National Center for Education Statistics

U.S. Department of Education
Institute of Education Sciences
NCES 2003–012

Third International Mathematics and Science Study 1999 Video Study Technical Report

Volume 1: Mathematics

Technical Report

September 2003

Jennifer Jacobs
Helen Garnier
Ronald Gallimore
Hilary Hollingsworth
Karen Bogard Givvin
Keith Rust
Takako Kawanaka
Margaret Smith
Diana Wearne
Alfred Manaster
Wallace Etterbeek
James Hiebert
James Stigler
LessonLab

Patrick Gonzales
Project Officer
National Center for Education Statistics

U.S. Department of Education
Rod Paige, *Secretary*

Institute of Education Science
Grover J. Whitehurst, *Director*

National Center for Education Statistics
Valena Plisko, *Associate Commissioner*

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street, NW
Washington, DC 20006

September 2003

The NCES World Wide Web Home Page is***: http://nces.ed.gov***
The NCES World Wide Electronic Catalog is***: http://nces.ed.gov/pubsearch***

Suggested Citation
U.S. Department of Education, National Center for Education Statistics. *Third International Mathematics and Science Study 1999 Video Study Technical Report Volume I: Mathematics,* NCES (2003-012), by Jennifer Jacobs, Helen Garnier, Ronald Gallimore, Hilary Hollingsworth, Karen Bogard Givvin, Keith Rust, Takako Kawanaka, Margaret Smith, Diana Wearne, Alfred Manaster, Wallace Etterbeek, James Hiebert, and James Stigler. Washington, DC: 2003.

For ordering information on this report, write:
U.S. Department of Education, ED Pubs
P.O. Box 1398
Jessup, MD 20794-1398

or call toll free 1-877-4ED-PUBS or go to the Internet: ***http://nces.ed.gov/timss***

Content Contact:
TIMSS Customer Service: (202) 502-7421 fax: (202) 502-7455
Email: timss@ed.gov

Foreword

This first volume of the Third International Mathematics and Science Study (TIMSS) 1999 Video Study Technical Report focuses on every aspect of the planning, implementation, processing, analysis, and reporting of the mathematics components of the TIMSS 1999 Video Study. The report is intended to serve as a record of the actions and documentation of outcomes, to be used in interpreting the results and as a reference for future studies.

The TIMSS 1999 Video Study is a complex and ambitious study conducted under the aegis of the International Association for the Evaluation of Educational Achievement (IEA) and managed by the U.S. Department of Education's National Center for Education Statistics in cooperation with its study partner, the National Science Foundation. Over a period of four years, the study researchers collected, transcribed, translated, coded, and analyzed hundreds of hours of videotapes of eighth-grade mathematics lessons in the seven participating countries. The design of the study built on the foundations established by the first TIMSS 1995 Video Study, but was improved and carried out through a collaborative process that involved individuals around the globe.

Each of the chapters of this report, and the appendices, focuses on critical steps taken in the planning and implementation of the study, from its initial design to how the data was analyzed. One of the more complex tasks of the study was the development of a coding system that addressed critical questions and was applicable to each country's unique education system. The resulting coding system for the mathematics videos is discussed in detail in this report with the aim of making the system available for review, improvement, and possible application to future studies.

This report follows the first release of data focusing on the eighth-grade mathematics lessons made available to the public in March 2003. Additional reports are planned that will focus on the results of analyses of the eighth-grade science videos collected as part of the study, and a comparison of eighth-grade mathematics teaching in the United States based on the videos collected for the 1995 and 1999 studies. A second volume of the technical report that focuses on the science videos will be released soon after the first science report is released, expected in early 2004.

Patrick Gonzales
Project Officer
National Center for Education Statistics

September 2003

Acknowledgments

The authors of this report wish to recognize the many people who contributed to making this report possible. In addition to the many contributors listed in appendix I of this report, the authors wish to thank the following individuals for their helpful comments: Patrick Gonzales, William Hussar, Ralph Lee, Valena Plisko, and Marilyn Seastrom of the National Center for Education Statistics; Michael Martin of Boston College; and Linda Shafer of the Education Statistics Services Institute.

The TIMSS 1999 Video Study was conducted by LessonLab, Inc. under contract to the National Center for Education Statistics, U.S. Department of Education. The U.S. National Science Foundation and the participating countries provided additional funding for the study.

Table of Contents

List of Tables and Figures

Chapter 1. Introduction

1.1 Introduction

The Third International Mathematics and Science Study (TIMSS) 1999 Video Study examined classroom teaching practices through in-depth analysis of videotapes of eighth-grade mathematics and science lessons. An update and expansion of the 1995 TIMSS Video Study, the TIMSS 1999 Video Study investigated nationally representative samples of classroom lessons from relatively high achieving countries. The Video Studies were designed to supplement the information obtained through the TIMSS 1995 and 1999 mathematics and science assessments.

The TIMSS 1999 Video Study was funded primarily by the National Center for Education Statistics (NCES) in the Institute of Education Sciences (IES) of the U.S. Department of Education, and the National Science Foundation. It was conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA), based in Amsterdam, the Netherlands. Support for the project also was provided by each participating country through the services of a collaborator who guided the sampling and recruiting of participating teachers. In addition, Australia and Switzerland contributed direct support for data collection and processing of their respective sample of lessons.

This report presents the technical aspects of collecting videotapes of mathematics lessons for the TIMSS 1999 Video Study. A parallel technical report on the science lessons will be released separately.

1.2 Goals

The broad goal of the mathematics portion of the TIMSS 1999 Video Study was to describe and investigate teaching practices in eighth-grade mathematics in a variety of countries including several countries with varying cultural traditions and with high mathematics achievement, as assessed through TIMSS 1995. The participating countries were Australia, the Czech Republic, Hong Kong SAR[1], the Netherlands, Switzerland, and the United States. Japan, which participated in the science portion of the TIMSS 1999 Video Study, did not participate in the 1999 data collection for the mathematics portion. However, the Japanese data collected in the TIMSS 1995 Video Study were reanalyzed and were included in many of the results presented in *Teaching Mathematics in Seven Countries: Results from the TIMSS 1999 Video Study* (Hiebert et al. 2003).

[1] For convenience, in this report Hong Kong SAR is referred to as a country. Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

In addition to the broad goal of describing mathematics teaching in seven countries, the TIMSS 1999 Video Study had the following research objectives:

- To develop objective, observational measures of classroom instruction to serve as appropriate quantitative indicators of teaching practices in each country.
- To compare teaching practices among countries and identify lesson features that are similar or different across countries.
- To describe patterns of teaching practices within each country.

1.3 Design of the Study

The TIMSS 1999 Video Study was designed to describe eighth-grade mathematics and science teaching in each participating country. The method employed was the video survey (Stigler, Gallimore, and Hiebert 2000). Video surveys allow researchers to integrate the qualitative and quantitative study of classroom teaching across cultures, increasing their chances of capturing not only universal quantitative indicators but culturally-particular qualitative categories as well. Video surveys combine videotaping with national probability sampling. Qualitative analyses of video can be validated against a national sample of videos. Quantitative analyses are rendered more interpretable by being efficiently linked to specific video examples of the categories coded.

Some of the challenges of studying teaching using video include creating standardized camera procedures, minimizing observer effects, and maintaining acceptable levels of coding reliability. As this report will describe, a detailed data collection protocol was developed and tested (see chapter 2) and several questionnaire items assessed teachers' perceived degree of bias due to the videocamera (see chapter 5).

Other challenges of video studies have to do with sampling strategies. To provide national-level pictures of teaching, the study videotaped each teacher once, teaching a single classroom lesson. It should be clear that taping only one lesson per teacher shapes the kinds of conclusions that can be drawn about instruction. Teaching involves more than constructing and implementing lessons. It also involves weaving together multiple lessons into units that stretch out over days and weeks. Inferences about the full range of teaching practices and dynamics that might appear in a unit cannot necessarily be made, even at the aggregate level, based on examining a single lesson per teacher. Consequently, the interpretive frame of the TIMSS 1999 Video Sudy is properly restricted to national-level descriptions and comparisons.

Another sampling issue concerns the way in which content is sampled. Eighth-grade mathematics courses are composed of different topics, and teaching might look different for different topics. Decisions about how to sample depend, again, on the goal of the study. To get a nationally-representative picture of eighth-grade mathematics teaching, the best procedure is to randomly select lessons across the school year. Different countries use different curricula and move through different sets of topics. The only reasonable way to deal with this variation is to sample steadily across the school year and to randomly select lessons at each point.

It might appear desirable to control for content by sampling the same topics in the curriculum in each country, but this turns out to be virtually impossible. Different curricula and different teachers across countries define topics so uniquely that the resulting samples become less rather than more comparable. If the researchers' goal is to compare the teaching of particular topics, and if the topics are selected and defined so there is a shared understanding of the material to be taught, then controlling for topic is a reasonable approach. But such a study would have a different goal than the one reported here.

The fact that images of teachers and students appear on the tapes makes it more difficult than usual to protect the confidentiality of study participants. This continues to be a serious issue when the data set is used for secondary analyses. The question is what procedures to establish to allow continued access to video data by researchers interested in secondary analysis (Arafeh and McLaughlin 2002). One option is to disguise the participants by blurring their faces on the video. This can be accomplished with modern-day digital video editing tools, but it is expensive at present to do so for an entire data set. A more practical approach, and the one employed for this study, is to define special access procedures that will protect the confidentiality of participants while still making the videos available as part of a restricted-use data set.

1.4 Overview of the Technical Report

This report provides a full description of the methods used to conduct the TIMSS 1999 Video Study. Chapter 2 discusses the field test process, including the development of videotaping procedures. In chapter 3 there is a full description of the sampling approach implemented in each country. Chapter 4 details how the data were collected, processed, and managed. Chapter 5 describes the questionnaires collected from both teachers and students in the videotaped lessons, including how they were developed and coded. Chapters 6 and 7 provide details about the codes applied to the video data by a team of international coders as well as several specialist groups. Lastly, in chapter 8, information is provided regarding the weights and variance estimates used in the data analyses. There are also numerous appendices to this report, including the questionnaires and manuals used for data collection, transcription, and coding.

Chapter 2. Field Test Study

2.1 Overview

Prior to the initiation of the TIMSS 1999 Video Study, a field test study was funded by NCES and conducted by LessonLab during May through August 1998. This chapter provides a brief overview of the field test study and focuses on the preliminary analysis of results relevant to the mathematics portion of the study.

In May 1998 when the field test study commenced, final decisions regarding the participating countries in the TIMSS 1999 Video Study had not yet been made. Some of the countries that collaborated in the field test study were therefore different from those in the final sample. The countries that participated in the field test study were: Australia, the Czech Republic, Japan, Luxembourg, the Netherlands, Switzerland, and the United States.

The goals of the field test were:

- To modify the methods and procedures of data collection and processing; and,
- To collect samples of videotapes for use in the development and refinement of data coding.

Although the methods and procedures of data collection developed in the TIMSS 1995 Video Study were useful in investigating mathematics classrooms in Germany, Japan, and the United States, a need for modifications for the TIMSS 1999 Video Study was identified for two main reasons. First, it was unknown whether the same data collection methods would work in science classrooms. Second, there was a desire to better document student work processes during lessons. Due to time constraints, field test data collection commenced as modified videotaping procedures were being developed. Several different cameras and filming procedures were tested in the field study before final equipment decisions were made and a data collection instruction manual was written.

In addition, modified teacher and student questionnaires were not yet ready during the field test study period. Therefore, the TIMSS 1995 Video Study teacher questionnaire was used, with some minor adjustments so that it would apply to science as well as mathematics teachers.

Below is a list of tasks that were carried out in the field test study:

- Selected video equipment;
- Developed/modified data collection methods and procedures;
- Developed/modified data processing methods and procedures;
- Videotaped mathematics and science classrooms from potential participating countries; and
- Reviewed the field test data to generate ideas for the development of a coding scheme.

The following sections of this chapter describe how these tasks were carried out and what was learned from them. Note that analyses of the field test videotapes were based on a relatively small sample from each country (see section 2.5). Therefore, the purpose of the analyses was to refine data collection, processing, and coding methods for the main data collection period. As such, conclusions reached through the field test, particularly as related to the typicality of events, were intended as preliminary hypotheses that could be further investigated in the full sample of videotapes.

2.2 Selecting Video Equipment

In the TIMSS 1995 Video Study, one videographer filmed all the mathematics lessons in each country using a single SONY Hi-8 camcorder. Although the Hi-8 camcorder produced high quality videos, for the TIMSS 1999 Video Study digital camcorders were used to achieve even higher quality videos.

In the 1995 Video Study, using one camera to film the lesson gave little freedom for the videographer to film activities not involving the teacher, such as student-student interaction. As a result, almost no information was available regarding students' work processes when they were completing assignments at their seats. This limitation was perceived as a problem in the 1995 Video Study, and therefore, a decision was made to use two cameras to film each lesson in the 1999 Video Study; one camera would be operated by the videographer and the other would be used as a stationary camera to capture the whole-class view at all times.

For the field test data collection, the SONY DX200 professional digital camcorder was selected to be the main camera (i.e., operated by the videographer) and the Canon Optura mini DV camcorder to be the second camera (i.e., used as a stationary camera).

In the field study, each teacher wore a wireless microphone, and both cameras were equipped with zoom microphones. For the wireless microphone, the Lectrosonic omni-directional lavaliere microphone, transmitter, and receiver were selected, which produced better sound quality and were more durable than the microphone used in the TIMSS 1995 Video Study. Two different zoom microphones were used. For the main camera, the Sennheiser K6P was selected, which is a high quality professional mic. For the stationary camera, the Canon ZM100 zoom mic was selected.

One downside of the Optura camcorder was that it required an external audio mixer to combine the audio from the teachers' wireless microphone with the zoom microphone mounted on the main camera. The Studio Pro XLR mixer was selected, which was a lightweight mixer that could be placed between the camera and the tripod.

Different tripods were used for each camera. For the main camera, the Mathews THM20 fluid head tripod was selected, and for the stationary camera, the Promaster 6400 photography tripod was selected.

The SONY DX200 produced high quality videos. However, based on videographers' experiences in the field study, using two Canon Optura mini DV camcorders for data collection in the main study was agreed to be more feasible. Not only were the SONY camcorders larger and heavier than the Optura camcorders, they required a heavier tripod and larger videotapes.

Videographers needed to travel around the country via airplane, trains, or car, in a variety of weather conditions. Then, once in the schools, a single videographer had to transport all the equipment to the classroom in order to be ready for the class, often under less than ideal conditions. Therefore, the SONY camcorders were deemed less suitable for the main study data collection in comparison to the Optura camcorders.

2.3 Developing Data Collection Procedures

Two video cameras were used in the TIMSS 1999 Video Study; one of the cameras was manually operated and the other was stationary. The primary role of the operated camera was to document the teacher, while the second camera was intended to provide supplementary footage, mainly documenting the students.

Regarding the positioning of the second, stationary camera, two options were considered:

1. place it in the front corner to capture the entire classroom; or
2. focus it on a few students.

The advantages of the first option were that it would be easy to know where to position the stationary camera, and it would provide a general view of what happened in the classroom. Additionally, since the stationary camera would maintain a wide shot of the entire lesson, the videographer would be free to occasionally take the operated camera off the teacher and document students' work processes. In other words, most likely the teacher would remain in view of the stationary camera, so his/her activities would still be recorded. One disadvantage of this option was that it would not provide long-term, close-up information about students' work processes.

The second option would provide detailed information regarding students' work processes, but only of a few students. For this option, an important question was "Who decides which students to focus on, and on what basis does that person decide?"

After evaluating advantages and disadvantages of these two options during the field test, the first option was selected for the 1999 Video Study.

Additionally, based on the field test data collection experience, the videotaping procedures developed in the 1995 Video Study were modified for the 1999 Video Study in three important ways: 1) including documentation of student work processes along with documentation of the teacher and the students, 2) positioning cameras to incorporate the second camera placement, and 3) pointing the teacher camera to film from the perspective of an ideal student and to keep track of the teacher. These modifications, along with other details regarding the data collection procedures implemented in the 1999 Video Study are presented in chapter 4, section 4.3.

2.4 Developing Data Processing Procedures

2.4.1 Digitizing and Storing Data on a Multimedia Database

In the TIMSS 1995 Video Study a system was developed to computerize the data and store them on multimedia database software (vPrism). This system was modified for the TIMSS 1999 Video Study to allow for more sophisticated editing and analyses of data.

The basic procedures that remained the same as the 1995 Study were: (1) video data would be digitized and compressed as MPEG-1 files; (2) supplementary materials would be scanned and converted to PDF files; (3) both MPEG video files and PDF document files would be stored on CD-ROM and on a computer server; (4) transcribers/translators would transcribe the lessons into English; (5) transcripts would be imported into vPrism and linked to the video files. One important new feature of the software was that each data unit (i.e., each lesson) in the database would have two video files—one from the teacher camera and one from the student camera. The development of this feature started as a field test study task.

2.4.2 Transcribing/Translating Lessons

In the TIMSS 1995 Video Study, U.S. lessons were transcribed in English, and German and Japanese lessons were translated and transcribed into English by bilingual staff. A protocol was developed to maintain the consistency of each transcription.

For the TIMSS 1999 Video Study, all lessons were also transcribed/translated into English. The original protocol was revised on the following points:

- Punctuation marks should be based on conventional English grammar rules; and,
- Each turn should not exceed three lines of text, as defined by the software.

Details are described in the TIMSS 1999 Video Study Transcription/Translation Manual in appendix A.

Because there would be over 1,000 mathematics and science lessons to transcribe/translate, LessonLab decided to explore the possibility of subcontracting these operations to professional transcribers/translators. This process was tested in the field study.

A U.S.-based company was subcontracted to transcribe and/or translate some of the field test data. After review, it was determined that the quality of their work was not satisfactory. Therefore other companies or individuals were sought to transcribe and/or translate the 1999 Video Study data.

In some countries, the initial transcriptions/translations were completed by individuals or companies in that country. With the aid of the National Research Coordinators, several freelance translators in the Czech Republic were recruited to do the first-pass translation/transcription. Each of these individuals specialized in one of four topics: mathematics, chemistry, biology, and physics. In the Netherlands, a company (Standby) was hired to do the first-pass translation/transcription. A member of the field test team met these translators in their countries, provided a brief orientation of the project, and went over the transcription protocol.

For the Australian and U.S. data, first-pass transcription was conducted by a company called Report Works. Most Swiss lessons were transcribed in their native language (i.e., French, German, or Italian) in Switzerland, and translations of a subset of lessons into English were completed by LessonLab.

These companies and individuals were deemed to produce high quality transcriptions/ translations, and were used for the remainder of the 1999 Video Study. (Hong Kong SAR lessons were translated/transcribed entirely by LessonLab.) All transcriptions/translations were reviewed by LessonLab, and time coding was completed at LessonLab. Additional details about the entire transcription/translation process can be found in chapter 4.

2.5 Collecting the Field Test Data

Three U.S.-based videographers were hired to videotape field test lessons in the seven countries. One of these videographers had collected the U.S. data for the TIMSS 1995 Video Study. Training took place at the University of California at Los Angeles and was based on the data collection procedures from the 1995 Video Study, with the modifications described above.

Table 2.1 shows the date of videotaping, country, location of schools, types of school, the subject and topic of lessons for the field test data. The number in parentheses indicates the number of lessons videotaped, by school type, subject, and topic.

Table 2.1. Field test tapings by country: 1998

Date (1998)	Country	School type	Subject (lessons)	Topic
May 18–22	Czech Republic	Basic (2) Integrated (1) Gymnasium (1)	Mathematics (5) Science (4)	Algebra (3) Geometry (2) Biology (2) Chemistry (1) Physics (1)
May 25–29	Switzerland	Lower track (1) Middle (1) Higher (1)	Mathematics (5) Science (4)	Algebra (2) Geometry (3) Biology (3) Physics (1)
June 2–4	Netherlands	Lower track (1) Middle track (1)	Mathematics (4) Science (4)	Algebra (4) Biology (1) Physics (3)
June 8–10	Luxembourg	Technical (1) Gymnasium (1)	Mathematics (4) Science (3)	Algebra (1) Geometry (3) Biology (2) Physics (1)
June 23–26	Australia	Government school (4)	Mathematics (4) Science (4)	Algebra (3) Geometry (1) Biology (2) Chemistry (1) Physics (1)
June 29-July 3	Japan	Private (1) Public (2)	Science (5)	Biology (5)
May 25-June 3	United States	Public (4)	Science (4)	Astronomy (1) Biology (1) Earth science (1) Physics (1)

NOTE: The number in parentheses indicates the number of lessons videotaped, by school type, subject, and topic.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

2.5.1 Collecting Videotapes

As described in sections 2.2 and 2.3, a single videographer filmed each lesson using two video cameras. The main camera was operated using procedures similar to those in the TIMSS 1995 Video Study, except that the camera was hand-held more often in order to capture students' work processes. A stationary camera maintained a wide angle shot on as many students as possible.

One of the challenges encountered during the field study was the difficulty in handling a large amount of camera equipment. The SONY DX200 was replaced with the Canon Optura for the main data collection primarily for this reason.

2.5.2 Collecting Additional Materials

In the 1995 Video Study, videographers were responsible for collecting all supplementary materials used or discussed in the lesson. These included:

- copies of pages of books, textbooks, or workbooks;
- copies of any written materials that were handed out to the students;
- copies of overheads that were projected;
- copies of worksheets or homework assignments that were discussed or given at the end of the lesson; and,
- copies of specific students' work that was discussed.

For the 1999 Video Study, the videotaped teacher was asked to gather all of these materials and send them back to their National Research Coordinator or to LessonLab, along with the completed teacher and student questionnaires. More information about these materials, including the questionnaire response rates, can be found in chapter 4.

2.5.3 Collecting Questionnaires

Modified teacher questionnaires were not ready when the field test data collection started. Therefore, the questionnaire developed in the 1995 Video Study was provided to the field test teachers. Because this questionnaire was only intended for mathematics teachers, some minor adjustments (e.g., topics covered in the videotaped lesson) were made so that it would apply to science teachers as well. The English version of the questionnaire was used for the teachers in Australia, the Netherlands[2], and the United States, and the German version was used in Luxembourg and Switzerland. The Japanese and Czech coordinators arranged the translation of the questionnaires into their languages and validated the quality of translation.

The teachers who participated in the field test study were provided with a copy of the questionnaire immediately after their lessons were filmed. They were instructed to return the completed questionnaire to the appropriate National Research Coordinator, who then sent it to LessonLab. Questionnaires were completed and returned by all teachers. They were used only to help the field test analysis team better understand the lessons.

The development of questionnaires for the TIMSS 1999 Video Study, including both the teacher and student questionnaires, is detailed in chapter 5.

2.6 Reviewing the Field Test Data

Field test data collection was completed in early July of 1998. Most of the videos were processed by the end of July, and were ready to be viewed. The plan was to spend August and September viewing the videos and generating ideas about coding development. The following sections describe how the field test data were reviewed and what was learned.

---

[2] In the main study, teachers in the Netherlands were provided with a version of the questionnaire in Dutch.

2.6.1 Field Test Analysis Team

By the end of June 1998, participation in the TIMSS 1999 Video Study was only certain for four of the field test countries: the Czech Republic, Japan, the Netherlands and the United States. Country representatives for the Czech Republic, the Netherlands, and the United States were recruited by the end of July. These individuals would later serve as country associates for the code development and analysis of the 1999 Video Study. The country associate for Japan was scheduled to be recruited in the summer of 1999 when the science data coding began, since Japan was only participating in the science portion of the study. Dr. Ineko Tsuchida, a researcher experienced in classroom research participated in the field test study data analysis including Japanese mathematics data collected as part of the TIMSS 1995 Video Study. The Czech associate was Svetlana Trubacova, a physics teacher from Slovakia. The Dutch associate was Karen Givvin, a second generation Dutch American with a Ph.D. in education from UCLA. The U.S. associate was Jennifer Jacobs, a Ph.D. student in developmental psychology at UCLA. Dr. Kass Hogan, a U.S. researcher experienced in science classroom research, was also invited to help analyze the field test videos.

Participation of Australia, Luxembourg, and Switzerland was still being negotiated, but it was decided that they would each send a representative to Los Angeles for several weeks to participate in preliminary video analyses. The Australian representative was Nick Scott, who has conducted research in mathematics education at the University of Melbourne. The Luxembourg representative was Dr. Jean-Paul Reeff, a psychologist and the international project manager at the Luxembourg Ministry of Education. The Swiss representative was Dr. Christine Pauli, a researcher of classroom instruction at the University of Zurich.

2.6.2 Reviewing Videos Individually

In order to elicit cultural beliefs and expectations about classroom instruction, the field test analysis team members were asked to first individually carry out the six tasks described below:

<u>Task 1</u>: View the videos from your own country and write down a brief description of each lesson. Do not yet share your observations with the other country representatives.

<u>Task 2</u>: View the lessons a second time and complete a five-column table for each lesson. The columns are: time, what the teacher is doing, what the students are doing, the mathematics or science content, and any additional comments.

<u>Task 3</u>: Describe some similarities and differences across the lessons.

<u>Task 4</u>: Select one mathematics and one science lesson that you believe is most typical of how these subjects are taught in your country. If you could interview the teacher of these lessons, how do you think he/she would answer the following questions: what does the teacher believe about the subject, what things should the students

be learning from the course, how do students best learn, and what is the teacher's role.

Task 5: View all the lessons from the other countries that have been selected as most typical. Write a brief description of each lesson and describe the most important similarities and differences between those lessons and the ones from your own country.

Task 6: Prepare a presentation of 15–30 minutes on what you have found.

The presentations were given on August 20 and 21, 1998 at the TIMSS Video Data Center meeting room in UCLA. Each representative spent about 30 minutes discussing the tasks and briefly describing the education system and instructional practices in their country. Each presentation was followed by a 15–30 minute discussion. All the presentations were videotaped and burned onto CD-ROM as a record.

2.6.3 Outcomes of the Tasks

Tables 2.2 through 2.8 describe the summary of outcomes that were generated by each of the country representatives. In particular, these tables present the country representatives' initial responses to the field test mathematics lessons from their own country.

Table 2.2. Australian country representative's initial responses to the Australian field test mathematics videos: 1998

Category	Response
Similarities across the three Australian mathematics lessons	▪ A brief introduction to put the lesson in the context of previous work. ▪ A period of direct instruction in the form of a demonstration of procedures. ▪ Students practicing the demonstrated procedures, and the teacher "working the room" attending to individual students or small groups. ▪ Students' self-correction of their work to monitor their own progress. ▪ Catering for individual differences by assigning differential amount of seatwork assignments.
Description of the lesson selected as most typical of Australian mathematics teaching	A traditional text-driven lesson on combining like terms. The lesson starts with some review of previous work (about 12 minutes) followed by a small teacher-led theoretical introduction then by student skill development. During seatwork the teacher visited each group of students. The justification given to the students for the work was that it would "all be on the test."
Inferred teachers' beliefs about mathematics and mathematics teaching	▪ Mathematics is a way of thinking and describing many phenomena in the world. ▪ There is a natural sequence that dictates what mathematics content should be taught and when it should be taught. ▪ Mathematics is best learnt through the practice of skills and procedures on a large number of examples. ▪ There is a set of facts that needs to be memorized to make practicing techniques for problem solving easier. ▪ Studying for tests is a helpful approach to consolidating understanding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Table 2.3. The Czech Republic country representative's initial responses to the Czech field test mathematics videos: 1998

Category	Response
Similarities across the three Czech mathematics lessons	▪ Lessons follow a traditional structure that consists of an extended review and an introduction to new topics. ▪ During the review, the teachers employ a variety of approaches but tend to "review with students" by soliciting a large amount of oral or written contribution from the students. ▪ During the introduction of new topics, the teachers explain the target topic but "use students' knowledge" by asking a number of questions.
Description of the lesson selected as most typical of Czech mathematics teaching	A geometry lesson in which the teacher explains at the beginning of the lesson what the topic is. Then the teacher "slowly and clearly explained every single step she was doing so students could understand everything and had enough time to make their own notes." The teacher also encourages students to discover new things, and the mathematical language used by the teacher and students is of a very high level.
Inferred teachers' beliefs about mathematics and mathematics teaching	▪ Students need to understand mathematics so they can use it in their everyday life. ▪ Students need to pay attention during class so they will understand the subject. They then need to practice. ▪ The best way to practice is to solve many different problems and to discuss their solutions. ▪ The teacher's role is basically to explain everything and to show how to use knowledge in solving many different problems. ▪ Explain everything in as much detail as possible and show many examples, (this) will increase the chance that students will learn.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Table 2.4. Japanese country representative's initial responses to the Japanese field test mathematics videos: 1998

Category	Response
Similarities across the three Japanese mathematics lessons	▪ Problem solving orientation, that is, the students had a set of problems to investigate or solve. ▪ The students were learning to apply a fundamental mathematical concept or formula into more complex situations or to prove the relationships among the characteristics of a geometric shape. ▪ Students' problem solving was limited only to the extent that they were expected to apply the mathematical notion or formula in different but related situations (i.e., not designed to promote hypothesis building).
Description of the lesson selected as most typical of Japanese mathematics teaching	An introductory lesson on geometry. The teacher spends a large amount of time exposing the students to the notion of a geometric shape. Then the teacher demonstrates how to prove that the vertically opposite angles created by two straight lines are equal.
Inferred teachers' beliefs about mathematics and mathematics teaching	▪ It is important to draw a connection between classroom mathematics and students' daily life. ▪ The students will learn mathematics by relating it to their daily life or the immediate surroundings. ▪ A certain level of teacher directive is important in the introduction of mathematics. ▪ Students need to attentively listen to teacher's explanation to grasp mathematical concepts or principles in order to accomplish the main goals in the lesson. ▪ The proper role of teachers is to be a model for students. ▪ It is important to use good questions in instruction especially to build a relationship with the students and to stimulate students' thinking.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Table 2.5. Luxembourg country representative's initial responses to the Luxembourg field test mathematics videos: 1998

Category	Response
Similarities across the three Luxembourg mathematics lessons	▪ Classroom arrangements are well suited for teacher-led instruction. ▪ Lesson goals are not made explicit. ▪ The teachers ask questions frequently to students, but the questions normally deal with small portions of knowledge, and answers are either obvious or the students can only guess. ▪ When students give wrong answers, the teacher simply proceeds to the next student or gives the answer himself. ▪ The teachers switch to the native language (Luxembourgish) in critical situations.
Description of the lesson selected as most typical of Luxembourg mathematics teaching	The lesson topic is on rays and segments. The teacher introduces the concept of rays through a series of examples. Teacher-led instruction: the teacher tries to systematically ask questions and tries to motivate students to produce partial solutions on the blackboard and on special teacher-prepared worksheets. The whole classroom situation is rather chaotic.
Inferred teachers' beliefs about mathematics and mathematics teaching	▪ Mathematics is a set of concepts and rules defining the relations between these concepts. ▪ Students should learn precise definitions of mathematical objects. ▪ Students learn best while carefully looking at the teacher's demonstrations at the blackboard, solving minor problems themselves in the classrooms and at home. ▪ The teacher's role is to carefully choose suitable examples that best illustrate the topics to be dealt with, to decompose the problem in a sequence of small steps, to prepare exercising sheets relating to the steps, and to guide students through the solution of a problem in the classroom.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Table 2.6. The Netherlands country representative's initial responses to the Netherlands field test mathematics videos: 1998

Category	Response
Similarities across the three Dutch mathematics lessons	▪ Working on small set of prepared problems. ▪ The teachers indicate what information is important and/or gives hints for those problems assigned where students may experience difficulty. ▪ Students work on assignments in pairs and ask the teacher questions when necessary. ▪ The teachers are highly active throughout lessons answering students' questions. They do so with what appears to be great patience and even pleasure.
Description of the lesson selected as most typical of Dutch mathematics teaching	The lesson is on linear equations. The teacher writes assignments on the board, provides an overview of the problems, then sends off the students to work in pairs on the assignments. The teacher returns to his desk at the front of the room, and students approach him there with questions. The teacher models problem solving using different strategies, identifying what works and what does not.
Inferred teachers' beliefs about mathematics and mathematics teaching	▪ Mathematics is a set of skills that can be used to understand everyday life experiences that involve numbers. Students will leave the class having learned how to think of these experiences in mathematical terms. ▪ Students best learn mathematics by first receiving a small amount of teacher instruction, and then solving problems alone or with their peers. If they have questions, they should feel comfortable asking the teacher for whatever degree of assistance they feel they need to understand the solution. ▪ Teachers need to know what students do and do not understand. Part of a teacher's role is to help students through the areas the teacher knows will be difficult.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Table 2.7. Swiss country representative's initial responses to the Swiss field test mathematics videos: 1998

Category	Response
Similarities across the three Swiss mathematics lessons	▪ The lessons follow a 3-step temporal structure: an introduction phase, in which the teacher establishes the students' attention and sets a positive climate, a main phase that starts with a goal or problem statement, and a closing phase that includes next lesson previews and homework assignments. ▪ Frequent use of teacher-directed classroom dialogue to co-construct a conceptual structure with the students. ▪ While assisting students individually during seatwork, the teachers tend not to tell the students what is right or wrong but try to scaffold instead. ▪ The emphasis is on anchoring conceptual knowledge through problem solving, often using practical, everyday problems. ▪ The teachers speak the official language (German) when talking publicly to the whole class and switch to the native language (Swiss German) when assisting students privately.
Description of the lesson selected as most typical of Swiss mathematics teaching	Two lessons were selected: an introduction lesson in which the teacher mainly guides the students through teacher-directed classroom discourse, and an application lesson in which a large portion of the lesson time is spent by the students working on problems independently of the teacher.
Inferred teachers' beliefs about mathematics and mathematics teaching	▪ One important goal of mathematics education would be to design learning situations that allow students to have increased confidence in their capacity to do mathematics. ▪ The teacher should provide each individual student with the appropriate support, help, or assistance, in correspondence to his/her needs. ▪ Students should master mathematical procedures, but not in a mindless way. Instead, they should understand, or at least be aware of, what they are doing when they operate with numbers and relations. ▪ It is important that … after the introduction of the procedures and concepts, there should be a phase of working through the conceptual structure, and that the students have many opportunities for practice. At the end, the students should be able to execute the procedure by themselves and to use it to solve applied problems.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Table 2.8. The U.S. country representative's initial responses to the United States field test mathematics videos: 1998

Category	Response
Similarities across the three U.S. mathematics lessons	▪ Most lesson time is spent working on problems as a whole class. ▪ The teacher guides the students through problems, providing some of the information needed to solve the problems (i.e., problems are broken down into pieces, and given students provide answers to the pieces of the problem). ▪ Lessons are focused on particular concepts the teacher wants students to learn.
Description of the lesson selected as most typical of U.S. mathematics teaching	A geometry lesson applying what students already know about "transformation" to coordinate graphs. The teacher uses problems to explain the conceptual points he wants to make. Throughout the lesson the teacher maintains a very animated personality, and continually asks for and incorporates the students' input. At the end of the lesson, the students work on a few of the same type of problems individually, and then share their answers.
Inferred teachers' beliefs about mathematics and mathematics teaching	▪ Mathematics is a set of concepts, as well as a vocabulary needed to talk about those concepts. ▪ Mathematics follows a set of rules and is logical, but not necessarily intuitive. Therefore, an expert is needed to explain the concepts and define the terms. ▪ Students learn by watching the teacher's demonstration of important concepts, and then by practicing on their own. ▪ Students need to be paying attention and following along as the teacher explains, but they also need some time for practicing on their own. ▪ The teacher should be an expert, and present the concepts and vocabulary clearly and in a meaningful sequence.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

2.6.4 Reviewing Videos as a Group

After reviewing and sharing individual observations, the field test analysis team worked together to create "meta-plans," that is, a list of their thoughts regarding each lesson. This idea was suggested by the Luxembourg representative, Dr. Reeff, and involved the following four steps:

Step 1: Lessons were discussed one at a time. Each person individually generated ideas about that lesson and wrote them down on notecards.

