NATIONAL CENTER FOR EDUCATION STATISTICS

Research and Development Report February 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

U.S. Department of Education
Office of Educational Research and Improvement NCES 99-074

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

James W. Stigler, Patrick Gonzales, Takako Kawanaka,
Steffen Knoll, and Ana Serrano

With the assistance of
Eric Derghazarian, Gundula Huber, Fumiko Ichioka, and Nicole Kersting

A Research and Development Report

U.S. Department of Education

Richard W. Riley
Secretary

Office of Educational Research and Improvement

C. Kent McGuire
Assistant Secretary

National Center for Education Statistics

Pascal D. Forgione, Jr.
Commissioner

Data Development and Longitudinal Studies Group

Martin E. Orland
Associate Commissioner

International Activities Program

Eugene Owen
Director

National Center for Education Statistics

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
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue, NW
Washington, DC 20208-5574
Contact: (202) 219-1333
e-mail: timss@ed.gov

Suggested Citation:
U.S. Department of Education. National Center for Education Statistics,
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 99-074, by James W. Stigler, Patrick Gonzales, Takako Kawanaka, Steffen Knoll, and Ana Serrano.
Washington, DC: U.S. Government Printing Office, 1999

February 1999

Available for downloading at http://nces.ed.gov/timss

Foreword

The Research and Development (R&D) series of the reports has been initiated:

1. To share studies and research that are developmental in nature. The results of such studies may be revised as the work continues and additional data become available.

2. To share results of studies that are, to some extent, on the "cutting-edge" of methodological developments. Emerging analytical approaches and new computer software development often permits new, and sometimes controversial, analysis to be done. By participating in "frontier research," we hope to contribute to the resolution of issues and improved analysis.

3. To participate in discussions of emerging issues of interest to educational researchers, statisticians, and the Federal statistical community in general. Such reports may document workshops and symposiums sponsored by NCES that address methodological and analytical issues or may share and discuss issues regarding NCES practice, procedures, and standards.

The common theme in all three goals is that these reports present results or discussions that do not reach definitive conclusions at this point in time, either because the data are tentative, the methodology is new and developing, or the topic is one on which there are divergent views. Therefore the techniques and inferences made from the data are tentative and are subject to revision. To facilitate the process of closure on the issues, we invite comment, criticism, and alternatives to what we have done. Such responses should be directed to:

Marilyn M. McMillen
Chief Statistician
Statistical Standards Program
National Center for Education Statistics
555 New Jersey Avenue NW
Washington, DC 20208-5654

Executive Summary

This report presents the methods and preliminary findings of the Videotape Classroom Study, a video survey of eighth-grade mathematics lessons in Germany, Japan, and the United States. This exploratory research project is part of the Third International Mathematics and Science Study (TIMSS). It is the first to collect videotaped records of classroom instruction—in any subject—from national probability samples.

OBJECTIVES

The Videotape Classroom Study had four goals:

- Provide a rich source of information regarding what goes on inside eighth-grade mathematics classes in the three countries.
- Develop objective observational measures of classroom instruction to serve as valid quantitative indicators, at a national level, of teaching practices in the three countries.
- Compare actual mathematics teaching methods in the United States and the other countries with those recommended in current reform documents and with teachers' perceptions of those recommendations.
- Assess the feasibility of applying videotape methodology in future wider-scale national and international surveys of classroom instructional practices.

SCOPE AND METHODS

The study sample included 231 eighth-grade mathematics classrooms: 100 in Germany, 50 in Japan, and 81 in the United States. The three samples were selected from among the schools and classrooms participating in the 1994-95 TIMSS assessments. They were designed as a nationally representative sample of eighth-grade students in the three countries, although, as will be explained later, some minor deviations arose.

One lesson was videotaped in each classroom at some point during the school year. The specific date for videotaping was determined in consultation with the school and the teacher in order to minimize conflicts with special events such as field trips or school holidays, and to minimize the videographers' travel expenses. Tapes were encoded and stored digitally on CD-ROM and were accessed and analyzed using multimedia database software developed especially for this project. All lessons were transcribed and then analyzed on a number of dimensions by teams of coders who were native speakers of the three languages. Analyses presented here are based on weighted data. The analyses focused on the content and organization of the lessons, as well as on the instructional practices used by teachers during the lessons.

FINDINGS

The video data are vast and will continue to provide rich analysis opportunities for researchers. The findings reported here, while preliminary, reveal a number of differences in instructional practices across the three cultures. These differences fall into four broad categories: (1) How lessons are structured and delivered; (2) What kind of mathematics is presented in the lesson; (3) What kind of mathematical thinking students are engaged in during the lesson; and (4) How teachers view reform.

How Lessons are Structured and Delivered

To understand how lessons are structured it is important first to know what teachers intend students to learn from the lessons. Information gathered from teachers in the video study indicate an important cross-cultural difference in lesson goals. Solving problems is the end goal for the U.S. and German teachers: How well students solve problems is the metric by which success is judged. In Japan, problem solving is assumed to play a different role. Understanding mathematics is the overarching goal; problem solving is merely the context in which understanding can best grow.

Following this difference in goals, we can begin to identify cultural differences in the scripts teachers in each country use to generate their lessons. These different scripts are probably based on different assumptions about the role of problem solving in the lesson, about the way students learn from instruction, and about what the proper role of the teacher should be.

Although the analyses are preliminary, there appears to be a clear distinction between the U.S. and German scripts, on one hand, and the Japanese script, on the other. U.S. and German lessons tend to have two phases: an initial acquisition phase and a subsequent application phase. In the acquisition phase, the teacher demonstrates and/or explains how to solve an example problem. The explanation might be purely procedural (as most often happens in the United States) or may include development of concepts (more often the case in Germany). Yet still, the goal in both countries is to teach students a method for solving the example problem(s). In the application phase, students practice solving examples on their own while the teacher helps individual students who are experiencing difficulty.

Japanese lessons appear to follow a different script. Whereas in German and U.S. lessons instruction comes first, followed by application, in Japanese lessons the order of activity is generally reversed. Problem solving comes first, followed by a time in which students reflect on the problem, share the solution methods they have generated, and jointly work to develop explicit understandings of the underlying mathematical concepts. Whereas students in the U.S. and German classrooms must follow the teacher as he or she leads them through the solution of example problems, the Japanese student has a different job: to invent his or her own solutions, then reflect on those solutions in an attempt to increase understanding.

In addition to these differences in goals and scripts, we also find differences in the coherence of lessons in the three countries. The greatest differences are between U.S. lessons and Japanese lessons. U.S. lessons are less coherent than Japanese lessons if coherence is defined by several criteria: U.S. lessons are more frequently interrupted, both from outside the classroom and within; U.S. lessons contain more topics—within the same lesson—than Japanese lessons; Japanese teachers are more likely to provide explicit links or connections between different parts of the same lesson.

What Kind of Mathematics is Presented

Looking beyond the flow of the lessons, we also find cross-cultural differences in the kind of mathematical content that is presented in the lessons. When viewed in comparison to the content of lessons in the 41 TIMSS countries, the average eighth-grade U.S. lesson in the video sample deals with mathematics at the seventh-grade level by international standards, whereas in Japan the average level is ninth-grade. The content of German lessons averages at the eighth-grade level.

The quality of the content also differs across countries. For example, most mathematics lessons include some mixture of concepts and applications of those concepts to solving problems. How concepts are presented, however, varies a great deal across countries. Concepts might simply be stated, as in "the Pythagorean theorem states that $a^2 + b^2 = c^2$," or they might be developed and derived over the course of the lesson. More than three-fourths of German and Japanese teachers develop concepts when they include them in their lessons, compared with about one-fifth of U.S. teachers. None of the U.S. lessons include proofs, whereas 10 percent of German lessons and 53 percent of Japanese lessons include proofs.

Finally, as part of the video study, an independent group of U.S. college mathematics teachers evaluated the quality of mathematical content in a sample of the video lessons. They based their judgments on a detailed written description of the content that was altered for each lesson to disguise the country of origin (deleting, for example, references to currency). They completed a number of in-depth analyses, the simplest of which involved making global judgments of the quality of each lesson's content on a three-point scale (Low, Medium, High). (Quality was judged according to several criteria, including the coherence of the mathematical concepts across different parts of the lesson, and the degree to which deductive reasoning was included.) Whereas 39 percent of the Japanese lessons and 28 percent of the German ones received the highest rating, none of the U.S. lessons received the highest rating. Eighty-nine percent of U.S. lessons received the lowest rating, compared with 11 percent of Japanese lessons.

The Kind of Mathematical Thinking in Which Students are Engaged

When we examine the kind of work students engage in during the lesson we find a strong resemblance between Germany and the United States, with Japan looking distinctly different. Three types of work were coded in the video study: Practicing Routine Procedures, Applying Concepts to Novel Situations, and Inventing New Solution Methods/Thinking. Ninety-six percent of student working time in Germany and 90 percent in the United States is spent in practicing routine procedures, compared with 41 percent in Japan. Japanese students spend 44 percent of their time inventing new solutions that require conceptual thinking about mathematics.

Teachers and Reform

A great deal of effort has been put into the reform of mathematics teaching in the United States in recent years. Numerous documents—examples include the National Council of Teachers of Mathematics *Curriculum and Evaluation Standards for School Mathematics* (1989) and the National Council of Teachers of Mathematics *Professional Standards for Teaching Mathematics* (1991)—encourage teachers to change the way they teach, and there is great agreement, at least among mathematics educators, as to what desirable instruction should look like. Although most of the current ideas stated in such documents are not operationalized to the extent that they could be directly coded, it is possible to view some of the indicators developed in the video study in relation to these current ideas.

When the video data are viewed in this way, Japanese teachers, in certain respects, come closer to implementing the spirit of current ideas advanced by U.S. reformers than do U.S. teachers. For example, Japanese lessons include high-level mathematics, a clear focus on thinking and problem solving, and an emphasis on students deriving alternative solution methods and explaining their thinking. In other respects, though, Japanese lessons do not follow such reform guidelines. They include more lecturing and demonstration than even the more traditional U.S. lessons, and we never observed calculators being used in a Japanese classroom.

and demonstration than even the more traditional U.S. lessons, and we never observed calculators being used in a Japanese classroom.

Regardless of whether Japanese classrooms share features of “reform” classrooms or not, it is quite clear that the typical U.S. classroom does not. Furthermore, the U.S. teachers, when asked if they were aware of current ideas about the best ways to teach mathematics, responded overwhelmingly in the affirmative. Seventy percent of the teachers claim to be implementing such ideas in the very lesson that we videotaped. When asked to justify these claims, the U.S. teachers refer most often to surface features, such as the use of manipulatives or cooperative groups, rather than to the key point of the reform recommendations, which is to focus lessons on high-level mathematical thought. Although some teachers appear to have changed these surface-level characteristics of their teaching, the data collected for this study suggest that these changes have not affected the deeper cultural scripts from which teachers work.

Key Points

Bearing in mind the preliminary nature of these findings, as well as the interpretations of the findings, we can, nevertheless, identify four key points:

- The content of U.S. mathematics classes requires less high-level thought than classes in Germany and Japan.
- U.S. mathematics teachers’ typical goal is to teach students how to do something, while Japanese teachers’ goal is to help them understand mathematical concepts.
- Japanese classes share many features called for by U.S. mathematics reforms, while U.S. classes are less likely to exhibit these features.
- Although most U.S. math teachers report familiarity with reform recommendations, relatively few apply the key points in their classrooms.

