of Indiana, and the Canadian provinces of Ontario and Quebec. Indiana, Ontario, and Quebec participated also at the fourth grade. Having also participated in 1999, Indiana has data at two points in time at eighth grade. Ontario and Quebec participated also in 1995 and 1999, and so have trend data across three points in time at both grade levels.

1.3 Student Populations

TIMSS 2003 had as its intended target population all students at the end of their eighth and fourth years of formal schooling in the participating countries. However, for comparability with previous TIMSS assessments, the formal definition for the eighth-grade population specified all students enrolled in the upper of the two adjacent grades that contained the largest proportion of 13-year-old students at the time of testing. This grade level was intended to represent eight years of schooling, counting from the first year of primary or elfi

four years of schooling, counting from the first year of primary or elementary schooling, and was the fourth grade in most countries.

1.4 Assessment Dates

TIMSS 2003 was administered near the end of the school year in each country. In countries in the Southern Hemisphere (where the school year typically ends in November or December) the assessment was conducted in October or November 2002. In the Northern Hemisphere, the school year typically ends in June; so in these countries the assessment was conducted in April, May, or June 2003.

1.5 Study Management and Organization

TIMSS 2003 was conducted under the auspices of the IEA. The study was directed by Michael O. Martin and Ina V.S. Mullis of the TIMSS & PIRLS International Study Center at Boston College, Lynch School of Education, where they also direct IEA's Progress in International Reading Literacy Study (PIRLS). The International Study Center was responsible for the design, development, and implementation of the study – including developing the assessment framework, assessment instruments, and survey procedures; ensuring quality in data collection; and analyzing and reporting the study results. Staff at the International Study Center worked closely with the organiza-

Exhibit 1.1 **Countries Participating in TIMSS 2003, 1999, and 1995** *(...Continued)*

Countries	Grade 8			Grade 4	
	2003	1999	1995	2003	1995
Jlol``l	●	●		●	
Kbaeboi^kap	●	●	●	●	●
Kbt Wb^i^ka	●	●	●	●	●
Klot^v	●		●	●	●
M^ibpafkf^k K^afli >rqe+	●				
Mefifmmfkbp	●	●		●	
Mloqrd^i			●		●
Oli^kf^	●	●	●		
Orppf^k Cbabo^qflk	●	●	●	●	
P^raf>o^f^	●				
P`lqi^ka	●		●	●	●
Pbof^	●				
Pfkd^mlob	●	●	●	●	●
Pils^h Obmr_if`	●	●	●		
Pilsbkf^	●	●	●	●	●
Plrqe >cof`^	●	●	●		
Pm^fk			●		
Ptbabk	●		●		
Ptfqwboi^ka			●		
Pvof^k >o^_ Obmr_if`''	●				
Qe^fi^ka		●	●		●
Orkfpf^	●	●		●	
Qrohbv		●			
Rkfqba Pq^qbp	●	●	●	●	●
Vbibk''				●	
Benchmarking Participants					
?^pnrb@lrkqov) Pm^fk	●				
Fkaf^k^Pq^qb) RP	●	●		●	
Lkq^oflMolsfk`b) @^k+'''	●	●	●	●	●
Nrb_b`Molsfk`b) @^k+'''	●	●	●	●	●

' >odbkqfk^ ^a j fkfpqboba qeb QFJPP /--0 a^q^ `liib`qflk lkb vb^o i^qb) ^ka afa klq p`lob ^ka mol`bpp fqp a^q^ fk qf j b clo fk`ir*pflk fk qefp obmloq+

''?b`^rpb qeb `e^o^`qbofpqf`p lc qebfo p^ j mibp ^ob klq `l j mibqbiv hklt k) ^`efbsb j bkq a^q^ clo Pvof^k >o^_ Obmr_if` ^ka Vb j bk ^ob mobpbkqba fk >mmbkafu C lc qeb Fkqbok^qflk^i obmloqp+

'''Lkq^ofl ^ka Nrb_b` m^oqf`fm^qba fk QFJPP .666 ^ka .662 ^p m^oq lc @^k^a^+

Each participating country appointed a National Research Coordinator (NRC) and a national center responsible for all aspects of TIMSS 2003 within that country. The TIMSS & PIRLS International Study Center organized meetings of the NRCs several times a year to review study materials and procedures, and to provide training in student sampling, constructed-response item scoring, and data entry and database construction.

ematics and science educators nominated by participating countries to advise on subject-matter issues in the assessment. Committee members also helped to develop tasks and items to assess problem solving and scientific inquiry.

Participating countries field-tested the items with representative samples of students, and all of the potential new items were again reviewed by the Science and Mathematics Item Review Committee as well as by NRCs. The resulting TIMSS 2003 eighth-grade assessment contained 383 items, 194 in mathematics and 189 in science. The fourth grade assessment contained 313 items, 161 in mathematics and 152 in science.