Step 2: The whole team shared and discussed the ideas.

Step 3: The whole team categorized and consolidated the ideas.

Step 4: Each person selected a) the five ideas he/she believed best characterized the lesson and b) the five ideas that were most different from lessons in the member's own country.

Meta-plans were created for all but two of the mathematics and science lessons denoted as "most typical" by the country representatives.

All of the meta-plans were created in the same manner. First, each country representative wrote down thoughts about the designated lesson on notecards—one thought per card. To the extent possible, these thoughts were to be objective descriptions or ideas about the lesson.

Next, each card was read, discussed, and categorized into one of the nine categories. These categories were loosely defined as lesson flow, content, lesson structure, teacher behavior, student behavior, climate, technical issues, comments, and missing events (see table 2.9). If two or more of the cards contained the same ideas, they were placed together.

Table 2.9. Meta-plan categories created to describe the field test lessons: 1998

Category	Definition
Lesson flow	Specific events that occurred in the lesson, listed in chronological order
Content	The mathematics or science content covered
Lesson structure	The general structure of the lesson
Teacher behavior	Noticeable behaviors by the teacher
Student behavior	Noticeable behaviors by the students
Climate	The lesson climate or atmosphere
Technical issues	Issues regarding the video quality
Comments	Personal opinions or judgments
Missing events	Events that were expected to but did not occur in the lesson

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Each country representative chose the five descriptions that best characterized the lesson, and marked these with a "C." They also chose the five descriptions that were most different from teaching in their own country, and marked these with a "D" (see table 2.10).

Table 2.10. Definitions of “characteristic” and “different” used in the meta-plans: 1998

Term	Definition
Characteristic (C)	A description that characterizes the lesson very well. For example, a very interesting feature of the lesson, or a feature that someone would very likely mention when giving a brief description of the lesson.
Different (D)	A description that is different from the teaching typically found in “your” country. Something that rarely occurs in lessons from “your” country.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Lastly, there was an extended discussion about each “typical” field test lesson that focused on the following three topics:

a) Descriptions that many representatives chose as characteristic of the lesson.

These descriptions represented a consensus among the country representatives regarding a particular lesson’s most critical features, and as such were considered likely candidates for further code development. Additionally, ideas designated by many team members as “characteristic” of a lesson but not “different” from lessons in other countries were likely to apply to several countries.

b) Descriptions that representatives chose as different from the teaching in their country.

These descriptions represented features that one or more team members regarded as different from the teaching typically found in their country. Such differences were important tools for discussion, as they were considered candidates for further code development to capture important distinctions across countries.

c) Descriptions that many representatives chose as both characteristic and different.

These descriptions represented consensus among the representatives regarding a particular lesson’s most critical features, and as such were candidates for further code development. However, since one or more representatives chose these ideas as “different,” they were likely to be unique to one (or possibly a few) countries.

The process in which the country representatives engaged, as described above, was a “bottom-up” process. That is, it revolved around extensive watching and discussion of individual lessons. The next step was to take a more “top down” and theoretical approach. The field study analysis team decided to once again employ the meta-plan technique to generate and share ideas regarding general coding strategies. Using the meta-plans from all of the individual lessons as guides, the team brainstormed several possible coding strategies for many of the categories, as well as for other categories. Country representatives then embarked on an extensive theoretical/literature review on topics in these categories.

2.7 Summary

2.7.1 What Was Modified Based on the Field Test

Table 2.11 describes the data collection and processing methods that were modified from the 1995 Video Study based on the field test experiences.

Table 2.11. Modifications to the data collection and processing methods made in the TIMSS 1999 Video Study: 1999

Item	TIMSS 1995 Video Study	TIMSS 1999 Video Study
Number of cameras used to film each lesson	One video camera was operated by the videographer: ▪ One SONY EVW300 three-chip professional Hi-8 camcorder ▪ One Bogen fluid head tripod ▪ One Hi-8 tapes	One video camera was operated by the videographer, another camera was used as a stationary camera: ▪ Two Canon Optura mini DV camcorders ▪ Mathews THM20 fluid head tripod, Promaster6400 photography tripod ▪ Two mini DV tapes
Camera positioning	Camera was placed at the side of the classroom one third of the way back from the front.	Main camera: Placed one third to one half of the way back from the front. Stationary camera: Placed high on the tripod along a sidewall near the front of the room.
Camera pointing	Principle 1: Document the perspective of an ideal student Principle 2: Document the teacher regardless of what the ideal student is doing.	Principle 1: Document the teacher. Principle 2: Document the students. Principle 3: Document the task.
Collection of additional materials used in the lesson	Videographers collected all materials.	The videotaped teacher sent copies of all the materials to LessonLab.
Data storage	One video file was linked to each lesson in the multimedia database software, vPrism.	Two video files were linked to each lesson in vPrism.

Table 2.11. Modifications to the data collection and processing methods made in the TIMSS 1999 Video Study: 1999—Continued

Item	TIMSS 1995 Video Study	TIMSS 1999 Video Study
Translation/transcription of data	Both first- and second-pass translation/transcription and timecoding were conducted at UCLA.	First-pass translation/ transcription was subcontracted to professional translators/transcribers for Australian, Czech, Dutch, and U.S. lessons. All second-pass transcription and timecoding were conducted at LessonLab.
Transcription protocol	▪ Turns at talk were marked when the speaker changed or when there was a gap in the talk. No limit for the length of a turn was set. ▪ Commas, exclamation points, semi-colons, and colons were not used. ▪ Names of classroom participants were transcribed without changes. ▪ The speaker code for a student was "S" at all times.	▪ Each turn should not exceed three lines of text. ▪ Punctuation marks were used based on conventional English grammar rules. ▪ Names of classroom participants were all changed to different names starting with the same letter. ▪ "SN" indicated a new student speaker, and "S" indicated the speaker was the same student as in the previous student turn.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

2.7.2 What Was Learned from the Analyses of the Field Test Data

2.7.2.1 Agreement in Describing the Lessons

One of the main goals in examining the field test data was to closely examine the biases of the cultural insiders. For this reason, the tasks performed by the field study team required them to individually watch the lessons and write about their impressions before engaging in any group discussions.

In preparation for the meta-planning sessions, the representatives from the seven countries summarized their ideas about each lesson. Again, the intention was that each representative would bring a unique perspective, and that there might be disagreements in how the representatives interpreted the lessons.

Sharing the individually generated lesson descriptions revealed a high level of agreement among the representatives (in a non-statistical sense). In almost all cases when a new idea was raised, other team members noted that they agreed with it or had themselves recorded a similar idea. The only disagreements among the representatives regarded the topic of "lesson climate." From the field test discussions, it appeared that views regarding climate were the most strongly affected by personal experience and/or cultural perspective.

*2.7.2.2 **Agreement on the Most Important Characteristics of the Lessons***

In addition to the team's general consensus on how to describe the field test lessons, there was also a good deal of agreement as to which descriptions were most "characteristic" of a given lesson. Even though each lesson could be described in many ways, the representatives largely agreed on which descriptions were the most critical for understanding (and ultimately coding) the lessons. Interestingly, for all lessons there was a high degree of agreement between the cultural insider (the representative from a given country) and the cultural outsiders (representatives from the other countries) in choosing which descriptions were most characteristic.

Tables 2.12 through 2.16 list descriptions that a majority of the team members considered to be very characteristic of the "typical" field test mathematics lessons from five countries. (Due to time constraints, the typical mathematics lessons from Australia and the United States were not included in this final analysis.) The preliminary theories that were developed regarding the typical components of these country's mathematics lessons would later be explored in the full dataset.

Table 2.12. Characteristics of the "typical" Czech field test mathematics lessons: 1998

Category	Characteristic
Content	An introduction and elaboration of concepts
Structure	Whole-class works together
Teacher behavior	Asks questions with different degrees of freedom
Student behavior	The student at the board has to talk through the solution

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Table 2.13. Characteristics of the "typical" Japanese field test mathematics lessons: 1998

Category	Characteristic
Lesson flow	Long illustration of new topic
Content	Logical thinking
	Wide spectrum of information connected and brought to the level of students
Teacher behavior	Meta comments
	Allows students a long time to think about questions

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Table 2.14. Characteristics of the "typical" Luxembourg field test mathematics lessons: 1998

Category	Characteristic
Lesson flow	Teacher-led instruction; one student solves a problem at the board
Lesson structure	Teacher centered
Teacher behavior	Asks questions, mostly of low cognitive level
Other	Students seem to have poor recall and understanding

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Table 2.15. Characteristics of the "typical" Netherlands field test mathematics lessons: 1998

Category	Characteristic
Lesson flow	Students work independently, teacher assists (tutoring)
Lesson structure	Students assist each other
Teacher behavior	Gives varied forms of assistance (e.g., direct telling, scaffolding[1])
	Almost no public explanation

[1]See Bruner (1966) for an explanation of scaffolding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Table 2.16. Characteristics of the "typical" Swiss field test mathematics lessons: 1998

Category	Characteristic
Lesson flow	Class starts with an opening problem that takes a long time
	Teacher demonstrates solutions
	Teacher assists students individually during seatwork
Content	Practice lesson
Lesson structure	Students work individually and in pairs
Student behavior	Students seem to be on task most of the time

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

2.7.2.3 Many Similar Coding Issues for Both Mathematics and Science Lessons

After developing meta-plans for all of the "typical" science lessons and most of the "typical" mathematics lessons, the field test analysis team noticed that many of the issues raised could be applied in a general sense to all of the data. That is, there seemed to be a number of general coding categories that could be appropriate for both the mathematics and science lessons. At a more specific level, these codes would have to be developed separately. However, watching and discussing both mathematics and science lessons led the team to consider many coding categories as applicable for both topics. Table 2.17 describes the list of the coding categories that resulted from these preliminary analyses of the field test data.

Table 2.17. Potential coding categories derived from analyses of the field test data: 1998

Category	Issue
Content	▪ Link between theory and “real life” ▪ Utility of knowledge ▪ Complexity of content (relatedness between topics, ideas) ▪ Content brought on level of students
Teacher content knowledge	▪ Whether teachers make mistakes in their explanations
Structure	▪ Elements of scientific method: inclusion and sequence ▪ Lesson closure (summary, recap, review, preview of the next lesson) ▪ Amount of individual vs. group work ▪ Ways or qualities of reviewing
Teacher behavior	▪ Structuring comments (lesson and topic)
Student behavior	▪ Time on task, student engagement ▪ Students’ applied activities
Discourse/interaction	▪ Teacher’s role during whole class discussion ▪ Teacher’s assistance during individual work time ▪ Students’ opportunities to communicate reasoning ▪ Amount of student-initiated inquiry during whole-class work and seatwork ▪ Explicit link between lessons, topics, and ideas
Climate	▪ How students’ errors were treated ▪ Teacher’s emphasis on performance, speed, mastery, and understanding
Classroom management	▪ Discipline talk ▪ Ability to make smooth transitions
Lesson characteristics	▪ Teacher controlled vs. student controlled learning and problem solving ▪ Coherence of lesson
Cognitive level	▪ Facilitation of transfer ▪ Work expected of students to think, do, and memorize ▪ Ways or qualities of practice ▪ Reflection on learning strategies ▪ Content knowledge (what to learn) ▪ Representations of phenomena/objects

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

2.7.3 Overall Summary

The field test for the TIMSS 1999 Video Study helped to generate improvements in data collection, processing, and analyses. Some of the most important modifications involved creating videotaping procedures for two cameras, updating the software to include two video tracks, generating transcription/translation protocols to incorporate five languages, and generating hypotheses and coding ideas to describe teaching in a wide range of countries.

The field test analysis team consisted of representatives for each of the countries participating in the field test study. These representatives spent several months in 1998 studying the data from all of the countries. As one task, they selected a "typical lesson" from their own country, and then viewed the "typical lessons" from all of the other countries. Structured group discussions about these lessons led to preliminary theories about the characteristics of instruction within each country as well as important differences in teaching across countries. These theories then paved the way for more intensive code development work to begin, as described in chapter 6.

Chapter 3. Sampling

3.1 Introduction

The TIMSS 1999 Video Study was designed to provide comparable information about nationally-representative samples of mathematics and science lessons in participating countries. To make the comparisons valid, it was necessary to devise a sampling design for each country that called for uniformity in sampling procedures but also allowed participating countries to account for differences in educational systems, as well as implementation limitations. In general, the sampling plan for the TIMSS 1999 Video Study followed the standards and procedures agreed to and implemented for the TIMSS 1999 Achievement Study. Most of the participating countries drew separate samples for the TIMSS 1999 Video Study than they did for the TIMSS 1999 Achievement Study, however. For this and other reasons, the TIMSS 1999 assessment data cannot be directly linked to the video database.

3.2 Selecting Countries to Participate in the Study

The TIMSS 1999 Video Study aimed to expand on the TIMSS 1995 Video Study by examining instruction in more countries, and in particular in more high-achieving countries. The selection of countries for inclusion in the TIMSS 1999 Video Study was based primarily on the results from the TIMSS 1995 mathematics assessment administered to eighth-grade students (NCES, 1996) with the aim of including countries that outperformed the United States. Since it was not operationally or financially possible to include all nations that outperformed the United States in mathematics in TIMSS 1995, the sponsors of the study invited four nations from Europe and Asia—the Czech Republic, Hong Kong SAR, Japan, and the Netherlands—to participate, along with the United States. In addition, Switzerland and Australia joined the study partly at their own expense. Japan agreed to participate in the science portion of the study only; the Japanese mathematics video data collected as part of the TIMSS 1995 Video Study were re-analyzed using the TIMSS 1999 Video Study coding system.

Table 3.1 lists the countries analyzed as part of the TIMSS 1999 Video Study along with their scores on the TIMSS 1995 and 1999 mathematics assessments. With the exception of Australia, the TIMSS 1999 assessment was administered after the 1999 Video Study sample. Thus, these assessment results were not used to select countries for the 1999 Video Study.

On the TIMSS 1995 mathematics assessment, eighth graders as a group in Japan and Hong Kong SAR were among the highest achieving students, and their results were not significantly different from one another. On average, students in the Czech Republic scored statistically below their peers in Japan but similar to those in Hong Kong SAR. Average scores in Switzerland and the Netherlands were similar to one another. The mathematics average for Australia was similar to that in the Netherlands. Students in the United States scored, on average, significantly lower than the other six countries.

Eighth graders in these countries continued to score significantly higher than their peers in the United States on the TIMSS 1999 mathematics assessment, except for students in the Czech

Republic whose scores were not significantly different than the scores for students in the United States in 1999. Switzerland did not participate in the TIMSS 1999 assessment.

Table 3.1. Average score on the TIMSS 1995 and TIMSS 1999 mathematics assessments, by country: 1995 and 1999

Country	TIMSS 1995 mathematics score		TIMSS 1999 mathematics score	
	Average	Standard error	Average	Standard error
Australia ◊	519	3.8	525	4.8
Czech Republic	546	4.5	520	4.2
Hong Kong SAR	569	6.1	582	4.3
Japan	581	1.6	579	1.7
Netherlands ◊	529	6.1	540	7.1
Switzerland	534	2.7	—	—
United States	492	4.7	502	4.0

◊ Nation did not meet international sampling and/or other guidelines in 1995. See Beaton et al. (1996) for details.
—Not available
NOTE: Rescaled TIMSS 1995 mathematics scores are reported here. Switzerland did not participate in the TIMSS 1999 assessment.
SOURCE: Gonzales, P., Calsyn, C., Jocelyn, L., Mak, K., Kastberg, D., Arafeh, S., Williams, T., and Tsen, W. (2000). *Pursuing Excellence: Comparisons of International Eighth-Grade Mathematics and Science Achievement from a U.S. Perspective, 1995 and 1999* (NCES 2001– 028). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

3.3 International Sampling Specifications

In general, the sampling plan for the TIMSS 1999 Video Study followed the standards and procedures agreed to and implemented for the TIMSS 1999 Achievement Study. The sampling plan proposed for the TIMSS 1999 Video Study internationally was a two-stage stratified cluster design. The first stage consisted of a stratified sample of schools, and the second stage consisted of a sample of mathematics and science lessons from the 8th-grade in the sampled schools. This relatively simple and cost-effective design was intended to produce national samples that would meet the analytical requirements necessary to allow estimates for classrooms and schools.

At a minimum, the sample design for each country had to provide a sampling precision at the lesson (or classroom) level equivalent to a simple random sample of 100 mathematics and 100 science 8th-grade lessons for Australia, the Czech Republic, the Netherlands, and the United States; 100 science 8th-grade lessons for Japan; 100 mathematics 8th-grade lessons for Hong Kong SAR; and 143 mathematics 8th-grade lessons for Switzerland[3]. There was no minimum sample size established for the study. Rather, each country sample was carefully scrutinized to determine acceptability.

[3] Switzerland wished to analyze its data by language group, resulting in a nationally-representative sample that is also statistically reliable for the French-, Italian-, and German-language areas.

In addition, an attempt was made to collect data across the school year, and thus be representative of the teaching that eighth-grade students received over an academic year. In order to achieve this, an average of 12–15 lessons were videotaped per month in each country. In Hong Kong SAR, an above average number of lessons were collected over a one-month period. To ensure that the data collected would be evenly distributed over the school year, nine lessons from this one-month period were randomly selected and omitted from the Hong Kong SAR sample. These lessons were replaced by others that were taught by the same teachers and videotaped at a later date.

3.3.1 The School Sampling Stage

Under the international sample design, the first sampling stage was the sampling of schools. The school sampling frame in principle included all schools in the country that had eligible students in 8th grade, the target grade. The school sample was required to be a Probability Proportionate to Size (PPS) sample. A PPS sample assigns probabilities of selection to each school proportional to the number of eligible students in the 8th-grade in schools countrywide. Westat[4] strongly recommended systematic sampling, with implicit and explicit stratification, to each of the participating countries. Systematic sampling was recommended because of its good properties with regard to lower sampling variance (when the implicit stratification structure is chosen well), and its relative simplicity, allowing for use by individual countries. Whether or not systematic sampling was used, the sample was required to be a scientific probability sample, selected using the techniques and principles of this method.

Under the proposed systematic sampling approach, all schools within the explicit stratum should be ordered by a set of school characteristics that become the implicit strata. The explicit strata were generally expected to be regions of the country or other similar subgroups for which an exact sample size was desired for each subgroup. The implicit strata were expected to be other school characteristics for which exact sample sizes within subgroups were not deemed necessary, but for which a small variability in sample size across subgroups was desired.

Once the final ordering of schools was determined, a sample was to be drawn for each explicit stratum by computing an aggregate measure of size where the measures of size are proportional to the school selection probabilities for each school on the ordered list. The first school's aggregate measure of size is equal to its measure of size, the second school's aggregate measure of size is equal to the summation of the first and second schools' measures of size, and so on, with the final school's aggregate measure of size equaling the total summation of all measures of size in the explicit stratum. A sampling interval was computed that is equal to the total measure of size for the explicit stratum divided by the sample size for the stratum.

A random number was chosen for the explicit stratum between 0 and the sampling interval. The school with the smallest aggregate measure of size greater than the random number would be selected. A stream of numbers was then generated by adding positive integer multiples of the sampling interval to the random number, until the total measure of size was exceeded. For each number in this stream, a school was to be selected by taking the school with the smallest aggregate measure of size greater than that number.

[4] Westat was contracted to guide the sampling and weighting procedures for the TIMSS 1999 Video Study.

If originally selected schools declined to participate, in some countries replacement schools were selected using the same procedure described above. In general, the original and replacement schools had very similar probabilities of selection into the initial sample. More information on the number of replacement schools used in each country is presented in sections 3.4 and 3.5 of this chapter. Additional details on the selection probabilities and weights assigned to the replacement schools in each country can be found in chapter 8.

3.3.2 The Classroom and Lesson Sampling Stage

The next stage following school selection was classroom selection within schools, and finally lesson selection. One mathematics and/or one science 8th-grade class per school was to be sampled, depending on the subject(s) to be studied in each country.[5] The classes were to be randomly selected from a list of eligible classes in each participating school. The classroom sampling design was to be an equal probability design with no subsampling of students in the classroom.

For schools in which both mathematics and science classes were to be videotaped, the mathematics classroom was selected first with each available mathematics classroom having an equal probability of being selected. However, science classes that were scheduled at the same time as the selected mathematics class were omitted, and the science classroom was then randomly selected from the remaining available classrooms.

One lesson from each selected mathematics and science classroom was then videotaped. The videotaping date was determined by the scheduler in each country, and was based on scheduling and operational convenience.

3.4 Selecting Samples Within Each Country

Within the guidelines specified above, each country developed its own sampling strategy. For example, in two countries the video sample was a subsample of the TIMSS 1995 or TIMSS 1999 Achievement Study schools.[6] Also, although most countries used replacement schools, some did not. All of the TIMSS 1999 Video Study countries were required to include at least 100 schools in their initial selection of schools; however some countries chose to include more for various reasons. Furthermore, although all countries had to obtain a systematic Probability Proportionate to Size (PPS) sample, they were allowed to define strata appropriate for their country.

In most countries, the school sample was selected by the national research coordinators. In the United States, Westat (a contracted research corporation) selected the school sample. In countries that used the same sample of schools as for the TIMSS 1999 Achievement Study, school samples were selected and checked by Statistics Canada. In all cases, countries provided

[5] Australia, the Czech Republic, Japan, the Netherlands, and the United States also collected video data on eighth-grade science lessons.
[6] For the German-language area of Switzerland, the video sample was a subsample of the TIMSS 1995 achievement school sample. For Hong Kong SAR most, but not all, of the video sample was a subsample of the TIMSS 1999 achievement school sample.

the relevant sampling variables to Westat, so that they could appropriately weight the school samples.

The national research coordinators were responsible for selecting the classroom sample in their country. LessonLab was responsible for selecting the classroom sample in the United States. Westat received information about the number of classes in each country, so that the classroom stages of sampling could be weighted correctly. Additional information on the weighted participation rates in each country is provided in chapter 8.

In all countries, the national research coordinator was responsible for securing and verifying that any consent required by law was obtained from teachers, students, and/or parents. In addition, the national research coordinator in each country determined the type of compensation that would be provided to participating teachers. In each country, teachers were provided locally appropriate monetary compensation, a book voucher, and/or a videotape of their lesson in return for participation.

The details of the sample selection in each country are provided below. For most countries, a table is provided describing the breakdown of the types of schools by source of funding and any other variables deemed pertinent by the national research coordinator. For more detailed information on the educational systems within each of the participating countries, see Robitaille (1997).

3.4.1 Australia Sample

Australian schools were sampled systematically from 13 explicit strata, which were defined by state/territory and metro/non-metro status. Within each stratum, the schools were sorted by sector (government, Catholic, and independent) and enrollment. A systematic Probability Proportionate to Size (PPS) sample of 100 schools was selected from the ordered list. The measure of size (MOS) was the estimated number of mathematics and science classes in the school. Sixty-one of the 100 originally sampled schools agreed to participate and 26 replacement schools were used, yielding a final sample size of 87 schools. Mathematics lessons were filmed in all of the selected classrooms.

Table 3.2 shows the breakdown of the type of schools in the Australian sample, based on source of funding.

Table 3.2. Number and percentage distribution of the Australian school sample, by source of funding: 1999

School sector type	Number	Percent
Total	87	100.0
Government	54	62.1
Catholic	16	18.4
Independent◊	17	19.5

◊Independent schools included Christian Community schools, non-Catholic religious schools, and others.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

3.4.2 The Czech Republic Sample

Schools in the Czech Republic were sampled systematically from two explicit strata: basic schools and gymnasia schools. Within each stratum, the schools were sorted and a systematic PPS sample was selected from the ordered list. Consistent with the distribution in the population, 90 schools were selected from the basic school stratum and 10 schools were selected from the gymnasia school stratum. The measure of size (MOS) was the number of students enrolled in the eighth grade. Eighty-nine out of 100 originally sampled schools agreed to participate and 11 replacement schools were used, yielding a final sample size of 100 schools. Mathematics lessons were filmed in all of the selected classrooms.

Table 3.3 shows the breakdown of the type of schools in the Czech Republic sample, based on source of funding as well as ability track.

Table 3.3. Number and percentage distribution of the Czech Republic school sample, by source of funding and ability track: 1999

School sector type	Number	Percent
Total	100	100.0
Funding		
State	98	98.0
Religious	1	1.0
Private	1	1.0
Ability track[1]		
Basic	90	90.0
Gymnasia	10	10.0

[1]In the Czech Republic there is a two-tiered school system at the lower secondary level (grades 6-9). At the time of data collection, basic schools (the lower tier) were attended by approximately 90 percent of lower secondary students. Student attending gymnasia schools (the upper tier) were required to pass an entrance examination to gain entrance to the school.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

3.4.3 Hong Kong SAR Sample

The videotaped schools in Hong Kong SAR represent a subset of schools that participated in the TIMSS 1999 Achievement Study. The TIMSS 1999 Achievement schools represent a systematic PPS sample with strata defined by source of funding of the schools (3 levels: government, aided, and private) and gender of students (2 types: co-educational and single-sex). Using an alphabetically-sorted list of the 180 schools in the Achievement Study sample, schools were chosen at a randomly selected interval. The school MOS was the reported 8th-grade enrollment. Sixty-three out of 100 originally sampled schools agreed to participate and 37 replacement schools were used, yielding a final sample size of 100 schools. Mathematics lessons were filmed in all of the selected classrooms.

During one month of data collection, the number of lessons taped in Hong Kong SAR far exceeded the monthly average. To ensure that the data collected would be evenly distributed over the school year, nine lessons from this one-month period were randomly selected and omitted from the sample. These lessons were replaced by others that were taught by the same teachers and videotaped at a later date. That is, the data collection period in Hong Kong SAR was extended by two months to incorporate the collection of these videotapes.

Table 3.4 shows the breakdown of the type of schools in the Hong Kong SAR sample, based on source of funding and management.

Table 3.4. Number and percentage distribution of the Hong Kong SAR school sample, by source of funding and management: 1999

School sector type	Number	Percent
Total	100	100.0
Government	8	8.0
Aided	85	85.0
Private	7	7.0

NOTE: Government schools are government funded and managed. Aided schools are government funded but managed by School Sponsoring Bodies.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

3.4.4 Japan Sample

The Japanese sample of mathematics lessons was collected in the TIMSS 1995 Video Study (see Stigler et al., 1999). The videotaped schools represent a subset of the 158 schools that participated in the TIMSS 1995 Achievement Study. Schools were sampled systematically from strata defined by size of community and size of school. Within each stratum, the schools were sorted and a systematic PPS sample was selected from the ordered list. One third of the schools in the TIMSS 1995 Achievement Study sample were randomly selected within each stratum for the 1995 Video Study.

Forty-eight out of 50 originally sampled schools agreed to participate and 2 replacement schools were used, yielding a final sample size of 50 schools. The sample size was reduced to 50 because collaborators at the Japanese National Institute for Educational Research (NIER) determined that 100 classrooms would create too great a burden for their country. This smaller sample size was deemed sufficient due to the homogeneity of the classrooms.

It is important to note that the Japanese sample size is considerably smaller than the sample sizes of all the countries participating in the TIMSS 1999 Video Study. In addition, videotaping was conducted largely over a condensed period of the school year and the Japanese sample is skewed toward geometry (Stigler et al., 1999). In the international report on the mathematics results (NCES 2003-013), it is noted that the Japanese sample contained lessons with high percentages of two-dimensional geometry problems relative to the other countries, which may be due to the topic sample (Hiebert et al., 2003). Therefore, where appropriate, analyses were run based only on lessons that contained two-dimensional geometry problems.

Videotaping was usually done in a different class from the one in which testing for the TIMSS 1995 Achievement Study was conducted. When there was a choice, the principal of each school chose the classroom to be filmed.

3.4.5 The Netherlands Sample

Schools in the Netherlands were sampled systematically from seven implicit strata based on ability tracking at the school level. A systematic PPS sample of 175 schools was drawn from a list sorted on the stratum and number of 8th-grade students (the Measure of Size). The sample was selected at the same time as the TIMSS 1999 Achievement Study sample. The same sampling interval was used for both samples, but the random number for the 1999 Video Study sample was the random number for the 1998–1999 Achievement Study sample plus 0.5. This was done to reduce the overlap between the two samples. However, significant overlap occurred in spite of this procedure. Sixty-nine of the 175 originally sampled schools had already been selected to participate in the TIMSS 1999 Achievement Study or were "first replacements" (the school following the sampled school on the sorted list for that study). Therefore, they were not considered eligible schools for the 1999 Video Study. An additional four schools were dropped randomly, and two other schools were ineligible, leaving 100 schools for the video sample. This constituted the number of schools considered to be in the original sample for the study.