These initial findings suggest a need for continued analysis of these data on eighth-grade mathematics practices. Caution should be exercised in generalizing to other subjects or grade levels.

Acknowledgments

A project as large as this one would not have been possible without the help of many people. A list, hopefully almost complete, of those who contributed to the project is included as appendix C. Aside from this, there are a number of people who deserve special mention. First, we would like to acknowledge our collaborators: Jürgen Baumert (Max Planck Institute on Education and Human Development, Berlin) and Rainer Lehmann (University of Hamburg) in Germany, and Toshio Sawada at the National Institute of Educational Research in Tokyo, Japan. We also wish to thank Clea Fernandez for her input during the early stages of the project; Michael and Johanna Neubrand for help in understanding German teaching practices; Alfred Manaster, Phillip Emig, Wallace Etterbeek, and Barbara Wells for their work on the mathematics content analyses; and Nancy Caldwell of Westat, who helped out in ways too numerous to mention. Shep Roey cheerfully ran many analyses many times, Lou Rizzo carefully developed the weights applied to the data, and Dave Kastberg supervised the final layout of the report (all of Westat); we greatly appreciate their help. The final report was greatly improved by the hard work of Ellen Bradburn and Christine Welch at the Education Statistics Services Institute. Finally, we would like to acknowledge the major contributions of Lois Peak (U.S. Department of Education), without whom this project would never have been done, and James Hiebert, who has improved the project at every step of the way, from coding and analysis to the writing of this report.

TABLE OF CONTENTS

List of Figures

Chapter 1. Introduction

The Third International Mathematics and Science Study (TIMSS) is the third in a series of international studies, conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA), which has assessed the mathematics achievement of students in different countries. The first two of these studies (Husen, 1967; McKnight, Crosswhite, Dossey, Kifer, Swafford, Travers, and Cooney, 1987) established that there were large cross-national differences in achievement and provided some information on contextual factors, such as curriculum, that could be related to the achievement differences.

In these prior studies, students from the United States scored low in comparison to other countries. Not enough was learned, however, about the contextual factors that might help to explain their relatively low performance. Finding out more about the instructional and cultural processes that are associated with achievement thus became a high priority in planning for the TIMSS.

In accordance with this priority, the National Center for Education Statistics (NCES) funded two studies to complement the main TIMSS study. Both of these studies focus on three countries: Germany, Japan, and the United States. The first involves comparative case studies of various aspects of the education systems of each country. The second is the Videotape Classroom Study.

The primary goal of the Videotape Classroom Study is to provide a rich source of information regarding what goes on inside eighth-grade mathematics classes in Germany, Japan, and the United States. We directed our attention to both teachers and students, seeking to describe the classes from both the perspective of teaching practices and that of the opportunities and experiences provided for students.

Aside from these general goals, the study had three additional objectives:

- To develop objective observational measures of classroom instruction to serve as quantitative indicators of teaching practices in the three countries;
- To compare actual mathematics teaching methods in the United States and the other countries with those recommended in current reform documents and with teachers' perceptions of those recommendations;
- To assess the feasibility of applying videotape methodology in future wider-scale national and international surveys of classroom instructional practices.

In this report we will provide a detailed account of the methods used in the study, as well as a preliminary look at the findings up to this point. We have only started to tap the vast wealth of information available in the videos we collected. But we have made great headway in solving the considerable logistical and methodological challenges presented by the study. This report relates what we have learned thus far.

In this introductory section we discuss what can be learned from classroom observation and the advantages offered by the use of video to collect such information. We also discuss the issues and problems that arise in the course of designing and carrying out a large-scale video survey, and we describe some of the approaches we have taken to meeting these challenges. In the Methods section we provide a detailed account of our methods. In subsequent sections we present results, first regarding the content of classroom instruction, then the organization and processes.

STUDYING PROCESSES OF CLASSROOM INSTRUCTION

This is the first large-scale study to collect videotaped records of classroom instruction in the mathematics classrooms of different countries. It also is the first study—for any grade level or subject matter—to attempt direct observation of instructional practices in a nationally representative sample of students within the United States. Thus this study constitutes an important new database and a new approach to data collection for NCES.

Chief among the factors associated with student achievement must surely be the processes of teaching and learning that transpire inside classrooms. Yet, until now there have been no observational data on instructional processes from a national sample of classrooms. In a series of papers commissioned by NCES in 1985, papers designed to set the agency's priorities for the next 10 years, the need for classroom process indicators was raised numerous times (Hall, Jaeger, Kearney, and Wiley, 1985). Cronin (1985), for example, expressed concern with the paucity of data that could document curricular breadth or the actual implementation of curricular reform in the classroom. Peterson (1985) cited a near complete lack of data on the quality of educational activities in the Nation's classrooms, or even on the time teachers devote to various instructional activities. Including such indicators in the future was a recommendation of the 1985 report.

Studies of classroom process can serve two broad purposes: First, they can result in indicators of classroom instruction that can then be used to develop and validate models of instructional quality. That is, we must understand the processes that relate instruction to learning if we are to be able to improve it. A second purpose of such studies is to monitor the implementation of instructional policies in classrooms. One example of such policies is contained in the National Council of Teachers of Mathematics (NCTM) *Professional Standards for Teaching Mathematics* (1991). The *Standards* represents one point of view on what instruction should look like in the classroom. Operationalizing this point of view in a system of classroom-based indicators would allow us to assess the degree to which the *Standards* are being implemented, and by coupling these indicators with performance measures, the effectiveness of the *Standards* as educational policy.

Despite the obvious value of studying classroom instruction, describing and measuring classroom processes, especially on a large scale, is difficult. To date, measures have been largely based on questionnaires in which teachers report on what happens in their own classrooms. Using questionnaires to measure classroom processes has both advantages and disadvantages, as we discuss here. Observations have different advantages and disadvantages. Although observation is a natural way to study classroom processes, it has generally been considered too difficult and labor intensive for large-scale studies. The methods described here, however, present an approach to overcoming this problem.

Advantages and Disadvantages of Questionnaires for Studying Classroom Processes

Most attempts to measure classroom processes on a large scale have used teacher questionnaires. Teachers have been asked, for example, to report on the percentage of time they spend in lecture or discussion, the degree to which problem solving is a focus in their mathematics classrooms, and so on. Questionnaires have numerous advantages: They are simple to administer to large numbers of respondents and usually can be easily transformed into data files that are ready for statistical analysis.

On the other hand, there are at least three major limitations in using questionnaires to study classroom instruction. First, the words researchers use to describe the complexities of classroom instruction may not be understood in the same way by teachers or in a consistent way across different teachers. The phrase "problem solving" is a good example. Many reformers of mathematics education call for problem solving to become the focus of the lesson. But different teachers interpret this phrase in different ways. One teacher may believe that working on word problems is synonymous with problem solving, even if the problems are so simple that students can solve one in 15 seconds. Another teacher may believe that a problem that can be solved in less than a full class period is not a real problem but only an exercise. Such inconsistency in the use of terms is common in the United States, where teachers have few opportunities to observe or be observed by other teachers in the classroom. It may be that because teacher training in the United States generally does not engage teachers in discussions of classroom instruction, and because teachers are often isolated from one another by the conditions under which they work, teachers do not develop shared referents for the words used to describe instruction. Thus, when teachers fill in questionnaires about their teaching practices, interpreting their responses is problematic.

A second problem with relying on questionnaire-based indicators of instruction concerns their accuracy in reporting processes that may, at least in part, be outside of their awareness. Teachers may be accurate reporters of what they planned for a lesson (e.g., what kind of demonstration they used to introduce the lesson) but inaccurate when asked to report on the aspects of teaching that can happen too quickly to be under the teacher's conscious control.

A third limitation of questionnaires is their static nature. Teachers can only answer the questions we as researchers thought to ask. An observer might notice something important just by being in the classroom. This problem is more serious in international research, where unfamiliarity with other nations' instructional approaches makes effective questionnaire design difficult.

Advantages and Disadvantages of Live Observations for Studying Classroom Processes

Having discussed some of the advantages and disadvantages involved in using questionnaires to study classroom processes, let us now discuss the advantages and disadvantages of using direct observational techniques. Direct observation overcomes some of the limitations identified for questionnaires: Observations allow behavioral categories to be defined objectively by the researcher, not independently by each respondent. They also enable researchers to study on-line implementation of instruction as well as the planned, structural aspects. Teachers themselves may be unaware of their behavior in the classroom, yet this same behavior could be easily accessible to the outside observer.

On the other hand, there are clear disadvantages of live observation as well. Just like questionnaires, observational coding schemes can act as blinders and may make it difficult to discover unanticipated aspects of instruction. The use of live observations also introduces significant training problems when used across large samples or, especially, across cultures. A great deal of effort is required to assure that different observers are recording behavior in comparable ways. In fact, when working in different cultures, it may be impossible to achieve high levels of comparability.

THE USE OF VIDEO FOR STUDYING CLASSROOM INSTRUCTION

Bearing in mind the limitations of questionnaires and of live observational coding schemes, especially in the context of cross-cultural research, it was decided to use video for the present study. Most researchers, on hearing the word "video," imagine a small-scale qualitative study. This study is anything but small: Large quantities of video were collected on national samples of teachers. In fact, one goal of this study was to explore video's feasibility for use in producing quantitative indicators based on large samples and on the combination of these quantitative indicators with qualitative information. In this section we will discuss the advantages and disadvantages of video over live observation in the study of classroom processes.

Enables Study of Complex Processes

Classrooms are complex environments, and instruction is a complex process. Live observers are necessarily limited in what they can observe, and this places constraints on the kinds of assessments they can do. Video provides a way to overcome this problem: Observers can code video in multiple passes, coding different dimensions of classroom process on each pass. On the first pass, for example, we coded the organization of the lesson; on the second, the use of instructional materials; and on the third, the patterns of discourse that characterize the classrooms of each country. It would have been impossible for a live observer to code all of these simultaneously.

Not only can coding be done in passes but it also can be done in slow motion. With video, for example, it is possible to watch the same sample of behavior multiple times, enabling coders to describe the behavior in great detail. This makes it possible to conduct far more sophisticated analyses than would be possible with live observers.

Increases Inter-Rater Reliability, Decreases Training Problems

Video also resolves problems of inter-rater reliability that are difficult to resolve in the context of live observations. The standard way to establish the reliability of observational measures is to send two observers to observe the same behavior, then compare the results of their coding. This is often inconvenient and is even infeasible for studies that are performed cross-culturally or in geographically distant locations. Using video to establish reliability means that the behavior can be brought to the observers instead of vice versa. Thus, in the context of a cross-cultural study, observers from different cultural and linguistic backgrounds can work collaboratively, in a controlled laboratory setting, to develop codes and establish their reliability using a common set of video data.

Using video also makes it far easier to train observers. With video, inter-rater reliability can be assessed not only between pairs of observers but between all observers and an expert "standard" observer. Disagreements can be resolved based on re-viewing the video, making such disagreements into a valuable training opportunity. And, the same segments of video can be used for training all observers, increasing the chances that coders will use categories in comparable ways.