Between one-third and two-fifths of the items at each grade level were in constructed-response format, requiring students to generate and write their own answers. Some constructed-response questions asked for short answers while others required extended responses with students showing their work or providing explanations for their answers. The remaining questions used a multiple-choice format. In scoring the items, correct answers to most questions were worth one point. However, responses to some constructed-response questions (particularly those requiring extended responses) were evaluated for partial credit, with a fully correct answer being awarded two points. The total number of score points available for analysis thus somewhat exceeds the number of items.

Not all of the items in the TIMSS 2003 assessment were newly developed for 2003. To ensure reliable measurement of trends over time, the assessment included also items that had been used in the 1995 and 1999 assessments. For example, of the 426 score points available in the entire 2003 mathematics and science assessment, 47 came from items used also in 1995, 102 from items used also in 1999, and 267 from items used for the first time in 2003. At fourth grade, 70 score points came from 1995 items, and the remaining 267 from new 2003 items.

Every effort was made eeMrt i n] e

1.9 Background Questionnaires

By gathering information about students' educational experiences together with their mathematics and science achievement on the TIMSS assessment, it is possible to identify factors or combinations of factors related to high achievement. As in previous assessments, TIMSS in 2003 administered a broad array of questionnaires to collect data on the educational context for student

been administered in compliance with international procedures, including the activities before the testing session, those during testing, and the

national data sets and appropriate linking among the many student, teacher, and school data files.

Throughout the process, the TIMSS 2003 data were checked and double-checked by the IEA Data Processing Center, the International Study Center, and the national centers. The national centers were contacted regularly and given multiple opportunities to review the data for their countries. In conjunction with the IEA Data Processing Center, the International Study

samples of students have to be linked. This being the case, IRT methodology was preferred by TIMSS for developing comparable estimates of performance

and the low benchmark was 400. Although the fourth- and eighth-grade scales are different, the same benchmark points were used at both grades. To enhance this reporting approach, TIMSS conducted a scale anchoring analysis to describe achievement of students at those four points on the scales. Scale anchoring is a way of describing students' performance at different points on a scale in terms of what they know and can do. It involves a statistical component, in which items that discriminate between successive points on the scale are identified, and a judgmental component, in which subject-matter experts examine the items and generalize to students' knowledge and understandings. Complementing this approach further, the TIMSS 2003 International Reports present examples of mathematics and science items that anchor at each of the benchmarks, and display student performance in each country on the example items.

TIMSS 2003 collected a wide array of information about the homes, schools, classrooms, and teachers of the participating students, as well as about the mathematics and science curriculum in each country. The TIMSS 2003 International Reports summarize much of this information, combining data into composite indices showing an association with achievement where appropriate. In particular, student mathematics and science achievement is described in relation to characteristics of the home, curriculum coverage, classroom instruction, and school environment.

Because the statistics presented in the international reports are estimates of national performance based on samples of students, rather than the values that could be calculated if every student in every country had 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 this report. The jackknife standard errors also include an error component due to variation among the five 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

References

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., O'Connor, K.M., Chrostowski, S.J., Gregory, K.D., Garden, R.A., and Smith, T.A. (2001), *J^qeb i ^qf`p ?bk`e* j ^ohfkd Obmloq QF J PP .666 ¬ Bfdeqe Do^ab7 >`efbsb j bkq clo R+P+ Pq^qbp ^ka Afpqof`qp fk ^k Fkqbok^qflk^i @lkqbuq*. Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Smith, T.A., Garden, R.A., Gregory, K.D., Gonzalez, E.J., Chrostowski, S.J, and O'Connor, K.M. (2003), *QF J PP >ppbpp i bkq Co^ j b t lohp ^ka Pmb`f^a`^qflkp /--0 %/ka Bafqflk&*, Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., and Chrostowski, S.J. (2004), *QF J PP /--0 Fkqbok^qflk^i J ^qeb j ^qf`p Obmloq7 Cfkafkdp col j FB>$p Qobkap fk Fkqbo* k^qflk^i J ^qeb j ^qf`p ^ka P`fbk`b Pqrav ^q qeb Clroqe ^ka Bfdeqe Do^abp*. Chestnut Hill, MA: Boston College.

Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., O'Connor, K.M., Chrostowski, S.J., Gregory, K.D., Smith, T.A., and Garden, R.A. (2001), *P`fbk`b ?bk`e i ^oh* fkd Obmloq QF J PP .666 ¬ Bfdeqe Do^ab7 >`efbsb j bkq clo R+P+ Pq^qbp ^ka Afpqof`qp fk ^k Fkqbok^qflk^i @lkqbuq*. Chestnut Hill, MA: Boston College.

Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., and Chrostowski, S.J. (2004), *QF J PP /--0 Fkqbok^qflk^i P`fbk`b Obmloq7 Cfkafkdp col j FB>$p Qobkap fk Fkqbok^qflk^i J ^qeb j ^qf`p ^ka P`fbk`b Pqrav ^q qeb Clroqe ^ka Bfdeqe Do^abp*. Chestnut Hill, MA: Boston College.