Paper Commissioned by the National Assessment Governing Board

Background Questions in TIMSS and PIRLS: An Overview

Ina V.S. Mullis

August 30, 2002

TIMSS and PIRLS International Study Center
Department of Educational Research, Measurement, and Evaluation
Lynch School of Education, Boston College

Background Questions in TIMSS and PIRLS: An Overview

Ina V.S. Mullis

INTRODUCTION TO TIMSS AND PIRLS

Both TIMSS (the Trends in International Mathematics and Science Study) and PIRLS (the Progress in International Reading Literacy Study) are research initiatives of the International Association for the Evaluation of Educational Achievement (IEA). IEA was founded in 1959 for the purpose of conducting comparative studies focusing on educational policies and practices in countries around the world. Headed by a permanent Secretariat in Amsterdam, the Netherlands, and supported by a Data Processing Center in Hamburg, Germany, IEA's membership includes about 70 countries. Two of its major studies, TIMSS and PIRLS, are directed by IEA's International Study Center at Boston College.

Torsten Husen, the founding chair of IEA, has said that IEA studies use the world as an "educational laboratory" within which the strengths and the weaknesses of educational practices can be assessed. In contrast to the United States, many countries have national or system-wide policies governing educational variables such as curriculum, teaching strategies, years of compulsory schooling, and tracking of students. Without much differentiation in the approaches used within a country, it is difficult to estimate the effectiveness of various policies and practices on educational outcomes. Comparisons across countries provide an opportunity to examine the impact on achievement of different educational approaches and additional insight into ones' own educational system. Practices that have become commonplace and accepted through repetition can be reconsidered and re-evaluated through the lens of alternative organizations or approaches used by others.

IEA has been measuring student achievement and collecting contextual information to facilitate student learning in mathematics and science for nearly 40 years. Mathematics first was assessed in 1964 and then again 1980-82, while science was assessed in 1970-71 and 1983-84. In 1990, the IEA General Assembly determined to assess mathematics and science together on a regular basis every four years to measure trends in student performance. The original TIMSS (the Third International mathematics and Science Study) was conducted in 1995, TIMSS-Repeat in 1999, and now TIMSS 2003 is underway (renamed the Trends in International Mathematics and Study). Conducted at the fourth and eighth grades, more than 50 countries are scheduled to participate in TIMSS 2003.

PIRLS was established to provide countries with information about student's achievement in the core curriculum area of reading to complement the mathematics and science data provided by TIMSS. PIRLS was conducted in 2001 to provide a 10-year trend back to IEA's Reading Literacy Study conducted in 1991. Beyond that, however, PIRLS is intended to be the first in a continuing cycle of regularly conducted international reading assessments. Thirty-five countries participated in PIRLS, which assesses children at the fourth grade, a point where students often have made the transition from learning to read and are now reading to learn.

PURPOSES AND VALUE OF COLLECTING INFORMATION ABOUT THE CONTEXTS FOR LEARNING

In order to understand levels of educational outcomes and differences in those levels for different students, it is important to understand the factors that influence student's educational experiences. These factors can be considered across an entire array of contexts or "settings" for learning from the student's home and community to the classroom and general social matrix in which schools are organized.

For any large-scale assessment such as TIMSS, PIRLS, or NAEP, a primary reason for collecting background information is to be able to describe the students being assessed. For example, background information is needed to fulfill NAEP's statutory requirement to report and analyze achievement data, whenever feasible, disaggregated by race or ethnicity, gender, socioeconomic status, disability, and limited English proficiency. Historically, one of NAEP's most important contributions has been documenting the performance gaps between advantaged and disadvantaged groups of students and how these discrepancies change over time.

It should be noted that to estimate accurately subpopulation results, NAEP uses each student's responses to the background questions themselves as well as the item parameter estimates and the student's responses to the cognitive items.[1]

Background information also is important to evaluate the potential for bias resulting from non-participation. That is, did the students absent or refusing to participate in the assessment appear to differ in major ways from the students that did participate, such that performance may have been artificially increased or decreased?

Another important reason for collecting background information is to inform educational policy by collecting descriptive information about the contexts for learning, sometimes described as opportunities to learn. Broadly, this involves the content that is officially specified in the curriculum, whether and how that content actually is taught, students propensity to learn, as well as a host of home and school supports that can enhance learning opportunities and the learning process. The assessment can provide information

[1] Sometimes referenced as conditioning, for more information about this methodology see Mislevy, R.J. (1991) Randomization-based inference about latent variables from complex samples. *Psychometrika, 56,* 177-196.

about whether schools and parents are implementing the practices most likely to help students learn, and, if not, the strategies, activities, and approaches that are being used. That is, background data can provide a picture of what is being done and how that coincides with what is thought to work best.

In conjunction with the descriptions of students, background information about educational settings and experiences also can reveal striking differences in how important aspects of education and educational resources are distributed among different groups of students. For example, do minority students have less access to science laboratory equipment than their counterparts? Do girls take less rigorous mathematics courses than boys?