Amenable to Post-Hoc Coding, Secondary Analysis

Most survey data sets lose their interest over time. Researchers decide what questions to ask and how to categorize responses based on theories that are prevalent at a given time. Video data, because they are "pre-quantitative," can be re-coded and analyzed as theories change over time, giving them a longer shelf life than other kinds of data. Researchers in the future may code videotapes of today for purposes completely different than those for which the tapes were originally collected.

Amenable to Coding from Multiple Perspectives

For similar reasons, video data are especially suited for coding from multiple disciplinary perspectives. Tapes of mathematics classes in different countries, for example, might be independently coded by psychologists, anthropologists, mathematicians, and educators. Not only is this cost effective, but it also facilitates valuable communication across disciplines. The most fruitful interdisciplinary discussions result when researchers from diverse backgrounds compare analyses based on a common, concrete referent.

Facilitates Integration of Qualitative and Quantitative Information

Video makes it possible to merge qualitative and quantitative analyses in a way not possible with other kinds of data. With live-observer coding schemes the qualitative and quantitative analyses are done sequentially: First, initial qualitative analyses lead to the construction of the coding scheme; then, implementation of the coding scheme leads to a re-evaluation of the qualitative analysis.

When video is available it is possible to move much more quickly between the two modes of analysis. Once a code is applied and a quantitative indicator produced, the researcher can go back and look again more closely at the video segments that have been categorized together. This kind of focused qualitative observation makes it possible to refine and improve the code, and may even provide the basis for a new code.

Provides Referents for Teachers' Descriptions

Mentioned earlier was the problem that teachers lack a set of shared referents for the words they use to describe classroom instruction. Video can, in the long run, provide teachers, as potential consumers of the research, with a set of such referents. Definitions of instructional quality and the indicators developed to assess instructional quality could be linked to a library of video examples that teachers can use in the course of their professional development. In the long run, a shared set of referents can lead to the development of more efficient and valid questionnaire-based indicators of instructional quality.

Facilitates Communication of the Results of Research

It is also possible, with video, to use concrete video examples in reporting research results. This gives consumers of the information a richer qualitative sense of what each category in the coding system means and a concrete basis for interpreting the quantitative research findings.

Provides a Source of New Ideas for How to Teach

Another advantage of video over other kinds of data is that it becomes a source of new ideas on how to teach. Because these new ideas are concrete and grounded in practice, they have immediate practical potential for teachers. Questionnaires and coding schemes can help us spot trends and relationships, but they can't demonstrate a new way of teaching the Pythagorean theorem.

Disadvantages

Despite all its advantages, video also has some disadvantages. At the very least, video raises a number of problematic issues that must be addressed if it is to yield accurate and valid information about classroom processes. In the next section we will discuss some of these issues and challenges.

ISSUES IN VIDEO RESEARCH

This section briefly discusses a number of issues that must be resolved in order to conduct meaningful video research.

Standardization of Camera Procedures

Left to their own devices, different videographers will photograph the same classroom lesson in different ways. One may focus on individual students, another may shoot wide shots in order to give the broadest possible picture of what is happening in the classroom. Yet another might focus on the teacher or on the blackboard. Because we want to study classroom instruction, not the videographers' camera habits, it is important to develop standardized procedures for using the camera and then to carefully train videographers to follow these procedures. This study has done so, and the procedures are described in the Methods section of this document.

The Problem of Observer Effects

What effect does the camera have on what happens in the classroom? Will students and teachers behave as usual with the camera present, or will we get a view that is biased in some way? Might a teacher, knowing that she is to be videotaped, even prepare a special lesson just for the occasion that is unrepresentative of her normal practices?

This problem is not unique to video studies. Questionnaires have the same potential for bias: Teachers' questionnaire responses, as well as their behavior, may be biased toward cultural norms. On the other hand, it may actually be easier to gauge the degree of bias in video studies than in questionnaire studies. Teachers who try to alter their behavior for the videotaping will likely show some evidence that this is the case. Students, for example, may look puzzled or may not be able to follow routines that are clearly new for them.

It also should be noted that changing the way a teacher teaches is notoriously difficult to do, as much of the literature on teacher development suggests. It is highly unlikely that teaching could be improved significantly simply by placing a camera in the room. On the other hand, teachers will obviously try to do an especially good job, and may do some extra preparation, for a lesson that is to be videotaped. We may, therefore, see a somewhat idealized version of what the teacher normally does in the classroom.

Minimizing Bias Due to Observer Effects

This study used three techniques for minimizing observer bias. First, instructions were standardized across teachers. The goal of the research was clearly communicated to the teacher in carefully written, standard instructions. Teachers were told that the goal was to videotape a typical lesson with typical defined as whatever they would have been doing had the videographer not shown up. Teachers were also explicitly asked to prepare for the target lesson just as they would for a typical lesson. (A copy of information given to teachers prior to the study is included as appendix A.)

Second, this study attempted to assess the degree to which bias occurred. After the videotaping, teachers were asked to fill out a questionnaire in which they rated, for example, the typicality of what we would see on the videotape, and describe in writing any aspect of the lesson they felt was not typical. We also asked teachers whether the lesson in the videotape was a stand-alone lesson or part of a sequence of lessons and to describe what they did yesterday and what they plan to do in tomorrow's lesson. Lessons described as stand-alone and as having little relation to the lessons on adjoining days would be suspect for being special lessons constructed for the purpose of the videotaping. In this study, however, lessons were rarely described in this way.

Finally, one must use common sense in deciding the kinds of indicators that may be susceptible to bias and taking this into account in interpreting the results of a study. It seems likely, for example, that students will try to be on their best behavior with a videographer present, and so we may not get a valid measure from video of the frequency with which teachers must discipline students. On the other hand, it is probably less likely that teachers use a different style of questioning while being videotaped than they would when the camera is not present. Some behaviors, such as the routines of classroom discourse, are so highly socialized as to be automatic and thus difficult to change.

Sampling and Validity

Observer effects are not the only threat to validity of video survey data. Sampling—of schools, teachers, class periods, lesson topics, and parts of the school year—is a major concern.

One key issue is the number of times any given teacher in the sample should be videotaped. This obviously will depend on the level of analysis to be used. If we need a valid and reliable picture of individual teachers, then we must tape the teacher multiple times, as teachers vary from day to day in the kind of lesson they teach, as well as in the success with which they implement the lesson. If we want a school-level picture, or a national-level picture, then we obviously can tape each teacher fewer times, provided we resist the temptation to view the resulting data as indicating anything reliable about the individual teacher.

On the other hand, taping each teacher once limits the kinds of generalizations we can make 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. If each teacher is taped once, it is not possible to study the dynamics of teaching over the course of a unit. Inferences about these dynamics cannot necessarily be made, even at the aggregate level, based on one-time observations.

Another sampling issue concerns representativeness of the sample across the school year. This is especially important in cross-national surveys where centralized curricula can lead to high correlations of particular topics with particular months of the year. In Japan, for example, the eighth-grade mathematics curriculum devotes the first half of the school year to algebra, the second half to geometry. Clearly, the curriculum would not be fairly represented by taping in only one of these two parts of the year.

Finally, although at first blush it may seem desirable to sample particular topics in the curriculum in order to make comparisons more valid, in practice this is virtually impossible. Especially across cultures, teachers may define topics so differently that the resulting samples become less rather than more comparable. Randomization appears to be the most practical approach to insuring the comparability of samples.

Confidentiality

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 when the data set is used for secondary analyses. An important issue, therefore, concerns how procedures can be established to allow continued access to video data by researchers interested in secondary analysis.

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 this for an entire data set. A more practical approach is to define special access procedures that will enable us to protect the confidentiality of participants while still making the videos available as part of a restricted-use data set.

Logistics

Contrary to traditional surveys, which require intensive and thorough preparation up front, the most daunting part of video surveys is in the data management and analysis phase. Information entered on questionnaires is more easily transformed into computer readable format than is the case for video images. Thus, it is necessary to find a means to index the contents of the hundreds of hours of tape that can be collected in a video survey. Otherwise, the labor involved in analyzing the tapes grows enormously.

Once data are indexed, there is still the problem of coding. Coding of videotapes is renowned as highly labor intensive. But there are strategies available for bringing the task under control. The present study has developed specialized computer software to help in this task. Emerging multimedia computing technologies will, over the next several years, revolutionize the conduct of video surveys, making them far more feasible than they have ever been in the past.

HARNESSING THE POWER OF THE ANECDOTE

Anecdotes and images are vivid and powerful tools for representing and communicating information. One picture, it is said, is worth a thousand words. On the other hand, anecdotes can be misleading and even completely unrepresentative of reality. Furthermore, research in cognitive psychology has shown that the human information processing system is easily misled by anecdotes, even in the face of contradictory and far more valid information (e.g., Nisbett and Ross, 1980). Methods of research design and inferential statistics were developed, in fact, specifically to protect us from being misled by anecdotes and experiences (Fisher, 1951).

A video survey, like the one being described here, provides one possible way to resolve this tension between anecdotes and statistics. Recognizing the power of video images, one can harness this power in two ways. First, discoveries made through qualitative analysis of the videos can be validated by statistical analysis of the whole set of videos. For example, while watching a video we might notice some interesting technique used by a Japanese teacher. If we only had one video, it would be hard to know what to make of this observation: Do Japanese teachers really use the technique more than U.S. teachers, or did we just happen to notice one powerful example in the Japanese data? Because we have a large sample of videos, we can turn our observation into a hypothesis that can be validated against the database.

In a complementary process, we might, after coding and quantitative analysis of the video data, discover a statistical relationship in the data. By returning to the actual video, we can find concrete images to attach to our discovery, giving us a means of further analysis and exploration, as well as a set of powerful images that can be used to communicate the statistical discovery we have made. Through this process we can uncover what the statistic means in practice.

Chapter 2. Methods

SAMPLING

Our goal was to collect national probability samples of eighth-grade mathematics students in Germany, Japan, and the United States. The final sample consisted of 100 lessons in Germany, 81 in the United States, and 50 in Japan. In addition, five "public use" tapes were collected in each country to serve as examples to help us communicate the results of the study. And, a subsample of 30 lessons in each country was selected from the final sample for in-depth analysis by a group of mathematicians and mathematics educators. We review each of these samples in more detail here.

All analyses reported here were done on the full sample of 231 lessons except the following, which used the subsample of 30 lessons in each country selected for analysis by the Math Content Group: (1) Analyses of the Math Content Group; (2) Some analyses of the use of the chalkboard; (3) Analyses of second-pass coding of discourse; and (4) Analyses of explicit linking within and across lessons. We have explicitly noted in the text whenever anything less than the full sample is used in an analysis.

The Main Video Sample

The main video samples were designed to be random subsamples of the TIMSS main study sample, which was selected according to the TIMSS sampling plan in each country. Our plan was to videotape 100 eighth-grade classrooms in Germany and the United States, and 50 in Japan. In the end, these targets were attained in Germany and Japan but not in the United States, where only 81 classrooms agreed to participate. The sample size in Japan was reduced to 50 primarily 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 reduction was further justified by the fact that certain characteristics of Japanese education (e.g., lack of tracking within or across schools, adherence to a national curriculum, and culturally more homogeneous population) led us to expect lower variability between classrooms in Japan.