Forty-nine originally sampled schools agreed to participate and 36 replacement schools were used, yielding a final sample size of 85 schools. Of the 36 replacement schools, in 27 cases the replacement was the initially designated first replacement school. The other 9 replacements were chosen from unused substitutes for other schools, including one case of a school that had previously been dropped. Mathematics lessons were filmed in 78 of the selected classrooms. (Science lessons only were filmed in 7 schools.)

Table 3.5 shows the breakdown of the type of schools in the Netherlands sample, based on source of funding as well as ability track.

Table 3.5. Number and percentage distribution of the Netherlands school sample, by source of funding and ability track: 1999

School sector type	Number	Percent
Total	78	100.0
Funding◊		
Public	25	32.1
Roman Catholic	24	30.8
Protestant-Christian	22	28.2
Non-denominational private	7	9.0
Not identified	3	3.8
Ability track		
(i)vbo	5	6.4
mavo	5	6.4
vbo/mavo	8	10.3
mavo/havo/vbo	20	25.6
havo/vwo	11	14.1
(i)vbo/avo/vwo	17	21.8
large (i)vbo/avo/vwo	12	15.4

◊Funding information is from teacher questionnaire responses. In the questionnaire, teachers identified schools with as many categories as applied. The total percent of all categories is therefore greater than 100 percent, and the sample size is greater than 78 schools.
NOTE: (i)vbo = (individual) junior vocational education; mavo = junior general education; havo = senior general education; vwo = pre-university education; avo = mavo and/or havo. Large (i)vbo/avo/vwo schools are those that serve all ability tracks and include a large number of students in the eighth grade. Because these schools had a different selection probability from those that were smaller, they were placed in a separate stratum.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

3.4.6 Switzerland Sample

Switzerland employed different sampling procedures for each of its three linguistic areas: German, French, and Italian.[7] Overall, the sampling of these three areas was proportionate to the student population: German (74.4 percent), French (21.8 percent), and Italian (3.8 percent). This distribution is based on estimates for the 9th-grade population of Switzerland from OECD (2001).

Switzerland analyzed its data by language group, resulting in a nationally-representative sample that is also statistically reliable for the French-, Italian-, and German-language regions.

[7] The country of Switzerland is divided into 26 cantons (the equivalent of provinces or states). Data for the German-language area were collected from 14 cantons, data for the French-language area was collected from 6 cantons, and data for the Italian-language area was collected from 1 canton. The language of instruction in the lessons from each canton corresponds to the language predominantly spoken in that canton.

Table 3.6 shows the breakdown of the type of schools in the Swiss sample based on source of funding, and Table 3.7 shows the breakdown based on ability track.

Table 3.6. Number and percentage distribution of the Swiss school sample, by language area and source of funding: 1999

School sector type	Number	Percent
Total	140	100.0
German-language area		
Public	72	51.4
Private religious	2	1.4
French-language area		
Public	39	27.9
Italian-language area		
Public	27	19.3

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Table 3.7. Number and percentage distribution of the Swiss school sample, by language area and ability track: 1999

Ability track	Number	Percent
German- language area	74	100.0
Basic requirements	26	35.1
Extended requirements	38	51.4
Highest requirements	10	13.5
French- language area	39	100.0
Basic requirements	5	12.8
Extended requirements	20	51.3
Highest requirements	14	35.9
Italian- language area◊	27	100.0
Basic requirements	9	33.3
Extended requirements	18	66.7

◊For the Swiss-Italian schools, ability tracking is within schools. That is, each school contains two tracks for mathematics classes.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

3.4.6.1 German-Language Area of Switzerland

Schools in the German- language area of Switzerland represent a subset of 133 schools that participated in (or were selected for) the TIMSS 1995 Achievement Study. Schools were sampled systematically from implicit strata. Schools were sorted on school type, stratum, canton, and number of 8th-grade students (the MOS), and a systematic PPS sample was selected from the ordered list. Fifty-one out of 75 originally sampled schools agreed to participate and 23 replacement schools were used, yielding a final sample size of 74 schools. Mathematics lessons were filmed in all of the selected classrooms.

3.4.6.2 French-Language Area of Switzerland

Classes in the French-language area of Switzerland were sampled systematically from class lists assembled by canton. Because these were selected from each canton by equal probability, the MOS value for each class is 1.

The sample allocation for each canton was based on the population proportion of that canton. Thirty-seven out of 41 originally sampled schools agreed to participate. Although 40 schools were initially sampled, one class in the sample was actually two classes. Because this was a random selection, and because the class division that affected this class is common-place in the population, it was decided to keep the “extra” class as a valid sample selection and increase the target sample size from 40 to 41 classes. The original 40 plus the “extra” class are considered the full complement of original selections for participation rate purposes.

Two replacement schools were used, yielding a final sample size of 39 schools. Mathematics lessons were filmed in all of the selected classrooms.

3.4.6.3 Italian-Language Area of Switzerland

In the Italian-language area of Switzerland all eligible schools in the population (i.e., all schools in the district of Ticino) were selected for the video study. Twenty-seven out of 35 of the schools in the population agreed to participate. Mathematics lessons were filmed in all of the selected classrooms.

3.4.7 United States Sample

3.4.7.1 United States Sample Selection Process

To select schools in the United States, first a sample of 52 geographic Primary Sample Units (PSUs) was selected from a frame of PSUs that represented the entire country. PSUs were defined to be counties or groups of counties. Ten of the PSUs, the ten largest Metropolitan Statistical Areas (MSAs), were included with certainty. Among the next largest 12 MSAs, a systematic random sample of 6 was selected. The remaining sample of PSUs consisted of a stratified sample of 36 other areas—18 MSAs, and 18 groups of counties from outside metropolitan areas. These 36 PSUs were stratified by geographic region, metropolitan/non-metro, and characteristics such as proportion of adults with college degrees and size of minority populations (the exact stratification characteristics varied by region and metro/non-metro status). The 36 PSUs were selected with probability proportional to population size as reported in the 1990 U.S. Population Census.

The PSU sample design is very similar to that used for the TIMSS 1995 Video Study, and is identical to that which was used for sampling schools for the TIMSS 1999 Achievement Study. In the TIMSS 1999 United States sample, different schools were used for the achievement and video study and, except for the 16 large metro areas, the PSUs for the two study samples were different.

A systematic PPS sample of 110 schools was chosen from the 13,261 schools that taught grade 8 in the selected PSUs. The schools were selected from a list sorted by region, urban/rural status (and, in the case of the 16 largest PSUs, central city/suburban status), type of school (public/private), and school size. Approximately two schools were selected from each PSU, but larger metropolitan PSUs may have had more than two schools selected, while from other PSUs a single school was selected.

The primary purpose for including the PSU stage of selection was to ensure that the sample for the TIMSS 1999 Video study was in different schools, and to a large extent in different school districts, from the TIMSS 1999 Achievement Study sample, the IEA Civics Education Study sample, and the National Assessment of Educational Progress Long-Term Trend and 2000 Field Test samples. All of these studies were taking place in national samples of grade 8 schools across the nation during the 1998–1999 school year.

Replacement schools were not used in the United States sample. Instead, an extra 10 schools were selected along with the 100 schools to increase the likelihood that 100 schools would participate in the study. It was likely that some selected schools would not be eligible. As it turned out, one school had ceased operations and another only had one student in the eighth-grade.

Eighty-nine out of the 108 eligible, originally sampled schools agreed to participate. Mathematics lessons were filmed in 83 of the selected classrooms. (Science lessons only were filmed in 6 schools.)

Table 3.8 shows the breakdown of the type of schools in the U.S. sample, based on source of funding.

Table 3.8. Number and percentage distribution of the United States sample, by source of funding: 1999

School sector type	Number	Percent
Total	83	100.0
Public	75	90.4
Private	8	9.6

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

3.4.7.2 United States Special Circumstances

A classroom in the United States was required to have at least three students to be eligible for selection into the lesson sample. Classrooms with fewer than three students could not be combined with an adjacent classroom. Two sampled schools were identified as ineligible after data collection efforts began. One school was closed permanently and another had only one student in the grade 8 mathematics and science classes.

When the available classes included a combination mathematics and science lesson which was team taught, one mathematics class was randomly selected from all classes teaching math, then a science class was selected following the procedures described above. In one case, the team-taught class was randomly selected and videotaped. This class fulfilled the requirement for one mathematics and one science lesson.

3.4.7.3 United States Refusals to Participate

The project director, other TIMSS 1999 Video Study senior staff members, and/or NCES contacted districts or schools that seemed disinclined to participate. They explained in detail the TIMSS 1999 Video Study, provided background information, answered questions, and discussed any concerns districts or schools had (see appendix B for copies of the information provided to U.S. superintendents, principals, and teachers). For some schools, the promise of a presentation about the video study by the project director was offered to encourage participation. In other cases, just providing reassurance of school and teacher confidentiality was enough to secure cooperation in the study.

Some districts and schools that had originally refused were later re-contacted and then accepted. In some cases, this was because a new superintendent or principal had come on board who had a more favorable opinion of the study. In other cases, this was simply because things were less hectic and they had more time to consider participating. In two cases, it was necessary to offer schools more money for compensation of lost time before they would agree to participate. Teachers in the United States were generally given $300 as compensation for their participation in the study, and a videotaped copy of their lesson.

Participation in the TIMSS 1999 Video Study in the United States required obtaining consent at three levels: district superintendent, school principal, and teachers. Of the sample of eligible respondents, 11 superintendents refused to allow the sampled school in their district to participate in the video study, 8 principals refused to allow their school to participate in the study. Furthermore, 14 mathematics teachers and 9 science teachers refused to participate. The tables below summarize the number of schools and teachers that accepted or refused.

Table 3.9. Participation status of United States schools, by type of school: 1999

Participation status	Total schools	Public schools	Private schools	Department of Defense schools
Number in original sample	110	96	13	1
Number accepted	89	81	8	0
Number refused	19	14	4	1
Number ineligible	2	1	1	0
U.S. school response rate – unweighted percentage	81	84	62	0

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Table 3.10. Participation status of United States teachers, by field of teacher: 1999

Participation status	Total mathematics teachers	Total science teachers	Total teachers
Number of original sample of teachers contacted	97	97	194
Number accepted	83	88	171
Number refused	14	9	23
U.S. teacher response rate – unweighted percentage	86	91	88

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

U.S. district superintendents' reasons for refusing to participate included the following:

- Superintendent did not want to be part of a video study;
- Committee of superintendents declined; they were not allowing any more research projects for the school this year;
- Superintendent said teachers had too many student assessments and tests to be able to accommodate an additional study; and
- No explanation provided.

U.S. principals' reasons for refusing to participate included the following:

- A private, religious school did not allow visitors on the campus;
- Principal said no teachers were interested;
- Principal said teachers were too busy ("this is not a good year to do something new"); and
- Principal said two teachers were new and were not comfortable being videotaped.

U.S. teachers' reasons for refusing to participate included the following:

- Teacher was not interested or not willing to participate;
- Teacher was too busy;
- Teacher was not comfortable being videotaped;
- Concern about consent forms; and

- No explanation provided.

3.4.7.4 United States Teacher Consent/Waivers

All participating teachers (and parents of the students in the class) were required to sign a consent form (see appendix C for a sample of the United States consent form). This consent/waiver released the usage of their lesson video for research purposes. In most instances, the school principal was relied upon to provide teachers with an introductory packet that described the study and invited them to participate. The consent/waiver form was contained in the packet, and was supposed to be signed and returned to LessonLab prior to the videotaping. Although most teachers confirmed their participation through the principal, very rarely did they return their consent/waiver in a timely fashion.

Often the videographer would attempt to collect the teacher's waiver on the day of the taping. In many of these instances, the teacher reassured the videographer that the waiver had already been signed and returned. However, after the fall of 1999, it was discovered that 28 lessons had been videotaped without the teachers' signed consent/waivers on file. After senior staff contacted these teachers and addressed any issues, all 28 teachers agreed to sign and return their waivers. All videotapes collected in the United States have the requisite consent/waivers on file.

3.5 Summary

All of the TIMSS 1999 Video Study countries were required to include at least 100 schools in their initial selection of schools. In this chapter, information was provided regarding the participation rates in each country, as well as their sampling strategies. Also presented were details regarding the nature of the participating schools within each country, including the types of schools in the sample by funding source and/or ability track.

Chapter 4. Data Collection and Processing

4.1 Data Collection

4.1.1 Nature of Data Collected

The primary focus of the data collection for this study was the videotaping of a full mathematics and/or science lesson in each sampled classroom. What counted as a lesson was determined by what was standard in each participating country.

Additional data were also collected to help understand the videotaped lesson more fully. This additional data included:

- A teacher questionnaire;
- A student questionnaire;
- Photocopies of text pages, worksheets, overhead transparencies, and other materials used in the lesson; and,
- A log sheet that videographers completed after each taping session.

4.1.2 Data Collection Schedule

The National Research Coordinators in each country were responsible for scheduling the videotaping and ensuring that taping was evenly distributed throughout the school year. LessonLab scheduled the videotaping of classrooms in the United States. As an added check, the receipt control system at LessonLab tracked the proportion of lessons that arrived from each country on a monthly basis, to ensure there was not a disproportionate number of tapes collected during any given month.

Most of the data collection took place in 1999. In some countries filming began in late-1998, and in other countries filming began in 1999. Data collected ended in either late-1999 or mid-2000 in order to sample lessons across the academic year in each country. Table 4.1 lists the start and end dates for data collection in the six participating countries.

Table 4.1. Data collection periods, by country: 1998–2000

Country	Data collection start date	Date collection end date
Australia	5/24/99	12/3/99
Czech Republic	11/26/98	10/20/99
Hong Kong SAR	3/11/99	12/13/99
Japan◊	11/94	3/95
Netherlands	12/8/98	3/22/00
Switzerland	5/1/99	7/31/00
United States	1/26/99	5/18/00

◊Japan did not participate in the mathematics portion of the TIMSS 1999 Video Study. Japanese data were collected as part of the TIMSS 1995 Video Study during the 1994-1995 school year. Information on Japanese data collection dates is from Stigler et al. (1999).
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

4.1.3 Number of Mathematics Lessons Filmed

Detailed sampling procedures are described in chapter 3. Table 4.2 provides the final sample size of mathematics lessons that were included in the study.

Table 4.2. Number of mathematics lessons included in the study, by country: 1999

Country	Number of mathematics lessons
Australia	87
Czech Republic	100
Hong Kong SAR	100
Japan[1]	50
Netherlands	78
Switzerland[2]	140
United States	83

[1]Japanese mathematics data collected in 1995.
[2]Seventy-four lessons were included from the German-language area of Switzerland, 39 lessons were included from the French-language area of Switzerland, and 27 lessons were included from the Italian-language area of Switzerland.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

In some countries, additional mathematics lessons were filmed but not included as part of the final sample for various reasons.

- In Hong Kong SAR, 9 additional lessons were filmed but not used in the study because they were not evenly distributed throughout the school year (see chapter 3, sections 3.3 and 3.4.3)

- In Switzerland, two lessons were filmed but then disqualified. In one case a seventh-grade class was inadvertently selected and filmed. Once the error was discovered, an eighth-grade class from the same school was selected and filmed. In another case a teacher decided to revoke her permission to participate in the study.
- In the United States, one lesson was filmed but not used in the study because it was not from a sampled teacher. Another lesson was excluded from the database because the sound and visual quality were too poor to be coded. The class was re-filmed at a later date.

4.1.4 Number of Questionnaires Collected

Each videotaped teacher was given a questionnaire to complete after the videotaped lesson, as was each student in the class. Response rates on the teacher questionnaire were between 96 and 100 percent, and response rates on the student questionnaire were between 83 and 97 percent, as shown in table 4.3. More complete information about the development and nature of the questionnaires is available in chapter 5.

Table 4.3. Teacher and student questionnaire response rates, by country: 1999

Country	Teacher questionnaire response rate (unweighted)		Student questionnaire response rate (unweighted)	
	Percentage	Sample size	Percentage◊	Sample size
Australia	100	87	83	1,942
Czech Republic	100	100	90	2,133
Hong Kong SAR	100	100	97	3,560
Netherlands	96	75	93	1,733
Switzerland	99	138	94	2,485
United States	100	83	89	1,623

◊Percentage represents the number of student questionnaires received out of the total number of students reported by the teacher to be enrolled in each class. Questionnaires were collected only from the students who were in attendance at the filming. Student questionnaires were received from all videotaped lessons except for 1 lesson in Australia, 4 lessons in the Netherlands, 3 lessons in Switzerland, and 3 lessons in the United States. All data are unweighted.

NOTE: Japan did not participate in the mathematics portion of the TIMSS 1999 Video Study.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

4.2 Data Receipt and Processing

4.2.1 Data Receipt

Data receipt was a collaborative effort between the data tracker, national research coordinators, and videographers. The following section describes the procedures that were put into place to ensure that all data was properly collected and accounted for.

All instruments (including teacher questionnaires, student responses, and instructions) were custom-produced for each participating country. Instruments were not produced or distributed until their translated versions met the approval of all parties. In Hong Kong SAR, for example, it was later agreed that both English and Cantonese questionnaires would be provided to teachers and students.

All videotapes, completed questionnaires, and ancillary materials were sent to LessonLab in Los Angeles for processing. Once final approval was granted and the data collection instruments were produced, the data tracker and national research coordinators agreed upon the best distribution and collection procedures to ensure that all data would arrive safely at LessonLab. These procedures took into consideration each country's geographical parameters, postal methods, and customs regulations.

Data collection procedures then were carefully reviewed with each videographer before entering the field. As part of the data collection process, all videographers were required to complete log sheets and assess each taping session; a copy is included as part of the TIMSS 1999 Video Study Data Collection Manual in appendix D. These log sheets provided the following information:

- The camera operator's name;
- The lesson's field identification number;
- Date and time that the lesson was taped;
- Whether any problems were encountered during videotaping; and,
- Any other comments by the videographer that might be useful for understanding the videotape.

Once the videotapes were received at LessonLab, the staff digitizer reviewed the videographer's log sheet and noted any technical problems they might have encountered. Such technical problems included:

- a classroom's air conditioner or poor acoustics interfered with the microphone's audio reception;
- the camera stopped operating due to battery failure;
- a brief time lapse while the videographer inserted an additional tape for a double-length lesson; and,
- a student accidentally unplugged the adapter cord causing the camera to stop.

In almost all cases, these difficulties were minor, and digital adjustments could be made to ensure their use in the study. Also, having two tapes of each lesson (from the teacher and student camera) helped to minimize the impact of these problems.

As data were collected from the field, the data tracker confirmed receipt of all data with each country's collaborator. Follow-ups were often needed with participating teachers in order to retrieve missing materials.

4.2.2 Data Processing

A sophisticated database management and tracking system was specially created for this study. This receipt control system was used to document the receipt and processing of video, teacher and student questionnaires, and other materials collected from the lessons. All processing of the data was entered into the system, and could be tracked daily. The major components of the system included:

- Document receipt and management component—for entering and maintaining receipt and process status of each videotape, questionnaire, and additional materials (e.g., videotape transcription/translation, digitizing, and coding status; questionnaire scanning and coding status; additional material scanning status);
- Process control component—for tracking the dates and status of data by a variety of classifications;
- Sample management component—for maintaining identification and links to teachers, schools, and countries;
- Management reporting component—which provides pre-specified reports;
- Database management component—which allows for importation, backing up, and archiving of all field data; and,
- Query processing component—for ad hoc queries.

Each lesson was identified by subject (mathematics or science), country, and classroom. This information was included in a bar code label that was applied to all the relevant materials for that lesson, including the two videotapes, the teacher questionnaire, each student questionnaire, and any other additional materials. Thus, the lesson, the teacher, and the students were linked by this unique identification number (LESSONID).

The LESSONID was then recorded in the receipt control tracking system. It contained information on the lesson subject, country, and teacher as shown in table 4.4 below.

Table 4.4. Information included in each LESSONID: 1999

Column	Identification	Code
1	Subject	M = Mathematics S = Science
2–3	Country	AU = Australia CZ = Czech Republic HK = Hong Kong SAR JP = Japan NL = Netherlands SW = Switzerland US = United States
4–6	Lesson	Number of lesson, range = 001–100

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

4.3 Videotaping Procedures in Classrooms

What is seen on video is dependent not only on what transpires in the classroom, but also on the way the camera is used to film the classroom. To achieve comparability across countries and classrooms, all camera operators must make similar filming decisions. Decisions on how to standardize filming also have an effect on what can be coded from the tapes. The TIMSS 1999 Video Study developed and used an explicit protocol for the collection of lesson videos and the operation of two cameras. Part of this protocol was developed during the field test study, details of which can be found in chapter 2, section 2.3.

4.3.1 Using Two Cameras

In the TIMSS 1995 Video Study, one camera was used to videotape lessons. However, in the TIMSS 1999 Video Study two cameras were used in order to obtain more detailed information on student behavior. One camera (the "teacher camera") focused primarily on the teacher, and was operated manually by a videographer. The videographer also used this camera to capture close-ups of the chalkboard or overhead screen, objects shown or used in the lesson, students' notebooks or worksheets during periods of private work, and teacher/student interactions during private work.

A second camera (the "student camera") was placed high on a tripod near the front of the room, positioned with a wide angle to include as many students as possible. The main goal of this camera was to capture students' interactions with the teacher and/or each other during the lesson. The student camera facilitated coding of the mathematics instruction, for example by reducing the number of inferences coders had to make about what students were doing in response to teacher talk and action, or to what student behaviors the teacher was referring.

The following statements guided camera operation and are excerpted from the TIMSS 1999 Video Study Data Collection Manual (see appendix D):

Two Camera Strategy

We are using two cameras in this study. The first camera will be operated by the videographer. It will be placed on a tripod, but will also be removed from the tripod whenever it is necessary to document the lesson. It will generally be placed between one third and one half of the way back from the front of the class, and will more often than not focus on the teacher and his or her zone of interaction. We will refer to this camera as the "teacher camera."

The second camera will be stationary. It typically will be placed up high on a tripod along a side wall near the front of the room, set to the widest shot possible, and used to capture as many students in the classroom as possible. We will refer to this as the "student camera."

The physical arrangement of classrooms and the activities that take place within them vary greatly. The videographer must decide where to place the cameras so that the documentation requirements outlined above can be met to the greatest possible extent. It is helpful, if possible, to talk with the teacher before the class begins to find out generally what is going to happen, and where the action will take place. The camera should be placed so that it can easily tape the main chalkboard or audiovisual device, the teacher, and some of the students in a single master shot. The position should also allow for easy panning to other areas of the classroom.

4.3.2 Videotaping Equipment

As described in chapter 2 section 2.2, Canon Optura mini DV camcorders were used for data collection based on their feasibility, as determined by the field test study. Two different tripods were used: the Mathews THM20 fluid head tripod was use for the teacher camera, and the Promaster 6400 photography tripod was used for the student camera.

Sound quality is another critical factor to take into account when studying classroom processes. Three microphones were used when filming each lesson: a wireless microphone attached to the teacher, a "shotgun" microphone set on top of the teacher camera, and a built-in microphone on the student camera. The wireless microphone included a Lectrosonic omni-directional lavaliere microphone, transmitter, and receiver. For the teacher camera, the Sennheiser K6P, a high quality professional microphone, was used. For the student camera, the Canon ZM100 zoom microphone was used.

Videographers carefully monitored the sound levels throughout the lesson, and made adjustments as necessary.

4.3.3 Basic Principles Guiding the Cameras

Camera operators were trained to film each classroom as if they were an observer watching the lesson. With regard to the teacher camera, they were instructed to carefully document the teachers' activities and behaviors during the lesson. For example, they were asked to capture what the teacher was doing and saying, and what information was being presented to the class.

Most of the time, the videographer decided where to focus the camera by taking the perspective of a model student in the class. In other words, if the teacher was lecturing, the camera focused on the teacher; if the teacher pointed to something on the chalkboard, the camera focused on the relevant portion of the chalkboard.

The teacher camera also filmed what students did and said during whole class instruction. Videographers focused mainly on the activities and behaviors of the students who were interacting with the teacher, but occasionally panned the classroom to see what other students were doing as well. During private work, the videographer was instructed to follow the teacher and document the activities and behaviors of the students who were interacting with the teacher. When possible, the videographer zoomed in on students' work to see how students were doing the assigned tasks.

The teacher camera was usually placed along a side wall in the classroom, about one third to half way from the front of the room. During whole class interaction, the camera was typically mounted on a tripod, but during private work it was often hand-held in order for the videographer to follow the teacher.

Videographers were trained to use a "master shot" most of the time, defined as the shot that gives the most encompassing view of the whole scene. This shot is thought to be the least subject to bias by both the videographer and the viewer, because it gives the most information about the context within which the action occurs. Also commonly used were "medium shots" and "close-ups" to capture information written on the chalkboard, overhead projector, or in students' notebooks.

The following three principles that guided the data collection procedures are excerpted from the TIMSS 1999 Video Study Data Collection Manual (see appendix D):

> <u>Document the Teacher</u>
>
> During lessons teachers engage in a variety of activities. For example, they explain concepts and procedures, pose problems, assign tasks, ask questions, write information on the chalkboard, walk around the classroom and assist individual students, etc. Because the main goal of this project is to study teaching in different countries, it is necessary that we thoroughly and carefully document the teacher's activities and behaviors during the lesson. Therefore, what the teacher is doing, what he/she is saying, and what information he/she is presenting to the class must be captured.
>
> <u>Document the Students</u>
>
> In order to understand what goes on in the classroom, it is important to know what the students are doing as well as what the teacher is doing. The main focus should be placed on the activities and behaviors of the students when they are interacting with the teacher, but what they do when they are not working with the teacher should also be documented from time to time. Although it is not possible to document everything that every student does or says, the goal here is to sample student behavior so that what is portrayed in the videotape is representative of what actually happened in the lesson.

Document the Tasks

During mathematics and science lessons, teachers assign various tasks to students. Normally the teacher presents the task to students clearly enough so that students understand what they are supposed to do, and it is usually not hard to see in the video what the assigned task is. Sometimes, however, students may actually engage in something that is not what the teacher intended. Also, if the class is broken into small groups, each group may be working on a different task.

In all cases, what we want to see in the video is the task that students are actually engaged in doing, whether or not it is what the teacher intended. To see clearly what students are doing it is often necessary to zoom in close enough to capture what at least a few of the students are working on.

4.3.4 What to Do in Common Situations

The 638 eighth-grade mathematics lessons in the sample had a wide variety of classroom arrangements and instructional situations, and a great deal of thought was given to handling these in a systematic way. Table 4.5 is an excerpt from the TIMSS 1999 Video Study Data Collection Manual (see appendix D) detailing relatively common teaching situations and how they should be filmed.

Table 4.5. Descriptions of common teaching situations and how they should be filmed: 1999

Descriptions of possible situations	What to do
• Teacher talks publicly • One student at the board works on a problem and talks publicly • Rest of the class works individually at their seats	• Focus on the teacher and the student at the board • Find a chance to document what other students are doing
• Teacher walks around assisting students privately and talks to the whole class from time to time • One student at the board works on a problem • Rest of the class works individually	• Document how the teacher instructs individual students • Document the student at the board and the information on the board when there is a chance
• Teacher stays at his/her desk assisting students privately • Rest of the class works on their own	• Document how the teacher instructs individual students (move close to them) • Document what other students are doing
• Students are in groups, and each group works on the same task • Teacher walks around assisting each group	• Document how the teacher assists each group (follow the teacher) • Document some groups when the teacher is not with them
• Students are in groups, and each group works on different tasks • Teacher walks around assisting each group	• Document how the teacher assists each group (follow the teacher) • Document the work of each group

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

4.3.5 Training Videographers

Videotaping procedures were field tested and modified prior to their implementation in the TIMSS 1999 Video Study. At least two videographers in each country were employed, and all completed a several day training session conducted by members of the TIMSS 1999 Video Study team. The training sessions involved learning about the goals of the study, getting familiar with the selected video equipment, going over the data collection manual (see appendix D) and practicing filming lessons in both mock and actual classrooms. In addition, videographers received regular feedback about their filming techniques, beginning with a review of their practice lessons.

To ensure that videographers were using the standardized procedures they had been taught, their first two mathematics and science lessons filmed were evaluated on six dimensions by specially trained country associates:[8] camera positioning, documenting the entire lesson, audio, documenting information, framing shots, and camera movements. These critiques were videotaped, sent to the videographers, and discussed as needed. Subsequent lessons by each videographer were also examined and evaluated on these same six dimensions, as discussed in the next section.

4.3.6 Monitoring Quality

Several procedures were used to monitor the quality of the videotapes. Prior to filming, videographers checked over their equipment, recharged their batteries, etc. Then, during filming videographers continually monitored the sound and video quality from each of the two cameras. After filming was completed, videographers checked the sound and visual quality of tape. Additionally, a specially trained technician checked the sound and visual quality before processing the tape. Very few tapes had major problems. Only one lesson had serious enough problems with the video and audio problems that it could not be used (see section 4.1.3 above). Other tapes with minor technical problems, such as low sound levels, could be adjusted digitally.

As a further quality control measure, lessons numbered 001 and then every tenth numbered lesson from each country (i.e., 001, 010, 020) was critiqued along the six dimensions listed above by specially trained country associates. Whenever problems were noted, they were discussed with the videographers. In all cases, the problems were minor (e.g., get closer shots of the blackboard, remember to film your digital watch when first starting the cameras so that footage from the teacher and student cameras can be easily synchronized) and only required the videographers to make minor adjustments in future videotaping sessions. Regular correspondence between the videographers and the mathematics code development team was maintained throughout the data collection period.

[8] Country associates were international representatives from each country, who comprised the code development team. For more details on the country associates, see chapter 6 and appendix H.

4.4 Constructing the Multimedia Database

4.4.1 Digitizing, Compressing, and Storing

Once data were collected they were sent for processing to LessonLab. The videotapes were sent to LessonLab directly by the videographers, immediately after they were filmed. The teacher and students' questionnaires were first sent from each teacher to the National Research Coordinator in their country, who then forwarded them to LessonLab. Teachers in the United States sent their questionnaires directly to LessonLab.

The two videotapes from each lesson (i.e., from the teacher and student cameras) were compressed into MPEG-1 format and stored on a video server. The questionnaires and supplementary materials (such as text pages, worksheets and tests) were scanned and stored digitally in PDF format. As a back-up measure, the video files and supplementary materials for each lesson were burned onto a single CD-ROM disk. Then, the videos and supplementary materials were entered into vPrism, a multimedia database software developed for the TIMSS 1995 Video Study and enhanced especially for the TIMSS 1999 Video Study.

English transcripts for each lesson were created and also entered into vPrism software, as described in the next section. One important feature of this software is that the transcripts can be linked by time code to the video. Viewers can watch either the teacher or student tape, and see a running English translation. They can also enter codes into the vPrism database, and directly access the supplementary materials. English transcripts were also created for each open-ended response in the teacher questionnaires and entered into Excel files for code development. Closed-ended responses on the teacher and student questionnaires were entered into separate Excel files.

4.4.2 Transcribing and Translating Lessons

Once the videotaped lessons were digitized, entered into the multimedia database, and made accessible through the network server, translators/transcribers (henceforth referred to as "transcribers") carefully reviewed each one and produced a full English transcript of all classroom interaction audible on the tape.

English transcripts linked to the video afforded code developers and coders the opportunity to view lessons from all of the countries during code development, coder training, and times when assistance was needed to make difficult coding decisions. Furthermore, the transcripts could be "exported," either in their entirety or in specified portions, so they could be used by specialists.

4.4.2.1 __Hiring Transcribers and Translators__

Transcribers were hired on the basis of their fluency in both English and in the language of the instruction being studied. For the TIMSS 1999 Video Study, the languages of instruction were Australian English, Cantonese, Czech, Dutch, Japanese, and American English. Most transcribers were educated in the country whose lessons they translated and transcribed. A mathematics or science background was also a strong determinant of hire.

4.4.2.2 <u>Developing and Training Standardized Procedures</u>

Each transcriber participated in a two-week training period, during which they learned the TIMSS 1999 Video Study transcription convention requirements, as well as the operation of the specialized vPrism transcription/timecoding software. For instance, transcribers were taught rules about how to indicate speakers, how to break speech into turns, how to use punctuation in a standardized manner, and how to translate technical terms in a consistent way. Details of these procedures may be found in the TIMSS 1999 Video Study Transcription and Translation Manual, included as appendix A.