DATA COLLECTION STRATEGIES

Considering the enormity and complexity of a country's educational system, especially one as diverse in approaches and population as the United States, it is a substantial task to even to develop a thoughtful organization scheme or framework for determining the most salient aspects of the educational process to study. What are the most important characteristics of the students themselves? What is it that students were supposed to have learned according to official policy statements? Did the curriculum include the material being assessed? If so, were students ever actually taught the material? Were the teachers well prepared to teach the topics? Did they use best practice? Did they have the necessary resources?

Based on the framework, decisions next need to be made about the ways to gather the background information. Across TIMSS, PIRLS, and NAEP, a number of innovative methods have been used to collect information about the contexts for student's learning. All have relied extensively on questionnaires administered to students, their teachers, and their school principals. Yet, it is possible to collect some information about school resources and student characteristics (e.g., age, gender) from school records. More interestingly, though are attempts at in-depths studies of materials and face-to-face interviews. For example,

TIMSS has used:

- An analysis of curriculum guides and instructional materials, primarily textbooks
- Videos of mathematics and science lessons
- Case studies based on interviews of school administrators, parents, students, and teachers
- Profiles or descriptions of the decision-making structure, organization, funding, and goals of the education systems (known as encyclopedias)
- National Curriculum Questionnaires to describe the topics students are supposed to have studied (by ability level if applicable).

PIRLS has included:

- A Parent Questionnaire addressed to parents or primary caregivers to investigate home literacy resources and parent's reading habits as well as their educational and economic backgrounds
- A pilot study using computers to collect information about student's familiarity with accessing and processing electronic text.

NAEP has:

- Collected examples of student work and assignments in reading and writing
- Interviewed students about their reading habits and instruction.

The drawback in using more innovative, in-depth approaches to background data is, of course, the expense both in terms of dollars and the time of the respondents as well as the feasibility in terms of even identifying and gaining the cooperation of the participants.

It needs to be emphasized, however, that the more traditional and less expensive methods of using student, teacher, and school questionnaires also are plagued by the same concerns about how to encourage high levels of participation. More and more, assessment respondents, especially teachers, are refusing to expend very much effort to provide background information. In some cases incentives can be used, but these can become very expensive and problematic, too. The approach more typically used is to try and ascertain the extent of the burden that will be tolerated by most respondents and tailor the background information to fit into that amount of effort.

As illustrated by the background efforts associated with TIMSS 2003 and PIRLS, the approach of having "the size of your eyes match the size of your stomach" severely curtails the extent of the background data collection effort.

Both TIMSS 2003 and PIRLS include:

- Student Questionnaire of 15 to 30 minutes for every student assessed
- Teacher Questionnaire of 30 minutes for the teacher of every student assessed
- School Questionnaire of 30 minutes for the principal or head administrator of every school with students assessed

In addition, PIRLS includes:

- Parent Questionnaire of 10 to 15 minutes for the parents of every student assessed
- Profile of context, organization, and goals of reading instruction (15 pages per country)

In addition to Student, Teacher, and School Questionnaires, TIMSS includes:

- Curriculum Questionnaire to be completed for mathematics (from an hour to a day depending on complexity and diversity of curriculum)

- Curriculum Questionnaire to be completed for science (from an hour to a day depending on complexity and diversity of curriculum)

From the perspective that typically only several questions can be asked per minute, the topics and information requested need to be carefully prioritized. The temptation often is to add just a few more questions, but it is best to continually envision being the busy person asked to donate half an hour of valuable time and keep the questionnaires as painless to complete as possible.

SETTING PRIORITIES AMONG BACKGROUND AREAS

The next section of the paper describes the TIMSS and PIRLS experiences in collecting different types of background information, including the importance attached to each, the method used to collect the data, and the role of the data in analysis and reporting. The seven educational areas discussed are:

- Curriculum
- Student Characteristics and Experiences
- Home-School Connection/Interaction
- School Environment
- Teacher Characteristics
- Classroom Resources
- Instructional Practices.

CURRICULUM

IEA studies, particularly in mathematics and science, are grounded in the attempt to assess what students have learned through their experiences in school. The conceptual framework for the TIMSS background data collection effort highlights the fact that a given curriculum may be viewed from three major perspectives—curriculum as intended, as implemented, and as attained.

For TIMSS, the **intended curriculum** consists of the mathematics and the science goals defined at the highest applicable level. For most countries, there is a national curriculum. Notable exceptions, of course, include the United States as well as Canada, Australia, and Germany where curriculum is generally set at the state (or provincial/canton) level. The intended curriculum is described in policy statements, regulations, curriculum guides, frameworks, and other official documents. It generally covers the content students should learn, the skills they should develop, and sometimes the attitudes to be developed. Depending on the country (or state), the spirit of the intended curriculum may be

reflected in teacher training courses, textbooks, curriculum guides or modules, resources (e.g., laboratory equipment, computers), and examinations.