The main TIMSS study focused on three separate age groups; the video sample was drawn from only one of these age groups, referred to as Population 2. Population 2 was defined as the pair of adjacent grades in each country which contained the largest percentage of 13-year-olds. In all three countries included in the video study, Population 2 was defined as grades seven and eight. NCES specifications for the study required that only eighth-grade classrooms be sampled for videotaping. According to the TIMSS international specifications, sampling in each country was accomplished by selecting schools, then classrooms within schools. Each country was required to sample a minimum of 150 schools and a minimum of one seventh- and one eighth-grade classroom within each school.

The selection of schools for the main TIMSS study followed a somewhat different procedure in each country. In the United States, schools were sampled from within primary sampling units (PSUs), geographically-defined units designed to reduce the costs of data collection. PSUs were stratified according to geographic region, metropolitan versus nonmetropolitan area, and various secondary strata defined by socioeconomic and demographic characteristics, then sampled with the probability of selection proportionate to the population of each PSU. Within each sampled PSU, schools were sampled with the probability of selection proportionate to the estimated number of students in the target grades. In Japan, schools were randomly selected from strata defined by size of community and size of school, with the probability of selection proportionate to the size of the population within each stratum. Germany followed a similar procedure but defined its strata by state and by type of school.

Further details regarding selection of the main TIMSS samples in each country can be obtained elsewhere (Foy, Rust, & Schleicher, 1996). Here, we describe how the subsamples were selected for the video study. Because specific details of sample selection and recruitment varied across the three countries, we describe each country's sample separately. (A discussion of weighted and unweighted response rates for each country can be found in appendix B.)

The U.S. Sample

The U.S. TIMSS sample for Population 2 consisted of a stratified random sample of 220 schools. Within each school, one seventh- and two eighth-grade classrooms were studied. One-half of these schools were randomly sampled to be part of the video study. Within each school, one eighth-grade classroom was randomly sampled to be videotaped.

Schools were selected for the video study as follows: First, Population 2 TIMSS schools were listed in the order in which they were originally sampled. Using this ordering, pairs of schools were generated. Within each pair one of the two schools was randomly sampled (with each school having an equal probability of being sampled). The unsampled school in the pair was reserved as a potential replacement for the sampled school. A total of 109 pairs were assigned, with one school unpaired, because one school of the original Population 2 sample of 220 schools had no eighth grade. The unpaired school was given a half chance of being selected. The final videotape sample size was 109. The unpaired school was not sampled.

Within each sampled school, one eighth-grade classroom was selected with equal probability from the two TIMSS eighth-grade classrooms in the school. There was no sorting or stratification of classrooms by level of mathematics taught. In the event that the sampled teacher refused to be videotaped, the classroom was never replaced by the other eighth-grade classroom in the same school. Instead, the entire school was replaced by its paired school.

Of the original 109 schools sampled, 100 were public and 9 were private. Forty schools, including one private school, refused to participate. The paired schools for 13 of these refusals were contacted, and 12 agreed to participate in the video study. Thus, the final video sample in the United States consisted of 73 public and 8 private schools. The high refusal rate among originally sampled U.S. classrooms should be kept in mind as a potential source of sampling bias.

Each teacher who participated in the study was awarded a $300 grant, its use to be determined "jointly by the teacher and the principal."

The German Sample

The German TIMSS sample for Population 2 consisted of a stratified random sample of 153 schools (of which 150 were eligible for participation) drawn from all states except Baden-Wuerthemberg. Sampling strata were defined by state, school type, distribution frequencies of each school type in each state, and classroom size. The random sampling of the schools was carried out by the Statistical Institutes of the German States. The four main participating school types were Gymnasium, Realschule, Hauptschule, and Integrierte Gesamtschule. Gymnasium is the highest academic track of schools. Gymnasium runs from 5th grade through 13th grade. Graduates of the Gymnasium are eligible to attend university. Realschule is the middle-level track. Realschule extends through 10th grade. Hauptschule is the lowest track, running through 9th grade. Graduates of Hauptschule are eligible to enter vocational schools. Integrierte Gesamtschule are relatively uncommon. In these schools, the three tracks are integrated into a single building, though the curricula and classes are still separate. A few schools, in the former East Germany, were not of these main types: Regelschule are combinations of Realschule and Hauptshule in a single building; Realschulklasse and Hauptschulklasse are special classes within schools that have modified curricula.

The schools for the video study were selected as follows: First, 100 schools were randomly sampled from the list of 153 schools originally sampled for the TIMSS study. Of these 100 schools, 15 refused to be videotaped. As schools declined, one of the main TIMSS replacement schools for the refusing school was selected to participate in its place. The breakdown of the final sample according to type of school is shown in figure 1. Within each school, the eighth-grade classroom that participated in the TIMSS assessments was selected for videotaping.

As in the United States, German teachers were paid a modest stipend for their participation.

Figure 1
German sample for the Videotape Classroom Study broken down by type of school

School Type	Mean
Gymnasium	34
Realschule	24
Hauptschule	23
Gesamtschule	9
Regelschule	3
Realschulklasse	4
Hauptschulklasse	3
Total	**100**

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

The Japanese Sample

The Japanese TIMSS sample for Population 2 consisted of 158 schools. One-hundred and fifty of the schools were public and eight were private.

Public schools were selected by stratified random sampling. First, two factors—size of the school (small, medium, large) and size of the city (small, medium, large)—were used to classify Population 2 schools. Small schools were defined as having 8 to 40 students enrolled in eighth grade, medium schools 41 to 160 students, and large schools, over 160 students. Small cities were defined as having a population of fewer than 50,000, medium cities between 50,000 and one million, and large cities one million or more. Because no school fell into the large school/small city stratum, sampling was based on eight strata. Schools were randomly selected from each stratum in proportion to the total number of schools in the stratum. Private schools were randomly selected from among the population of private schools in Japan.

One third of the schools in the TIMSS sample were then randomly selected within each stratum for the video study, yielding a sample size of 50. Of these 50 schools, two declined to participate. Each of these was replaced by randomly selecting another school within the same stratum.

One eighth-grade classroom was selected from each school. In the event the mathematics teacher assigned to this classroom declined to participate in the video study, the particular class/teacher was never replaced by another eighth-grade classroom in the same school. Instead, another school within the same stratum replaced the entire school.

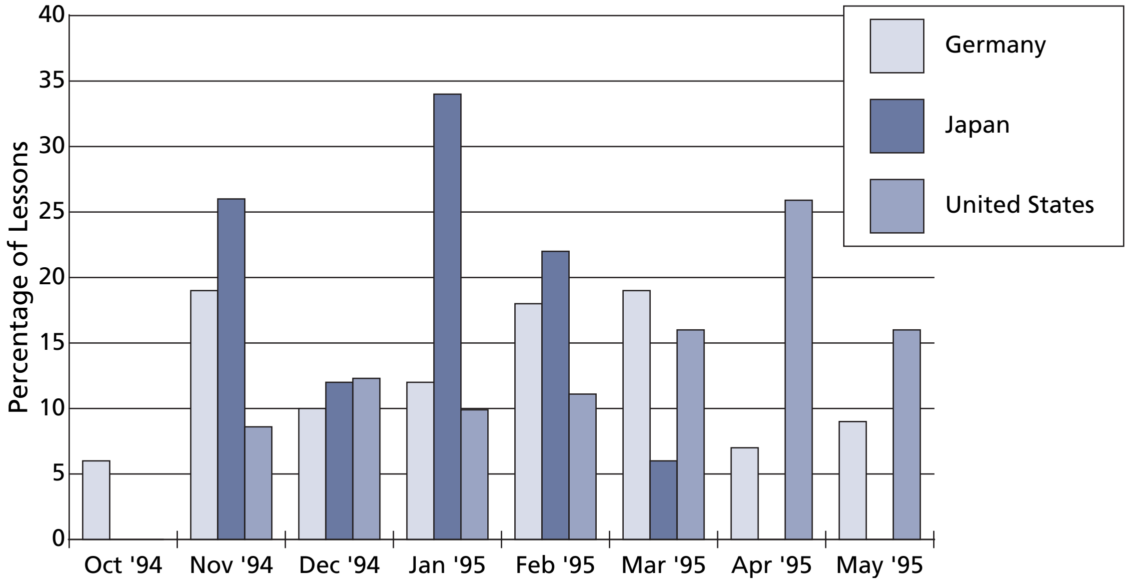
Of the schools that participated in the videotape study, all but one school participated in the main TIMSS study as well. However, during the selection of classrooms to be videotaped, a deviation from the original plan to test TIMSS classrooms arose. Because NIER did not want to overburden the teachers, videotaping was usually done in a different class from the one in which testing for the main study was conducted, unless there was only one eighth-grade classroom in the school. When there was a choice, the principal of each school chose the classroom in which the videotape study occurred. Although it is unlikely that there are significant student achievement differences between the main TIMSS classroom and the classroom chosen for the videotape study, it is possible that there are differences in teacher characteristics. It should be kept in mind that Japanese principals exercised discretion in the choice of classrooms to be videotaped.

Participating schools and teachers were offered a small token of appreciation for their participation by the U.S. government. Each teacher also received a videotape of his or her teaching.

Sampling Time in the School Year

Our goal was to spread the videotaping out evenly over the school year. In Germany and the United States we accomplished this goal by employing a single videographer in each country to tape over an 8-month period, from October 1994 through May 1995. Unfortunately, we were not able to implement the same plan in Japan. Because the school year begins in April in Japan, following a schedule analogous to the other two countries would have meant starting in June and taping through December. Unfortunately, this schedule was not possible due to the need to coordinate the videotaping with the test administration. The result was that videotaping in Japan had to be compressed primarily into a 4-month period, from November 1994 through February 1995, with a few lessons taped in March. The distribution of videotaping over time in each country is presented in figure 2.

Figure 2
Distribution of videotaping over time in each country



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Unfortunately, the consequence of taping only during the second half of the Japanese school year is made more problematic by the near-universal adherence in Japan to a national curriculum. In most eighth-grade classrooms in Japan, the first half of the school year is devoted to algebra, the second half to geometry. Thus, our sample in Japan is skewed toward geometry. Although this is a limitation of the study, we did try to diminish the problem by sampling five additional algebra lessons during the next Japanese school year. These lessons were included in the subsample for the math content group, described here, though they were not included in the main analyses.

Subsample for the Math Content Group

A subsample of 30 videotaped lessons was selected from each country for in-depth content analysis by a group of mathematicians and mathematics educators. (This group is described in more detail in appendix C.) This total of 90 tapes was selected as follows: first, lessons in the video study were categorized as being primarily geometry or primarily algebra (broadly defined to include advanced topics in arithmetic). Then, 15 algebra and 15 geometry tapes were chosen randomly from each country to constitute the subsample.

The 15 algebra tapes from Japan included 5 that were sampled later in an attempt to remedy the overrepresentation of geometry in the sample (see section above on "Sampling Time in the School Year"). These tapes were obtained by proportionally randomly choosing 5 schools of the 50 that participated in the video study, and then choosing a different teacher (i.e., one not already videotaped teaching geometry) to be videotaped teaching algebra.

Additional Tapes for Public Use

Because participants in the video study were guaranteed confidentiality, videotapes collected in the study cannot be shown publicly. However, because we believe that video examples will be extremely useful for communicating the results of the study, we decided to collect five tapes in each country that could be used for this purpose. For these tapes, we obtained written releases from the teachers and from the parents of students appearing in the tapes.