Each videotaped lesson was processed and reviewed by two transcribers prior to its final processing and review by the transcription manager. Every audible utterance by the teacher and students was translated into English from the original language by the "first-pass" transcriber, who reviewed the lesson in its entirety up to three times before passing it on to the "second-pass" transcriber.

Upon receiving the preliminary "first-pass" transcript, the "second-pass" transcriber compared the transcript line-by-line with the videotaped lesson, making any necessary corrections or additions. Once the transcript was fully corrected, the second-pass transcriber separated the utterances into segments that were no more than three typed lines in length, as defined by the software. This was done in preparation for timecoding the utterances, which enables the transcript to be linked to the video and displayed as English subtitles. Finally, the second-pass transcriber "time-coded" the transcript, linking the videotape (in hours, minutes, seconds and frames) with the corresponding in-point (i.e., start point) of the transcribed utterance.

4.4.2.3 <u>Quality Control Measures</u>

Having been reviewed in its entirety up to six times (three by each transcriber), each lesson was then submitted to the transcription manager, who performed an overall review of the transcript by checking the grammar, spelling, punctuation, formatting, and the accuracy of timecode/utterance synchronization. As an additional quality control measure, completed transcripts were selected at random and checked line-by-line by the transcription manager, with the assistance of the transcriber. Any concerns about transcription procedure were discussed with the individual transcriber and at monthly transcription department meetings.

4.5 Summary

This chapter covered a number of issues related to the collection and processing of the TIMSS 1999 Video Study mathematics data. Videotapes, questionnaire responses, and other supplementary materials were processed at LessonLab using a sophisticated database management and tracking system. Both videographers and transcribers followed well-defined protocols in order for videotaping and transcription/translation procedures to be standardized across countries. Specific quality control measures were in place to carefully monitor both groups. Video data and corresponding English transcripts were entered into vPrism, a multimedia database software developed for the TIMSS 1995 Video Study and enhanced especially for the

TIMSS 1999 Video Study. Coders used this software to watch the teacher and student tapes, see running English translations, access supplementary materials, and enter codes into the database.

Chapter 5. Questionnaire Data

5.1 Development of the Teacher and Student Questionnaires

5.1.1 Teacher Questionnaire

5.1.1.1 __Purpose of the Questionnaire__

The purpose of the teacher questionnaire was to elicit information that would provide important background for the analysis and interpretation of the videotapes. The information collected from teacher responses was used in two ways.

1) Coders used the information from the questionnaire to make better judgments about what they saw on the videotapes. For instance, sometimes coders needed to know the teacher's goal for a lesson in order to make sense of the activities that constituted the lesson. As another example, when coders segmented the lessons into periods of review, new, and practicing/applying new material, it helped coders to refer to questionnaire items where the teacher identified what students had previously been taught and what they were expected to learn from this lesson.

2) Information from several questionnaire items also was used to assess the typicality of the lesson captured on videotape. Although teachers were instructed not to prepare in any special way for this lesson, what transpired on the day of the taping was potentially not typical compared to what normally happens in a given classroom. An atypical lesson may have resulted from nervousness on the part of the teacher, excitement on the part of the students, or some special event not connected to the TIMSS 1999 Video Study. Furthermore, questionnaire responses might identify a sampling bias. For example, if teachers reported that the lessons were "stand-alone" lessons rather than part of a curricular series, this information could indicate an atypical lesson.

5.1.1.2 __Constructing the Mathematics Questionnaire__

Constructing the teacher questionnaire was a multi-step process that took place over several months. Two parallel versions of the teacher questionnaire were developed at the same time for the mathematics and science teachers participating in the TIMSS 1999 Video Study. The following section describes this process in detail for the mathematics version.

Step 1: Review of the TIMSS 1995 Video Study Questionnaire

The TIMSS 1995 Video Study teacher questionnaire was reviewed by project staff. Items that were particularly useful were highlighted, and those that did not provide the anticipated information were examined more closely. In addition, project members involved with coding and reporting data in the 1995 Video Study identified those questions that had produced the most helpful information. They then discussed the limitations of other questions that were less successful. Two questions guided the item analysis: (1) Did this item measure an issue that was important to retain in the TIMSS 1999 Video Study questionnaire? (2) If yes, does the item need to be rewritten to achieve the intended purpose?

The TIMSS 1995 Video Study teacher questionnaire was modestly revised and used in the TIMSS 1999 Video Study field test (see chapter 2).

Step 2: Review of Teacher Questionnaires from Other Studies

The teacher questionnaires used in the TIMSS 1995 Achievement Study, as well as other recent education research projects, also were reviewed. Items that fit into the designated domains were considered for inclusion in the TIMSS 1999 Video Study questionnaire.

A set of decision-making guidelines emerged during this process. The guiding questions were:

1) Does the item help to better understand the context of the videotaped lesson?
2) Does the item help to make judgments about the relationship of the videotaped lesson to current thinking about mathematics or science teaching?
3) Could the question be answered simply by looking at the videotape?

Step 3: Drafting and Revising the Questionnaire

Information generated from the discussions mentioned above was used to draft new sample items for the TIMSS 1999 Video Study teacher questionnaire. The new items were reviewed and a first draft of the revised teacher questionnaire was created. An additional guideline was added: Don't change an item on the TIMSS 1999 Video Study teacher questionnaire unless a strong case can be made for the necessity of a change.

The questionnaire development team and additional project members (including the Project Director and the Chief Analyst) reviewed the draft. The questionnaire was electronically sent out for review to collaborators in each of the participating countries. They were asked to provide feedback to the development team by July 6, 1998.

Questionnaires from the teachers who participated in the TIMSS 1999 Video Study field study test were translated into English (when necessary) and reviewed.

Based on these reviews, the team revised the questionnaire and created the "final" version.

Step 4: Creating and Translating the Final Version

A panel, consisting of NCES personnel and mathematics and science education experts, reviewed all items on the questionnaires for consistency, clarity, and utility. Suggested revisions from the National Research Coordinators, education experts, and NCES were incorporated into the final versions of the teacher questionnaires.

A panel at the Office of Management and Budget (OMB) also reviewed and approved the questionnaire. The OMB review panel suggested randomly sorting the teacher attitude items in question 57 and reversing the direction of several items in this question. Revisions were made and reported to OMB (see section 5.1.3).

National adaptations of the questionnaire were made, according to the requests of each participating country (see section 5.1.4)

5.1.1.3 __Questionnaire Item Justification__

The final version of the questionnaire asked mathematics teachers to provide additional information about the videotaped lesson, their background and experience, attitudes, and professional development. A copy of the questionnaire provided to U.S. teachers is included as appendix E. The questionnaire included the seven domains listed below. In this section, a rationale is provided for each domain.

- The videotaped lesson;
- The larger unit or sequence of lessons;
- The typicality of the videotaped lesson;
- Ideas that guide teaching;
- Educational background, teaching background, and teaching load;
- School characteristics; and
- Attitudes about teaching.

The Videotaped Lesson

This section of the questionnaire was designed to gather contextual information about the lesson recorded on videotape. Some information necessary for understanding the lesson might not have been evident from simply watching the videotape. These background items were collected in this section of the questionnaire.

<u>Content of the Lesson</u>
Item 1: Knowing the teacher's definition of the content of the lesson facilitates interpretation of the tape. This information is especially helpful when the teacher has content goals in mind that might not be immediately obvious to the coder.

Items 2 and 3: These questions elicit the sources of influence on the content of the videotaped lesson. Item 2 asks if there is an external document or textbook that played a major role in the teacher's decision to teach this content. Access to the relevant document could provide insight into how the teacher interpreted these materials and how they influenced the teaching of the lesson. Item 3 requests the name of such documents.

Item 4: This question elicits important information about the sources of influence on the teacher's lesson. These sources may influence the teacher's ways of understanding and representing the content as well as providing him/her with ideas about pedagogical strategies. In addition, the question (especially in combination with other questions) provides glimpses into the teacher's tendency to collaborate with colleagues and the teacher's ways of thinking about students.

Item 5: This question serves two purposes. First, it provides an outline of the content of the lesson from the teacher's perspective. Secondly, it clarifies which content is new to students and which is review. This is important in making coding decisions about the nature of the lesson activities (e.g., whether an activity contains new information).

Intended Student Learning

Items 6 and 7: Knowing the teacher's intended goal facilitates interpretation of the tape (item 6). The goal of the lesson also may explain differences in observed instruction. Item 7 provides teachers opportunity to highlight portions of lessons they considered problematic and explain why. Coders might use this information to understand the lesson.

Item 8: This question about the teacher's perception of resource limitations gives teachers a chance to express which additional resources would have improved the lesson. In addition, the item enables cross-national comparisons of perceived resource needs.

Teacher Planning

Item 9: This question helps assess typicality of planning for taped lesson. Although the teachers are instructed to plan and teach the lesson just as they would normally do, some teachers may put in extra time planning this lesson.

Items 10 to 14: The ability levels of the students will not be known from the videotape. Thus, items 10 and 11 help us learn whether the teacher put students in groups according to ability or other reasons. These items give a rough indication of the mix of students working together in the small groups. Also, because schools in different jurisdictions may or may not use "tracking" which we cannot infer from the tapes, questions 13 and 14 will help us identify such practices. Teaching techniques may differ according to ability level of the students; these questions alerts coders to any special quality of this particular group of students and their abilities.

Items 15 to 18: These items will indicate what kinds of preparation students have had for the videotaped lesson. A classroom activity may serve a different purpose for students who are already familiar with materials used in the lesson than it would for students who are seeing the material for the first time.

Assessment

Items 19 and 20: Assessment tasks provide important windows into teacher thinking. In particular, assessment tasks reflect a great deal about the kind of learning that is valued by the teacher (factual, conceptual, procedural, etc.). The assessment also can help us evaluate the alignment among the teacher's goals, the teacher's instructional practices in the lesson, and the assessment.

The Larger Unit or Sequence of Lessons

The questions in this section asked the teacher to place the videotaped lesson in the context of a larger unit or sequence of lessons. These questions were important for three reasons: (1) standards documents in mathematics education describe good teaching as connected and as developing student conceptual understanding across time; (2) the data could be used to make judgments about the teacher's views about teaching; and (3) in combination with the videotaped lesson, the data could enable the construction of a more complex view of teaching. Thus, teacher responses to these questions provided data on teaching that supplemented data provided by the videotape alone (without incurring the enormous cost of videotaping additional lessons).

Items 21 through 25: Placing the videotaped lesson in the context of a sequence of lessons helps clarify the teacher's goals and purposes before and beyond the videotaped lesson. Is the content development in this lesson closely linked to other lessons? How does the teacher think about content and the development of ideas over time? How long are the sequences?

Item 26: This item helps assess the typicality of the video lesson and places the video lesson in a broader context, and provides insights into the teacher's thinking about effective mathematics teaching. Requesting teachers to describe the lesson in words commonly used in her/his nation provides information on cultural differences in types of lessons in each nation.

The Typicality of the Videotaped Lesson

This section of the questionnaire is designed to gather information about the typicality of the lesson that was videotaped. The study will generate descriptions of mathematics teaching that are deemed typical in each country. It is important to know whether the lessons that are videotaped are indeed typical.

Items 27 to 31: These questions address the important issue of whether the instruction recorded on the videotapes is judged as typical by the teacher. Typicality ratings are elicited for teaching methods and student participation. The teacher will also be asked to describe any aspects of the lesson that were not typical. Analysis will examine differences in judged typicality across countries. National portraits of what is marked as atypical will be used to moderate interpretation of findings.

Item 32: This item is designed to assess the effect that being videotaped had on the teacher.

Ideas that Guide Teaching

This section of the questionnaire was designed to provide insights into the teacher's knowledge and personal views of good mathematics teaching.

Item 33: This item that identifies teachers' broadest instructional goals for the school year provides a measure of teacher's knowledge and attitudes toward current thinking about mathematics, and her/his own teaching philosophies.

Items 34 to 37: These questions were designed to assess teachers' response to and awareness of current ideas about how to teach mathematics in the classroom. The teacher's self-rating is complemented by information about how they acquired this information and their list of familiar documents.

Items 38 and 39: This item asks teachers to describe a particular part of the lesson in relation to reform concepts that provides information on how teachers define these concepts.

Items 40 and 41: These questions serve as an indicator of the teacher's involvement in professional development activities that are consistent with peer collaboration and observation recommended in standards and reform documents.

Educational background, teaching background and teaching load

Items 42 to 51: These items inquire about the teacher's pre-service and subsequent preparation for teaching and for teaching specific subject matters.

Item 52: This item asks teachers to identify how much time is spent preparing to teach and doing other school-related work.

School Characteristics

Item 53 to 56: These questions ask for a basic description of the school including size, type, how students are admitted, number of teachers of mathematics or science, and grade levels. Teacher responses indicate whether or not the school has any special status that might contribute to the nature of the observed teaching. For example, students at a magnet school might receive a different kind of mathematics instruction or have access to more resources than a traditional school.

Attitudes about Teaching

This section provides information on the teacher's attitudes towards teaching mathematics. The items suggest ways in which the teacher thinks about her/his work, the students, and mathematics. It is important to examine the satisfaction of teachers since this factor might be associated with differences among teachers.

Items 57a, b, n to p: These items explore the teacher's attitudes towards teaching mathematics

Items 57k to n, q: These items probe the teacher's attitudes towards and interest in mathematics.

Items 57b, m, r to x: These items examine the teacher's attitudes towards students.

Items 57d to h: These items probe the teacher's attitudes towards professional development and growth.

Items 57c, f, i to k: These items explore the teacher's feelings of satisfaction with working conditions.

Items 57k, y, z, aa, bb: These items explore the teacher's feelings of being appreciated and respected.

Items 58 and 59: NCES added these items to assess knowledge of and participation in implementing the National Council of Teachers of Mathematics (NCTM) standards.

5.1.2 Student Questionnaire

The student questionnaire was designed to elicit basic demographic characteristics of the students (such as age and gender), the home environment, and educational expectations of students participating in the videotaped lesson. The United States student questionnaire consisted of 12 closed-ended questions. A copy of this questionnaire is included as appendix F. Contingent upon receiving NCES approval, each country could revise their student questionnaire to make the items nationally appropriate. Australia, Hong Kong SAR, and Switzerland included the full set of questions, the Czech Republic included 6 of the questions, and the Netherlands included 11 of the questions.

5.1.3 Approval of Questionnaires

The first versions of the TIMSS 1999 Video Study teacher and student questionnaires were designed to provide an opportunity for individual countries to make modifications to some questions or response options in order to include the appropriate wording or options most consistent with their own national education systems. These versions of the teacher and student questionnaires were approved by NCES and the OMB review panel on November 16, 1998. Each country revised the questionnaires as needed. These national adaptations of the questionnaires then were reviewed by the national research coordinators and the country associates, and requested revisions were sent to NCES for approval. Data collection in a country did not begin until approval of that country's teacher and student questionnaires was received. The following table presents the dates final versions of the questionnaires were approved for the participating countries.

Table 5.1. Dates of approval for national versions of questionnaires, by country: 1998–1999

Country	Date of Approval
Australia	4/12/99
Czech Republic	11/16/98
Hong Kong SAR	1/15/99
Netherlands	11/16/98
Switzerland	12/17/98
United States	11/16/98

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

5.1.4 National Modifications of the Questionnaires

Most items in the teacher and student questionnaires were common to the questionnaires of all participating countries. Table 5.2 indicates the number of open- and closed-ended questions in each country's teacher questionnaire.

Table 5.2. Number of items in the teacher questionnaire, by country: 1999

Country[◊]	Open-ended questions	Closed-ended questions
Australia	27	31
Czech Republic	25	32
Hong Kong SAR	26	32
Netherlands	23	32
Switzerland: German-speaking	25	29
Switzerland: French-speaking	25	32
Switzerland: Italian-speaking	23	31
United States	27	32

[◊]Japanese mathematics teacher questionnaire data were collected as part of the TIMSS 1995 Video Study. Those results can be found in Stigler et al (1999).
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

There are four general types of options for adapting the questionnaires to the purposes of each participating country. The first of these was a translation option in which countries were asked to translate terms and expressions into the local idiom if they thought it necessary. For the most part, these were minor wording changes to such things as the format for recording dates. Translations of the questionnaires were completed in each participating country.

The second type of option required that a country consider the nature of a concept defined internationally and then develop country-specific names for the items, or even country-specific indicators for a particular concept. The specification of the name of the national curriculum guide is an example of the first kind. Developing country-specific indicators for the multi-item measure of family wealth exemplifies the second kind of translation task in this option category.

Another option that allowed for the inclusion of questions in the national questionnaires was encouraged but not obligatory. The last option was for a country to include questions with particular national relevance.

Table 5.3. Modifications and additions to teacher and student questionnaire items, by country: 1999

Country	Translations or minor word changes	Modified with country-specific names	National options added	Questions not applicable nationally
Teacher questionnaire:				
Australia	8F, 8I	4L, 4M, 13, 14, 42, 51, 54, 55	57CC, 57DD	2B, 4N
Czech Republic	—	42, 54	—	44 to 47
Hong Kong SAR	—	4L, 4M, 42, 54	—	
Netherlands	—	54, 55	—	13, 14
Switzerland		42, 51, 54	—	2B, 4N, 13, 14, 44 to 47
United States	—	—	58, 59A-D	—
Student questionnaire:				
Australia	—	3 to 5, 8 to 11	13	—
Czech Republic	—	8, 9	—	3 to 6, 10, 11
Hong Kong SAR	—	4 to 6, 8 to 11	—	3
Netherlands	—	4 to 6, 8 to 11	—	3
Switzerland	—	3 to 6, 8 to 11	—	—
United States	—	—	—	—

— Not applicable
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

5.2 Coding of Open-Ended Items in the Teacher Questionnaire

5.2.1 Code Development

The teacher questionnaires consisted of both closed-ended (or forced-choice) and open-ended (or free-response) items. Open-ended questions were appropriate for this study because of its cross-cultural nature, which made it especially difficult to anticipate the possible range of teachers' responses. Teachers were expected to spend approximately 45–50 percent of the time for filling out the questionnaires responding to the open-ended questions.

The 32 open-ended items on the teacher questionnaire required development of quantitative codes to analyze the responses. Teachers' answers to these questions were translated into English by coders who were bilingual in English and one of the other relevant languages. Coding of the data then was carried out using the English translations by a team headed by the Chief Analyst of the TIMSS 1999 Video Study.

The open-ended questions were partitioned into two types: short-answer questions and extended-response questions. Short-answer items required teachers to provide a brief response to a question. For example, "What materials are you aware of that describe current ideas about the teaching and learning of mathematics?" Extended-response items required teachers to provide a more lengthy and detailed response to a question. For example, "What was the main thing you wanted students to learn from the videotaped lesson?"

Separate codes for each open-ended item were developed using a four-phase process. First, before examining teachers' responses, categories of anticipated responses were developed based on current research in mathematics teaching and learning and advice from subject matter specialists. This part of the process helped the code developers (1) form a common interpretation of the question, (2) identify categories that may not be provided in the teachers' responses, and (3) address culturally specific issues, such as the meanings of phrases used in the different countries.

Second, categories were further developed based on the responses from the first 10 mathematics teacher questionnaires received from each country. Teachers' actual responses were used in the code development process because they allowed codes to reflect the variety of comments possible as well as teachers' interpretations of the questions. The process of within country and then across country category development was selected so that the categories created would retain responses unique to a country.

Third, codes were created using the categories generated in the preceding two phases considering frequencies of responses, the cultural significance of a code, and the importance of a category in understanding teachers' beliefs and goals. Comparing categories in this way ensured that the codes for the free-response items reflected the different educational systems of the study as well as current understandings of teaching and learning.

Fourth, the codes were checked for reliability. Using these results, the codes were further revised and then applied to the remainder of the questionnaires.

5.2.2 Reliability

Codes developed for the free-response items are described in detail in the TIMSS 1999 Video Study Mathematics Teacher Questionnaire Coding Manual (see appendix G). Inter-rater reliability was established on all of the open-ended items that were coded. For each item, two coders independently coded 10 randomly selected lessons from each country. An 85 percent inter-rater reliability criterion was used. If an 85 percent level was not achieved initially, discrepancies were discussed and necessary modifications were made to the code definition. Reliability was then attempted on a different, randomly selected set of lessons. This procedure is similar to reliability procedures used in the TIMSS 1995 Achievement Study to code students' responses to the open-ended assessment tasks (Mullis et al. 1998: B-32).

Table 5.4 lists the reliability scores for each of the open-ended questionnaire items that were coded. In each case, reliability was calculated as the percentage of agreement between coders.

Table 5.4. Reliability estimates for eighth-grade mathematics teacher questionnaire open-ended response codes: 1999

Teacher questionnaire item	Item reliability (percent)
Name of the videotaped course	98
Other subject matter content of the videotaped lesson (TQ1u)	97
Other materials used when planning this lesson (TQ4o)	100
Ideas that were mainly review and new to students (TQ5)	86
Main thing students should learn from the videotaped lesson (TQ6)	89
Nature of the class size limitation (TQ8d)	100
Basis by which students were assigned to groups (TQ11)	98
What students were expected to do for homework (TQ16)	98
How students will be assessed (TQ20)	98
What was different from how you normally teach (TQ28)	94
How you hear about current ideas (TQ36)	92
Part of the lesson that exemplified current ideas (TQ39)	87
Teaching certification (TQ43)	88
Undergraduate major (TQ44)	89
Undergraduate minor (TQ45)	97
Graduate major (TQ46)	95
Graduate minor (TQ47)	100
How are students admitted to the school (TQ55)	98

NOTE: Inter-rater agreement was calculated as the number of agreements divided by the sum of the number of agreements and disagreements.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

5.3 Questionnaire Analyses

As presented in the international report on the mathematics results (Hiebert et al., 2003), teacher responses to the questionnaires were used in the following ways:

1. *Computation of national-level univariate statistics.* Distributions by country provided information on the basic characteristics of the sample.

2. *Help interpret observed classroom data.* Coders used qualitative analyses of selected questionnaire items to enhance interpretations of the videotape.

Most of the analyses of teacher responses to the questionnaire that are presented in *Mathematics Teaching in Seven Countries: Results from the TIMSS 1999 Video Study* (Hiebert et al. 2003) include comparisons of means or distributions across six countries for questionnaire data. In all cases, the lesson was the unit of analysis. Analyses were conducted in two stages. First, means or distributions were compared across all available countries using either one-way ANOVA or Pearson Chi-square procedures. Variables coded dichotomously were usually analyzed using ANOVA, with asymptotic approximations.

Next, for each analysis that was significant overall, pairwise comparisons were computed and significance determined by Bonferroni adjustment. The Bonferroni adjustment was made assuming all combinations of pairwise comparisons. For continuous variables, Student's *t* values were computed on each pairwise contrast. Student's *t* was computed as the difference between the two sample means divided by the standard error of the difference. Determination that a pairwise contrast was statistically significant with $p<.05$ was made consulting the Bonferroni *t* tables published by Bailey (1977). For categorical variables, the Bonferroni Chi-square tables published in Bailey (1977) were used.

A significance level criterion of .05 was used for all analyses. All differences discussed in *Mathematics Teaching in Seven Countries: Results from the TIMSS 1999 Video Study* (Hiebert et al. 2003) met at least this level of significance, unless otherwise stated. Terms such as "less," "more," "greater," "higher," or "lower," for example, are applied only to statistically significant comparisons.

All tests were two-tailed. Statistical tests were conducted using unrounded estimates and standard errors, which also were computed for each estimate.

The analyses reported in *Mathematics Teaching in Seven Countries: Results from the TIMSS 1999 Video Study* (Hiebert et al. 2003) were conducted using data weighted with survey weights, which were calculated specifically for the classrooms in the TIMSS 1999 Video Study (see chapter 8 for a more detailed description of weighting procedures).

5.4 Summary

To help understand and interpret the videotaped lessons, questionnaires were collected from teachers and students in each lesson. The teacher questionnaire was designed to elicit information about the professional background of the teacher, the nature of the mathematics course in which the lesson was filmed, the context and goal of the filmed lesson, and the teacher's perceptions of its typicality. The construction of this questionnaire was an elaborate process, and justifications for each item are reported in this chapter.

The student questionnaire was designed to elicit basic demographic characteristics of the students, their home environment, and their educational expectations. Both the teacher and student questionnaires were approved by a review panel, and then country appropriate versions were created under the direction of the national research coordinators in each country.

The teacher questionnaire contained a number of open-ended items, for which a coding scheme was developed and applied. Reliability statistics are presented.

Chapter 6. Coding Video Data I: The International Mathematics Team

This chapter describes the coding of the video data by the International Mathematics Coding Team. First, background is provided on the personnel involved, including the code development team, advisory groups, and the coders. Next, details on the code development process are provided, along with information about each code. Methods used to train coders, measure reliability, and ensure quality control are also described.

6.1 Coding Personnel

6.1.1 Code Development Team

An international team was assembled to develop codes to apply to the TIMSS 1999 Video Study mathematics data. The team consisted of country associates (bilingual representatives from each country) and was directed by a mathematics educator. The mathematics code development team worked closely with two advisory groups: a group of national research coordinators representing each of the countries in the study, and a steering committee consisting of five, North American, mathematics educators. Refer to appendix H for a list of the code development team members, the national research coordinators, and the steering committee members.

6.1.1.1 Mathematics Code Development Team

Each country participating in the project was represented by a country associate, who was fluent in the language and well versed in the cultural background of the country. The country associates served as representatives for their countries, providing reminders of the diversity of instruction and challenging the coding system to account for them. Furthermore, the representatives provided an "insider's" interpretation of events in the videotaped classrooms. Thus, the impressions of both cultural insiders and outsiders were considered when developing codes. Additionally, the country associates helped to hire and manage coders for their country, and could aid them in making coding decisions that might involve cultural or linguistic nuances.

The country associate team was headed by a mathematics coordinator who directed the code development effort, analyses, and reporting of data. The associate director of the 1999 Video Study also played an active role in the mathematics code development team by participating in conceptualizing and defining codes, and guiding analyses and reporting of the data.

As a group, the mathematics code development team was responsible for creating and overseeing the coding process. The team discussed coding ideas, created code definitions, wrote a coding manual, gathered examples and practice materials, designed a coder training program, trained coders and established reliability, organized quality control measures, consulted on difficult coding decisions, and managed the analyses and write-up of the data.

6.1.1.2 National Research Coordinators

A national research coordinator was designated for each of the participating countries. These coordinators were all from academic or research institutions in their own country, and were also involved in the TIMSS 1995 and/or TIMSS 1999 Achievement Studies. As national research coordinators of the 1999 Video Study they played several roles. On an operational level, they organized the data collection in their country (i.e., designing a sampling procedure tailored to their country, selecting schools, modifying the questionnaires for teachers and students in their country, contacting teachers, and scheduling videotaping). They served also as advisors throughout the study. Meetings including the mathematics code development team and national research coordinators were held at least once each year throughout the project to discuss the progress made to date and to gather input on pertinent tasks such as developing research questions, defining specific codes, and reviewing the data.

Several national research coordinators made independent visits to LessonLab during the life of the project, and contributed to the ongoing code development process. Additionally, the coordinators served as hosts to country associates when the latter visited their "home" country for meetings with educators and teachers.

The coordinators also occasionally convened groups of experts in their country, to perform tasks requested by the code development team. These experts were individuals identified as being particularly knowledgeable about mathematics and education in their country and as being interested in cross-cultural video research. For example, experts were asked to help in the development of hypothetical teaching models in each country (see chapter 2) and to review particular code definitions.

6.1.1.3 Steering Committee

A North American mathematics steering committee was convened, composed of a diverse group of individuals who represented a cross-section of interests within mathematics education. All members of the steering committee were based in the United States or Canada. The steering committee met with the mathematics code development team yearly; these meetings sometimes overlapped with the national research coordinator annual meetings. Steering committee members reviewed and commented on research priorities, identified research questions, provided input on code definitions, and reviewed tables and drafts of the final report.

6.1.2 Coders

6.1.2.1 International Coding Team

Most of the videotape coding was conducted by an international group of specially trained coders at LessonLab. Similar to membership on the mathematics code development team, members of the international video coding team represented all of the participating countries. They were fluently bilingual so they could watch the lessons in their original language, and not rely heavily on the English transcripts. In almost all cases, coders were born and raised in the country whose lessons they coded. Many had a particular interest in education, teaching, and/or mathematics.

In general, two videotape coders from each country were employed. Collaboration between coders, particularly those from the same country, was encouraged. Coders also interacted closely with the country associates throughout the coding period.

For all countries except Switzerland, hiring, training and coding took place at LessonLab. Swiss training and coding took place in Zurich, Switzerland. There was frequent communication between the Swiss coders and the LessonLab team, and in particular with the Swiss country associate who was based at LessonLab. Daily or weekly electronic and telephonic communication was used to ensure equivalence between the Swiss and LessonLab operations. For most training sessions, the Swiss country associate traveled to Zurich, explained the codes to the coders there, and led them through the initial reliability process. In addition, several members of the Swiss research team traveled to the United States and spent an extended period of time at LessonLab participating in the code development process.

6.1.2.2 Specialist Coders

Most of the codes presented in the report *Teaching Mathematics in Seven Countries: Results from the TIMSS 1999 Video Study* (Hiebert et al., 2003) were applied by members of the international video coding team. As noted above, these individuals were cultural insiders and fluent in the language of the lessons they coded. However, not all of them were experts in mathematics or teaching. Therefore, several specialist coding teams with different areas of expertise were employed to create and apply codes regarding the mathematical nature of the content, the pedagogy, and the discourse. Several of these specialist teams made use of coding and tables prepared by the international video coding team. The work of the specialist coding teams will be discussed in the next chapter.

6.2 Code Development Process

6.2.1 Developing a Coding Scheme

The mathematics code development team, with the aid of national research coordinators and the steering committee, developed a guiding set of research questions and a framework for constructing individual codes. Strategies for code development were sensitive to the twin goals conceived early in the project: to describe the nature of teaching within each country, and to compare teaching across all countries. Although code development strategies for achieving these goals were not conflicting, they required somewhat different approaches.

Both strategies outlined above were implemented by constructing individual codes that reliably captured important features and segments of lessons. To begin the process, the mathematics code development team consulted instruments and coding protocols used in previous studies of teaching, including the TIMSS 1995 Video Study, and textbooks and curriculum materials from each participating country. Codes that would answer research questions regarding the nature of teaching within each country and/or the differences and similarities in teaching across countries were defined, piloted, and refined. As this work revealed new insights into teaching within and across countries, the set of research questions was revised and new codes were suggested.

To capture the nature of teaching within each country, the mathematics code development team began with the conclusion of the TIMSS 1995 Video Study—that there are unique cultural patterns of teaching mathematics in each country. At the beginning of the TIMSS 1999 Video Study, cultural “insiders” (including the country associates, the national research coordinators, and mathematics educators) developed hypotheses about specific instructional patterns that might be found in eighth-grade mathematics classrooms in their country. These hypotheses took the form of “country models” (see section 6.2.2 below) and were continually revisited to ensure that each country's perspective on teaching was considered as individual codes were constructed.

6.2.2 Field Test and Constructing Tentative Country Models

In early 1998, at least four mathematics field test lessons were collected in each country. These videotapes of eighth-grade mathematics classrooms provided an initial opportunity to observe teaching in the different countries in the sample. An international group of representatives[9] met together for an entire summer, viewing and reflecting on these tapes. They followed a structured protocol throughout this period, with the intention of generating hypotheses that could later be tested by quantitative analyses of the full data set (see chapter 2 for more details on the field test study.)

These discussions yielded six dimensions that the representatives agreed framed classroom practice and were of interest across countries and lessons: purpose, classroom routine, actions of participants, content, classroom talk, and classroom climate. The dimensions were then used to create country models—holistic representations of a “typical” mathematics lesson in each country. The hypothesized country models were developed in collaboration with the national research coordinators, steering committee members, and other colleagues in each country, and refined over a period of several months.

The goal was to retain an “insider perspective,” and faithfully represent the critical features of teaching in each country in the coding system. The country models served two purposes toward this end. First, the models provided a basis on which to identify key, universal variables for quantitative coding. Second, they described a larger context that might be useful in interpreting the coding results.

---

[9] Most of these representatives continued in the role of code developer.

The hypothesized country models for Australia, the Czech Republic, Hong Kong SAR, the Netherlands, Switzerland, and the United States are presented in tables 6.2 through 6.7. Table 6.1 provides a key to the symbols used in these models. A country model was not created for Japan.