In the original TIMSS conducted in 1995, the intended curriculum was studied through a very detailed and labor intensive analysis of prescribed textbooks, curriculum guides, examinations, and official policy statements. For TIMSS 1999 and TIMSS 2003, a high level of country participation depended on reducing the burden substantially, essentially to a single questionnaire. For example, for TIMSS 2003 each country will complete a questionnaire for each subject and grade (a total of four maximum – grades 4 and 8 in mathematics and science). The questionnaire is based on the content and process topics specified in the assessment framework, with each country providing information about the percentage of students that should have been taught the topic and the grades at which the topic was covered. The questionnaire provides very interesting information about cross-country variations in curriculum and is given high priority. Obviously, if a topic is not even in the intended curriculum then students are not likely to have learned it, and the poor performance can be accepted or the policy changed.

Method of Data Collection in TIMSS 2003 and PIRLS 2001: Curriculum Questionnaires completed by each country for TIMSS 2003. Not included for PIRLS 2001.

Relationship with Student Achievement: Generally, countries with more rigorous mathematics and science curricula have higher performance.

Priority: High for TIMSS 2003.

Collecting data on the **implemented curriculum**—what topics teachers actually teach—is a more complicated matter. TIMSS relies on teacher reports to provide this vital information. However, teachers of the assessed students may not be teaching particular topics in the grade in question because students have learned the material in an earlier grade. Or perhaps the teacher is not responsible for teaching the topic, because he or she teaches a special course such as biology so does not cover physics.

On the other hand, teachers may not cover a topic because it is in the curriculum at a later grade or simply not included at all. If teachers do cover a topic, the duration or difficulty can vary. Also, many things can affect curriculum coverage, such as resources, school climate, teacher's preparation, and classroom characteristics. Still, in as much as is possible, TIMSS does ask teachers about whether they have taught the various topics specified in the framework. This section of the teacher questionnaire is given high priority, but with less enthusiasm because the burden is high. The teacher protests are strong; the response rates unacceptably low in some countries, and the data difficult to interpret in some instances. Nevertheless, it cannot be assumed that students will learn everything they have been taught so an effort is made to help determine whether low performance is a result of omission (teachers report no coverage) or of ineffective teaching/learning.

In PIRLS 2001, school principals are asked at which grade a set of 12 reading skills and strategies receive a major emphasis in instruction. The reading skills and strategies

ranged from knowing letters of the alphabet to describing style and structure of text. Because reading curricula seem to be similar across primary schools, this approach was relatively successful.

Method of Data Collection in TIMSS 2003 and PIRLS 2001: Teacher Mathematics and Science Questionnaires for TIMSS 2003. School Questionnaire for PIRLS 2001.

Relationship with Student Achievement: Difficult to disentangle, but interesting to know what topics teachers and schools report emphasizing.

Priority: High for TIMSS 2003 and PIRLS 2001.

STUDENT CHARACTERISTICS AND EXPERIENCES

To meet basic reporting requirements, it is vital to collect fundamental information about the **demographic characteristics** of individual students, including gender, race or ethnicity, disabilities, limited English proficiency, and socioeconomic status. Related variables include parents' education level, items in the home (e.g., books, computer), and sometimes language spoken in the home. Geographic location (e.g., region of the country or state), type of community (e.g., rural, urban), and type of school (e.g., public, private) also may be considered here.

In various studies, including NAEP, and at various times, this information has been asked of the student's themselves (gender, race/ethnicity, socioeconomic indicators), obtained from school records (gender, race/ethnicity, disabilities, limited English proficiency), provided by observation (gender and race/ethnicity), and collected from parents (socioeconomic level). Considering the importance of this information, it is unfortunate that data about such demographic variables is very difficult to collect accurately and interpret appropriately. The data provided may not reflect the intended question for reasons having to do with definitions, perceptions, and respondents simply not knowing the information. Enormous amounts of time and resources have been and will continue to be devoted to collecting valid and reliable background data in this area.

Method of Data Collection in TIMSS 2003 and PIRLS 2001: School records and Student Questionnaires for TIMSS 2003. School records, Parent and Student Questionnaires for PIRLS 2001.

Relationship with Student Achievement: Strongly related to student achievement.

Priority: High for TIMSS 2003 and PIRLS 2001.

Even though it is clear that students' effort, attitudes, and personal interests can impact achievement, collecting information about **students' attitudes** may or may not be given high priority depending on the subject area being assessed. For example, student's confidence in their ability to learn and do mathematics is highly related to their achievement. TIMSS 2003 has two attitude scales used for both mathematics and science. One concerns student's self-concept in each subject, respectively. The other concerns student's value of learning and doing well in each subject respectively.

PIRLS 2001 includes two student attitude scales, one on students' positive attitudes towards reading and one on their reading self-concept.

Method of Data Collection in TIMSS 2003 and PIRLS 2001: Student Questionnaires for TIMSS 2003 and PIRLS 2001.

Relationship with Student Achievement: Strongly related to student achievement.

Priority: High for TIMSS 2003 and PIRLS 2001.