It is not easy to find teachers who will agree to being videotaped for public viewing; one cannot simply select such teachers at random. In the United States, we relied on networks of friends and contacts to identify teachers for public taping, hoping that the teachers who agreed to participate would be representative of those included in the large study sample. German and U.S. public use tapes were not included in the analyses presented. In Japan, in accordance with the preference of our collaborators, public use tapes were selected from among the main study tapes, and permission to use the tapes publicly was secured after the fact.

OVERVIEW OF PROCEDURES

We primarily collected two kinds of data in the video study: videotapes and questionnaires. We also collected supplementary materials (e.g., copies of textbook pages or worksheets) deemed helpful for understanding the lesson. Each classroom was videotaped once on a date convenient for the teacher. One complete lesson—as defined by the teacher—was videotaped in each classroom.

Teachers were initially contacted by a project coordinator in each country who explained the goals of the study and scheduled the date and time for videotaping. Because teachers knew when the taping was to take place, we knew they would attempt to prepare in some way for the event. In order to cut down somewhat on the variability in preparation methods across teachers we gave teachers in each country a common set of instructions. Teachers were told the following:

> Our goal is to see what typically happens in [U.S. or German or Japanese] mathematics classrooms, so we really want to see exactly what you would have done had we not been videotaping. Although you will be contacted ahead of time, and you will know the exact date and time that your classroom will be videotaped, we ask that you not make any special preparations for this class. So please, do not make special materials, or plan special lessons, that would not typify what normally occurs in your classroom. Also, please do not prepare your students in any special way for this class. Do not, for example, practice the lesson ahead of time with your students.

On the appointed day the videographer arrived at the school and videotaped the lesson. After the taping each teacher was given a questionnaire and an envelope in which to return it. The purpose of the questionnaire was to assess how typical the lesson was according to the teacher and to gather contextual information important for understanding the contents of the videotape. Both taping procedures and questionnaire contents are described in more detail.

Field Test

All procedures were tested in a field test, which was conducted in spring 1994. For the field test we collected nine videotapes from each country, together with all of the supplementary data. In addition to testing procedures of data collection, field test tapes were used to help in development of coding and analysis methods, as described more fully here. A full report of the field test may be found in Stigler & Fernandez (1995).

VIDEOTAPING IN CLASSROOMS

The success of any video survey will hinge on the quality, informativeness, and comparability of the tapes collected. What we see on a videotape results not only from what transpires in the classroom but also from the way the camera is used. If our aim is to compare certain aspects of instruction, then we must make sure that these aspects are clearly captured on all the tapes. In addition, we want to make sure that we are comparing classroom instruction and not camera habits. There are many decisions that must be made by the camera operator; if these are not made in a standardized manner, then the resulting tapes will not be comparable across classrooms or countries.

We developed procedures for camera use in collaboration with Scott Rankin, an experienced videographer who had worked with us in previous projects and who, therefore, was familiar with the challenges of documenting classroom instruction. Our goal was to develop a set of general principles and rules of thumb that would be relatively simple for our videographers to learn, yet comprehensive enough to apply in any classroom situation.

We should note at the outset that we decided to use one camera per taping instead of two, which made it impossible to see all of the students in a class. This constraint was based on budget considerations, although it also simplified the process of coding and analysis. Consequently this study did not collect detailed information on student behavior.

In the following sections we describe the procedures used for videotaping classroom instruction; our method for training videographers to use the procedures consistently; and an evaluation of the success of our training by comparing camera use across our three videographers.

Basic Principles for Documenting Classroom Lessons

Because we wanted to see each lesson in its entirety, all videotaping was done in real time: The camera was turned on at the beginning of the class, and not turned off until the lesson was over. This means that we can study the duration of classroom activities by measuring their length on the videotape. Obviously, this would not be possible if there were any gaps in the recording.

Classrooms are complex environments where much is going on at any given time; it is impossible to document everything, particularly when only one camera is used. We decided on two principles to guide videographers in their choices of where to point the camera. These principles yield a comprehensive view of the lesson being taped.

Principle #1: Document the perspective of an ideal student. Assume the perspective of an ideal student in the class, then point the camera toward that which should be the focus of the ideal student at any given time. An ideal student is one who is always attentive to the lesson at hand and always occupied with the learning tasks assigned by the teacher. An ideal student will attend to individual work when assigned to work alone, will attend to the teacher when he or she addresses the class, and will attend to peers when they ask questions or present their work or ideas to the whole class. In other words we chose to point the camera so as to capture the experience of a student who is paying attention to the lesson as it unfolds. In cases where different students in the same class are engaged in different activities, the ideal student is assumed to be doing whatever the majority of students are doing.

Principle #2: Document the teacher. Regardless of what the ideal student is doing, be certain to capture everything that the teacher is doing to instruct the class. Usually the two principles are in agreement: Whenever the ideal student is attending to the teacher, both principles would have the camera pointed at the teacher. However there are times when the two principles are in conflict. Take, for example, a case where the majority of students are doing seatwork while the teacher is working privately with two students at the board. The ideal student would be focused on his or her work, not on the teacher. In situations like this one the videographer must go beyond these two basic principles in order to determine where to point the camera.

The Exceptions: Three Difficult Situations

We have identified three common situations where the principles alone cannot guide choices about what to capture on the videotape. These situations are: (1) when the ideal student would be focused on something other than the teacher, (2) when two speakers who are having a conversation will not fit in a single shot, and (3) when a speaker and an object being discussed will not fit in a single shot. We developed a set of guidelines so that videographers will choose similar (i.e., comparable) shots when faced with each of these situations and so that these shots will contain a maximum amount of useful information. In the rest of this section we present a more detailed discussion of these three situations and how we chose to film them.

Situation #1: When the ideal student is not watching the teacher. As already mentioned, there are times when the ideal student should be attending to something other than the teacher. This most often occurs when students are given a task to work on individually or in small groups. Teachers can use this time in different ways. Sometimes they will walk around the class and monitor students' work. This is ideal from the videographers' point of view because by following the teacher with the camera one can also get a sense of what students are doing. In some instances, however, a problem arises because the teacher does not circulate through the class but rather stays at the board or her desk. In such cases the camera would need to be pointed in two different directions (toward the teacher and toward the students) in order to capture both the teacher and the focus of the ideal student.

Videographers were instructed to handle such situations by alternating between these two points of view. They were told to slowly do a sweep of the classroom by panning away from the teacher and then panning back to the teacher so as to document what the students are doing. After this sweep they were told to keep the camera directed at the teacher unless the nature of the students' activity changes in any significant way (e.g. new materials are introduced or they break into groups). If the students' activity were to change, videographers were instructed to carry out another sweep of the students, then return to the teacher.

Situation #2: When two speakers will not fit in a single shot. A second difficult situation occurs when the teacher is conversing with a student (or a student is conversing with another student) and the two speakers are far enough apart that they do not fit in a single camera shot. This often occurs when a teacher calls on a student seated in the back of the room, then proceeds to talk back and forth to the student.

In this case videographers were instructed to move the shot from speaker to speaker as they take turns talking. An exception to this rule occurs when one of the speaker's turns of speech are so brief that there is no time to shift the camera before the turn is over. In this case the camera should be kept on the person doing the most talking.

Situation #3: When the speaker and the object being discussed will not fit in a single shot. Another difficult situation occurs when a speaker and an object he or she is discussing will not both fit into a single camera shot. This happens frequently, for example, when someone is talking about things written on the chalkboard or about concrete representations of a mathematical situation or concept.

In this kind of situation videographers were told to document the object for long enough to provide the visual information needed to make sense of the talk, then to keep the shot on the speaker. For example, if the teacher is talking about a problem on the blackboard, the videographer should first tape the problem, then move to the teacher.

There is one important exception to this rule. Sometimes it is not sufficient to briefly see the object and then move to the speaker because the talk will make no sense unless one is seeing the object as it is being talked about. For example, if the speaker is pointing to specific features of the object as he or she talks, and if the pointing must be seen in order to understand the talk, then the rule is that the camera must stay on the object so that the talk can be understood.

How Close to Frame the Shot

Aside from making sure that videographers point their cameras at comparable things, we also wanted to make sure that their shots are framed in comparable ways. An extreme close up of the teacher talking would provide a very different sense of the action taking place than a wide shot where the teacher is seen in the context of the classroom.

We decided that in general we wanted the widest shot possible, a shot professional videographers call the "Master of Scene" (MOS) or, more simply, the "master shot." From an aesthetic point of view closer shots often look better. However, the MOS provides more contextual information and thus was judged more appropriate for our purposes. The master shot also is less prone to bias because it does not artificially focus the viewer in on whatever aspect of the lesson the videographer judged to be most interesting.

Sometimes, however, there is crucial information that cannot be captured in a master shot. Common examples include objects being discussed during the lesson or things written on the blackboard. In such instances the camera should zoom in close enough to capture this information. In other words, although our preferred view of the classroom is the MOS, a closer shot must be used when it is needed to understand what is going on. Videographers were told to hold close shots long enough to enable a viewer to read or form a mental image of the information.

Moving from Shot to Shot

Finally, having devised guidelines for what to include in the shot, we also needed some rules for how to move from shot to shot. This, too, must be done in a standardized way if the tapes are to be fully comparable.

The guidelines we gave to the videographers were based on principles of good camera work. We taught them how to compose shots and execute camera movements in ways that follow basic cinematographic conventions and fundamentals of good composition. Aside from wanting them to follow the same conventions, we wanted them to carry out good camera work. Bad camera work calls attention to itself and distracts the viewer from the contents of the tape.

Training Videographers

In order to make sure that the rules were applied correctly and reliably we had to work intensively with the videographers. Each videographer participated in two training sessions, both of which were conducted by our professional videographer, Mr. Rankin. The first training session lasted 9 days in the spring of 1994, after which each videographer was sent out to collect 10 practice tapes for a field test. The second training session lasted 5 days and was held in the early fall of 1994. Following this second training session videographers were given a test, then when they passed, sent off to collect the data.

We designed the training sessions with two goals in mind: First, we wanted to teach the videographers our camera use rules to the point that they could follow them by second nature. In an actual taping situation videographers would have to make rapid decisions about where to point the camera without time for reflection. Second, we wanted the videographers to learn and practice the fundamental skills of camera use. These skills include, for example, changing from one camera angle to another quickly without losing a focused image, tracking moving objects without having the object leave the shot, and moving rapidly back and forth from close-ups to master shots while ending up centered on the shot that needs to be captured.

The first training session was devoted to five activities: learning to use the equipment, practicing basic principles of good camera work, presentation and discussion of the standardized rules for taping classrooms, practice taping in mock classrooms, and practice taping in real classrooms. Activities in the second training session included reviewing and discussing the rules, critiquing practice tapes, and more practice taping in mock classrooms. A monitor hooked to the camera during the training sessions allowed videographers to rotate between practicing with the camera and watching/critiquing their peers in collaboration with the instructor.

We would like to pause here and insert a helpful hint for others contemplating this kind of work. One has two alternatives in deciding who to hire and train as a video survey videographer: One can hire scientists (i.e., educational researchers) and train them to take good pictures, or one can hire artists (i.e., photographers) and teach them the importance of following standardized rules for camera use. The latter proved far easier, and the pictures are much more aesthetically pleasing.