Table 6.1. Key to symbols and acronyms used in hypothesized country models

Symbol/Acronym	Meaning
T	Teacher
S	Student
Ss	Students
HW	Homework
BB	Blackboard
‖: :‖	Segment may repeat

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Figure 6.1. Hypothesized country model for Australia

	Review	Introduction of New Material	Assignment of Task	Practice/Application & Re-instruction			Conclusion
				practice/ application	reassignment of task	practice/application	
Purpose	reinforce knowledge; check/correct/ review homework; re-instruct	acquisition of knowledge;	assignment of task	application of knowledge	assignment of task	application of knowledge	reinforce knowledge
Classroom Routine	review of relevant material previously worked on	presentation of new material	assignment of task	completion of task	assignment of task	completion of task	summary of new material; assignment of homework
Actions of Participants	T – [at front] ask Ss questions; elicit/embellish responses; demonstrate examples on BB	T – [at front] provides information asking some Ss questions and using examples on BB	T – [at front] describes text book/ worksheet task	T – [roams room] provides assistance to Ss as needed and observes Ss progress on set task	T – [at front] re-explains text book/ worksheet task	T – [roams room] provides assistance to Ss as needed and observes Ss progress on set task	T – [at front] provides information & asks Ss questions
	Ss – [in seats] respond to & ask T questions; listen to T explanations, watch demonstrations	Ss – [in seats] listen to T explanations and respond to T questions	Ss – [in seats] listen to T descriptions	Ss – [in seats] work individually or in pairs on task	Ss – [in seats] listen to T descriptions	Ss – [in seats] work individually or in pairs on task	Ss – [in seats] listen to T descriptions; respond to & ask T questions
Content	related to previous lesson	definitions/ examples building on ideas previously worked on	description of task; focus on text/worksheet problems	text/worksheet problems	description of task; focus on text/ worksheet problems	text/worksheet problems	text/worksheet problems; homework problems
Classroom Talk	T talks most; Ss one-word responses	Mix of T/S talk although discussion clearly T directed	T provides direct instructions	mix of T/S and S/S talk – including explanations & questions	T provides direct instructions	mix of T/S and S/S talk – including explanations & questions	mix of T & T/Ss talk – including explanations & questions
Climate	somewhat informal - relaxed yet focused						

NOTE: Refer to table 6.1 for a key to the symbols and acronyms used in this table.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Figure 6.2. Hypothesized country model for the Czech Republic

Purpose	Review			Constructing New Knowledge			Practice	
	Evaluating	**Securing old knowledge**	**Re-instruction**	**Activating old knowledge**	**Constructing new topics**	**Formulating the new information**	**Practice "working - through"**	**Using knowledge in different problems**
Classroom Routine	oral exam test homework	set of problems; homework	dialogue	Experiment, solving problems, demonstration, dialogue	dialogue		dialogue, solving problems	solving problems
Actions of Participants	T – giving grade	T – gives individual help	T – explaining procedure			T – writing notes at the board		
	Ss – solving problems at the board	Ss – at the board		Ss – answering questions, solving problems at the board			Ss – solving problem; more then one student solving one problem	
Content	content probably from unit			special problems prepared in special order, solutions are very visible, strong connection with new topics	step by step solving problem, solutions very visible	mathematical statements and definitions; something new that students don't know		stronger connection with real life
Classroom Talk	answering questions; fast pace		T talk most	T-S dialogue	T talks most of the time; slow pace	T talks most of the time		more mathematically open questions
Climate	few mistakes allowed Ss very quiet serious atmosphere	mistakes are not graded but not expected, Ss talk loudly					more mistakes allowed	

NOTE: Refer to table 6.1 for a key to the symbols and acronyms used in this table.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Figure 6.3. Hypothesized country model for Hong Kong SAR

	Review	Instruction	Consolidation		
Purpose	To review material and prepare for the present lesson	To introduce and explain new concepts and/or skills	To practice the skills learned		
Classroom Routine	T – goes over relevant material learned in the past, sometimes through asking Ss questions	T – introduces a new topic T – explains the new concepts/skills T – shows one or more worked examples	Seat-Work T – assigns seat-work Ss – work on seat-work T – helps individual Ss	Evaluation T – asks some Ss to work on the board T – discusses the work on the board with Ss	Homework T – assigns homework Ss – start doing homework
Actions of Participants	T – talks at the blackboard	T – explains at the blackboard T – works examples on the blackboard	T – talks at blackboard T – walks around the class	T – discusses Ss' work on the board	T – talks at blackboard
	Ss – listen in their seats Ss – answer questions from their seats	Ss – listen and/or copy notes at their seats	Ss – listen and then work in their seats	Ss – listen in their seats Some Ss work on the board	Ss – listen in their seats
Content	Usually low demand of the cognitive processes	Higher demand in the cognitive processes Definitions/proofs/examples Heavy reliance on textbook	Medium demand on the cognitive processes Select exercises Focus on procedures or skills		
Classroom Talk	T – talks most of the time Pace relatively fast Convergent questions by T Conversation evaluation	T – talks most of the time Pace relatively slow Mostly convergent questions and some divergent questions Less evaluative	Some informal S talk (with each other) Pace relatively slow		
Climate	Serious Relatively quiet Mistakes less acceptable	Serious Relatively quiet Mistakes more acceptable	Less serious Less quiet Mistakes more acceptable		

NOTE: Refer to table 6.1 for a key to the symbols and acronyms used in this table.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Figure 6.4. Hypothesized country model for the Netherlands

Purpose	**Re-instruction**		**Instruction**		**Assignment of Task**	**Students Attempt Problems**
Classroom Routine	**Going Over Old Assignment**		**Presenting New Material**		**Assignment of Task**	**Student Problem Solving** Continued work on old assignment and/or initial efforts on new assignment
Actions of the Participants	*Option 1* Completion of each problem as a class	*Option 2* T gives hints for selected problems	*Option 1* T verbalizes	*Option 2* Complete reliance on text	T – writes assignment on BB (or may give verbally)	If "Re-instruction" follows Option 2, Ss first work on the old assignment, then work on the new T – available to answer S-initiated questions T – gives mostly procedural assistance T – generally provides answers freely; doesn't require much S input T – may give semi-public assistance (at front of room) or private assistance (at Ss desks)
	T – goes through assignment, problem by problem at the front of the class, with or w/o use of the BB; Emphasis is on procedures	T – provides partial assistance (e.g., hints) on selected problems at the front of the class; T – provides answers on paper (e.g., answer sheet, access to T manual); Emphasis is on procedures	T – verbalizes text presentation and/or points to selected features of the text presentation	None		
	Ss – follow along at their desks, respond to T questions, ask clarifying questions	Ss – follow along at their desks; Very low S involvement	Ss – listen to T at their seats	Ss – read about new topic(s) from the text, at their desks	Ss – write assignment into their agendas	Ss – work in pairs at their desks and ask T for assistance when necessary, either at their desks or at the front of the room
Content	Small number of multi-part problems from the text; Assignment given yesterday and worked on as HW; Generally one solution method provided		Heavy reliance on text; new material presented within the context of a task/problem		Small number of multi-part problems from the text (~5) to be continued tonight as HW; Ss only need to find one solution method (any one solution is O.K.)	
	Problems are in a real world context (might be considered "application"), situations vary across tasks, T rarely solicits errors					
Classroom Talk	*Option 1* T asks Ss questions and rephrases Ss' responses	*Option 2* T briefly gives partial information on selected problems; Ss rarely ask questions. Less S talk than in *Option 1*	*Option 1* Direct instruction	*Option 2* None	Direct instruction; T verbalizes the assignment as written on the BB	S-S talk regarding assignment; 1-on-1 (or 2 to 3-on-1) private, S-T conversations initiated by S, but then dominated by T
	Low level of evaluation/low concern for assessment					
Climate		High level of S freedom and responsibility		High level of S freedom and responsibility		Moderate level of noise is accepted by T
	High error tolerance by the T, T-S relationship is relaxed					

NOTE: Refer to table 6.1 for a key to the symbols and acronyms used in this table.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Figure 6.5. Hypothesized country model for Switzerland: Classroom patterns of Swiss mathematics lessons *with* introduction of new knowledge

Purpose	Opening	Construction of new cognitive structure	Working-through	Practice (automatization, rehearsal)	Practice (automatization, rehearsal)
Classroom Routine	Collecting homework, informal talk	Interactive instruction[1] T – presents 'real action' T – models problem solving	Interactive instruction Problem solving	Ss write or read at their desks Interactive instruction	
Actions of Participants		T – asks questions and explains, demonstrates procedure, or states a problem…	....	....	
		Ss – answer questions, observe T, imitate, act, solve problems; work as a whole class	(See Notes)[3] Ss – work as a whole class	Ss – individual, group, or pair work	
Content		New concept is introduced in a step by step fashion, starting from Ss previous knowledge and/or their everyday experience Goal: Ss understand the concept (on their level of knowledge); Usefulness of concept (for further learning, and as a tool for everyday practice) emphasized; Visualization (*Anschauung*) is important; New information is reinforced (presented at board or textbook in a standardized fashion); Relationship between tasks: no set (often: Problem-like situation)	Sequence of carefully selected tasks related to new topic Relationship between tasks: Set 2[4]	Collection of tasks related to new topic Relationship between tasks: Set 1	.... Relationship between tasks: Set 1, 2, …
Classroom Talk		*Lehrgespraech*[2] (Interactive instruction; long wait-time, Ss expected to actively participate in construction process)	Interactive instruction	T-S-dialogue	
Climate					

Figure 6.5. Hypothesized country model for Switzerland: Classroom patterns of Swiss mathematics lessons *with* introduction of new knowledge—Continued

[1]The introduction phase may include some further actions that may be embedded in the interactive instruction, such as teacher presentation, or modeling or "real actions."
[2]Most frequently a new topic (concept) might be co-constructed by means of interactive instruction (*Lehrgespraech*). The means of guidance are primarily teacher questions and hints. The procedure is oriented toward the Socratic dialogue. The teacher questions serve two main purposes: (1) to guide and initiate students' thinking (e.g., propose a certain point of view, or perspective on a problem), and (2) to diagnose students' actual understanding. An important feature of quality of a *Lehrgespraech* is the need for sufficient wait-time after the teacher's questions.
[3]Reform 1: In reform-oriented classrooms another pattern of introduction lessons might be expected: (1) student independent problem solving in pairs, groups or individually (inventing procedures for solving new, open problems, discovering principles, regularities, and so on); (2) discussion of the different approaches and negotiating an accepted approach. This approach (influenced by scholars of mathematics didactics in Germany and the Netherlands) is presently recommended in teacher education and professional development. (It is unclear if this is observable at the eighth grade level.)
[4]Not all students always solve the same tasks (individualization of instruction).
NOTE: Refer to table 6.1 for a key to the symbols and acronyms used in this table.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Figure 6.6. Hypothesized country model for Switzerland: Classroom patterns of Swiss mathematics lessons *without* introduction of new knowledge

Purpose	**Opening**	**Working-through or practice** goal: understanding and/or proficiency	**Practice** goal: understanding and/or proficiency	**Re-Instruction, sharing**	**Practice** goal: understanding and/or proficiency	**Re-Instruction, sharing**	**Using knowledge in different situations/to solve diff. problems**
Classroom Routine	Collection of homework, informal talk	Interactive instruction	Ss solve tasks[3]	Share and check Ss' solutions (*Besprechung*) - interactive instruction - S presentation - discussion	Ss solve tasks	Share and check Ss' solutions (*Besprechung*) - interactive instruction - S presentation - discussion	problem solving interactive instruction
Actions of the Participants		Classwork[1]	Ss – individual, group, or pair work	Classwork	Ss – individual, group, or pair work	Classwork	Ss – individual, group, or pair work
Content		Topic: introduced in a previous lesson T may start with short review of topic, and solve some examples of tasks Relationships between tasks: no set, or Set 1 or Set 2[2]	Relationships between tasks: Set 1 or Set 2	Relationships between tasks: Set 1 or Set 2	Progression to more demanding tasks, finally: to demanding application problems (possibly not in the same lesson, but later)[4] Relationships between tasks: Set 1 or Set 2	Relationships between tasks: Set 1 or Set 2	Character of tasks: Given new situations but connection to mathematics concepts is not obvious Relationships between tasks: Set 2 or no set
Classroom Talk		Interactive instruction	T-S-dialogue, and/or S-S-conversation	Interactive instruction/Discussion	T-S-dialogue, and/or S-S-conversation	Interactive instruction/Discussion	T-S-dialogue, S-S-conversation, discussion…
Climate							

Figure 6.6. Hypothesized country model for Switzerland: Classroom patterns of Swiss mathematics lessons *without* introduction of new knowledge—Continued

[1]The sequence of activity units varies, and does not always start with a classwork phase.
Reform 2: In some reform classrooms there will be no or almost no classwork phase and each student may be proceeding through a weekly assigned collection of learning tasks (arranged in collaboration with the teacher; individualized instruction). As with Reform 1, it is not clear if and how many teachers are in fact practicing this reform model of instruction (which is recommended in teacher development) at the eighth-grade level.
[2]Not all students always solve the same tasks (individualization of instruction).
[3]As a general pattern an alternation between students solving tasks at their own and of sharing/checking/re-instruction based on students' work in a classwork sequence may be expected, but the duration of and total amount of the phases is not predictable. The first unit may provide some special kinds of tasks (warm up, or a motivating starting task). In most cases, the teacher will vary the social structure (e.g., classwork – individual work – classwork – pair work – and so on).
[4]There is a progression from easier to more demanding tasks over the entire learning phase; usually the progression leads to application problems (most often, applied story problems).
NOTE: Refer to table 6.1 for a key to the symbols and acronyms used in this table.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Figure 6.7. Hypothesized country model for the United States

Purpose	**Review of Previously Learned Material** ***A***			**Acquisition of Knowledge** ***B***	**Practice & Re-instruction** ***C***	
	Assess/ evaluate	Assess/ Evaluate, re-instruct, secure knowledge	Secure knowledge, activate knowledge	• •		• •
Classroom Routine	Quiz *A1*	Checking homework *A2*	Warm-up/ Brief review *A3*	Presenting New Material *B*	Solving Problems (Not for homework OR for homework) *C1*	*C2*
Actions of the participants	T – tells or solicits answers T – at the front	T – tells or solicits answers T – may work through difficult problems T – at the front	T – tells or solicits answers T – may work through problems T – at the front	Information provided mostly by T T – tells students when, why, and how to use certain procedures T – asks short-answer questions T – may do an example problem T – at the front	T – Ss practice through example problems T – at the front	T – walks around the room T – provides assistance to Ss who raise their hands
	Ss – Students take quiz Ss – provide or check their answers Ss – at their seats	Ss – provide or check their answers Ss – at their seats Ss – may put their answers on the board	Ss – complete problem(s) Ss – provide or check their answers Ss – at their seats	Ss – listen & answer T's questions Ss – may work on an activity, as explicitly instructed by the T Ss – at their seats	Ss – help the teacher do the problems Ss – at their seats	Ss – work individually or in small groups at their seats Ss – may state their answers as a class
Content	Content related to previous lesson	Content related to previous lesson	Content may or may not be closely related to the new topic	Simple rules or definitions stated by T, focus is mostly on procedure (little reflection on concepts)	More problems very similar to what the T has just shown	More problems very similar to what the T has just shown
Classroom Talk	Known-answer questions, relatively quick pace, more S turns, T evaluates, recitation?[1]	Known-answer questions, relatively quick pace, more student turns, T evaluates	Known-answer questions, relatively quick pace, more student turns, T evaluates	Fewer student turns; Direct instruction and lectures are possible	Recitation, more S turns, direct instruction?	T-S dialogue, S-S dialogue (private talk)
Climate	T wants correct answers					Friendly atmosphere

Figure 6.7. Hypothesized country model for the United States—Continued

[1]Recitation = A series of short, known-answer questions posed by the teacher, to solicit correct answers from students. Consists mainly of Initiation-Response-Evaluation sequences.

NOTE: Refer to table 6.1 for a key to the symbols and acronyms used in this table.

An alternative U.S. classroom pattern occasionally exists that does not resemble this model. These are considered “reform” mathematics lessons. They typically consist of an open-ended problem posed by the teacher, a long period of seatwork during which the students work on the problem, and then a period of “sharing” when the students provide their answers and the teacher summarizes the key points.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

6.2.3 Deciding What to Code

Classroom lessons are filled with many activities and human interactions, more than can be described even when analyzing only a few lessons. The challenge is compounded when 638 mathematics lessons from seven countries are to be analyzed. Decisions must be made to focus the analysis and reduce the complexity. The mathematics code development team sharpened its focus by setting priorities among the six dimensions of classroom practice presented in the country models. Because this was a study of mathematics teaching, not generic teaching, the content dimension emerged as a dimension of special interest. Research on teaching and learning, results gleaned from the TIMSS 1995 Video Study, the field test lesson reviews, the nature of the country models, and the suggestions of the steering committee, all reinforced the initial focus on content.

The mathematics code development team quickly discovered that content in most eighth- grade mathematics lessons is carried through working on problems. Again, this consensus was reinforced through reading the research literature and discussions with the national research coordinators and other cultural insiders. Mathematics is taught in all participating countries largely through the use of problems (Hiebert et al. 2003).

Using the mathematical problem as a primary code provided a window into other questions of interest, such as what kind of mathematics was presented, who did most of the mathematical work, and what kind of work was done by students and by teachers. Segmenting lessons into mathematical problems paved the way to examine important aspects of the learning opportunities provided for the students.

6.2.4 Coverage and Occurrence Codes

An initial coding challenge was how to extract the mathematical problems from each lesson in order to examine them in greater detail. Problems usually were embedded in a variety of contextual elements and it was necessary to dissect the lesson in various ways in order to see how the problems were situated. The mathematics code development team ascertained that two kinds of codes would be useful in this process: coverage codes and occurrence codes.

Coverage codes parsed the entire lesson, or a specified part of the lesson, into non-overlapping segments. Every moment of the lesson, or specified part, was "covered" by one of the mutually exclusive and exhaustive categories. For example, a mathematics lesson could be segmented into periods of time when there was either: 1) no mathematical work, 2) mathematical organization or management, or 3) mathematical work. Then, the mathematical work time could be segmented into either working on problems or not working on problems.

Occurrence codes were used to identify the occurrence of a particular event, either within the lesson or within a specified part of the lesson. For example, a mathematics lesson might or might not have contained a goal statement. Similarly, a mathematical problem might or might not have been related to a real world context, or involved physical materials. Codes such as these were developed to describe how often events of interest occurred within lessons, problems, and other such segments.

6.2.5 Creating a Code Development Procedure

To ensure that each country associate provided input into the development of codes, a 6-step process was established (see table 6.2). This process both distributed the work across team members and encouraged their feedback and support.

Table 6.2. Six-step mathematics code development process

Step	Action
1	Full group of code developers held an initial discussion of particular research goals and questions, and generated ideas for relevant codes
2	Subgroup developed a preliminary proposal for a code, with alternatives
3	Full group discussed the alternatives, and made a decision about which option(s) to pursue
4	Subgroup developed a revised proposal, including definitions and examples
5	Full group tried out the code on sample lessons
6	Full group shared their results, revised the definition, and entered it into the coding manual

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

The first step in the process was for the entire mathematics code development team to review particular research questions and generate ideas for codes that might answer those questions. Alternatively, the team sometimes considered codes that had been used in other studies or were suggested by viewing the videotapes, and discussed whether they might address one or more research questions. These brain-storming sessions helped to determine the general nature of the codes and allowed everyone in the group to contribute ideas.

The second step in the code development process was to establish a subgroup that would meet and discuss more specific details regarding a code. The subgroup's task was to write a preliminary proposal for the code, which might include several alternative definitions with rationales and video examples. Then, the subgroup shared their ideas and proposal with the full group. Other team members often generated new ideas and raised questions to push the development of the code further along. Next, the subgroup re-worked their initial proposal, wrote revised definitions, and looked for examples. Typically, they would suggest a lesson or lesson segments on which the full group could try out the code.

The subgroup and full group meetings would alternate as long as necessary, until a coding definition was agreed upon by the entire mathematics code development team. That definition was then entered into the coding manual, along with illustrative examples (and occasionally counter-examples). A copy of the TIMSS 1999 Video Study Mathematics Coding Manual is included as appendix I. The manual served as a tool for training coders, as well as a shared reference throughout the coding process. During training, coders sometimes suggested improvements to the codes. Those agreed upon by the code development group were then incorporated in the manual. Once the definitions were finalized, a strict reliability procedure was implemented for each code, as discussed in section 6.5.

6.3 Applying the Coding Scheme

Applying codes requires paying close attention to specific details in a lesson. Often coders were asked to note when a certain activity began or ended, and then describe the activity as one of several possible types. In order to reduce information processing demands, enable high inter-coder reliability, and ensure continued high quality coding, the codes were separated into "passes," with only a subset of codes applied during each pass. A pass involved viewing the entire lesson (and re-viewing critical portions) and marking all relevant segments. Altogether, 45 codes were applied in seven coding passes.

Most of the codes in the first few passes were coverage codes, and segmented the entire lesson into meaningful chunks that later could be studied in more detail. In Pass 1, coders marked the beginning and end of the lesson, and then divided the lesson into periods of public and private interaction. Passes 2 and 3 involved dividing the lesson into periods of time when mathematical problems were and were not worked on. Additionally, coders had to note the beginning and ending time of each problem, and write down the problem statement and problem solution.

The fourth pass was comprised of occurrence codes for specific events that might occur during the lesson, such as outside interruptions, goal statements, and lesson summaries. In the fifth and sixth passes, numerous questions were asked about each mathematical problem that had been identified. For example, was the problem connected to the real world, how many solutions were presented publicly, and was the problem worked on or discussed by the class for more than 45 seconds.

Pass 6 also involved a series of questions about periods of time marked as private interaction, such as what kind of problems were students assigned to work on, and did they work individually or in groups. Another set of codes in Pass 6 explored whether particular resources were used during the lesson, such as computers and calculators. Finally, in Pass 7, coders divided each lesson into segments according to their purpose— addressing previously learned content, introducing new content, or practicing and applying new content.

In the sections that follow, the codes in each pass are described and defined. The code definitions provided in this chapter are simplified, partial definitions. Complete definitions of the codes, as applied to the video data, can be found in the TIMSS 1999 Video Study Mathematics Coding Manual, included as appendix I. Where appropriate, examples and rationales are provided.

6.3.1 Pass 1: Beginning and End of Lesson; Classroom Interaction

The first coding pass contained two codes: time of lesson (LES) and classroom interaction (CI). Time of lesson involved marking the beginning and ending of each lesson, defined as the first and last public talk by the teacher that appeared to require all students' attention. The talk did not have to be about mathematics, but it should have signaled the point when a good student would recognize the lesson as starting or finishing.

LES was used to establish the duration of each mathematics lesson, and it enabled the calculation of “percent of lesson” variables. Furthermore, all subsequent coding had to occur within the defined lesson boundaries.

Marking classroom interaction (CI) patterns required coders to consider each moment of lesson time and decide which of five mutually exclusive and exhaustive categories it best fit. The five categories were: 1) entirely public interaction, 2) public information provided by the teacher, but optional for student use, 3) public information provided by a student, but optional for student use, 4) public and private work both apparent; subgroups of students varied in classroom interaction pattern, and 5) entirely private interaction.

The CI code builds on the coding of “classwork,” “seatwork,” and “classwork/seatwork combination” in the TIMSS 1995 Video Study. Due to the more varied interaction patterns found in the TIMSS 1999 Video Study lessons, creating additional categories was deemed necessary. Another, more subtle, change in this code was the focus on how teachers and students interacted, rather than how they were organized. The revised name of the code and coding categories reflects this emphasis.

6.3.2 Pass 2: Content Activity Coding

Pass 2 contained a single coverage code, content activity (CC), with 13 mutually exclusive and exhaustive categories. This code described the content activities in the lesson using very general terms. For example, whether mathematics was being conducted, and if so whether the mathematics was presented through a problem or a non-problem segment.

The coding categories were: non-mathematical work, mathematical organization, independent problem, concurrent problem set-up, concurrent problem seatwork, concurrent problem classwork, concurrent problem mixed activity, answered only problem, interrupting a problem, non-problem segement, break, and technical difficulty.

The CC code contained one of the most important segmentations: marking the in- and out-points of mathematical problems. The definition of a mathematical problem was intentionally generous, so that most borderline cases would likely be included as problems. The main requirement of a problem was that a mathematical operation must be necessary in order to arrive at the intended answer, and the problem needed to require some degree of thought by an eighth-grade student. The mathematics code development team’s rationale for such a broad definition was that mathematical problems would be explored in depth by a variety of additional codes by groups of coders each looking into different aspects of the problems.

Within the CC code, mathematical problems were broken down into three types: independent, concurrent, and answered only. For problems worked on one at a time (that is, independent problems) it was relatively straightforward to discern how much lesson time was devoted to each problem. On the other hand, for problems that were assigned as a set and then worked on privately (that is, concurrent problems) it was unknown how long students in the lesson spent working on each problem. However, time spent on concurrent problems as a set could be ascertained. Furthermore, concurrent problems could be described in terms of the following four work phases: set-up, seatwork, classwork, and a mixture of classwork and seatwork.

Answered only problems were defined as problems that had been completed by students prior to the videotaped lesson, and for which only answers were shared. These were typically either homework problems or problems worked on in an earlier lesson.

Non-problem segments were periods of time that contained mathematical information, but not problems. For example, the teacher might have presented a new concept, connected mathematical ideas to the real world, or discussed some historical background. These non-problem segments were coded in greater detail in later passes.

Break and technical difficulty segments were identified in rare cases when students were given an official break (such as during double lessons), or when the lesson could not be coded due to a temporary loss of video footage or audio.

6.3.3 Pass 3: Concurrent Problems

Because concurrent problems were treated as a set in Pass 2, the code concurrent problem (CP) was created for Pass 3. Coders marked the in- and out-point of each CP, and numbered them sequentially. Marking the approximate in and out-points allowed for further examination of each CP by subsequent codes. As noted above, concurrent problems by definition shared some private working on time; therefore the amount of time spent on individual CPs was not computed.

As part of Pass 3, coders also created a lesson table for each video. Lesson tables displayed all of the coding in Passes 1–3, along with the “problem statement” and “target result” for each independent, concurrent, and answered only problem. The problem statement described the task to be completed, and the target result was the answer or solution to the problem statement. These tables served a number of purposes: they acted as quick reference guides to each lesson, they were used in the development process for later codes, and they enabled problems to be further coded by specialist coding teams.[10]

6.3.4 Pass 4: Content Occurrence Codes

Pass 4 was comprised of six occurrence codes having to do with general content issues in the lesson. Coders marked whether each event happened in the lesson or not, and if so how many times. They also noted the in-point of each event (and in some cases the out-point as well). The six codes were: assignment of homework, goal statement, historical background, outside interruption, summary of lesson, and real life connection or application in non-problem segments.

[10] A subset of these lesson tables were expanded and then coded by the Mathematics Quality Analysis Group, described in chapter 7.

The assignment of homework (AH) code indicated whether teachers gave students homework to complete for a future lesson. AH, in combination with another code about homework in Pass 5, provides information about whether or not homework was assigned and how much discussion there was about homework problems during the lesson.

Making a goal statement (GS) is one way for an instructor to tell his or her students what will be covered in the lesson. It can serve as an advanced organizer, and help students know what mathematics the teacher intends to cover. In order for a GS to be coded, the teacher had to note the specific topic that students were expected to learn from the entire lesson or from a large portion of the lesson.

Linking mathematical content to its historical background (HB) is one kind of connection teachers can provide in their lessons. This kind of linking sets the mathematics in context, and can help connect different subject areas. For example, the teacher might note that a Greek man named Pythagoras was the originator of a mathematical theorem. HB was coded whenever the teacher and/or students made such a connection.

Noting the occurrence of outside interruptions (OI) was essentially a replication of a TIMSS 1995 Video Code of the same name. The definition was slightly revised in this study to make application to a larger sample possible. Generally speaking, the 1999 Video Study definition is somewhat more inclusive than the 1995 Video Study definition.

Similar to goal statements, summaries of lessons (SL) informed students as to what mathematics they were expected to learn from the lesson. These summaries might organize the mathematical information presented in the lesson and highlight the most important concepts.

The last code in Pass 4 looked specifically at non-problem segments (as defined in Pass 2), and noted whether they contained a real life connection or application. This code, in combination with several others in later passes (such as real world connections within problems and the use of real world objects during the lesson), indicates how often teachers linked the mathematical material to students' experiences outside the classroom.

6.3.5 Pass 5: Problem-Level Codes

Fifteen codes in Pass 5 examined more closely the mathematical problems identified in each lesson. Since these codes were applied by an international team comprised of native speakers who were not necessarily mathematics experts, they emphasized the pedagogy surrounding the problems rather than their content. Other codes aimed at uncovering the mathematics content in a more precise manner were developed and applied by specialist teams, as described in chapter 7.

The TIMSS 1995 Video Study suggested that there was a difference in how much time countries spent going over homework problems from the previous night, or starting on homework problems due for the next lesson. In this study, a code was included to denote whether each mathematical problem was a homework or non-homework problem (H). Those designated as homework were further described as either previously assigned or assigned for a future lesson.

Based on this coding, it was possible to determine exactly what proportion of problems were homework, and estimate the amount of time spent on such problems.

Through initial viewings of the lessons, the mathematics code development team noticed that teachers sometimes assigned a large set of problems, but allocated subsets of these problems to particular students. For example, the teacher might have divided the class into groups, and assigned each group a different worksheet to complete. Therefore, a code was created to identify how many students (HS) each problem was intended for, or more specifically, whether it was intended for the entire class or not.

To complement the HS code, a required or optional (RO) code was developed to determine which problems were required of students and which were designated as optional. For example, a teacher might have required students to complete the first five problems on a worksheet, but allowed the next 10 problems to be optional. The HS and RO codes provide information on the exact mathematical problems students were given the opportunity to work on in the lesson, as well as some of the pedagogical techniques teachers used to assign these problems.

The next two codes in Pass 5 explored the degree to which problems were presented to students in a real world context. Problem context (PC) asked whether each problem was set up with mathematical language or symbols only, or with something more than numbers and symbols. For example, the problem statement may have been given in the form of a story. Real life connection (RLC) ascertained whether a reference to real life was contained in the problem set-up and, if not, whether such a reference occurred as the class worked on the problem. These codes, together with the Pass 2 code real life connection in non-problem segments (RLNP), indicate how often the material presented in the lesson was explicitly connected to the outside world.

Three codes were created that describe various forms of representation that may have been used when working on each problem. Like problem context (PC) and real life connection (RLC), these codes explored the degree to which mathematics was presented in a context that utilized other forms of representation besides numbers or mathematical symbols. For example, the problem may have contained a graph (GR), table (TA), or drawing or diagram (DD).

Of growing interest in many countries is the use of physical materials during mathematics lessons. A code was developed in the TIMSS 1995 Video Study to determine the number of lessons in which manipulatives were used, and by whom. In the TIMSS 1999 Video Study, this code was expanded to capture whether physical materials (PM) were manipulated during each mathematical problem, and if so whether they were used by the teacher, students, or both. Common physical materials found in mathematics classes included measuring instruments, geometric solids, and cut-out plane figures. Additional codes in Pass 6 helped to specify the type of resources that were used during the lesson.

Another code applied to problems was the degree of choice students were given in selecting a solution method (SC). For example, students might have been given an “open choice” to use any solution method they liked, a “limited choice” of several identified options, or no choice. (A related Pass 6 code was the number of solution methods presented publicly.)

If students did have some choice in selecting a solution method, the problem was further coded to specify whether it met several other criteria. For example, problems were coded for whether at least two methods presented, with at least one of the methods critiqued or discussed at length. This code, labeled facilitating exploration (FE), captured a very specific method of solving problems which is of interest to some mathematics educators.

Based on the results of the TIMSS 1995 Video Study, the code development team did not expect to find numerous instances of proofs, verifications, or derivations (PVD) in eighth-grade mathematics lessons. The PVD code was essentially a replication of the “proofs” code from the previous video study. Due to the larger number of lessons and coders in the TIMSS 1999 study, however, an expanded coding definition was written. To further ensure consistency across coders, each potential PVD was checked by a mathematics expert (who was familiar with the coding definition) before it was marked as such.