Similarly, **students' habits and activities** may have particular importance for a given subject area, for example, the amount of out-of-school reading most likely would be of high interest in connection with the reading assessment. In general, though, collecting information about students' attitudes and out-of-school experiences and activities seems of lower priority than some other contextual areas. One problem is related to the questionable validity of responses, since there is no readily feasible way to check accuracy. The question sets required to establish attitude scales or long lists about all the activities students might have done simply add to burden and often have little explanatory or policy relevant impact. Nevertheless, TIMSS 2003 asked students about a variety of nine activities they can spend time on out of school, including television and videos, computer games, talk with friends, jobs at home, paid job, sports, reading for enjoyment, internet, and homework. PIRLS 2001 took a different approach, focusing only on activities related to reading. It asked students about a set of seven literacy activities, e.g., talking with friends about reading and reading for enjoyment. It also asked students about what they read, e.g., newspapers, magazines, and different types of books.

Method of Data Collection in TIMSS 2003 and PIRLS 2001: Student Questionnaire for TIMSS 2003 and PIRLS 2001.

Relationship with Student Achievement: Variable for TIMSS but literacy activities and amount of out of school reading positively related to achievement in PIRLS.

Priority: Moderate for TIMSS 2003, high for PIRLS 2001.

HOME-SCHOOL CONNECTION/INTERACTION

As part of the background data collection effort, it is possible to gather information about **home support for student academic** achievement and whether parents are engaging in the types of activities that encourage student learning. Especially in the early years, parent and caregiver's involvement in children's schooling may be key to academic development. In particular, involved parents can reinforce the importance of learning to read and do mathematics, encourage and monitor the completion of homework assignments, and generally express interest as well as provide praise and support. Information also can be gathered about parent/teacher conferences and how often parents participate in school activities, such as concerts, sports, and fundraising. Initial results suggest this may be an area of emerging promise and interest.

Method of Data Collection in TIMSS 2003 and PIRLS 2001: School Questionnaire for TIMSS 2003. Parent, Teacher, and School Questionnaires for PIRLS 2001.

Relationship with Student Achievement: Analyses not completed.

Priority: High for PIRLS 2001, moderate for TIMSS 2003.

SCHOOL ENVIRONMENT

TIMSS and PIRLS routinely collect background data about the availability of **school resources**, both of a general nature such as having good heating and lighting as well as related to specific assessment subjects such as having science laboratories or remedial reading specialists. While it is possible to have high academic achievement with limited resources, it is unusual, and studies in the United States indicate that schools in poor, urban areas often are in disrepair structurally. TIMSS and PIRLS collect information on school resources by asking school principals about approximately 12 to 20 different types of resources, covering aspects of facilities, materials, equipment, and staff.

Method of Data Collection in TIMSS 2003 and PIRLS 2001: School Questionnaire for TIMSS 2003 and PIRLS 2001.

Relationship with Student Achievement: Generally, more resources relate to higher achievement.

Priority: Moderate for TIMSS 2003 and PIRLS 2001.

Sadly, **school safety** is an emerging area of concern and policy relevance for TIMSS and PIRLS. Across the cycles of TIMSS and PIRLS, students' reports of incidences of being bullied and generally intimidated have increased even at the younger grades. As might be expected, students in safer environments have higher performance than students in more volatile environments. Including sets of questions about school safety in the student and school questionnaires has been given high priority in recent assessments. PIRLS 2001 included a set of six questions in the Student Questionnaire and a set of 13 questions in the School Questionnaire. TIMSS 2003 included a set of five questions in the Student Questionnaire and 13 questions in the School Questionnaire.

Method of Data Collection in TIMSS 2003 and PIRLS 2001: Student and School Questionnaires for TIMSS 2003 and PIRLS 2001.

Relationship with Student Achievement: Generally, fewer problems relate to higher performance.

Priority: High for TIMSS 2003 and PIRLS 2001.

Another area worth consideration relates to the school's **academic orientation** and goals. Research on effective schools suggests that successful schools identify, communicate, and work towards good work habits, discipline, and academic excellence. The teachers and students feel motivated and supported in their instructional and learning activities. Developing sets of questions in this area is new to TIMSS, but was included in the

School Questionnaire (set of eight questions), Teacher Questionnaires (same eight questions), and Student Questionnaire (four questions). The PIRLS 2001 School Questionnaire included five questions.

Method of Data Collection in TIMSS 2003 and PIRLS 2001: School, Teacher, and School Questionnaires for TIMSS 2003. School Questionnaire only for PIRLS 2001.

Relationship with Student Achievement: Analyses not yet completed.

Priority: Moderate for TIMSS 2003 and PIRLS 2001.

TEACHER CHARACTERISTICS

Demographic information about teachers, including gender, age, and experience usually have been included in TIMSS and PIRLS. Policy relevant issues may include the aging of the teaching force in certain geographic or subject areas, the limited number of minority teachers to serve as mentors and role models for minority children, or that urban areas have a disproportionate number of inexperienced teachers. Such information is collected via the teacher questionnaire and given relatively high priority.

Method of Data Collection in TIMSS 2003 and PIRLS 2001: Teacher Questionnaires for TIMSS 2003 and PIRLS 2001.

Relationship with Student Achievement: Weak, but of interest for policy analysis.

Priority: Moderate for TIMSS 2003 and PIRLS 2001.