Evaluating the Comparability of Camera Use

At the end of the second training session we gave the videographers a test to measure and document how well they had internalized all they had been taught. A 7-minute mock lesson was created that covered many of the situations videographers needed to know how to handle. The lesson was taught three times, each one identical to the others, and was taped each time by one of the three videographers. The resulting tapes were analyzed and evaluated to make sure that our videographers would shoot lessons in a standardized manner.

To evaluate the videographers' performance on the test we first produced a description of how the test lesson should have been videotaped. We listed the 22 events that took place in the lesson, then determined how each event should be taped given the procedures we had developed.

Once we had a description of how the test-lesson should have been taped, we evaluated each videographer's performance against this ideal. We used a three-point scale to score how well they taped each of the 22 lesson events. They were given a score of zero if they broke any of the rules that they needed to take into account; for example, if they did not zoom in to capture information that they were supposed to capture, or if they pointed the camera at the wrong thing, they would be given a score of zero. They were given a score of one if they showed an understanding of the rule they needed to carry out but did not apply it in a timely fashion. For example, if they needed to zoom in and capture what the teacher was pointing to but reacted too slowly and missed this information, or if they let the teacher walk around the class for a while before they decided to follow her, they would receive a score of one. They were given a score of two if they applied the rules exactly as we had predicted they should.

The scores obtained were all in a similar range and also were relatively high. The German videographer received a score of 35 out of a possible total of 44. The Japanese videographer received a score of 36 and the U.S. videographer a score of 43. In addition, of the 66 events scored for the three videographers only four were rated a zero (which means that a rule was actually broken only four times). Two of these zeroes were obtained by the German and two by the Japanese videographer. This means that no videographer ever showed more than two rule breeches for the entire test.

The test lesson tapes were also used to evaluate the quality of each videographer's camera work. First we generated a list of possible flaws that a videographer might produce. Our list included the following flaws:

- Cropping shots too tightly (e.g., cutting off part of someone's head)
- Cropping shots too wide (e.g., too much head room)
- Zooming in/out and then having to reframe the shot
- Zooming in/out and then having to refocus the shot
- Panning while zoomed in tightly
- Jerky or awkward camera movement during zooms or pans
- Losing from the frame the object that is being tracked
- Unnecessary camera movement
- Bad coordination between zooms and pans
- Very unbalanced composition

We used this list to score each videographer's performance on a four-point scale for each of the 22 events in the test lesson. Videographers were given a score of three on an event if we could find no flaw in their camera work. They received a score of two if one flaw could be found, a score of one if two flaws could be found, and a score of zero if at least three flaws could be found.

All videographers obtained scores that were within a similar range and judged to be satisfactorily high. The Japanese videographer received a score of 51 out of a possible total of 66. The German videographer received a score of 52, and the U.S. videographer a score of 60.

Both evaluations of the test confirmed our informal impression that camera standardization had been reached by the end of the training.

Videographers were in the field for a prolonged period of time. We worried, therefore, that they might slowly forget what they were taught or develop bad habits. In order to make sure that they continued using the camera correctly, every 10th tape that came in from the field was evaluated using a scoring system similar to the one described. Videographers were given feedback about how they were doing. In particular, they were immediately informed if they had, in any way, drifted away from the standards we knew they were able to follow. In actuality, this almost never happened.

Some Notes on Equipment

The quality of the data depends to a great extent on the quality of the equipment used in collecting the data. Thus, we wanted high quality cameras that would produce excellent images, and high quality microphones that would enable us to hear most of what goes on in the classroom. At the same time, we needed equipment that could be operated by a single videographer.

The camera we selected was a Sony EVW-300 three-chip professional Hi-8 camcorder. Each camera was mounted on a Bogen fluid-head tripod. (Tripods that are not fluid head will produce jerky camera movements.) A small LCD monitor was mounted on the camera to help operators view what they were taping. Sound was collected using two microphones, one a radio microphone worn by the teacher, the second a shotgun zoom microphone mounted on the camera. This equipment was used both for data collection and for training videographers.

TEACHER QUESTIONNAIRE

A complete copy of the English version of the teacher questionnaire can be found in appendix D. The purpose of the teacher questionnaire was to elicit information that would aid us in analysis and interpretation of the videotapes. Items for the questionnaire were generated by project personnel in consultation with persons working on the main TIMSS questionnaire, questionnaire design specialists from Westat,[1] mathematics educators, and classroom teachers. Questions were edited and selected to yield a questionnaire that could be completed by teachers in approximately 20-30 minutes.

The questionnaire was translated and back-translated into German and Japanese and then pilot-tested on teachers participating in the field test. German, Japanese, and U.S. collaborators discussed the responses from the field test, and based on these discussions the questionnaire was revised.

The final translation of the questionnaire was painstakingly reviewed, question by question, by a group of German, Japanese, and U.S. researchers, each of whom was bilingual in two of the three languages. Questions that were judged too difficult to translate accurately were dropped from the questionnaire.

The resulting questionnaire consisted of three parts with a total of 28 questions. In Part A we asked questions about the lesson that we videotaped, and about how the class was constituted and who the students were. In Part B we asked the teachers to compare what happened in the videotaped lesson with what would typically transpire in their classroom. In Part C we asked teachers to describe what they know about current ideas on mathematics teaching and learning, and asked them to evaluate their own teaching in the videotape in light of these current ideas.

The information collected in the questionnaire served three purposes. First, it helped us to assess the quality and comparability of our samples across the three countries. Although teachers were instructed not to prepare in any special way for the videotaping, we were still aware that what we saw on the videotape might not be typical of what normally happens in a given classroom. Teachers thus were asked to rate the typicality of the videotaped lesson, and these ratings were compared across countries. Similarly, we were able to assess the comparability of the samples across the three countries along several other dimensions as well. For example, whether a lesson dealt with new material or review material might be expected to influence the kind of teaching technique used. Knowing the percentage of lessons in each country that were new versus review helped us to judge the comparability of the samples.

[1] Westat, of Rockville, Maryland, was the general contractor that conducted the U.S. TIMSS and the Videotape Classroom Study.

A second purpose for the questionnaire was to provide coders with information that would help them interpret what they saw on the videotapes. For example, it is often necessary to know the teacher's goal for a lesson in order to make sense of the activities that constitute the lesson, and so we asked the teacher to state his or her goal for the lesson. Similarly, to interpret the meaning a specific question has for students it is often helpful to know whether the question probes new material or reviews previously learned information. Again, teachers were asked to categorize the content of the lesson in this way on the questionnaire.

Third, the questionnaire responses did, in some cases, enter directly into the analyses—statistical and qualitative—of the videotapes. This occurred in several ways. First, questionnaire responses were entered into correlational analyses within each country to help us relate contextual factors to variations in classroom instruction. Second, by asking teachers to comment on the lesson that was videotaped we were able to learn more about how teachers interpret the language of reform in mathematics education. For example, if a teacher told us that her lesson was focused on problem solving, we could look at the video to see what she meant by the term "problem solving."

Response rates on the questionnaire were high: In Germany, 91 percent of the teachers whom we videotaped returned their questionnaires, in Japan, 94 percent, and in the United States, 97.5 percent.

CONSTRUCTING THE MULTIMEDIA DATABASE

Once the tapes were collected, they were sent to project headquarters at University of California, Los Angeles (UCLA) for transcription, coding, and analysis. The first step in this process was to digitize the video and store it in a multimedia database, together with scanned images of supplementary materials. The videotapes were then transcribed, and the transcript was linked by time codes to the video. This multimedia database was then accessed, coded, and analyzed using the multimedia database software system that we developed for this project.

Digitizing, Compression, and Storage on CD-ROM

To facilitate the processing of such large quantities of video data, we decided to digitize all of the video and supplementary materials, which allowed them to be stored, accessed, and analyzed by computer. Each videotape was digitized, compressed, and stored on CD-ROM disks, one lesson per disk. We then designed and built a multimedia database software application that would enable us to organize, transcribe, code, and analyze the digital video.

Digital video offers several advantages over videotape for use in video surveys. First, the resulting files are far more durable and long lasting than videotape. CD-ROM disks are assumed to last for at least 100 years, as opposed to a much shorter lifespan for videotape. Digital video files also can be copied without any loss in quality, which again is not true for videotapes. And, digital files will not wear out or degrade with repeated playing and re-playing of parts of the video. Digital video also enables random, instantaneous access to any location on the video, a feature that makes possible far more sophisticated analyses than are possible with videotape. For example, when coding a category of behavior it is possible to quickly review the actual video segments that have been marked for that category. This rapid retrieval and viewing of coded segments makes it much easier to notice inconsistencies in coding, or to discover new patterns of behavior, that would not be possible without such rapid access.

As videotapes arrived in Los Angeles they were digitized and compressed into MPEG-1 format on a large hard disk. Text pages, worksheets, and other supplementary materials collected by the videographers were digitized on a flatbed scanner and stored in PICT format on the same hard disk drive as the accompanying videotape. All files for each lesson were then burned onto a single CD-ROM disk.

Transcription/Translation of Lessons

Transcription of videotapes is essential for coding and analysis. Without a transcript, coders have difficulty hearing, much less interpreting, the complex flow of events that stream past in a classroom lesson. It also is possible to code some aspects of instruction directly from the transcript, without viewing the video at all. We thus transcribed, as accurately as possible, the words spoken by both the teacher and the students in each lesson and, for German and Japanese lessons, translated the transcriptions into English.

We had several reasons for translating the German and Japanese tapes into English. First, translations were used for training coders from different cultures to apply codes in a comparable way, and for establishing independent inter-rater reliability of codes across coders from different cultures. Even though a translation is never perfect, agreement between coders working with a translation can give us a rough estimate of how reliable a code is. A second purpose for the translation was to aid us in multilanguage searches of the database. If we want to locate, for example, times when a teacher discussed the concept of area we can search using the English word "area." Finally, having the lessons translated allows members of the research team not fluent in German or Japanese to view and understand lessons taught in those languages.

Procedures were developed to ensure that all transcriptions/translations were carried out in a standardized manner. For example, transcribers were given 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. Using the multimedia database software developed for this project, coders had instant access to the video as they worked with the linked transcripts, and so could easily retrieve the context needed for interpreting the transcript. It was therefore not necessary to transcribe the contextual information generally needed for understanding written transcripts. By the same token, translations of the German and Japanese lessons did not have to be perfect, as all coding was done by native speakers of the language being coded. The translations served as a guide, but not as the actual foundation for coding. Coders did not rely on translations to make subtle judgments about the contents of the video.

Videotapes were transcribed and translated by teams of transcribers fluent in the language they were transcribing. Some members of the German and Japanese teams were native speakers of those languages, while others were native speakers of English but fluent in German or Japanese. Each tape was transcribed/translated in two passes. One person worked on the first pass transcription/translation of a tape and then a different person was assigned to review the work. A hard copy of the first pass transcription/translation was printed out, and the reviewer marked any points of disagreement on this copy. The two individuals then met, discussed all the proposed revisions, and came to an agreement about what the final version should be. In the extremely rare cases in which disagreements could not be resolved, a third party was consulted.

The last step in the transcription/translation process was to time code the tapes (i.e., to mark the exact point at which each utterance begins).