Two codes in Pass 5 provide information about whether the correct answer to a problem was provided publicly. Sometimes independent problems were started by the class but not completed or, more frequently, students were assigned a set of concurrent problems and none or only some of them were solved publicly. One code explored the number of target results presented publicly for each problem (NTR), and the other explored whether the target result was presented in different forms (DFTR). For example, a problem might have two correct answers, with both presented publicly. Or, a problem might have one correct answer, presented publicly in multiple forms (such as a decimal and a fraction).

The mathematics code development team also developed a code to separate longer problems from shorter ones. After much discussion and viewing of problems, forty-five seconds was agreed upon as a criterion to distinguish problems worked on for a relatively long time from briefer problems. Thus all problems were coded to determine whether they were greater than or equal to forty-five seconds or whether they were less than forty-five seconds (LWO).

The LWO and NTR codes not only provided important information about the nature of the problems in the lessons, they also served as gatekeepers for other codes by narrowing down the number of problems that need to be examined. For example, problems worked on for less than forty-five seconds or that did not have a target result publicly presented were excluded from the coding of facilitating exploration (FE – discussed in this section) and problem summary (PSM – discussed in Section 6.3.6). Excluding such problems made sense for such codes and reduced the workload of coders.

6.3.6 Pass 6: Resources, Private Work, and Non-Problem Segments

Pass 6 contained an assortment of codes related to various aspects of the lessons. The first set of codes had to do with resources in the classrooms. Coders marked whether any of the following were used during each lesson: chalkboard (CH), projector (PRO), television or video (TV), textbooks or worksheets (TXW), special mathematical materials (SMM; e.g., rulers, graph paper, or base-ten blocks), real world objects (RWO; e.g., maps or dice), calculators (CALC; further classified as either regular or graphing), and computers (COMP).

Some of the inspiration for these codes was drawn from the TIMSS 1995 Video Study. One of the often cited findings from that study is the extent to which chalkboards and overhead projectors were used in the United States as compared to Japan. For the TIMSS 1999 Video Study, the code developers wanted to explore the use of these and similar classroom resources, such as textbooks. Further, given the interest many of the participating countries had in contextualizing mathematics, usage of mathematical materials and real world objects was also noted. (In Pass 5 such objects were coded when they were used to solve a problem, however they were sometimes used in non-problem segments. Therefore, this set of resource codes was applied to each lesson as a unit.) In addition, a number of countries participating in the 1999 Video Study expressed an interest in knowing how often calculators and computers were used in the videotaped lessons.

Two additional codes about problems were included as part of Pass 6. One was similar to a code developed in the TIMSS 1995 Video Study regarding alternative solution methods. In that study, coders marked the largest number of alternative solution methods presented for any identified task. In the TIMSS 1999 Video Study a similar code was applied to each mathematical problem, and noted whether more than one solution method was publicly presented (MSM). If so, coders specified whether the students suggested any of the solution methods.

The problem summary (PSM) code was applied only to problems longer than 45 seconds that had a publicly presented target result (correct answer). This code ascertained whether the teacher summarized the major steps or critical rule involved in the problem. In many ways this code complements the goal statement (GS) and summary of lesson (SL) codes in Pass 4, as it identifies instances when the teacher emphasized the important mathematics that students were expected to learn from the videotaped lesson.

Four codes were developed as part of Pass 6 to classify the non-problem (NP) segments marked in Pass 2. These codes, along with the Pass 4 real life non-problem (RLNP) code, provide information about what mathematics the students were engaged in when they were not working on problems. Each NP segment was classified as containing at least one of the following: contextual information (CON), a mathematical concept, theory, or idea (CTI), a mathematical activity (AC), or the teacher discussing a homework assignment or test (HT).

All of the remaining codes in this pass had to do with the work students completed privately, at their seats. The code private work assignment (PWA) drew heavily on the TIMSS 1995 Video Study "performance expectations" code. Performance expectations referred to the kind of tasks students worked on during seatwork, such as practicing routine procedures, inventing new solutions, or applying concepts in new situations. The PWA code involved somewhat broader and more concrete categories. That is, coders determined whether the assignment involved using steps students were already familiar with to solve problems, or if it required something more. In other words, this code distinguished between assignments in which students used entirely known procedures from those in which students had to do something new. Cues from the lesson along with teachers' questionnaire responses helped coders to determine whether the mathematical concept(s) and solution method(s) were known to students before they started the assignment.

The last four codes applied only to portions of lessons marked in Pass 1 as "private interaction." Coders noted whether the majority of students worked individually, in pairs, or in groups, and marked any shifts in their organization. The organization of students code (OS) was very similar to the TIMSS 1995 Video Study coding of whether students worked by themselves or in a group during seatwork; however, the OS code differentiated between working in pairs from working in groups of three or more.

Several codes were developed to capture what teachers did while their students worked privately. For example, teachers might have spent this time displaying mathematical information on the board or overhead projector (DI). Such information could be intended for students to use as they worked on their assignment, or it could be in preparation for an upcoming public, whole class segment. Teachers might also have spent their time engaging in an administrative activity that was unrelated to the students' current assignment (AA). For example, they could have taken roll or checked to make sure students completed their homework.

Teachers often made public announcements (PA) during private work. Such announcements appeared to be intended for all students to hear, and could either provide information related to the current assignment or they could be entirely unrelated to the current assignment. Announcements related to the assignment were further classified as containing either mathematical or organizational information. Unrelated announcements (such as disciplinary comments) could be considered interruptions to students' work time.

6.3.7 Pass 7: Purpose

The last code applied by the international coding team provided information about the lesson purpose (P). This was a coverage code, meaning that every moment of each lesson had to be segmented into one of three mutually exclusive and exhaustive purpose types, which could shift as the lesson progressed.

The purpose code was developed through a somewhat different and more elaborated process than most of the codes described above. The intention was to create a simple, universal code with purpose categories flexible enough to fit each country. From their experience developing country models early in the study, the mathematics code development team knew that purpose segments were relatively easy to classify within countries, but much harder to agree upon across countries. In order to develop a purpose code that would satisfactorily represent the pedagogy in each of the participating nations, country teams were assembled. These teams, which included the national research coordinators, country associates, and coders, worked together to name and define purpose categories appropriate for their country. Then, all of the teams met to discuss each country's ideas. Finally, the mathematics code development team agreed on three categories that incorporated all of the suggestions and fit well for each country.

The purpose code contained three categories: 1) addressing content introduced in a previous lesson, 2) introducing new content, and 3) practicing, applying, or consolidating new content introduced in the current lesson. Defining the boundaries between these three categories for any particular lesson required a great deal of cultural knowledge. Therefore, reliability for this code was established only between coders of the same country, and not across countries as was done for the other codes (see section 6.5 on reliability and quality control).

Although reliability was not established between coders from different countries, in many important ways the training, reliability, and application of the purpose code were the same as that of other codes. For example, coders were trained as a single group with equal access to instruction from the country associates and questions from fellow coders. When discussing and practicing the purpose code, lessons from all countries were used. Furthermore when coders applied this code to their designated set of lessons, they were encouraged to discuss difficult coding decisions with their fellow coders—regardless of country—just as they had done for prior codes. For the purposes of establishing reliability, however, it seemed most appropriate to pair coders from the same country, since marking the exact boundaries between purposes often required understanding cultural nuances in the lesson climate and language.

The purpose code provided information about the nature of the mathematics at different points in the lesson, and helped to place the content in a sequential context. For example, teachers might shift topics as they move from a review to a new phase. Fewer and longer problems might be worked on during the new phase. Looking at codes together in this manner could paint a more detailed portrait of the lesson videos and may be useful for creating broad descriptions of teaching in each country.

6.4 Coder Training

As described above, codes were developed, practiced and applied in passes. Once definitions were completed for each code in a pass, training materials were created and a reliability procedure was developed.

Training for each coding pass involved three stages: introduction, practice, and reliability. First, coders were provided with the coding manual, which contained carefully worded definitions for each code, as well as notes and examples. Coders and country associates met to introduce and discuss each code in the pass, including the definitions and accompanying notes. For most codes, video examples were shown of each coding category. Coders frequently raised questions about the rationale and purpose behind the codes, or requested further clarification of the definitions. The country associates sometimes used the coders' input to make minor revisions to the coding manual.

After learning the definitions and watching examples from a particular coding pass, coders were given the opportunity to practice applying the codes. In these practice sessions coders were provided with a select set of lessons, or portions of lessons, usually representing all the countries in the sample. Coders were instructed to work individually to apply the codes, and then compare their coding to an answer key. To create these answer keys, each country associate individually coded the lessons. Then the mathematics code development team met and reached consensus on the appropriate coding. Once they finished practicing, coders and country associates would meet again to discuss any problems or concerns that arose.[11]

Throughout the training process, coders were encouraged to make suggestions for improving the code definitions. For some of the later coding passes, particularly Passes 6 and 7, coders played a substantially more active role in assisting the mathematics code development team to create code definitions and train their colleagues. In particular, coders and country associates formed subgroups to test code definitions, assemble practice materials, and train other coders.

Once coders and country associates felt comfortable with the codes, and confident that they could apply them reliably, coders took an initial reliability test. Details of the initial reliability procedure and calculations are discussed in section 6.5.1 below. After establishing reliability on the codes in a pass, coders applied them to lessons from their country. Various additional quality control measures were put in place to ensure reliable and valid coding and data entry. For example, mid-point reliability was calculated for each code once coders completed at least half of their assigned lessons.

Occasionally coders did not reach an acceptable level of initial reliability on some codes in a pass. On two of these codes (lesson duration and content activity), coding definitions were then modified by the code development team, coders were re-trained, and they established reliability using a new set of lessons. On two other codes (elaborated problems and teacher assistance during private work), coders could not establish an acceptable level of reliability even after re-training and re-testing. Therefore these codes were dropped.

Coders were each responsible for a particular number of lessons, and coding was done individually. However, collaboration among coders was encouraged, especially among coders from the same country. Also, country associates were available to help with questions and difficult lessons. When coders came across lessons that were particularly hard to code, the entire mathematics code development team met to watch them and determine how to accurately apply the codes. These decisions were then explained in writing and distributed to all coders (see the TIMSS 1999 Video Study Mathematics Coding Manual included as appendix I.).

6.5 Reliability and Quality Control

Coders established initial reliability for all codes prior to their implementation. After they finished coding approximately half of their assigned set of lessons (in most cases about 40–50 lessons), coders established midpoint reliability. The minimum acceptable reliability score for each code was 85 percent (averaged across coders). Individual coders or coder pairs had to reach at least 80 percent reliability on each code.[12]

[11] Lessons or portions of lessons that were coded by the country associates and then by coders as "practice" were considered coded. Therefore, the coders assigned to those particular lessons simply had to enter the coding into the appropriate software.

[12] The minimum acceptable reliability score for all codes (across coders and countries) was 85 percent. For coders and countries, the minimum acceptable reliability score was 80 percent. That is, the reliability of an individual coder OR the average of all coders within a particular country was occasionally between 80–85 percent. In these cases clarification was provided as necessary, but re-testing for reliability was not deemed appropriate.

Reliability was computed either as agreement between coders and a master document, or as inter-rater agreement between pairs of coders. In all cases, reliability statistics were calculated based on a "percent correct" approach (Bakeman and Gottman 1997). A master refers to a lesson or part of a lesson coded by consensus by the mathematics code development team. To create a master, the country associates independently coded the same lesson and then met to compare their coding and discuss disagreements until consensus was achieved. Masters often were used to establish initial reliability, particularly in the early passes. Inter-rater agreement between coders typically was used to establish midpoint reliability. Inter-rater agreement was also used to establish initial reliability in some of the later passes, for which coders helped to develop coding definitions.

The formula used, in all cases, to compute reliability was:

Number of Agreements ÷ (Number of Agreements + Number of Disagreements).

This formula was used regardless of whether reliability was established between coders and a master document, or as inter-rater agreement. What counted as an agreement or disagreement depended on the specific nature of each code, and is explained in detail in sections 6.5.1 and 6.5.2. Note that when codes required timing and categorization decisions, both were taken into account as either agreements or disagreements.

Table 6.3 lists the initial and midpoint reliability scores for each code, averaged across coders. Since the computation of reliability for codes differed somewhat, the specific procedures used to calculate initial and midpoint reliability for each code are presented in sections 6.5.1 and 6.5.2.

Table 6.3. Initial and midpoint reliability statistics for each code applied by the International Coding Team, by code: 1999

Pass	Code	Initial reliability[1] (percent)	Midpoint reliability[2] (percent)
1	Lesson (LES)	93	99
1	Classroom interaction (CI)	94	92
2	Content activity (CC)	90	87
3	Concurrent problem (CP)	94	90
4	Assignment of homework (AH)	99	93
4	Goal statement (GS)	99	89
4	Historical background (HB)	100	100
4	Outside interruption (OI)	96	96
4	Summary of lesson (SL)	98	99
4	Real life within non-problem (RLNP)	98	96
5	Homework (H)	99	98
5	How many students (HS)	98	100
5	Required or optional (RO)	98	100
5	Problem context (PC)	97	92
5	Real life connection (RLC)	98	100
5	Graphs (GR)	97	98
5	Tables (TA)	99	98
5	Drawings/diagrams (DD)	97	94
5	Physical materials (PM)	95	97
5	Student choice of solution method (SC)	90	93
5	Proof/verification/derivation (PVD)	99	97
5	Number of target results (NTR)	96	94
5	Number of different forms of the target result (DFTR)	92	94
5	Length of working on (LWO)	95	94
5	Facilitating exploration (FE)	96	95
6	Chalkboard (CH)	96	100
6	Projector (PRO)	98	100
6	Television or video (TV)	100	100
6	Textbook or worksheets (TXW)	98	98
6	Special mathematical materials (SMM)	92	93
6	Real-world objects (RWO)	98	100
6	Calculators (CALC)	98	95
6	Computers (COMP)	100	98
6	Multiple solution methods (MSM)	99	98
6	Problem summary (PSM)	97	95

Table 6.3. Initial and midpoint reliability statistics for each code applied by the International Coding Team, by code: 1999—Continued

Pass	Code	Initial reliability[1] (percent)	Midpoint reliability[2] (percent)
6	Contextual information (CON)	92	91
6	Mathematical concept/theory/idea (CTI)	92	94
6	Activity (AC)	97	97
6	Announcing or clarifying homework or test (HT)	95	98
6	Private work assignment (PWA)	93	98
6	Display information (DI)	96	89
6	Administrative activity (AA)	93	95
6	Organization of students (OS)	96	96
6	Public announcements (PA)	86	86
7	Purpose (P)	87	94

[1]Initial reliability refers to reliability established on a designated set of lessons before coders began work on their assigned lessons.
[2]Midpoint reliability refers to reliability established on a designated set of lessons after coders completed approximately half of their assigned lessons.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

6.5.1 Initial Reliability

Most frequently, initial reliability was determined by comparing coders' individual markings of lessons to masters of those lessons, as described above. This method is considered a rigorous and cost-effective alternative to inter-coder reliability (Bakeman and Gottman 1997).

For some codes, standard, inter-coder testing was deemed the most appropriate method of determining reliability. For example, the videotape coders played a large role in the development of the Pass 6 codes, and the mathematics code development team did not consider themselves expert enough to create masters for these codes. Additionally, only within-country reliability was established for the purpose code, and it would not have been appropriate to calculate reliability against the international code development team (see section 6.3.7).

A percent agreement reliability statistic was computed for each coder by dividing the number of agreements by the sum of agreements and disagreements (Bakeman and Gottman 1997). Then, average reliability was calculated across coders and across countries for each code. In cases where coders did not reach the set reliability standard, they were re-trained and re-tested using a new set of lessons. Codes were dropped if 85 percent reliability could not be achieved (or if individual coders could not reach at least 80 percent reliability) (see section 6.4).

Some codes required coders to indicate a time. In these cases, coders' time markings had to fall within a predetermined margin of error. This margin of error varied depending on the nature of the code, ranging from 10 seconds to 2 minutes. Rationales for each code's margin of error are provided in the sections that follow.

Exact agreement was required for codes that had categorical coding options. In other words, if a code had four possible coding categories, coders had to select the same coding category as the master. In some cases, coders had to both mark a time (i.e., note the in- and/or out-point of a particular event) and designate a coding category. Reliability for the coding category was calculated only if coders marked the time within the given margin of error.

Reliability calculations differed somewhat depending on the nature of each code. A detailed explanation of how the initial reliability was computed for each code is provided below.

6.5.1.1 ***Initial Reliability for Lesson***

Coders watched six videos (from six countries) and marked the in- and out-points of each lesson. Coders' markings were compared against master lessons. Each in- and out-point had to be within 30 seconds of the master to be counted as an agreement. A 30-second margin of error was deemed appropriate and reasonable for this code based on the notion that eighth-grade mathematics lessons typically last about 30–50 minutes.

6.5.1.2 ***Initial Reliability for Classroom Interaction***

Coders watched two lessons (one from their own country plus one from another country), and noted all shifts and categories for classroom interaction. Coders' markings were compared against master lessons. Each in- and out-point had to be within 20 seconds of the master to be counted as an agreement. A 20-second margin of error was deemed appropriate and reasonable for this code because the definition of classroom interaction required that these segments last at least 1 minute in order to be coded as such. Each categorization had to be exactly the same as the master to be counted as an agreement.

If coders marked the in- or out-points incorrectly, this was counted as a disagreement. However, they were then told the correct time(s) and given the opportunity to adjust their labels. This two-step process was deemed reasonable because categorizations were dependent on the placement of classroom interaction shifts.

6.5.1.3 ***Initial Reliability for Content Activity***

Coders watched three lessons (one from their own country plus two from other countries), and noted all shifts and categories for content activity. Coders' markings were compared against master lessons. Each in- and out-point had to be within 10 seconds of the master to be counted as an agreement. A 10-second margin of error was deemed appropriate and reasonable for this code because certain content activity categories were required to last at least 20 seconds in order to be counted as such.[13] Each categorization had to be exactly the same as the master to be counted as an agreement.

If coders marked the in- or out-points incorrectly, this was counted as a disagreement. However, they were then told the correct time(s) and given the opportunity to adjust their labels. This two-step process was deemed reasonable because categorizations were dependent on the placement of content activity shifts.

6.5.1.4 Initial Reliability for Concurrent Problems

Coders watched three lessons (one from their own country plus two from other countries), and noted all of the concurrent problems. Coders' markings were compared against master lessons. Each in- and out-point had to be within 10 seconds of the master marking to be counted as an agreement. A 10-second margin of error was deemed appropriate and reasonable for this code because other types of problems—categories of content coverage—had a 10-second margin. The number of concurrent problems marked by coders was compared to the number of concurrent problems in the master.

6.5.1.5 Initial Reliability for Pass 4 (6 Occurrence Codes)

Coders watched 14 lesson segments (two per country), and noted the appearance(s) of any occurrence code. Coders' markings were compared against master lessons. Reliability was calculated separately for each code.

For each lesson segment, an agreement was counted if the coder marked an occurrence with an in-point within 1 minute of the master. A 1-minute margin of error was deemed appropriate and reasonable because the unit of analysis for occurrence codes was the lesson. That is, analyses were most likely to explore whether a particular event occurred within a given lesson (not how many times it occurred). A disagreement was counted if the coder omitted an occurrence marked on the master, or marked an occurrence not on the master.

6.5.1.6 Initial Reliability for Pass 5 (15 Codes about Problems)

Coders watched 44 mathematical problems (at least four per country, as marked in Passes 2 and 3), and applied the 15 problem-level codes to all problems. Coders' markings were compared against master lessons. Reliability was calculated separately for each code.

For each problem, an agreement was counted on a given code if the coder marked exactly the same coding category as the master, and a disagreement was counted if the coder marked a coding category different from the master. Time was not included in the reliability calculations for these codes because the in- and out-points of each problem had previously been determined in Pass 2.

---

[13] A smaller margin of error was found to be problematic because teachers often pause between sentences or between activities. Therefore, a buffer time of a few seconds proved necessary to accommodate variation in how these pauses were treated by coders.

6.5.1.7 ***Initial Reliability for Pass 6 (19 Codes about Resources, Mathematical Problems, Non-Problem Segments, and Private Work)***

Reliability for all Pass 6 codes was determined by calculating the mean inter-rater agreement among pairs of coders. Coders were paired randomly; however, coders from the same country could not be grouped together.[14] For each coder, 3 lessons were randomly chosen from among their assigned set of lessons. Coders coded their selected lessons and their partner's selected lessons. Reliability was calculated separately for each code.

Most codes in Pass 6 only required coders to make categorization decisions. An agreement was counted if both coders marked the same coding category, and a disagreement was counted if coders marked different coding categories. Time was not included in the reliability calculations for the resource codes because they applied to the entire lesson (e.g., was a blackboard used or not). For the codes applied to mathematical problems and non-problem segments, time was not included in the reliability calculations because the in- and out-points had previously been determined in Pass 2.

The codes about private work—organization of students (OS) and public announcement (PA) —required coders to both mark a time and designate a coding category. The shifts or in-points had to be within 20 seconds of one another in order to be calculated as an agreement. This period of time was deemed appropriate and reasonable because both of these codes represented relatively short periods of lesson time. Each categorization had to be exactly the same to be counted as an agreement.

6.5.1.8 ***Initial Reliability for Purpose***

Reliability for purpose (P) was determined by calculating the mean inter-rater agreement among pairs of coders from the same country. Coders watched four lessons (two from each coder's assigned set of lessons plus two from another country), and noted all shifts and categories for purpose. For each lesson, a coder's markings were compared to his/her coding partner's markings. Each in- and out-point had to be within 2 minutes of his/her partner's marks to be counted as an agreement. A 2-minute margin of error was deemed appropriate and reasonable for this code because purpose segments were generally long (approximately 17 minutes, on average across countries), and noting their beginning and end points often required a great deal of inference. Each categorization had to be exactly the same to be counted as an agreement.

6.5.2 Midpoint Reliability

Midpoint reliability for the code lesson duration (LES) was determined by comparing coders' marking of lessons to master lessons. For all other codes, midpoint reliability was determined by calculating the mean inter-rater agreement among pairs of coders. By halfway through the coding process, coders were considered to be more expert in the code definitions and applications than the mathematics code development team. Therefore, in general, the most appropriate assessment of their reliability was a comparison with other coders.

---

[14] The Swiss coders established only within country inter-rater reliability because they were trained separately from the other videotape coders and carried out their coding in Switzerland rather than at LessonLab.

Coder pairs were always randomly assigned according to the following conditions: 1) coders could not be from the same country,[15] and 2) coders could not have the same partner for initial and midpoint reliability. Lessons were selected for each coder by randomly choosing from among their seven most recently coded lessons. Coders reviewed their selected lessons and coded their partner's lessons. In the process, coders were instructed to consult the coding manual and keep notes regarding the "implicit" rules they applied. That way if disagreements arose, the coder pairs could support their decisions. After their reliability scores were calculated, coder pairs were encouraged to resolve such coding differences on their own, seeking help from other coders and country associates as needed.

Reliability calculations differed somewhat depending on the nature of each code. A detailed explanation of how the midpoint reliability was calculated for each code is provided in the sections that follow.

6.5.2.1 ***Midpoint Reliability for Lesson***

Midpoint reliability for lesson duration (LES) was calculated in exactly the same way as initial reliability.

Midpoint Reliability for Classroom Interaction

Inter-rater midpoint reliability was established for classroom interaction (CI). For each lesson, a coder's markings were compared to his/her coding partner's markings. Each in- and out-point had to be within 40 seconds of his/her partner's marks to be counted as an agreement. This was the same as allowing a 20-second margin of error when comparing an individual coder's markings against a master lesson, as used to determine initial reliability. (Coders could mark a segment 20 seconds before or after the master, thus the range between pairs of coders could be up to 40 seconds.) Other calculations were exactly the same as for initial reliability.

6.5.2.2 ***Midpoint Reliability for Content Activity***

Inter-rater midpoint reliability was established for content activity (CC). For each lesson, a coder's markings were compared to his/her coding partner's markings. Each in- and out-point had to be within 20 seconds of his/her partner's marks to be counted as an agreement. This was the same as allowing a 10-second margin of error when comparing an individual coder's markings against a master lesson, as used to determine initial reliability. (Coders could mark a segment 10 seconds before or after the master, thus the range between pairs of coders could be up to 20 seconds.). Other calculations were exactly the same as for initial reliability, except that coders watched two lessons (one from their own country and one from their partner's country) rather than three lessons.

6.5.2.3 ***Midpoint Reliability for Concurrent Problem***

[15] There were two exceptions: within country inter-rater reliability was determined for the Swiss coders on all codes (for reasons explained in an earlier footnote), and for all coders on the Purpose code.

Inter-rater midpoint reliability was established for concurrent problem (CP). For each lesson, a coder's markings were compared to his/her coding partner's markings. Each in- and out-point had to be within 20 seconds of his/her partner's marks to be counted as an agreement. This was the same as allowing a 10- second margin of error when comparing an individual coder’s markings against a master lesson, as used to determine initial reliability. (Coders could mark a segment 10 seconds before or after the master, thus the range between pairs of coders could be up to 20 seconds.) Other calculations were exactly the same as for initial reliability, except that coders watched two lessons (one from their own country and one from their partner’s country) rather than three lessons.

6.5.2.4 ***Midpoint Reliability for Pass 4 (6 Occurrence Codes)***

Inter-rater midpoint reliability was established for all Pass 4 codes. For each lesson, a coder's markings were compared to his/her coding partner's markings. Each in- and out-point had to be within 2 minutes of his/her partner's marks to be counted as an agreement.This was the same as allowing a 1-minute margin of error when comparing an individual coder’s markings against a master lesson, as used to determine initial reliability. (Coders could mark a segment 1 minute before or after the master, thus the range between pairs of coders could be up to 2 minutes.) Other calculations were exactly the same as for initial reliability, except that coders watched 16 lesson segments (eight from their own country plus eight from another country) rather than 14 segments.

6.5.2.5 ***Midpoint Reliability for Pass 5 (15 Codes Regarding Problems)***

Inter-rater midpoint reliability was established for all Pass 5 codes. Other calculations were exactly the same as for initial reliability, except that coders watched four lessons (two from their own country plus two from another country), and applied each code to all of the mathematical problems in these lessons rather than watching 44 mathematical problems.

6.5.2.6 ***Midpoint Reliability for Pass 6 (19 Codes Regarding Resources, Problems, Non-Problem Segments, and Private Work)***

Midpoint reliability for all Pass 6 codes was calculated in exactly the same way as initial reliability.

6.5.2.7 ***Midpoint Reliability for Purpose***

Midpoint reliability for purpose (P) was calculated in exactly the same way as initial reliability.

6.5.3 Other Quality Control Measures

A variety of additional quality control measures were put in place to ensure accurate coding. These measures included: 1) discussing difficulties in coding lessons reliably with the mathematics code development team and/or other coders, 2) checking the first two lessons coded by each coder, either by a country associate or by another coder, and 3) discussing hard-to-code lessons with country associates and/or other coders.

6.5.4 Data Entry, Cleaning, and Statistical Analyses

Most codes were entered directly into the multimedia database, so that the videotapes and English transcripts could be linked directly with specific codes. The data then were exported either in spreadsheet format for statistical analyses, or in table format for further study by specialist coding groups. In some cases, where the vPrism software was not conducive for particular types of coding, codes were entered into an Excel spreadsheet.

Codes from Passes 1–4 were entered directly into a vPrism database. Codes from Pass 5 were entered into an Excel database. For Pass 6, all codes were entered into vPrism except the two regarding problems, which were entered into Excel. Pass 7 coding was entered into vPrism.

A data cleaning process was put in place for both the vPrism and Excel databases. For the vPrism data, coders first recorded their coding decisions in writing onto printed lesson transcripts. Then they entered this information into vPrism. Lastly, coders exported the vPrism data for each lesson and compared it to their markings on the transcripts. In this way, data entry errors were immediately noted and corrected. In addition, errors detected through preliminary data analyses were examined and corrected. For example, coding that was outside of a possible range was detected and extreme outliers on particular codes were studied.

For the Excel data, coders first recorded their coding decisions in writing onto a printed spreadsheet for each lesson. Then they entered this information into Excel. Every tenth lesson was checked for accuracy, and errors were corrected.

Once they were cleaned, all of the data were aggregated to the lesson level, with each coding pass in a separate datafile. The full sample and replicate weights were then appended to each file. Finally, statistical analyses were run using the weighted data in Wesvar and/or SPSS.

6.6 Conclusion

In summary, the mathematics code development team created 45 codes that were applied to the video data in seven passes by an international team of coders. Initial and midpoint statistics were computed on each code, using the percent correct procedure described in Bakeman and Gottman (1997), and in all cases exceed 85 percent.

Chapter 7. Coding Video Data II: Specialists

Most codes were applied to the video data by a team of international coders, who were cultural insiders and fluent in the language of the lessons they coded. However, not all of them were experts in mathematics or teaching. Therefore, several specialist coding teams with different areas of expertise were employed to create and apply codes regarding the mathematical nature of the content, the pedagogy, and the discourse.

7.1 Mathematics Problem Analysis Group

The mathematics problem analysis group was comprised of individuals with expertise in mathematics and mathematics education. The group was directed by Diana Wearne (University of Delaware) and included Margaret Smith (University of Iowa) and Eric Sisofo (University of Delaware). They developed and applied a series of codes to all of the mathematical problems in the videotaped lessons. The group worked from written records of the lessons that listed the statement for every problem the students were asked to solve and a solution of the problems presented during the lesson.

7.1.1 Coding for Topic

The purpose of the topic code was to assist in describing the mathematics that students were encountering during each lesson. The mathematics problem analysis group constructed a comprehensive, detailed, and structured list of mathematical topics covered in eighth grade in all participating countries. Initially, the list was created by reviewing textbooks and national/state/regional curriculum guidelines from each country and watching lesson videos. The list was refined and expanded during the coding process. Whenever the latter occurred, the group members conferred and had to agree to the designation of a new topic code—that it described a situation for which no topic code existed—before it was added to the list.

All problems worked on in the lesson (i.e., coded as “independent problems” or “concurrent problems” by the international coding team), were assigned a mathematics topic code. For example, a problem could be assigned the topic of determining the surface area of a given three-dimensional object, finding the mean of a distribution, or graphing a linear function.

The final topic code list was fairly specific and consisted of 564 codes. The 22 broad categories of topic codes are described below. These categories were mutually exclusive.

Within each of the 22 broad categories, one or more subcategories were used to classify problems as “applications”—that is, problems that required students to apply procedures they have learned in one context in order to solve problems presented in a different context. Using these categorizations, the mathematics problem analysis group was able to identify how many problems were applications.

Applications might, or might not, be presented in real-life settings. The following problem is an example of real-life application: “A rectangular shaped garden is twice as long as it is wide. If the length of the fence enclosing the garden is 24 meters, what are the dimensions of the garden?" Non-real life applications include problems such as, "The sum of three consecutive integers is 240. Find the integers."

Category 1. Whole Numbers/Number Theory

This category includes operations with whole numbers including ordering; properties of the operations; factors; integer exponents; roots when the result is a whole number; arithmetic and geometric sequences and series; and applications and proofs associated with these topics. This category includes 33 topic codes.

Category 2. Fractions and Decimals

This category includes operations with fractions and decimals; order of fractions and decimals; properties of the operations; equivalent/improper/complex fractions; creating a representation of the numbers; translating between decimal and common fraction form; raising to powers and finding roots; significant digits; rational and irrational numbers; and applications and proofs associated with these topics. This category includes 35 topic codes.

Category 3. Ratio, Proportion, and Percent

This category includes ratio; proportion; percent; trigonometric ratios defined in a right triangle; relationships among the trigonometric ratios; inverse proportion and variation; and applications and proofs associated with these topics. There are 27 topic codes in this category.

Category 4. Integers

This category includes operations with integers; properties of operations; models for integers; ordering; exponents (positive and negative integers, fractional); scientific notation; and applications associated with these topics. There are 20 topic codes in this category.

Category 5. Geometry: Angles

This category includes classification of angles; relationships among angles of particular triangles; relationships among interior and exterior angles of a triangle; angles associated with parallel lines; angles associated with a circle; the sum of the measures of the angles of a polygon; and applications and proofs associated with any of these topics. There are 35 topic codes in this category.

Category 6. Geometry: Triangles and Lines in a Two-Dimensional Plane (excluding area and perimeter)

This category includes parallel lines; classification of triangles; relationships among sides of certain triangles; relationships among interior and exterior angles of a triangle; Pythagorean Theorem; congruent triangles; similar polygons; and applications and proofs related to these topics. This category includes 42 topic codes.