The emerging area given the highest priority in TIMSS, however, relates to mathematics and science **teacher's education and professional development**. Many countries are focusing on improving teacher education and setting higher standards for certification such as passing an exam or completing a probation period. Also, the amount of teaching "out-of-field" has become a matter of concern. TIMSS 2003 has several sets of newly developed questions in the teacher questionnaires designed specifically to address teachers' formal education, certification, preparation, assignments, and professional development. School principals are asked about opportunities for professional development and teachers' qualifications.

Method of Data Collection in TIMSS 2003 and PIRLS 2001: School and Teacher Questionnaires for TIMSS 2003. Teacher Questionnaire only for PIRLS 2001.

Relationship with Student Achievement: Analyses not yet completed.

Priority: High for TIMSS 2003, moderate for PIRLS 2001.

For TIMSS 2003, **teacher attitudes** have only a modest role. The mathematics and science teachers, respectively, are asked about whether they feel their subject areas are dynamic, evolving, useful fields (set of approximately nine questions) as well about their attitudes towards particular teaching approaches (e.g., memorization).

Method of Data Collection in TIMSS 2003 and PIRLS 2001: Teacher Questionnaires for TIMSS 2003.

Relationship with Student Achievement: Weak, but of interest for policy analysis.

Priority: Moderate for TIMSS 2003.

CLASSROOM RESOURCES

Collecting extensive information about **resources** became central in PIRLS, since ensuring that students have ready access to books by way of classroom reading corners and classroom libraries is considered crucial to effective reading instruction. There also was an emphasis on availability of computers and the Internet for reading and writing. Related questions dealt with the availability of specialists to deal with students with reading difficulties. In TIMSS, teachers and principals were asked about access to calculators, computers, and the Internet, in particular. Teachers were asked about how resource shortages limit their teaching (set of eight questions).

Method of Data Collection in TIMSS 2003 and PIRLS 2001: Teacher and School Questionnaires for TIMSS 2003 and PIRLS 2001.

Relationship with Student Achievement: Analyses not yet completed for PIRLS 2001. Difficult to disentangle for TIMSS.

Priority: High for PIRLS 2001, moderate for TIMSS 2003.

INSTRUCTIONAL PRACTICES

High priority always is placed on using large-scale assessments such as TIMSS, PIRLS, and NAEP to identify instructional practices that relate to high achievement. However, this often meets with mixed success for several reasons. One problem highlighted in the TIMSS videos of mathematics classrooms has to do with teacher's attention to form over function, such as putting students in groups to do individual work or providing substantial amounts of class time for homework. Thus, the strategies deemed to be effective are reported as being used, but in actuality are not being implemented in the ways envisioned to enhance learning. Another problem relates to teachers more frequent use of particular strategies with higher- or lower-performing students. For example, mathematics teachers are more likely to use higher-order problem solving activities with higher-performing students, and drill-and-practice strategies with lower-performing students. This may serve to exacerbate performance differences, since a reinforcing, spiraling phenomena may be occurring whereby higher-performing students get more enriching activities and improved achievement, while lower-performing students receive more tedious tasks lowering their interest, motivation, and future achievement.

Interestingly, according to their reports, teachers tend to be relatively more eclectic in the use of instructional approaches than might be suggested by literature recommending particular approaches (e.g., phonics versus whole language). Finally, since views on

"what works" evolve and change over time it sometimes is difficult to be able to report timely data about best practice.

Despite these complexities, it is extremely important for large-scale studies such as TIMSS, PIRLS, and NAEP to collect information about instructional practices. The information helps to ascertain the extent to which current research recommendations are being put into practice and to capture what teachers actually are doing. For example, it seems from the TIMSS videotapes that, despite recommendations by the National Council for Teachers of Mathematics for more emphasis on problem-solving and communication, many mathematics teachers appear to be teaching as they themselves were taught. As depicted in the TIMSS videos, mathematics classes primarily involve homework review, listening to lecture-style presentations and clarifications of previously discussed material, students working problems independently, and taking tests or quizzes.

The TIMSS 2003 Teacher Questionnaires and, particularly, those for PIRLS 2001 include a number of questions about instructional practices. An important area is the amount of **instructional time** devoted to each subject area. Providing ample instructional time is a necessary but not sufficient condition for high achievement. For example, in TIMSS, the higher performing countries generally devote more instructional time to mathematics and science than the lower performing countries. On the other hand, some countries that devote the most time do not have the highest achievement. For example, the United States devotes a comparatively high amount of time to mathematics instruction, and has relatively mediocre achievement. In PIRLS 2001, the amount of instructional time needed to be asked in the context of the reading curriculum. Related questions concerned the emphasis on direct reading instruction compared to a more integrated approach of reading as part of language study and literature or even across the curriculum.

Method of Data Collection in TIMSS 2003 and PIRLS 2001: Teacher Questionnaires for TIMSS 2003 and PIRLS 2001.

Relationship with Student Achievement: Difficult to interpret for TIMSS, since it appears to depend on how effectively the time is used. Analyses not yet completed for PIRLS 2001.

Priority: High for PIRLS 2001 and TIMSS 2003.