DEVELOPING CODES

Deciding What to Code

In deciding what to code we had to keep two goals in mind: First, we wanted to code aspects of instruction that relate to our developing construct of instructional quality; Second, we wanted the codes we used to provide a valid picture of instruction in three different cultures.

For the first goal, we sought ideas of what to code from the research literature on the teaching and learning of mathematics, and from reform documents—such as the NCTM *Professional Standards for Teaching Mathematics*—that make clear recommendations about how mathematics ought to be taught. We wanted to code both the structural aspects of instruction (i.e., those things that the teacher most likely planned ahead of time), and the on-line aspects of instruction, (i.e., the processes that unfold as the lesson progresses).

The dimensions of instruction we judged most important included the following:

- *The nature of the work environment.* How many students are in the class? Do they work in groups or individually? How are the desks arranged? Do they have access to books and other materials? Is the class interrupted frequently? Do the lessons stay on course, or do they meander into irrelevant talk?
- *The nature of the work that students are engaged in.* How much time is devoted to skills, problem solving, and deepening of conceptual understanding? How advanced is the curriculum? How coherent is the content across the lesson? What is the level of mathematics in which students are engaged?
- *The methods teachers use for engaging students in work.* How do teachers structure lessons? How do teachers set up for seatwork, and how do they evaluate the products of seatwork? What is the teacher's role during seatwork? What kinds of discourse do teachers engage in during classwork? What kinds of performance expectations do teachers convey to students about the nature of mathematics?

Our second goal was to accurately portray instruction in Germany, Japan, and the United States. Toward this end, we were concerned that our descriptions of classrooms in other countries make sense from within those cultures, and not just from the U.S. point of view. One of the major opportunities of this study, after all, is that we may discover approaches to mathematics teaching in other cultures that we would not discover looking in our culture alone. We wanted to be sure that if different cultural scripts underlie instruction in each country, we would have a way to discover these scripts.

For this reason, we also sought coding ideas from the tapes themselves. In a field test, in May 1994, we collected nine tapes from each country. We convened a team of six code developers—two from Germany, two from Japan, and two from the United States—to spend the summer watching and discussing the contents of the tapes in order to develop a deep understanding of how teachers construct and implement lessons in each country.

The process was a straightforward one: We would watch a tape, discuss it, and then watch another. As we worked our way through the 27 tapes we began to generate hypotheses about what the key cross-cultural differences might be. These hypotheses formed the basis of codes (i.e., objective procedures that could be used to describe the videotapes quantitatively). We also developed some hypotheses about general scripts that describe the overall process of a lesson and devised ways to validate these scripts against the video data.

Developing Coding Procedures

Once we had developed a list of what to code we began developing the specific coding procedures. Coding procedures were developed by a group of four code developers, all of whom had participated in the initial viewing and discussion of the 27 field test tapes. One of the developers was from Germany (Knoll), one from Japan (Kawanaka), and one from the United States (Serrano). Each of these three were doctoral students in either psychology or education, and all had classroom teaching experience. The fourth member of the team was a doctoral student in applied linguistics (Gonzales), also from the United States, who helped us work through the technical issues involved in coding classroom discourse.

The coding development group first viewed field test tapes and a definition of the category to be coded was proposed. Each member of the group then attempted to apply the definition to field-test tapes from their country. Difficulties were brought back to the group and definitions were revised and refined. This process was repeated until all members of the group were satisfied with the definitions and procedures, and in agreement with the coding of each instance.

Once codes were developed, coders were trained to implement the codes. Coders, like the code developers, came from Germany, Japan, and the United States. In order to reduce the likelihood that subtle contextual cues would be missed or misinterpreted, coders only coded tapes from their country, except for purposes of training or the assessment of inter-rater reliability. Training was comprised of several activities. Code developers used group meetings to present definitions and discuss procedures for coding. Coders then practiced coding on the field test tapes. Results of practice coding were brought back to the group for discussion and any disagreements were resolved. This process was repeated until coders from each country were applying the codes in a consistent way.

Before beginning to code the main study tapes, a formal reliability assessment was conducted to insure independent agreement across coders at a level of at least 80 percent for each judgment. Reliability was assessed by comparing each coder's results with a standard produced by the coding development team. If reliability could not be established at the 80 percent level, the code was either dropped or sent back for revision. For the reliability assessment, coders worked with tapes from all countries, relying on English translations when necessary. We reasoned that reliability established across coders from different cultural backgrounds would be a low estimate of the actual reliability achieved among coders coding only tapes from their native countries. This also enabled us to make sure that coders from different countries were applying the codes in a comparable way.

Throughout this process we endeavored to be strategic. For example, just having collected 100 hours of video does not mean that all 100 hours must be analyzed. Depending on the frequency of what is being coded, it may be possible to time sample or event sample. It is also important to divide coding tasks into passes through the data in order to lessen the load on coders. This increases reliability and speeds up coding.

Implementation of Codes Using the Software

The code module of our software enables coders to view synchronized video and transcript on their computer screen. On-screen controls allow them to move instantly to the point in the video that corresponds to the highlighted transcript record, or to the point in the transcript that is closest in time to the current frame of video. Depending on the code, codes can be marked either as time codes in the video or as highlighting in the transcript.

FIRST-PASS CODING: THE LESSON TABLES

We have found that it is useful to have an intermediate representation of each lesson that can serve to guide coders as they try to comprehend a lesson, and that can be coded itself. For this purpose, our first step in coding the lessons was to construct a table that maps out the lesson along several dimensions. Each of these is defined in more detail as we present the results of the study, but a general idea of what they are is useful at this point:

- *Organization of class.* Each videotape is divided into three segments: pre-lesson activities (Pre-LA), lesson, and post-lesson activities (Post-LA). The lesson needs to be defined in this way because lesson is the basic unit of analysis in the study.
- *Outside interruption.* Interruptions from outside the class that take up time during the lesson (e.g., announcements over the public address system) are marked on the tables as well.
- *Organization of interaction.* The lesson is divided into periods of classwork (CW), periods of seatwork (SW), and periods of mixed organization. Seatwork segments are characterized as being Individual (students working on their own, individually), Group, or Mixed.
- *Activity segments.* Each classwork and seatwork segment is further divided, exhaustively, into activity segments according to changes in pedagogical function. We defined four major categories of activities: Setting Up, Working On, Sharing, and Teacher Talk/Demonstration. (Each of these was divided into subcategories, which are defined more completely in Chapter 4.)
- *Mathematical content of the lesson.* The mathematical content of the lesson is described in detail. Content is marked, for analytical purposes, into units which are noted on the table: Tasks, Situations, Principles/Properties/Definitions, Teacher Alternative Solution Methods [TASM] and Student Generated Solution Methods [SGSM]. (A more detailed description of each can be found below.) In addition, frames from the video are digitized and included in the table to help illustrate the flow and content of the lesson.

An example of what the resulting tables look like is shown in figure 3, which represents one of the Japanese lessons in our sample (JP-012).[2] The table contains five columns. The first column indicates the time code at which each segment begins as well as the corresponding page number from the printed lesson transcript. The second column shows the segmentation of the lesson by organization of interaction, the third by activity. The fourth and fifth columns show the symbolic description of the content and the concrete description of the content, respectively. Rows with lines between them show segment boundaries. Seatwork segments are shaded gray. The acronyms used refer to the coding categories described.

[2] Throughout this report, individual lessons from the sample are referred to by ID numbers, which include country of origin (GR, JP, US) followed by the lesson ID. In addition, all names used in excerpts from lesson tables and transcripts have been changed to pseudonyms.

Figure 3

Example of first-pass coding table for Japanese lesson (JP-012)

ID: JP012
Topic: 1.4.1 Transformations
Materials: Chalkboard; computer

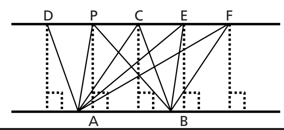
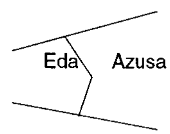
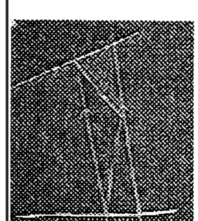
Page # (Time)	Organization	Activity	Content Symbol	Description of Content
1 (00:01)	Pre-LA			
1 (00:27)	CW	Working On: T/S/PPD	PPD1	(Computer) The triangles between two parallel lines have the same areas. 
1 (01:26)		Setting Up: M	S1	(Chalkboard) There is Eda's land. There is Azusa's land. And these two peoples' border line is bent but we want to make it straight. 
3 (03:34)			T1 [A]	Try thinking about the methods of changing this shape without changing the area.
3 (04:05)	SW: I	Working On: T/S/PPD		
4 (07:04)	SW: G			"People who have come up with an idea for now work with Mr. Azuma, and people who want to discuss it with your friends, you can do so. And for now I have placed some hint cards up here so people who want to refer to those, please go ahead."
16 (19:20)	CW	Sharing: T/S	SGSM1	 First you make a triangle. Then you draw a line parallel to the base of triangle. Since the areas of triangles between two parallel lines are the same we can draw a line here. [See the diagram]

Figure 3 Continued
Example of first-pass coding table for Japanese lesson (JP-012)

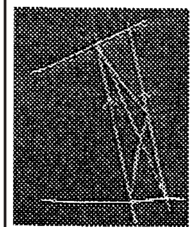
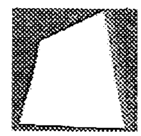
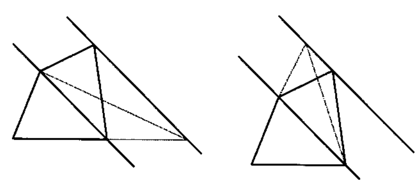
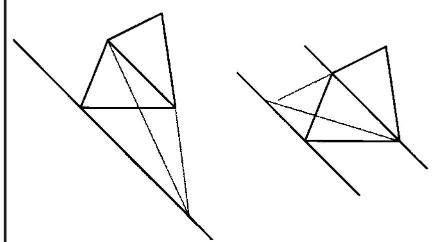
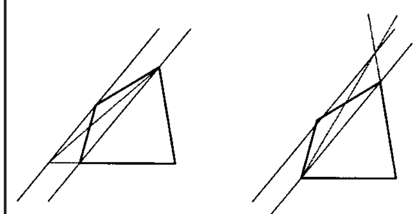
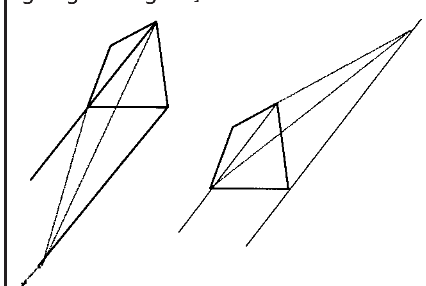
Page # (Time)	Organization	Activity	Content Symbol	Description of Content
16 (19:20)	CW	Sharing: T/S	SGSM2	 We make a triangle and draw a line parallel to the base of the triangle by fitting it with the apex. Since the length of the base doesn't change and the height in between the parallel lines doesn't change. So whereever you draw it the area doesn't change with the triangle that we got first.
19 (22:57)		Setting Up: Phys/Dir	S2	(Chalkboard) 
19 (23:25)			T2 [A]	Without changing the area please try making it into a triangle.
19 (23:39)	SW: I	Working On: T/S/PPD		
21 (26:47)	SW: SG			"Then people who are done please go to Mr. Azuma again. And people who want hints I will leave hint cards here so please look at them and try doing it. Itís also fine to do it with your friends.î (Hint cards unidentified.)
37 (46:11)	CW	Sharing: T/S		"We will make them ABCD."
			SGSM1	[Draw a diagonal line AC and make a triangle by drawing a parallel line going through D]
			SGSM2	