Category 7. Geometry: Quadrilaterals and other N-Gons (excluding perimeter and area)

This category includes definitions of various quadrilaterals; theorems relating to parallelograms including specific parallelograms; exterior and interior angles of regular polygons; and applications and proofs related to these topics. This category includes 28 topic codes.

Category 8. Geometry: Perimeter and Area of Figures in a Two-Dimensional Plane

This category includes finding perimeter and area of polygons and circles; finding areas of sectors; computing arc lengths; developing these procedures; Hero's formula; and applications associated with these topics. There are 28 topic codes in this category.

Category 9. Geometry: Three Dimensional Figures: Descriptions

This category includes descriptions of three-dimensional figures; categorizing the figures; constructing or using nets of figures; Euler's formula; and applications and proofs (not involving computing area or volume) associated with these topics. There are 26 topic codes in this category.

Category 10. Geometry: Three-Dimensional Figures: Surface Area

This category includes defining surface area; developing procedures for computing surface areas of prisms, cylinders, cones, pyramids, and spheres; and applications involving surface areas of three-dimensional figures. There are 15 topic codes in this category.

Category 11. Geometry: Three-Dimensional Figures: Volume

This category includes defining volume; developing procedures for computing volume of prisms, cylinders, cones, pyramids, and spheres; and applications involving volumes of three dimensional figures. There are 16 topic codes in this category.

Category 12. Geometry: Geometric Transformation

This category includes defining various transformations (translation, rotation, reflection) and applications involving single and multiple transformations. There are 12 topic codes in this category.

Category 13. Geometry: Constructions

This category includes constructing perpendicular and parallel lines; angle bisectors; angles (including specific angles such as 60° angle); constructing triangles under given conditions (e.g., all possible triangles given an angle and two sides of a triangle); constructing quadrilaterals under given conditions, constructing tangents to circles; inscribing certain polygons in a circle; and dividing lines into specific ratios. Also included is the use of computer software. There are 26 topic codes in this category.

Category 14. Statistics/Probability: Graphical Representations of Data

This category includes constructing and interpreting bar graphs, circle graphs, line graphs, histograms, scatter plots, stem and leaf plots, and frequency polygons. Also included are gathering data and selecting the appropriate graph; recognizing bias in a sample; and recognizing misuse of graphs. Thee are 21 topic codes in this category.

Category 15. Statistics/Probability: Statistics

This category includes defining measures of central tendency and measures of dispersions; and determining procedures for computing these measures. Also included are applications involving selecting the appropriate measure; determining the measure; and identifying misinterpretation and misuses of these measures. Other coded topics include standard deviation; various distributions (e.g., normal, skewed); and identifying a representative sample. There are 27 topic codes in this category.

Category 16. Statistics/Probability: Probability

This category includes definitions and applications of theoretical and empirical probability. Also included are definition and applications of complementary, independent, and dependent events. Other topics included are expected outcome; odds; conditional probability; and advanced counting principles. There are 27 topic codes in this category.

Category 17. Algebra: Linear Functions: Simplifying Expressions and Solving Equations.

This category includes simplifying algebraic expressions; solving linear equations; applications involving linear equations; solving pairs of linear equations and their related applications; solving linear inequalities and related applications; and solving absolute value equations and inequalities. This category includes 49 topic codes.

Category 18. Algebra: Linear Functions: Graphs

This category includes plotting points; determining slope and y-intercepts from the associated linear function or from the graph; determining slopes of parallel and perpendicular lines; determining the domain and range of functions; graphing linear functions; representing a graph with a linear function; solving pairs of equations graphically; and responding to questions based on a situation and its associated graph. Also included are graphing linear inequalities; graphing absolute value equalities and inequalities; determining the procedure for computing the distance between two points and computing these distances; and determining the co-ordinates of the mid-point of a line segment. This category includes 39 topic codes.

Category 19. Algebra: Quadratic Functions and Other Non-linear, Non-trigonometric Functions

This category includes operations with quadratic functions; factoring quadratic functions; real and irrational numbers; operations with complex numbers; developing the quadratic formula; the quadratic discriminant; solving quadratic equations; solving applications involving quadratic equations; and direct and inverse variation. Also included are solving higher order equations. This category includes 21 topic codes.

Category 20. Algebra: Graphing Non-Linear, Non-Trigonometric Functions

This category includes graphing equations by plotting points; estimating solutions to quadratic equations from the graph; applications based on the graph; graphing exponential functions; graphing conic sections given by name; graphing higher powers; and applications based on the graphs of higher powers. This category includes 13 topic codes.

Category 21. Trigonometry

This category includes definitions of trigonometric functions (both based on the right triangle and the unit circle); applications involving trigonometric functions; graphing trigonometric functions; and finding values for specific angles (e.g., $\sin 30^\circ$). Also included are proving trigonometric identities and solving equations based on identities. There are 14 topic codes associated with this category.

Category 22. Miscellaneous Topics

This category includes Venn diagrams and applications using Venn diagrams; properties of real numbers; operations using other bases or alternative algorithms (e.g., lattice multiplication); and logic problems. There are 10 topic codes for this category.

7.1.2 Coding for Complexity

Procedural complexity was judged primarily on the number of steps leading to a solution and the number of sub-problems which must be completed in order to solve the original problem. It should be emphasized that this code was related to the procedural and not the conceptual complexity of the problem.

Three coding categories were developed:

Low Complexity

General guidelines for assigning this category were as follows: The solution process for the problem required four or fewer steps/decisions and no sub-problems. If a student represented a situation with an equation, this constituted one step. If the problem required obtaining information from a graph or table when the exact information was provided in the table/graph, this would designated as a low procedural complexity problem—assuming the problem did not require more than four steps to resolution. Examples of low procedural complexity problems are: (a) solve the equation $3(x + 2) + 5 = 7$ and (b) find the surface area of a right cylinder, given the height and the radius of the base.

Moderate Complexity

These problems required one or more of the following: (a) solving a sub-problem, (b) more than four steps/decisions are necessary to solve the problem, (c) the need to extrapolate from quantities in a table or graph. Examples of moderate procedural complexity problems are: (a) $3(x - 2) = 7 - (x + 4)$ and (b) determine the height of a right circular cylinder given the radius and the surface area.

High Complexity

These problems required at least two sub-problems. For example, (a) construct a set of 20 scores such that the mean and median differ by one, and (b) compare the surface areas of a sphere and a right circular cylinder which contain the same volume.

7.1.3 Coding for Relationship

The relationship among problems was judged by examining each problem and a preceding problem in the lesson. For example, a problem could require the same solution procedures as a previous problem, it could require some important additional operations, or it could be totally unrelated to any preceding problem.

An initial set of four coding categories was suggested by the mathematics code development team. After coding a number of lessons, the members of the mathematics problem analysis group found it necessary to add another seven categories in order to fully capture the relationship among the various problems in the lesson.

The 11 relationship categories are described below:

Repetition [R]

This category indicates the problem was exactly or mostly the same as the preceding problem. The numbers or algebraic expressions may have been different, but the procedures were the same.

Repetition [RR]

This category indicates the problem was exactly or mostly the same as one of the preceding problems in the lesson.

Dependent [D]

This category indicates the solution to the previous problem is necessary to solve the current problem.

Dependent [DD]

This category indicates the solution to one of the previous problems in the lessons was necessary to solve the current problem.

Extension [EX]

This category indicates the problem required many of the same operations as the preceding problem plus some important additional operations. The category also includes cases where the problem was a generalization of previous problems.

Extension [EEX]

This category indicates the problem required many of the same operations as a previous problem in the lesson plus some important additional operations.

Simplification [S]

This category was assigned when the problem illustrated a simpler example of the previous problem or was used to provide emphasis (e.g., to compare $a + a$ with $a \cdot a$).

Elaboration [E]

This category indicates the problem was similar to the previous problem but used a different set of operations (e.g., solving the problem another way).

Thematic Connection 1 [T1]

This category indicates the problem required operations that were much different than the first. However the mathematics topic was similar. This code was only used when there was a mathematical thematic connection and no other relationship applied (e.g., finding the mean and the median of a set of numbers)

Thematic Connection 2 [T2]

This category indicates the connection was with the scenario. The code was only used when no other relationship applied and a thematic connection was apparent.

Unrelated [U]

This category was assigned when the problem required operations much different than other problems in the lessons and neither of the thematic codes applied.

Since initial reliability scores for this code were below 85 percent, the team agreed to collapse some of the coding categories. R and RR were collapsed, D and DD were collapsed, and EX and EEX were collapsed, resulting in 8 final categories of relationships.

7.1.4 Reliability

The members of this group each established reliability with the director by coding a randomly selected set of lessons from each country. They computed initial reliability as well as reliability after approximately two-thirds of the lessons had been coded. Their percent agreement was above 85 percent for each of the three codes at both time points.

Initial reliability was computed on a set of 33 randomly selected lessons. The set included five lessons from each of six countries (Australia, the Czech Republic, Hong Kong SAR, the Netherlands, Switzerland, and the United States), and three lessons from Japan. Altogether the 33 lessons contained 747 problems. This meant there were 747 topic and procedural complexity codes and 713 relationship codes (since the initial problem in each lesson was not assigned a relationship code).

The director of the mathematics problem analysis group prepared a “master” for each lesson. Table 7.1 lists the percentage agreement for each code between each of the two coders and the director. Also noted is the average percentage agreement for each code with the two coders’ scores combined.

Table 7.1. The mathematics problem analysis group’s initial reliability scores: 1999

Coders	Topic (percent agreement)	Procedural complexity (percent agreement)	Relationship (percent agreement)
Coder 1	90	83	87
Coder 2	88	90	88
Combined	89	87	88

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Midpoint reliability was established after approximately two-thirds of the lessons had been coded. Again, the director prepared a “master” for each lesson. However, the two coders each coded different lessons. Coder 1 coded 10 lessons and coder 2 coded 8 lessons. Coder 1’s lessons included 237 problems (237 topic and procedural complexity codes, and 227 relationship codes). Coder 2’s lessons included 135 problems (135 topic and procedural complexity codes, and 127 relationship codes). Their reliability is presented in table 7.2

Table 7.2. The mathematics problem analysis group's midpoint reliability scores: 1999

Coders	Topic (percent agreement)	Procedural complexity (percent agreement)	Relationship (percent agreement)
Coder 1	92	92	84
Coder 2	87	88	87
Combined	90	90	88

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

7.2 Mathematics Quality Analysis Group

A second specialist team possessed special expertise in mathematics and teaching mathematics at the postsecondary level. The mathematics quality analysis group was directed by Alfred Manaster (University of California, San Diego) and included Phillip Emig (California State University, Northridge), Wallace Etterbeek (Sacramento State University), and Barbara Wells (University of California, Los Angeles). The same team previously was commissioned to develop and apply codes for the TIMSS 1995 Video Study. The mathematics quality analysis group reviewed country-blind written records for a randomly selected subset of 20 lessons from each country except Japan. Japan was not included because the group already had analyzed a subsample of the Japanese lessons as part of the TIMSS 1995 Video Study which meant, among other things, that country blindness could not be ensured (see Stigler et al. (1999) and Manaster (1998) for a report of the group's findings in the 1995 Study).

7.2.1 Developing Extended Lesson Tables

Specially trained members of the international video coding team created extended written records for each lesson in this subset. These records contained substantially more detail about the mathematics and pedagogy in the lesson than those used by the mathematics problem analysis group. These 120 tables all followed the same format: they included details about the classroom interaction, the nature of the mathematical problems worked on during class time, descriptions of time periods during which problems were not worked on, mathematical generalizations, labels, links, goal statements, lesson summaries, and other information deemed relevant to understanding the content covered during the lesson. Most tables were accompanied by an appendix that contained screen-shots from the video[16], mainly including graphics that provided information about the statement or solution of a problem or other mathematical assertion.

The tables were "country-blind," with all indicators that might reveal the country removed. For example, "pesos" and "centavos" were used as units of currency, proper names were changed to those deemed neutral to Americans, and lessons were identified only by an arbitrarily assigned ID number. The mathematics quality analysis group worked solely from these written records, and had no access to the video data.

[16] These screen shots did not include pictures of teachers or students, or any other information which could be used to identify the country.

7.2.2 Constructing Timelines

The group's first step in analyzing these tables was to divide each lesson into segments and construct a timeline that reflected the flow of the mathematics in the lesson. The timeline began with the first presentation of material or discussion directly related to mathematics. A new segment started when there appeared to be a significant shift in the mathematics content. Each segment was described according to the mathematics presented, and also indicated how the content was treated—for example by students working individually or in groups, or by a public presentation of problems.

A draft timeline and description of each segment was prepared by at least one of the four group members, and then reviewed and revised until there was unanimous agreement by the entire team. Considerable discussion sometimes occurred prior to agreement on the segment divisions and the description of content within each segment.

7.2.3 Developing and Applying the Coding Scheme

Based on the extended lesson tables and timelines, the mathematics quality analysis group created and applied a coding scheme to describe both the segments and the lessons as whole units. The scheme was reviewed by mathematics experts in each country and then revised based on the feedback received.

The group applied their coding scheme by studying the written records of the lessons and reaching consensus about each judgment. In the sections below each code is described in more detail, following the order in which they were discussed in the international report.

7.2.3.1 Content Level

The group rated the content level of each lesson on a scale from 1 (elementary) to 5 (advanced), considering the curriculum covered over the span of the lesson. A score of 3 (moderate) was given to lessons that included content often encountered by students just prior to the standard topics of late elementary or early secondary algebra. One rating was assigned to each lesson based on the rating that best described the content of the lesson, taken as a whole.

7.2.3.2 Procedural, Conceptual, or Notational Mathematics

Each lesson segment was classified as containing procedural, conceptual, and/or notational mathematics. Segments might have contained one or more of these types of mathematics, or none of them.

Lesson segments containing procedural mathematics included those in which problems were solved by executing procedures that appeared to be known by the students. They were also coded when a procedure was presented without much explanation.

In conceptual segments the mathematical concepts, ideas, or procedures were developed. For example, a procedure might have been introduced as the outgrowth of an underlying mathematical property. Segments of conceptual mathematics might have included examples and explanations for why things work like they do. Often the development and first application of a solution procedure was coded as conceptual, whereas subsequent occurrences of the method were coded as procedural.

Segments were characterized as notational when a mathematical definition was presented, or when notational conventions commonly used in mathematical activity were discussed.

7.2.3.3 Mathematical Reasoning: Deductive, Developing a Rationale, Generalizations, and Counter-examples

The mathematics quality analysis group found few instances of mathematical reasoning in the TIMSS 1995 Video Study data. Therefore, for the TIMSS 1999 Video Study, they elaborated and sharpened their coding scheme in an attempt to identify a variety of special reasoning forms that might be present in eighth-grade mathematics lesson. They still required the reasoning to be explicit in order to be marked as such. An exception to this general rule was made when the nature of the problem being solved required reasoning for its solution.

Several kinds of reasoning were recorded whenever they were seen explicitly in a segment of the lesson. Deductive reasoning refers to the derivation of a conclusion from stated assumptions using a logical chain of inferences. There was no requirement that the derivation be formal (e.g., a formal proof), but there was usually an accompanying explanation.

Developing a rationale was coded when there was an explanation or motivation, in broad mathematical terms, of a mathematical assertion or procedure. This type of reasoning was less systematic or precise than deductive reasoning. For example, teachers might show that the rules for adding and subtracting integers are logical extensions of those for adding and subtracting whole numbers, and that these more general rules work for all numbers. When such explanations took a systematic logical form, they were coded as deductive reasoning; when they took a less systematic or precise form, they were coded as developing a rationale.

Generalizations were marked when several examples led to the formulation of an assertion about their shared properties. This process is similar to what many people call inductive reasoning. Generalizations might involve, for example, graphing several linear equations such as $y = 2x + 3$, $2y = x – 2$, and $y=-4x$, and making an assertion about the role played by the numbers in these equations in determining the position and slope of the associated lines.

Segments were coded as containing a counter-example whenever an example was used to show that an assertion cannot be true. For instance, suppose someone claims that the area of a rectangle gets larger whenever the perimeter gets larger. A counter-example would be a rectangle whose perimeter becomes larger but the area does not become larger.

7.2.3.4 <u>Overall Judgment of Mathematical Quality: Coherence, Presentation, Student Engagement, and Overall Quality</u>

The mathematics quality analysis group made four judgments about the mathematical quality in the lessons, using a 5-point rating scale for each judgment. First they rated the lessons on coherence. That is, how well the mathematical components of the lesson were interrelated, ranging from “fragmented” to “thematic.” A rating of 1 indicated that the lesson had multiple unrelated themes or topics, and a rating of 5 indicated that the lesson had a central theme that progressed saliently through the whole lesson.

Presentation ratings were based on the extent to which the mathematics was developed over the course of the lesson, on a scale ranging from “undeveloped” to “fully developed.” The rating depended upon the extent to which mathematical reasons and justifications were provided for the mathematical results presented or used in the lesson and the quality of these mathematical arguments. This judgment also took into account whether links were made between known material and less familiar material, and whether mathematical errors were made by the teacher. The lowest rating was applied to lessons that were descriptive or routinely algorithmic with little mathematical justification provided for why things work like they do. The highest rating was applied to lessons in which concepts and procedures were mathematically motivated, supported, and justified.

The group also examined the likely extent of students being actively engaged with meaningful mathematics during each lesson. The scale ranged from very unlikely to very likely. A rating of very unlikely indicated a lesson in which students were asked to work on very few problems and those problems did not appear to stimulate reflection on mathematical concepts or procedures. A rating of very likely indicated a lesson in which students were expected to work actively on, and make progress solving, problems that appeared to raise interesting mathematical questions for them and then to discuss their solutions with the class.

The last judgment made by the mathematics quality analysis group concerned the overall quality of the lesson. This judgment took into account the three codes described above— coherence, presentation, and student engagement—and was defined as the opportunities that the lesson provided for students to construct important mathematical understandings. The rating scale ranged from low to high.

7.3 Problem Implementation Analysis Team

The problem implementation analysis team was directed by Margaret Smith (University of Iowa) and included Christopher S. Hlas (University of Iowa). They analyzed a subset of mathematical problems and examined 1) the types of mathematical thought processes implied by the problem statement and 2) whether or not those mathematical processes were publicly addressed in the completion of the problem. That is, the team explored whether the assumptions about the kinds of mathematics students would participate in—based on the kinds of problems they were assigned—were realized in the completion of those problems.

Mathematical processes in the problem statement and completion of the problem were determined by examining the types of mathematical thinking and reasoning typically associated with the problem statement and those made explicit during public discussion of the problem. For example, the team determined whether the problem asked students to conjecture or reason, or whether the problem statement was one typically associated with the execution of a mathematical procedure. Then they explored whether the problem was completed in a way that reflected the problem statement, or whether there was evidence of mathematical reasoning not implied by the problem statement.

Using the video data, translated transcripts, and the same lesson tables provided to the mathematics problem analysis group (see section 7.1), the problem implementation analysis team analyzed only those problems that were publicly completed during the videotaped lesson (that is, those independent and concurrent problems for which a target result was publicly presented, as described in chapter 6). Problems had to be publicly completed in order for the group to validly code "problem implementation." Furthermore, the group did not analyze data from Switzerland, since most of the Swiss transcripts were not translated into English.

Reliability was established by comparing a random set of 10 lessons from each country coded by the director of the group with one outside coder. Reliability of at least 85 percent was achieved for both codes across all countries.

Coding by the problem implementation analysis group was carried out in two phases. In the first phase, coders identified the nature of each problem statement that was assigned to students. In the second phase, coders identified the problem implementation type for each problem completed publicly during the videotaped lesson.

7.3.1 Coding Problem Statement Types

Problem statements for all independent and concurrent problems were identified in lesson tables created by the international coding group (see chapter 6). One of three mutually exclusive and exhaustive coding categories was applied to each problem statement, which identified the main mathematical process objective implied by the problem statement. These categories were assigned without consideration of the information or problems that had previously been presented in the lesson. It was assumed that such contextual information would be captured in the problem implementation code (National Assessment Governing Board 1999; Robitaille 1995; Schmidt et al. 1997; U.S. Department of Education 1999).

7.3.1.1 Problem Statement—Using Procedures

Problem statements coded as "using procedures" were those typically associated with routine algorithms such as calculations, symbol manipulation, and practicing of formulae. These problems are generally associated with following a routine process or set of "steps." This category did not imply that there were no mathematical decisions to be made, but rather that the decisions assumed a set path—such as in a computer decision-making scheme. For example, in a problem such as "Solve for x in the equation $3x+4=2x-1$" students could make decisions about how to rearrange the equation that lead to a certain routine path (i.e., a student could choose to add 1 to both sides or subtract 2x from both sides.) Other examples of problem statements coded as using procedures are the following: “Given two sides of a rectangle find the area” and “Calculate the length of the hypotenuse of a right triangle given the length of two sides.”

Not all using procedures problems were decontextualized. For instance, the following “application” problem would be coded as using procedures: “A summer parks recreation program has space for 60 campers. On the first day of enrollment 32 campers enrolled. Reduce 32/60 to lowest terms to find out what fraction of the space is filled.” The key aspect was that the problem asked students to complete a problem typically associated with routine procedure.

7.3.1.2 __Problem Statement—Making Connections__

Problem statements coded as making connections were those that asked students to engage in special forms of mathematical reasoning such as conjecturing, generalizing, and verifying. They were situations that asked students to think about mathematical concepts, develop mathematical ideas, or extend concepts and ideas.

As noted above, application or contextualized problem statements were not necessarily coded as making connections. Conversely, making connections problem statements did not need to be contextualized. An example of a decontextualized making connections problem statement would be if students were given an equation and then asked to determine the effect of changing one of the coefficients on the corresponding graph (e.g., “What if instead of y=3x+4 I had a negative three, so it was y=-3x+4. Would that do anything to my graph?”).

Some other examples of making connections problem statements included those that asked students to find a pattern, describe a relationship, generalize, compare results and methods, find examples of a mathematical principle, or write a problem with given conditions.

7.3.1.3 __Problem Statement—Stating Concepts__

Problem statements coded as stating concepts asked students to recall information regarding a mathematical definition, formula, or property. These problems typically had one step in which the recall of such information was needed to fit the example to a definition or property. Examples include: “Plot the point (3, 2) on a coordinate plane” and “Draw a polygon that is not convex.”

7.3.2 Coding Problem Implementation Types

After coding the types of problem statements in a lesson, coders identified the way in which each problem was completed, or implemented, during the videotaped lesson. Because the problem implementation code was designed to capture the types of mathematical processes made explicit during the lesson, only communication that was available to the whole class was used for these coding decisions. Coders relied upon all public verbal and non-verbal communication occurring after the problem statement and through the completion of the problem to apply this code.

When coding for problem implementation, coders were asked to consider the underlying mathematical concepts of the problem and to think about the connections that were discussed during the lesson. There were no "cue" words that coders could depend upon to identify the type of implementation. For example a teacher asking a student to "explain why" did not necessarily mean that there would be a discussion of mathematical principles and relationships. Additionally, coders could not consider time as indicative of the type of mathematical processes that transpired. For example, a class might have taken a long time to work through a problem, the discussion of which consisted solely of listing of the procedures that were used.

Coders needed to rely heavily on their knowledge of mathematics when considering the nature of the mathematical processes made explicit during the implementation. To help coders to make coding decisions, two rules of thumb were created:

> Rule 1: When unsure of the code you would like to apply, pretend that the problem is being completed in a different country and decide what code you would assign if it took place in that country. It will be useful to consider using a country that appears different from the one coding. For example, if coding a lesson in the United States, consider the code you would apply if the conversation were taking place in a Japanese lesson.
>
> Rule 2: When unsure about how explicit the nature of the mathematical talk is, assume that you are a student struggling to understand the intended concept. Consider the types of communication available for you, as the student, to try to make sense of it.

Using these rules, one of four mutually exclusive and exhaustive coding categories was applied to each problem implementation, as described in the sections below.

7.3.2.1 ***Problem Implementation—Giving Results Only***

Problem implementations were marked as giving results only when the public talk about the problem centered solely on the statement of the final result. The only additional talk that may have occurred was a statement of the problem. No intermediary steps or connections were discussed or shown publicly.

This implementation type was applied to problems where the teacher read off the solutions, or "cycled" through students trying to solicit the correct answer. It was also possible that the final result was displayed on the board or overhead. If such a display included the steps used to complete the problem then it was not coded as giving results only.

7.3.2.2 ***Problem Implementation—Using Procedures***

The problem implementation was coded as using procedures when the routine execution of an algorithm was used to work on and complete a problem. Generally speaking, in this type of problem students and teacher talked only about how to progress to find the answer, such as stating the steps taken along the way.

Some "why" questions may have been addressed during the execution of the problem, but in using procedures implementations the responses to such questions included only descriptions of how to complete the problem or the "rule" that was being followed, rather than focusing on the underlying mathematical concepts. For example, the teacher might have responded to a why question by saying "You need to divide 5 on both sides because whatever you do to one side you do to the other," rather than addressing what it means for two groups to contain equal quantities or some other underlying mathematical concept.

7.3.2.3 ***Problem Implementation—Making Connections***

Problem implementations were coded as making connections when the completion of such problems included mathematically rich discussions. Such discussions might focus on mathematical relationships, and include descriptions of properties and concepts containing mathematical justifications that were not stated as rules but as logically necessary consequences. If applicable, relationships between examples and principles might be demonstrated. Moreover, these mathematical ideas and relationships needed to be made explicit for all members of the class to see and think about the connections.

Some examples of making connections problem implementations included: describing connections between multiple representations (i.e., pictorial and numeric), making and justifying generalizations, comparing the mathematics of different solution methods, and considering why a particular process was mathematically appropriate.

7.3.2.4 ***Problem Implementation—Stating Concepts***

A problem implementation was coded as stating concepts if, during the completion of the problem, the class alluded to a mathematical concept but did not provide any descriptions of mathematical relationships or note why the concept was appropriate for the given situation. In other words, the mathematics discussed included more than a statement of the steps that were followed, but there was no description of how the example was related to underlying mathematical concepts.

Stating connections problem implementations included: stating a formula without addressing why the formula was appropriate for the given example, stating a property as justification (i.e., distributive property), and citing a definition without describing its relationship to the given problem.

Early coding attempts revealed a tendency to apply this code more frequently than others, likely because it was a "middle ground." To help to avoid this error in coding, coders were asked to identify what mathematical concept(s) were addressed and then to identify what additional explanation would have been necessary for the problem implementation to be coded as making connections. Making these explicit allowed coders to identify features of the implementation that differentiated stating concepts implementations from using procedures or making connections implementations.

7.3.3 Examples

To help describe how problems were coded and to show how different implementations of the same problem statement might look, some examples are provided. The first example shows how a using procedures problem statement could be implemented as either using procedures or as making connections. The second example shows how a making connections problem statement could be implemented as making connection or as using procedures. These examples are based on problems that occurred during the videotaped lessons.

7.3.3.1 __Using Procedures Problem Statement Implemented as Using Procedures__

The following problem would be coded as a using procedures problem statement with a using procedures implementation:

Solve for x in the equation 2x + 3 = x - 5.
A student describes the steps used to arrive at the result of x=-8.
Student: *First I subtracted x from both sides and got 2x + 3 - x = x - 5 -x and then I combined like terms and got x + 3 = -5. So then I subtracted 3 from both sides and got x = -5 - 3 which is equal to -8.*

7.3.3.2 __Using Procedures Problem Statement Implemented as Making Connections__

The same problem statement as in the example above (*Solve for x in the equation 2x + 3 = x – 5)* would have a making connections implementation in the following scenario:

The class discusses the fact that this equation has one solution (i.e., x = -8). They compare this equation with those equations that are identities—for which all values of x hold true. Such a discussion would lead the implementation to be considered as making connections because the class is trying to connect ideas about what makes an equation have one solution rather than an infinite number of solutions.

Teacher: *What did you do?*
Student 1: *First I tried to see if all values of x would solve the problem.*
Teacher: *How did you do that?*

Student 1: *Well, first I tried some values and found that they did not always come out equal if I plugged them in for x. So, 2 times 1 is 2, plus 3 is 5, but 1 minus 5 is negative 4 not positive 5.*

Teacher: *So, do you know for sure?*
Student 2: *I tried to make the two sides look the same. I multiplied x and negative 5 by negative 2 and got 2x - 10 but -10 is not +3 so they were not the same.*
Student 3: *Well, Student 1 would have to try all the numbers which would take a very long time. But Student 2 does not because she could not find one number that would make the two sides equal when multiplying.*

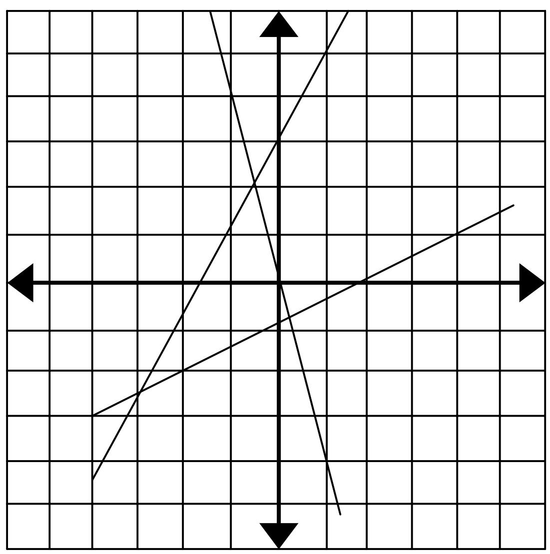
Student 4: *I just went and tried to make x equal to something, not just x equal to x. So, I worked the problem and got x equal to -8. If I had not gotten x equal to -8 but got something like 2 equals 2 or something like that, then I would know that anything I put in for x could be used because both sides of the equation are saying the same thing. Like when we did 2x - 6 = 2(x-3) and we found that we could use any number for x and always get both sides to equal each other. Then we had to see what was making both sides of the equation the same thing. But in this problem since I got a value for x that means it can only be that one value.*

7.3.3.3 <u>Making Connections Problem Statement Implemented as Making Connections</u>

The following problem would be coded as a making connections problem statement with a making connections implementation:

Using the equations y=2x+3, 2y=x-2, and y=-4x, examine the role of the numbers in determining the position and slope of the associated lines.

After constructing the graph, students discuss different aspects of the lines and how these aspects relate to the graphs of each line.



As part of this discussion, students try to make sense of the connection between an equation written in y=mx+b form to its graphical representation. After considering the role the constant plays in determining where the line crosses the y-axis, the class begins to discuss the relationship between the equation, the graph, and substituting 0 for x. They resolve that it is okay to generalize that in y=mx+b form the constant determines where the line crosses the y-axis because substituting zero for x in this form of the equation is the same thing. The second part of the discussion transpires in a similar manner and looks at the effect the sign of "m" has on the position of a line. This discussion involves examining the role of the sign in the equation as well as in the graph.

This type of implementation would be coded as making connections because the class connects different representations, discusses the effects that changes have on these representations, and considers ways to generalize and justify these generalizations. (Sample dialogue was not provided for this example because it would be complex and lengthy.)

7.3.3.4 <u>Making Connections Problem Statement Implemented as Using Procedures</u>

The same problem statement as in the example above (*Using the equations y=2x+3, 2y=x-2, and y=-4x, examine the role of the numbers in determining the position and slope of the associated lines)* would have a using procedures implementation in the following scenario:

A student graphs each of the lines on the chalkboard, and then discusses the steps he used.

Student: *The first one is already in y = mx + b form, so I started at 3 on the y-axis and went up 2 and over 1 because m is 2. Then I connected the points. For the next one I had to divide both sides by 2 so I got y = 1/2 x - 1. I started at -1 and went up 1 and over 2 this time because m is ½, not 2, here. For the last one there was no b, so I plugged in 0 for x and got 0 for y, then I plugged in 1 for x and got negative 4 for y, and then I plugged in -1 for x and got 4 for y. I plotted those three points then connected them.*

Teacher: *Okay, so those are your three lines. Any questions?. . .No? Okay, next lines.*

7.3.4 Reliability

Two coders independently coded 10 lessons from each country, which is at least 10 percent of the lessons from each country. These lessons were randomly selected from those lessons that included at least one problem that was completed publicly during the lesson. Reliability was calculated by comparing the number of agreements with the total number of independent problems (IPs) and concurrent (sets of) problems (CPs) with the at least one target result presented publicly in the videotaped lesson. Average inter-rater agreement for problem statements and implementations is shown in table 7.3.