To facilitate instruction, teachers can adopt a variety of approaches to **classroom organization**. Notwithstanding that class size can affect such decisions, whole-class, small-group, and independent work can all be effective. In reading, grouping students by ability has been a common practice. In PIRLS 2001, teachers were asked about differentiation of instruction by ability groups. Both students and teachers were asked about the frequency of instructional strategies involving individual students, small groups, or the whole class. In TIMSS 2003, the teachers were asked how often they asked students to do problems on their own, work together in small groups, or listen to whole-class lecture style presentations. The students were asked essentially the same set of questions.

Method of Data Collection in TIMSS 2003 and PIRLS 2001: Teacher and Student Questionnaires for TIMSS 2003 and PIRLS 2001.

Relationship with Student Achievement: Analyses not yet completed for PIRLS 2001. Weak for TIMSS.

Priority: High for PIRLS 2001, moderate for TIMSS 2003.

Teachers also were queried about using various types of **instructional materials**. In PIRLS 2001, the teachers were asked about how often students read different types of texts (e.g., literary, informational, and documents) and about the use of different reading materials (e.g., books, newspapers, and magazines). The reading teachers also were asked about their use of different media. In TIMSS 2003, the mathematics and science teachers are asked if they use a textbook and, if so, whether it is the primary basis for their lessons or a supplementary resource.

Method of Data Collection in TIMSS 2003 and PIRLS 2001: Teacher Questionnaires for TIMSS 2003 and PIRLS 2001.

Relationship with Student Achievement: Analyses not yet completed for PIRLS 2001. Weak for TIMSS since textbook use is almost universal.

Priority: High for PIRLS 2001, moderate for TIMSS 2003.

In PIRLS 2001, teachers were asked two sets of questions about **instructional strategies and activities**. One series of eight questions was about different ways to help students develop reading comprehension (e.g., identifying main ideas, comparing with experiences, making generalizations and inferences). Another series of eight questions concerned ways students were asked to respond to what they had read (e.g., workbooks, discussion, drawings, or plays). Students also were asked the same set of eight questions about how they responded to what they have read. In TIMSS 2003, only moderate emphasis was placed on collecting data about instructional strategies. Both students and their mathematics teachers were asked about a variety of nine approaches, some having to do with content and some with strategies (e.g., fractions, problem solving, relating material to daily lives, and asking students to explain their answers). Similarly, students and their science teachers were asked about a variety of eleven approaches, half of which were related to conducting experiments or investigations.

Method of Data Collection in TIMSS 2003 and PIRLS 2001: Teacher and Student Questionnaires for TIMSS 2003 and PIRLS 2001.

Relationship with Student Achievement: Analyses not yet completed for PIRLS 2001. Weak for TIMSS but of interest for policy analysis.

Priority: High for PIRLS 2001, moderate for TIMSS 2003.

Given its potential for consolidating student learning and extending student learning time, **homework**, including the frequency, amount, and amount of class time spent on reviewing and beginning new assignments, has received extensive attention in TIMSS.

While there has been some reduction in sets of questions about homework, it still is considered high priority and included in the student and teacher questionnaires. Students are asked about the frequency of receiving mathematics and science homework and how much time they typically spend on their homework assignments. As part of a related issue, they also are asked about extra lessons or tutoring. Teachers are asked not only about the frequency and length of homework assignments, but also about the kinds of homework they give. Additionally, teachers are asked five questions on how they follow up on homework assignments. Since reading is not uniformly taught as a separate subject at grade four, less emphasis is given to questions about homework in PIRLS 2001. However, both teachers and students were asked about the frequency and length of homework assignments involving reading.

Method of Data Collection in TIMSS 2003 and PIRLS 2001: Teacher and Student Questionnaires for TIMSS 2003 and PIRLS 2001.

Relationship with Student Achievement: In TIMSS, a modest positive relationship exists in some countries at the eighth grade. At fourth grade, because homework is often assigned for remedial purposes, the relationship is more difficult to interpret. For PIRLS 2001, the analyses have not yet been completed, but a similar phenomenon is anticipated.

Priority: Moderate for TIMSS 2003 and PIRLS 2001.

In TIMSS, extensive data has been collected about the types and frequency of **assessment** and testing used by teachers. The area is not given a high priority but always seems to survive at some level. In TIMSS 2003, teachers are asked about the frequency of testing, the item formats used in their tests (e.g., proportion of constructed response versus objective items), and their emphasis on questions assessing higher-order cognitive processing (i.e., applications, patterns, and justifications). Questions about item format are relevant to TIMSS because countries vary in their familiarity with multiple-choice items.

In PIRLS 2001, teachers were asked about the emphasis they place on diagnostic tests, classroom tests, national examinations, commercial tests, and their own professional opinions when evaluating student progress in reading. They also are asked how often they use eight different assessment approaches (e.g., multiple-choice questions, reading aloud, oral questioning, and oral reports). Finally, they are asked about using portfolios (collection of samples of students' work, a reading log, etc.). In both TIMSS 2003 and PIRLS 2001, students simply are asked how often they have a quiz or test.

Method of Data Collection in TIMSS 2003 and PIRLS 2001: Teacher and Student Questionnaires for TIMSS 2003 and PIRLS 2001.

Relationship with Student Achievement: Weak for TIMSS, although in some countries, including the United States, it appears that lower-performing students may be tested more frequently. For PIRLS 2001, the analyses have not yet been completed.

Priority: Moderate for TIMSS 2003 and PIRLS 2001.

DEVELOPING THE TIMSS AND PIRLS BACKGROUND QUESTIONNAIRES

Since IEA studies have a rich tradition of trying to understand the student contexts for learning, considerable effort generally is devoted to developing a conceptual framework for data collection and the instruments and procedures themselves.[2] The *TIMSS Assessment Frameworks and Specifications 2003* not only describes the mathematics and science content to be assessed in future assessments, but also describes the contextual factors associated with students' learning in mathematics and science that will be investigated.[3]

Developing the *TIMSS Assessment Frameworks and Specifications 2003* began by updating the *Curriculum Frameworks for Mathematics and Science*[4] used as the basis for the 1995 and 1999 assessment. This process involved widespread participation and reviews by educators around the world.

An international panel of mathematics and science education and testing experts provided guidance for the general form the assessment frameworks should take. The U.S. National Science Foundation provided support for the meetings and the work of the expert panel. Using an iterative process, successive drafts were presented for comment and review by National Research Coordinators, national committees, and expert panel members. The National Research Coordinators are responsible for implementing the study in their countries and work with the international project staff to ensure that the study is responsive to their concerns, both policy-oriented and practical.

TIMSS 2003 collects a range of information about the context for learning in mathematics and science. The contextual framework encompasses five broad areas:

- Curriculum
- Schools
- Teachers and their preparation
- Classroom activities and characteristics
- Students

---

[2] For information about the original TIMSS conceptual framework and how it was put into operation, see Robitaille, D.F. & Garden, R.A. (1996). Design of the study. In D.F. Robitaille & R.A. Garden, (Eds.), *Research Questions & Study Design: TIMSS Monograph No. 2*, Vancouver, B.C.: Pacific Educational Press.

[3] Mullis, I.V.S., Martin, M.O., Smith, T.A., Garden, R.A., Gregory, K.D., Gonzalez, E.J., Chrostowski, S.J., & O'Cnnor, K.M. (2001). *TIMSS Assessment Frameworks and Specifications 2003*. Chestnut Hill, MA: Boston College, International Study Center.

[4] Robitaille, D.F. et al. (1993). *Curriculum Frameworks for Mathematics and Science: TIMSS Monograph No.1*, Vancouver, B.C.: Pacific Educational Press.

In particular, TIMSS 2003 examines the curricular goals of the educational system and how the system is organized to attain those goals; the educational resources and facilities provided; the teaching force and how it is educated, equipped, and supported; classroom activities and characteristics; home support and involvement; and the knowledge, attitudes, and predispositions that students and teachers themselves bring to the educational enterprise. The contextual framework identifies the major characteristics of the educational and social contexts that will be studied with a view to improving student learning.

The *Framework and Specifications for PIRLS Assessment 2001*[5] is intended as a blueprint for IEA's future work in assessing reading literacy. The framework grew from a collaborative process involving many individuals and groups – notably the PIRLS Reading Development Group and the National Research Coordinators of the almost 40 countries that planned to participate in PIRLS 2001. All told, the framework underwent several iterations in response to the comments and interests of the PIRLS countries and the reading research community and embodies the ideas of many individuals and organizations around the world. Support for developing the PIRLS framework was provided by the National Center for Education Statistics of the U.S. Department of Education and the participating countries. The framework for the PIRLS questionnaires consists of a conceptual model relating reading outcomes – students' reading literacy achievements and attitudes – to home, school, and national and community contexts. The factors within the home, school, and national and community contexts addressed by PIRLS are:

- National and Community Contexts
    - Demographics and resources
    - Governance and organization of education system
    - Curriculum characteristics and policies
- Home Contexts
    - Activities fostering reading literacy
    - Language in the home
    - Home resources
    - Home-school connection
    - Students' out-of-school literacy activities
- School Contexts
    - School environment and resources

[5] Campbell, J.R., Kelly, D.L., Mullis, I.V.S., Martin, M.O., & Sainsbury, M. (2001). *Framework and Specifications for PIRLS Assessment 2001*. Chestnut Hill, MA: Boston College, International Study Center.

- Teacher training and preparation
- Classroom environment and structure
- Instructional strategies and activities
- Instructional materials and technology

To provide special assistance in developing the background questionnaires, special committees were established for both TIMSS and PIRLS. The TIMSS Questionnaire Item Review Committee consisted of eight policy and assessment specialists representing eight different participating countries. Similarly, the PIRLS Questionnaire Development Group, comprising six of the PIRLS National Research Coordinators, was instrumental in the design of the PIRLS questionnaires.