Table 7.3. Average inter-rater agreement in coding for problem statement and problem implementation types, by country: 1999

Country	Agreement on problem statements (percent)	Agreement on problem implementations (percent)
Australia	87	90
Czech Republic	90	91
Hong Kong SAR	92	90
Japan	95	93
Netherlands	87	89
United States	90	88
Average reliability	90	90

NOTE: Inter-rater agreement was calculated as the number of agreements divided by the sum of the number of agreements and disagreements.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

7.4 Text Analysis Group

The text analysis group used designated portions of the mathematics lesson transcripts to conduct various discourse analyses. The group was directed by Bruce Lambert (University of Illinois at Chicago) and included Clement Yu (University of Illinois at Chicago), David Lewis, Fang Liu (University of Illinois at Chicago), Rodica Waivio (University of Illinois at Chicago) and Sam Mansukhani (University of Southern California). The group employed computer software to conduct quantitative analyses of classroom talk during periods of public interaction. Word-based or lexical features were used to analyze the teacher and student talk in the mathematics lessons.

Because of resource limitations, computer-assisted analyses were applied to English translations of lesson transcripts. In the case of the Czech Republic, Japan, and the Netherlands all lessons were translated from the respective native languages, and in the case of Hong Kong SAR, 34 percent of the lessons were conducted in English, so only 66 percent were translated. English translations of Swiss lessons were not available.

Transcriber/translators were fluent in both English and their native language, educated at least through 8th grade in the country whose lessons they translated, and had completed two-weeks training in the procedures detailed in the TIMSS 1999 Video Study Transcription and Translation Manual. A glossary of terms to was developed to help standardize translation within each country.

The translation and transcription of lesson videos was organized and supervised by David Olsher (University of California at Los Angeles), Wendy Klein (University of California at Los Angeles), Lindsey Engle (University of California at Los Angeles), Don Favareau (University of California at Los Angeles), and Susan Reese (LessonLab).

7.5 Conclusion

In addition to the team of international coders described in chapter 6, four specialist groups were enlisted to analyze portions of the TIMSS 1999 Video Study mathematics data. The mathematics problem analysis group and the problem implementation analysis group studied the mathematical problems in the lessons, and the mathematics quality analysis group made judgments about the nature of the mathematics presented in the lessons. Each of these three groups was comprised of individuals with particular expertise in mathematics, mathematics education, and teaching. A fourth group, the text analysis group, created and implemented specially designed software to study the nature of the classroom talk in the lessons.

Chapter 8. Weighting and Variance Estimation

8.1 Introduction

As described in chapter 3, the samples of classrooms for the study were selected using two-stage probability sampling methods. The first stage of selection was the sample of schools. For each subject area (mathematics and science) the second stage involved the random selection of one eighth-grade classroom. Some countries participated for only one subject area, so that one classroom was selected from the eighth-grade classes in that subject area.

To make valid inferences from the data, it was necessary to account for the features of the sample design in the analysis. There were two components to this process. The first was to incorporate into the analysis survey weights that reflected the selection mechanism (in particular the selection probabilities used to draw the samples) and also any nonresponse at the school or classroom level. These survey weights were added so that the estimates from the data would be unbiased as estimates of the relevant parameters in the full population of classes.

The second feature that needed to be accounted for was the effect of the design on the sampling variances of the estimates. Usually in a two-stage design, there is concern about the effects of clustering the data within first-stage sampling units. Because in the TIMSS 1999 Video Study only one classroom was selected from each school, per subject, the important feature that must be accounted for was the stratification employed as part of the school sampling process. For the United States it was also important to reflect the slight clustering of the school sample within the selected geographic primary sampling units (see chapter 3). This was achieved by using the jackknife procedure, which could be implemented in data analyses by utilizing a set of 50 jackknife replicate weights.

This chapter includes the following information:

- the procedure for applying base weights to the sampled classes, reflecting the probability of selection;
- the procedure for conducting nonresponse adjustments to the weights;
- the jackknife replication variance estimation procedure and how it was implemented;
- how the survey weights and replicate weights should be used for analyzing the data; and,
- the response rates for the study.

Flowcharts that describe the detailed steps for weighting the data for each country are included in appendix J. These charts show the order in which the various steps were implemented in each country and the number of records processed at each step. Although each country required a unique approach to weighting, common features applied. The output from the school sampling for each country was obtained from the country representative. This was used to create initial school weights, giving the reciprocal of the selection probability of the school, and also to establish the pattern for the jackknife replicate weights. Then these weights were adjusted for school nonresponse, where required. In most cases Westat proprietary SAS (Statistical Analysis Software) macros for creating jackknife replicate weights and carrying out nonresponse adjustments were used for this purpose, as is reflected in the flowcharts by reference to "REP_BWGT.MAC", "REP_PREP.MAC", and "COLL_ADJ.MAC".

The material in this chapter covers the weighting procedures for the TIMSS 1999 Video Study. The weighting procedures for the TIMSS 1995 Video Study, conducted in Germany, Japan, and the United States, are described in a separate report (Rizzo 1996).

8.2 Classroom Base Weights

Classroom base weights were calculated from two components: school selection probabilities and classroom selection probabilities. In all countries except the United States, the school selection probabilities were based on the probability of each school in the school sample. Classroom selection probabilities were based on the probability, within each school, of the selected classroom.

8.2.1 School Selection Probabilities

Classroom base weights were created by Westat, based on information about the school sampling process provided by the national research coordinators in each country. Such information either included the probability of selection of each school in the sample, or enough detail so that the probability could be readily determined. The selection probability for school, i, was denoted as P_i.

In most countries replacement schools were used to replace selected schools that did not participate. The exceptions were the Italian-speaking area of Switzerland (where all schools were included in the study) and the United States. When replacement schools were used, they were assigned the selection probability that was associated with the replacement school itself. That is, each school was assigned the probability that it would have been selected in the initial sample (although, of course, it was not selected initially). In most cases the original and replacements were very similar schools, and in particular they were similar in size. This meant that the original and replacement schools generally had very similar probabilities of selection to the initial sample.

For the United States the first stage of sample selection consisted of selecting 52 geographic primary sampling units (PSUs). These PSUs were selected with probabilities proportional to population size, with the ten largest metropolitan areas in the country selected with certainty. Then from an aggregate list of schools within the 52 PSUs, a sample of 110 schools was selected. The school selection probability of each of these 110 schools was, therefore, the product of two probabilities: 1) the PSU selection probability, and 2) the school within PSU selection probability.

The school within PSU probabilities were constructed such that when the two probability components are multiplied together, the school selection probability looks just as it would if the sample had been drawn directly from the entire list of schools in the country. Thus the introduction of this additional stage of sampling had no real impact on the base weights assigned to the schools. It did, however, affect the sampling variability of the study estimates. This was therefore reflected in the method of estimating sampling variances via the jackknife procedure, as described in section 8.4.

8.2.2 Classroom Selection Probabilities

One classroom (per subject area) was selected from each school. The classrooms within a school were each given an equal chance of selection. Thus if the number of classes for a subject area in school i was C_i, the classroom selection probability of the selected classroom was $\frac{1}{C_i}$.

8.2.3 Classroom Base Weights

The base weight for each classroom was the reciprocal of the product of the school selection and classroom selection probabilities. That is, for a classroom selected from school i, the base weight, BW_i, was calculated as:

$$BW_i = \frac{C_i}{P_i}.$$

The classroom base weights have the following property: had all schools participated (or been successfully replaced), then the sum of these weights across the entire sample within the country would give an unbiased estimate of the total number of classrooms in a country (or close to an unbiased estimate when replacement schools were used). This property also holds true for subpopulations within a country, such as those defined by type of school or geographic region.

Thus in the absence of nonresponse, these classroom base weights are a mechanism to provide valid generalizations from the sample to the national population. They correct any imbalance that may have arisen in the sample, either as the result of intentional oversampling of some kinds of schools or due to imperfections in the information about the size of a school available at the time of sampling.

In the Czech Republic and Hong Kong SAR, there was 100 percent response once replacement schools were taken into account. Therefore the base weights have the property described above. In the other countries, however, nonresponse adjustments were needed to ensure that the results from data analyses would be close to unbiased.

8.3 Nonresponse Adjustments

This section describes the procedure for creating nonresponse adjustments to compensate for cases where a sampled school had one or more eligible classes but none was videotaped.

First, schools were grouped into cells. The principles in forming cells were that: a) schools within the same cell should be somewhat similar with respect to characteristics that might relate to the phenomena being studied; b) there were at least six responding schools (i.e., the selected classroom was videotaped) in each cell; and c) as many cells could be formed as were reasonable given constraints a) and b).

The idea behind nonresponse adjustments was to compensate for missing data from nonresponding schools by increasing the weights of similar responding schools. Principles a) and c) above were aimed at making the schools that receive such weight adjustments as similar to the nonresponding schools as possible. If such an effort were carried to too great an extreme, however, the beneficial effects of reducing nonresponse bias could be outweighed by the increase in sampling variance that results from assigning different weights to different classes. Principle b) above addressed this concern.

The nonresponse cells were generally based on the sampling stratification variables. There were two reasons for this. The sampling strata were often chosen for the sample design because they were known or thought to be related to the study outcomes. Thus they also make good characteristics for forming nonresponse adjustment cells. The second reason was that the stratification variables were known for the nonresponding schools, but there was little other relevant information available about them. Table 8.1 presents the variables used to form nonresponse adjustment cells and the number of cells created for each country.

Table 8.1. Variables used to form nonresponse adjustment cells and the number of cells created, by country: 1999

Country	Variables used to define nonresponse adjustment cells	Number of nonresponse adjustment cells	Maximum nonresponse adjustment
Australia	Explicit sampling strata	8 (4 of these cells had no nonresponse)	1.57
Czech Republic	No nonresponse after replacement	—	—
Hong Kong SAR	No nonresponse after replacement	—	—
Japan	Explicit sampling strata	4 (2 of these cells had no nonresponse)	1.31
Netherlands	Explicit sampling strata	7 (1 of these cells had no nonresponse for math; 2 for science)	1.42 (mathematics) 1.42 (science)
Switzerland			
French-speaking	Canton[1]	2 (1 of these cells had no nonresponse)	1.48
German-speaking	Nonresponse adjustments made by the national research coordinator	—	—
Italian-speaking	School level and size	3	1.44
United States	Urban/suburban/rural (derived from type of location)	3	1.41 (mathematics) 1.34 (science)

— Not applicable
[1]A canton is the equivalent of a province or state.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Within each nonresponse adjustment cell, a nonresponse adjustment factor was calculated as:

$$NRF = \frac{\sum_{i \in \left\{ \begin{array}{c} \text{eligible} \\ \text{sampled} \\ \text{schools} \end{array} \right\}} BW_i}{\sum_{i \in \left\{ \begin{array}{c} \text{responding} \\ \text{schools} \end{array} \right\}} BW_i}$$

The final weight for the classroom selected from school i (FW_i) was given as the product of the classroom base weight, BW_i, and the nonresponse adjustment factor for the cell to which the school belonged, NRF_i. That is:

$$FW_i = BW_i \times NRF_i$$

Note that in the Netherlands and the United States, the nonresponse adjustments sometimes varied by subject area. This variation was due to the fact that in some schools in these countries the selected class was videotaped for one subject but not the other. In addition, in some schools in the United States the number of eighth-grade mathematics and science classes were not the same.

8.4 Variance Estimation using the Jackknife Technique

Sampling variances were computed for each country using the jackknife technique. This technique takes into account the design used to select the classroom samples as well as the effect on sampling variance due to the nonresponse adjustments. Nonresponse adjustments were computed in order to mitigate against any nonresponse bias. However, since these adjustments involved calculating ratios of sample estimates within cells and then applying these ratios to the weights, they also have an impact on the sampling variances of estimates derived from the study. The variance estimates obtained via the jackknife approach reflect this appropriately.

The general jackknife technique was implemented as follows: The selected schools were sorted in the order in which they were sampled. That is, they were sorted by explicit sample stratum and then within that sort they were arranged in the order that they were in prior to the systematic selection. Then successive schools were paired.

A single jackknife replicate was created by dropping one of the schools from the sample and doubling the contribution of the complementary pair member of the dropped school. Then the statistic of interest was re-estimated using the modified data set so created. This process was repeated by successively dropping one member (chosen at random) from each of the pairs of schools. For this study the typical design was to select 100 schools, giving rise to 50 pairs of schools. Thus in this way 50 replicate estimates could be derived corresponding to each estimate made from the full data set.

If in general T jackknife replicates are formed, numbered by t = 1,2, …., T, then the appropriate formula for the variance of an estimate, $\hat{X}$, is given by

$$\text{var}\left(\hat{X}\right)=\sum_{t=1}^{T}\left(\hat{X}_{(t)}-\hat{X}\right)^2$$

where the sum from $t=1$ to T is over the replicate estimates, and $\hat{X}_{(t)}$ denotes the estimate of X derived from replicate t.

In practice the jackknife replication procedure is most straightforwardly implemented by creating a set of separate weight variables, one corresponding to each replicate. The weight variable was constructed by setting to zero the replicate weight of the school that was dropped for the replicate in question and giving its complementary school a replicate weight that was double its base weight. All the other schools got a replicate weight for that particular replicate that was the same as its base weight. Thus if 50 replicates were formed, a given school would have 49 of its replicate weights equal to the base weight with the fiftieth being either zero or twice the base weight.

Once these replicate weights have been created from the base weights, the nonresponse adjustment procedures are applied to the full set of replicate weights, just as for the full sample base weight. Thus the final replicate weights for each school may vary somewhat from being equal to the school's full sample weight, or double that weight, or zero, because of different nonresponse adjustment factors calculated for each replicate.

For example, the full sample nonresponse adjustment that applies to a particular school might have a value of 1.1. When the nonresponse adjustment is recalculated for a given replicate (the first, say), the nonresponse adjustment calculated for that school for that replicate might be 1.09, for example. Thus once the nonresponse adjustments are applied to the base weights, for the full sample and each replicate, the pattern of replicate weights will no longer follow the simple relationship to the full sample weight (that applied to the replicate base weights) whereby each replicate weight was either equal to its full sample weight, double that weight, or zero. In this way the replicate weights are able to reflect the impact of the nonresponse adjustments on sampling variance.

Some countries did not have 50 pairs of schools. In those cases replicate pairs were formed as described above, and the unused replicate weights were filled out with values equal to the full sample weight. It can be seen from the above formula that one can add an arbitrary number of additional replicate weights, all equal to the full sample weight, without changing the variance estimate. This was done so that all countries would have the same number of replicate weights on the file. This was needed to make analyses that involved multiple countries practical to carry out.

In the United States a modified approach was needed to reflect the use of the PSU stage of sample selection. For schools from within the ten certainty PSUs, the procedure was as described above. For schools in the other 42 PSUs, 21 replicates were formed by pairing the selected PSUs, again by considering the stratification and sample selection ordering. This resulted in 36 pairs within the United States—21 pairs of PSUs and 15 pairs of schools from within the 10 certainty PSUs. The replicate for a pair of PSUs was formed by deleting all the sampled schools from one of the PSUs and doubling the base weights of all the classes from its paired PSU. Then a final set of an additional 14 replicate weights were created by giving each class the full sample weight for each.

The jackknife technique is described in detail in Wolter (1985) and summarized in Rust (1985), and Rust and Rao (1996). Theoretical properties are summarized in Shao (1996). This jackknife approach is essentially the same as that used in 1994–1995 TIMSS, the TIMSS 1995 Video Study, and TIMSS 1999.

8.5 Using the Weights in Data Analyses

As mentioned earlier, valid population inference using the TIMSS 1999 Video Study data required the use of the full sample weights for parameter estimation and the replicate weights for sampling variance estimation.

For estimating parameters, each variable value from a classroom in the data file should be associated with its full sample weight for all statistics. Thus, to estimate the population mean of variable, *X*, measured for each classroom in the sample, the appropriate formula is:

$$\hat{\bar{X}} = \left(\sum_i FW_i \times X_i \right) \Big/ \left(\sum_i FW_i \right),$$

where X_i is the value of *X* for school *i*.

If estimating the median (or other quartiles), it is the median of the empirical distribution where each class contributes to the distribution in proportion to its value of FW_i. When complex analysis, such as linear regression, are carried out, again each unit should be weighted by FW_i to carry out the analysis.

To obtain appropriate estimates of sampling error, as measured by the estimated standard error of a parameter estimate, the 50 jackknife replicate weights included with the data should be used following the approach described in Section 8.4.

Both the weighting and variance estimation can be carried out using standard statistical software (such as SAS or STATA), or specialized statistical software such as WesVar (Westat 2000) or SUDAAN (version 8 only) (Research Triangle Institute 2001). These specialized programs read in the full sample weights and the 50 replicate weights and automatically apply the approaches to parameter estimation and jackknife replicated variance estimation that are described here.

Most general statistical software can readily apply the full sample weights to arrive at unbiased parameter estimates. However, appropriate standard error estimates cannot be routinely obtained by such software. One must write specific routines to carry out the calculations described in Section 8.4. Because the formula for the jackknife variance estimator takes the same form no matter what the parameter estimator looks like, this is feasible. However, most analysts are likely to find that they can more readily and surely derive appropriate standard error estimates using WesVar or SUDAAN.

The use of the replicate design based on paired schools means that statistical tests on the data should be conducted assuming that the degrees of freedom available for variance estimation is equal to half the number of classrooms in the data. This compares to the standard situation where the number of classrooms would be used as the number of degrees of freedom.

When data from several countries are combined, in general 50 degrees of freedom should be assumed in any analyses. This is because there are only 50 replicate weights on the file no matter how many countries' data are being combined for the analysis.

When conducting analyses that combine data from several countries, it is important to note that, in the absence of any special steps to the contrary, the countries contribute to the combined estimate in proportion to the number of grade 8 classes in the county. Thus the United States will dominate any combined mean, for example. In the case of the mathematics data collected in the TIMSS 1995 Video Study, the situation was exacerbated greatly by the fact that the weights summed to the number of grade 8 students in the country, rather than the number of grade 8 classes (as was the case with TIMSS 1999 Video Study data). Thus a simple combination of the 1999 mathematics data with the 1995 data from Japan will have an overall mean that is dominated by the data from Japan.

8.6 Weighted Participation Rates

This section describes the procedures used to calculate the TIMSS 1999 Video Study's weighted participation rates. A participation rate reflects the proportion of total sampled eligible cases from which data were obtained. In the TIMSS 1999 Video Study, the participation rate indicates the percentage of sampled schools for which videotapes were completed. These rates are presented by country and with the rate components in tables 8.2 and 8.3.

Unweighted participation rates, computed using the actual numbers of schools, reflect the success of the operational aspects of the study (i.e., getting schools to participate). Participation rates weighted to reflect the probability of being selected into the sample describe the success of the study in terms of the population of schools to be represented.

Participation rates were computed both before and after replacement. The participation rate before replacement identifies the proportion of originally sampled schools that participated; the participation rates after replacement gives the percentage of all schools sampled (including original and replacement schools) that participated.

Table 8.2. Mathematics participation rates before replacement, by country: 1995 and 1999

Country	Weighted school participation, before replacement (percentage)	Weighted numerator, before replacement[1]	Weighted denominator, before replacement[2]	Unweighted school participation, before replacement (percentage)	Unweighted numerator, before replacement[3]	Unweighted denominator, before replacement[4]
Australia	61	5,839	9,586	59	59	100
Czech Republic	89	110,877	124,583	89	89	100
Hong Kong SAR	63	49,950	79,286	63	63	100
Japan[5]	96	1,507,288	1,573,369	96	48	50
Netherlands	50	54,454	108,501	50	49	98
Switzerland[6]	71			76	114	151
French-speaking	90	799	886	90	37	41
German-speaking	64	26,089	40,506	67	50	75
Italian-speaking	77	27	35	77	27	35
United States	76	2,105,483	2,755,605	77	83	108

[1]The weighted numerator is the sum of the sampling weights of all the participating schools in the sample.
[2]The weighted denominator is the sum of the sampling weights of all the eligible schools in the sample.
[3]The unweighted numerator is the number of participating schools in the sample.
[4]The unweighted denominator is the number of eligible schools in the sample.
[5]Japanese mathematics videos were collected for the TIMSS 1995 Video Study.
[6] The weighted overall Switzerland participation rates incorporate the student population distribution of the three language-speaking regions of the country: German (74.4 percent), French (21.8 percent), and Italian (3.8 percent). This distribution is based on estimates for the 9th-grade population of Switzerland from OECD (2001). *Knowledge and skills for life: First results from the OECD Programme for International Student Assessment (PISA) 2000*. Organization for Economic Co-operation and Development: Paris. The weights of the separate parts are not analogous. Each area used a different definition for the first-stage sampling measure of size.
NOTE: For Australia, the Czech Republic, and the Netherlands, these figures represent the participation rates for the combined mathematics and science samples.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

Table 8.3. Mathematics participation rates after replacement, by country: 1995 and 1999

Country	Weighted school participation, after replacement (percentage)	Weighted numerator, after replacement[1]	Weighted denominator, after replacement[2]	Unweighted school participation, after replacement (percentage)	Unweighted numerator, after replacement[3]	Unweighted denominator, after replacement[4]
Australia	85	8,127	9,586	85	85	100
Czech Republic	100	124,583	124,583	100	100	100
Hong Kong SAR	100	79,286	79,286	100	100	100
Japan[5]	100	1,573,369	1,573,369	100	50	50
Netherlands	85	92,339	108,501	87	85	98
Switzerland[6]	97			93	140	151
French-speaking	95	842	886	95	39	41
German-speaking	99	40,054	40,506	99	74	75
Italian-speaking	77	27	35	77	27	35
United States	76	2,105,483	2,755,605	77	83	108

[1]The weighted numerator is the sum of the sampling weights of all the participating schools in the sample.
[2]The weighted denominator is the sum of the sampling weights of all the eligible schools in the sample.
[3]The unweighted numerator is the number of participating schools in the sample.
[4]The unweighted denominator is the number of eligible schools in the sample.
[5]Japanese mathematics videos were collected for the TIMSS 1995 Video Study. The response rates after replacement for Japan differ from that reported previously (e.g., Stigler et al. 1999). This is because the procedure for calculating response rates after replacement has been revised to correspond with the method used in the TIMSS 1995 and TIMSS 1999 Achievement Studies.
[6] The weighted overall Switzerland participation rates incorporate the student population distribution of the three language-speaking regions of the country: German (74.4 percent), French (21.8 percent), and Italian (3.8 percent). This distribution is based on estimates for the 9th-grade population of Switzerland from OECD (2001). *Knowledge and skills for life: First results from the OECD Programme for International Student Assessment (PISA) 2000*. Organization for Economic Co-operation and Development: Paris. The weights of the separate parts are not analogous. Each area used a different definition for the first-stage sampling measure of size.
NOTE: For Australia, the Czech Republic, and the Netherlands, these figures represent the participation rates for the combined mathematics and science samples.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

8.6.1 General procedure for weighted participation rate calculations

In general each country's weighted school participation rate was the sum of base weight times the measure of size for all eligible participating sampled schools divided by the combined sum of the base weight times the measure of size of both the eligible participating schools and the eligible refusing schools. Ineligible and excluded schools were not included in the calculations. The basic formulae are:

$$\text{Weighted response rate, before replacement} = 100 \times \frac{\sum_{i \in \left\{\substack{\text{responding} \\ \text{original} \\ \text{schools}}\right\}} MOS_i / P_i}{\sum_{i \in \left\{\substack{\text{eligible} \\ \text{original} \\ \text{schools}}\right\}} MOS_i / P_i}$$

$$\text{Weighted response rate, after replacement} = 100 \times \frac{\sum_{i \in \left\{\substack{\text{all} \\ \text{responding} \\ \text{schools}}\right\}} MOS_i / P_i}{\sum_{i \in \left\{\substack{\text{responding schools} \\ +\text{ refusing originals} \\ \text{not replaced}}\right\}} MOS_i / P_i}$$

The base weights used in the participation rate calculations were those derived directly from the sampling probabilities, prior to any adjustments for school refusals. They were not the final weights delivered to LessonLab but were contained within them, as those final weights consisted of the base weights adjusted to compensate for patterns of nonresponse.

8.6.2 Country-specific procedures

Each country provided a unique measure of size variable. Furthermore, there were possible sample design differences among countries which could potentially affect the way in which an eligible participating school was represented in the participation rate calculation. Therefore, the rate calculation methodology by country is provided for full disclosure and documentation completeness. Table 8.4 shows, by country, the name of the variable that was used to derive the measure of size (MOS) that was used, together with the school base weights, to derive the participation rates.

Table 8.4. Variables used for participation rate calculations, by country: 1999

Country	Measure of size (MOS)
Australia	Number of grade 8 classes
Czech Republic	Number of grade 8 students
Hong Kong	Number of grade 8 students
Japan - science	Number of grade 8 students
Netherlands	Number of grade 8 students
Switzerland	
French-speaking	1 (this is a factor, not a variable)
German-speaking	Number of grade 8 students
Italian-speaking	1 (this is a factor, not a variable)
United States	Number of grade 8 students

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

In the French-speaking area of Switzerland, classes were sampled directly from class lists assembled by canton. Since these were selected from each canton by equal probability, the MOS value for each class was 1. In the Italian-speaking area of Switzerland, all classes were included in the study; so again the MOS value for each class was 1.

The sample for the German-speaking area of Switzerland was a subsample of the TIMSS 1995 schools selected with equal probability (75 desired video schools from 133 schools of unconfirmed participation or selection status in the 1994–1995 TIMSS). School weights were provided to Westat after nonresponse and other adjustments were incorporated, but the TIMSS 1995 participation and replacement status were not readily available to incorporate into the TIMSS 1999 Video Study participation rate calculations. The sampling interval from TIMSS 1995 and the school measure of size was provided with the adjusted school base weights. This allowed for the TIMSS 1999 Video Study participation rates to be based on a derived TIMSS 1995 school base weight, sclwt_0, where sclwt_0=(smplint[17]/MOS2), which removed all adjustment factors and redistributed the base weights back to the refusing original video schools. The sampling rate for the video study (75/133) was not included in the provided school base weights since the video schools were selected with equal probability. It was also not included in the participation rates.

8.7 Summary

Analyses on the TIMSS 1999 Video Study data were conducted using data weighted with survey weights. These weights were calculated specifically for the classrooms in this study. This chapter described how the classroom base weights were calculated and what adjustments were made for nonrespondent selected schools. In addition, the jackknife technique was explained, along with a description of how the weights are intended to be used by anyone wishing to conduct analyses using this data.

---

[17] Smplint is the school sampling interval used to select the TIMSS 1995 Swiss sample.

References

American Association for the Advancement of Science, Project 2061 (1993). *Benchmarks for Science Literacy*. New York: Oxford University Press.

Arafeh, S. and McLaughlin, M. (2002). *Legal and Ethnical Issues in the Use of Video in Education Research* (NCES 2002–01). Washington, DC: NCES. Available at http://nces.ed.gov

Bailey, B.J.R. (1977). Tables of the Bonferroni t statistic. *Journal of the American Statistical Association*, 72, 469-478.

Bakeman, R., and Gottman, J.M. (1997). *Observing Interaction: An Introduction to Sequential Analysis.* Second Edition. Cambridge: Cambridge University Press.

Beaton, A., Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Kelly, D.L., and Smith, T.A. (1996). *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College.

Bruner, J.S. (1966). *Toward a theory of instruction*. Cambridge, MA: Harvard University Press.

Gallimore, R. (1996). Classrooms are just another cultural activity. In D. Speece and B. Keogh (Eds.), *Research on Classroom Ecologies: Implications for Children with Learning Disability* (pp. 229–250). Mahwah, NJ: Lawrence Erlbaum.

Gonzales, P., Calsyn, C., Jocelyn, L., Mak, K., Kastberg, D., Arafeh, S., Williams, T., and Tsen, W. (2000). *Pursuing Excellence: Comparisons of International Eighth-Grade Mathematics and Science Achievement from a U.S. Perspective, 1995 and 1999* (NCES 2001–028). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K.B., Hollingsworth, H., Jacobs, J., Chui, A.M., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeek, W., Manaster, C., and Stigler, J. (2003). *Teaching Mathematics in Seven Countries: Results from the TIMSS 1999 Video Study* (NCES 2003-013). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Manaster, A. B. (1998). Some characteristics of eighth grade mathematics classes in the TIMSS videotape study. *American Mathematical Monthly*, 105, 793–805.

Mullis, I.V.S, Jones, C., and Garden, R.A. (1996). Training for free response scoring and administration of performance assessment. In M.O. Martin and D.L. Kelly (Eds.), *Third International Mathematics and Science Study Technical Report, Volume 1: Design and Development.* Chestnut Hill, MA: Boston College.

Mullis, I.V.S., and Martin, M.O. (1998). Item analysis and review. In M.O. Martin and D.L. Kelly (Eds.) *Third International Mathematics and Science Study*

Technical Report, Volume II: Implementation and Analysis Primary and Middle School Years (Population 1 and Population 2). Chestnut Hill, MA: Boston College.

National Academy of Sciences (1996). *National Science Education Standards*. Washington, D.C.: National Academy Press.

National Assessment Governing Board (1999). *Mathematics Framework for the 1996 and 2000 National Assessment of Educational Progress.* Washington, DC: National Assessment Governing Board.

OECD (2001). *Knowledge and skills for life: First results from the OECD Programme for International Student Assessment (PISA) 2000*. Organization for Economic Co-operation and Development: Paris.

Research Triangle Institute (2001). *SUDAAN User's Manual, Release 8.0*. Research Triangle Park, NC: Research Triangle Institute.

Richardson, V. and Placier, P. (2001). Teacher change. In V. Richardson, (Ed.), *Handbook of Research on Teaching* (4th ed., pp. 905–947). Washington, DC: American Educational Research Association.

Rizzo, L. (1996). Report on classroom sampling weights for the TIMSS Videotape Study. Project report prepared for UCLA by Westat, March 13, 1996.

Robitaille, D.F. (Ed.). (1997). *National Contexts for Mathematics and Science Education: An Encyclopedia of the Education Systems Participating in TIMSS.* Vancouver: Pacific Educational Press.

Robitaille, D. F. (1995). *Mathematics Textbooks: A Comparative Study of Grade 8 Texts*. Vancouver, Canada: Pacific Education Press.

Rust, K. (1985). Variance estimation for complex estimators in sample surveys. *Journal of Official Statistics*, *1, (4),* pp. 381–397.

Rust, K.F., and Rao, J.N.K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, *5, (3),* pp. 283-310.

Schmidt, W. H., McKnight, C. C., Valverde, G. A., Houang, R. T., and Wiley, D. E. (1997). *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Mathematics*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Shao, J. (1996). Resampling methods in sample surveys (with discussion). *Statistics*, Vol. 27, pp. 203-254.

Stigler, J. W., Gallimore, R., and Hiebert, J. (2000). Using video surveys to compare classrooms and teaching across cultures: Examples and lessons from the TIMSS video studies. *Educational Psychologist*, 35(2), 87–100.

Stigler, J. W., Gonzales, P., Kawanaka, T., Knoll, S., and Serrano, A. (1999). *The TIMSS Videotape Classroom Study: Methods and Findings from an Exploratory Research Project on Eighth-Grade Mathematics Instruction in Germany, Japan, and the United States* (NCES 1999-074). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Stigler, J. W., and Hiebert, J. (1999). *The Teaching Gap: Best Ideas from the World's Teachers for Improving Education in the Classroom*. New York: Free Press.

Westat (2000). *WesVar 4.0 User's Guide*. Rockville, MD: Westat.

Wolter, K.M. (1985). *Introduction to Variance Estimation.* New York: Springer-Verlag.

U.S. Department of Education, Office of Educational Research and Improvement. National Center for Education Statistics. (1996). *Pursuing Excellence: A Study of U.S. Eighth-Grade Mathematics and Science Teaching, Learning, Curriculum, and Achievement in International Context: Initial Findings from the Third International Mathematics and Science Study* (NCES 97–198). Washington, DC: U.S. Government Printing Office.

U.S. Department of Education, Office of Educational Research and Improvement. National Center for Education Statistics. (1999). *The NAEP 1996 Technical Report* (NCES 1999–452). By Allen, N.L., Carlson, J.E., and Zelenak, C.A. Washington, DC: National Center for Education Statistics.

This page left intentionally blank.