The TIMSS background questionnaires were developed through a collaborative process involving the TIMSS International Study Center, the National Research Coordinators, the Questionnaire Item Review Committee, and the IEA Data Processing Center. The process began with the National Research Coordinators of the more than 50 countries planning to participate in TIMSS 2003 thoroughly reviewing the questionnaires used in 1995 and 1999. Since many of the National Research Coordinators had participated in previous iterations of TIMSS, they were very familiar with the utility and analytic potential of the many questions in the background instruments. A number of questions were deleted and then a serious discussion held concerning emerging areas of policy-relevant interest. The background questionnaire development process included a series of reviews of draft instruments, a field test of the questionnaires in 38 countries, a review of field-test data, and revision of the field-test questionnaires. The final questionnaires were reviewed by the Questionnaire Item Review Committee and the national research coordinators.

The TIMSS 2003 questionnaires were developed with a focus on feasibility and increased utility. There was unanimous sentiment that response burden should be kept to the absolute minimum, in order to maximize participation and data quality. The instruments were designed to collect information about the most important factors related to students' contexts for learning mathematics and science at fourth and eighth grades.

The PIRLS questionnaires were developed through a collaborative process involving the PIRLS International Study Center, the National Research Coordinators, the Questionnaire Development Group, the Reading Development Group, and the IEA Data Processing Center. The process began with an extensive review of the literature including the questionnaires and results from IEA's Reading Literacy Study conducted in 1991. The process included a series of reviews of draft instruments, a field test of five questionnaires in 30 countries, a review of field-test data, and revision of the field-test questionnaires. The final questionnaires were reviewed by all the committees and the national research coordinators.

In developing the PIRLS questionnaires, the aim was to create instruments that could be used to collect reliable information related to children's reading literacy achievement, as outlined in the framework, without unduly burdening students and schools. Altogether,

the instruments were intended to provide a picture of children's experiences from early language and literacy development to the time of the PIRLS assessment (fourth grade in most countries).

REPORTING TIMSS AND PIRLS BACKGROUND DATA

To increase the reliability and validity of the background measures, an effort was made throughout developing the TIMSS and PIRLS background questionnaires towards creating sets of items that would provide scales for reporting. A concerted effort was made to build on the procedures used in TIMSS 1999 to create summary indices from background data.

In an effort to summarize the information obtained from the background questionnaires concisely and focus attention on educationally-relevant support and practice, TIMSS combined information to form indices that were more global and reliable than the component questions. For example, indices were formed for students' home educational resources, students' attitudes towards mathematics or science, teachers' emphasis on reasoning and problem solving, teachers' confidence in their preparation to teach mathematics or science, and availability of school resources for mathematics or science instruction.[6]

According to the responses of students, their teachers, or their schools to particular background questions, students were placed in a "high," "medium," or "low" category for the index, with a high level being set so that it corresponded to conditions or activities generally associated with higher academic achievement. For example, a three-level index of home educational resources was constructed from students responses to three questions: number of books in the home, educational aids in the home (computer, study desk/table for own use, and dictionary), and parents' education. Students were assigned to the high level if they reported having more than 100 books, having all three educational aids, and that at least one parent finished university. Students at the low level reported having 25 or fewer books in the home, not all three educational aids, and that neither parent had completed high school. Students with all other response combinations were assigned to the middle category.

For each index, the percentage of students in each category for each country was presented together with the average achievement for those students.[7] For example, the TIMSS Index of Home Educational Resources showed that students with more home

---

[6] For a complete list of indices, see Smith, T.A., (2000). Reporting questionnaire data. In *TIMSS 1999 Technical Report*, M.O. Martin, K.D. Gregory, & S.E. Stemler (Eds.), Chestnut Hill, MA: Boston College, International Study Center.

[7] As an example, the TIMSS 1999 Mathematics Benchmarking data for the Index of Home Educational Resources can be seen in Exhibit 4.1 (page 112) of Mullis, I.V.S. et al. (2001). *Mathematics Benchmarking Report TIMSS 1999 – Eighth Grade: Achievement for U.S. States and Districts in an International Context.* Chestnut Hill, MA: Boston College, International Study Center.

resources had higher mathematics achievement. Benchmarking jurisdictions with the greatest percentages of students with home resources were among the top-performing jurisdictions, and those with the lowest achievement were four urban districts that also had the lowest percentages with high levels of home resources.

END NOTES

Considering the complexity of the educational enterprise in the United States and the extreme importance of minimizing respondent burden in NAEP, the National Assessment Governing Board is facing a challenging task in assuming responsibility for NAEP's background data collection instruments and procedures. In attempts to provide valid and reliable background measures, as well as broad coverage of important factors that can help explain achievement levels and lead to improved performance, assessment and advisory committees to endeavors such as TIMSS, PIRLS, and NAEP inevitably recommend more data collection than can be reasonably accommodated by students, teachers, and schools.

Determining sets of questions that can be reliably interpreted and answered by respondents is very difficult in and of itself, considering the pitfalls of self-report data. The task becomes even more daunting when one considers the extreme importance of choosing wisely from the enticing smorgasbord of possibilities only the small number of such questions that may be actually administered. Nevertheless, given many years of experience and several examples to draw on, e.g., previous NAEP assessments, TIMSS, PIRLS, and SASS, the National Assessment Governing Board is more than equal to the challenge.