Figure 3 Continued
Example of first-pass coding table for Japanese lesson (JP-012)

Page # (Time)	Organization	Activity	Content Symbol	Description of Content
37 (46:11)	CW	Sharing: T/S	SGSM3	[Draw a diagonal line AC and make a triangle by drawing a parallel line going through B]
			SGSM4	
			SGSM5	[Draw a diagonal line BD and make a triangle by drawing a parallel line going through A]
			SGSM6	
			SGSM7	[Draw a diagonal line BD and make a triangle by drawing a parallel line going through C]
			SGSM8	
39 (48:58) 40 (49:47)		Working On: HW	HS1 HT1	Pentagon ABCDE. "Let's try making the pentagon into a triangle ... I'll make that then a homework."
40 (50:25-50:45)	Post-LA			

NOTE: Abbreviations used in the table: Pre-LA=Pre-Lesson Activities; CW=Classwork; T/S/PPD=Task/Situation/Principle Property Definition; PPD1=first Principle/Property/Definition in the lesson; Setting Up: M=Setting Up/Mathematical; S1=first situation in the lesson; T1=first task in the lesson; Ti[A]=initial, key, or target work of Task i; SW:I=Individual Seatwork; SW:G=Seatwork in Groups; SW:SG=Seatwork in Small Groups; Sharing: T/S=Sharing Task/Situation; SGSM1=first Student Generated Alternative Solution Method in the lesson; Setting Up: Phys/Dir=Setting Up: Physical/Directional; HS1=first Homework Situation in the lesson; HT1=first Homework Task in the lesson; HW=Homework; Post-LA=Post-lesson activities.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

We used these first-pass tables for two purposes. First, they were used by subsequent coders to get oriented to the contents of the videotapes. Often it takes a great deal of time for coders to figure out what is happening in a lesson. The tables ease the way, providing an overview of the structure and content of each lesson.

A second use for the tables is as objects of coding themselves. Some aspects of the lesson can be coded from the tables without even going back to the videotapes. Examples of such codes include TIMSS content category, nature of tasks and situations, and changes in mathematical complexity over the course of the lesson.

METHODS FOR DESCRIBING MATHEMATICAL CONTENT

There are many possible ways of describing mathematical content. One can describe content at a general level in terms of topics (e.g., ratio and proportion, or linear functions); in terms of categories such as "concepts" and "applications"; in terms of the specific tasks and situations assigned to students; or in terms of performance expectations. We attempted in this study to use all of these techniques.

The bases of our content descriptions are found in the first-pass coding tables we made for each lesson. We presented an example of these tables in the First-Pass Coding section. Here, we will give a more detailed description of how we constructed the content descriptions for these tables.

As coders watched the video, they first produced a written description of the content in concrete terms. The following example (figure 4), excerpted from the lesson table presented in the First-Pass Coding section (JP-012), illustrates the different kinds of information recorded in the content description. The example shows one task and one situation that the teacher presents to the students, a hint provided by the teacher to the students during Seatwork, and a student's solution method to the problem.

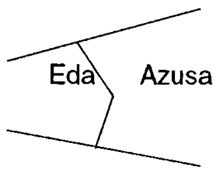
Figure 4

Excerpt from the content description column of the lesson table for JP-012

Description of Content

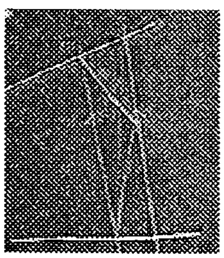
(Chalkboard)

There is Eda's land. There is Azusa's land. And these two people's border line is bent but we want to make it straight.



Try thinking about the methods of changing this shape without changing the area.

"People who have come up with an idea for now work with Mr. Azuma, and people who want to discuss it with your friends, you can do so. And for now I have placed some hint cards up here so people who want to refer to those, please go ahead."



First you make a triangle.
Then you draw a line to the base of the triangle.
Since the areas of triangles between two parallel lines are the same we can draw a line here.
[See the diagram]

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Once the concrete description is recorded, coders produced symbolic category descriptions of the content. Categorizing the content serves two functions: First, it helps guide the coders as they struggle to determine the proper level of detail to include in the description; second, it is useful for the analysis of content. We used five mutually exclusive categories for describing content:

- *Situation (S)*—The mathematical environment in which tasks are accomplished. (For example, real-world scenarios, word problems, and equations could all be the situations within which tasks are performed.)
- *Task (T)*—The mathematical goal or operation to be performed on a situation. (For example, "Try thinking about the methods of changing this shape without changing the area" is a task performed within the situation defined by the particular shape that is presented.)

- *Teacher Alternative Solution Method (TASM)*—An alternative method for solving a problem. A first method must be presented within the same lesson in order for there to be an alternative method coded.
- *Student Generated Solution Method (SGSM)*—A solution method generated and then presented by a student.
- *Principles, Properties, and/or Definitions (PPD)*—Mathematical information that is not contained in tasks and situations.

Each of these codes are numbered in order to keep track of the content as the lesson unfolds. A change in number signifies a shift to a new event. Numbers of tasks and situations are linked. For example, the notation T1-S1 represents a specific task and situation combination. If the same task is assigned for multiple situations (as might happen, for example, with a worksheet containing a number of similar exercises), the notation might be as follows: T1, S1-1, S1-2, S1-3. The first number after the "S" refers to the task (T1) that is being performed on each situation. The second number (1, 2, 3) refers to the number of different situations. A situation related to more than one task would be expressed similarly: S1, T1-1, T1-2, T1-3.

Relatedness of tasks/situations to previous events (e.g., Teacher Alternative Solution Methods) is expressed in symbolic form with parentheses. For example, T3-S3 (TASM2) means that task and situation "3" is directly related to the previous Teacher Alternative Solution Method 2.

THE MATH CONTENT GROUP

As mentioned in the introduction, one advantage of collecting video data is that they can be analyzed from multiple perspectives. In this report, we include some analyses by an independent group of researchers who are experts in mathematics and mathematics teaching. There were four members of this group (see names and brief biographical descriptions in appendix C). One had taught mathematics primarily at the high school level, one primarily at the college level but with extensive experience also teaching high school students, one primarily at the college level, and one at both college and graduate levels.

The group was assigned the task of analyzing the mathematical content of each lesson based only on information contained in the lesson tables. The group worked with a subsample of 90 lessons, 30 from each country. In order to reduce the possibility of bias, tables were disguised so that the group would not know the country of origin. This typically required only minor changes in such details as the names of persons and currencies. Lesson tables were identified only by an arbitrarily assigned ID code; relation of this code to country was not revealed until all coding was complete. Thus, although the group was denied the additional information that would have been provided by looking at the videotapes, the goal was to make possible a blind analysis of content. The analytic tools for describing the lesson content were developed by this group and will be described when we present the results of their analyses.

Coding of Discourse

Language is one of the key tools teachers and students use for instruction and learning. Consequently, focusing on how language is used in the classroom has the potential to enrich our analysis of instructional processes (see, for example, Bellack, 1966; Cazden, 1988; Mehan, 1979). It is also true that reformers of mathematics education have focused a great deal of attention on changing the kind of discourse that goes on during mathematics lessons. Mathematics teachers and students, it is suggested, should use language in much the same way that mathematicians do: to explain, justify, conjecture, and elaborate on mathematical understandings (see, for example, Hiebert & Wearne, 1993; Lampert, 1991).

Coding of discourse is very labor intensive, and very difficult when working across three languages. We decided to code discourse in several passes, and to employ a sampling scheme to save resources. We based our coding system on previous work (e.g., see references in previous paragraph), and on analysis of the field-test tapes.

Public and Private Talk

Our first step in coding discourse was to make a distinction between public and private talk. Public talk was defined as talk intended for everyone to hear; private talk was talk intended only for the teacher or an individual student. When the teacher stopped at an individual student's desk to make comments on the students' work this was generally coded as private talk, regardless of whether others could hear what the teacher was saying. The important thing is that the talk was primarily intended for this individual student alone. On the other hand, if a teacher stopped in the middle of a classwork period to criticize the behavior of a student sitting in the back of the room, this was coded as public talk. In this case, everyone had to stop and listen, even if they were not the one being disciplined.

All further coding of discourse was done on public talk only. Because public talk was accessible to everyone, we assumed that it would provide the most valid representation of the discourse environment experienced by students in the classroom.

First-Pass Coding and the Sampling Study

Next, we divided all transcripts into utterances, which was the smallest unit of analysis used for describing discourse. An utterance was defined as a sentence or phrase that serves a single goal or function. Generally, utterances are small and most often correspond to a single turn in a classroom conversation.

Utterances were then coded into 12 mutually exclusive categories. Six of the categories were used to code teacher utterances: Elicitation, Direction, Information, Uptake, Teacher Response, and Provide Answer. Five categories were applied to student utterances: Response, Student Elicitation, Student Information, Student Direction, and Student Uptake. One category, Other, could be applied to both teacher and student utterances. Elicitations were further subdivided into five mutually exclusive categories: Content, Metacognitive, Interactional, Evaluation, and Other. And Content Elicitations were subcategorized as well. Definitions of each of these categories will be presented later, together with the results.

Although all lessons were coded with the first-pass categories in the lesson transcripts, we decided to enter only a sample of the codes into the computer for preliminary analysis.

Thirty codes were sampled from each lesson according to the following procedure. First, three time points were randomly selected from each lesson. Starting with the last time point sampled, we found the first code in the transcript to occur after the sampled time. From this point, we took the first 10 consecutive codes, excluding Other, that occurred during public talk. If private talk was encountered before 10 codes were found, we continued to sample after the period of private talk. If the end of the lesson was encountered before 10 codes were found, we sampled upward from the time point until 10 codes were found. The same procedure was repeated for the second and first of the three time points. In those cases, if working down in the lesson led us to overlap with codes sampled from a later time point, we reversed and sampled upward from the selected time point.

Two kinds of summary variables were used for the sampling study: (1) Average number of codes (out of 30) in each lesson that were of each category, and (2) Percentage of lessons that contained any codes of each category (within the 30 codes sampled).

Second-Pass Coding of Discourse

For second-pass coding of discourse we decided to work with a subsample of lessons. We chose to study the 30 lessons in each country that had been selected for analysis by the Math Content Group, in part because they were balanced in their representation of algebra and geometry. Before proceeding, however, we wanted to know how well the subsample of 30 lessons in each country represented the larger sample, specifically with regard to discourse. To answer this question we compared the subsample of 30 lessons in the Math Content Group sample with the rest of the lessons in each country on each of the discourse variables produced in the first-pass sampling study (presented earlier).

Each variable was analyzed using a two-way analysis of variance (ANOVA), with country and sample group as factors. On only one analysis did we find a significant effect of sample group. However, neither for this variable nor for any of the others did we find a significant Country x Sample interaction.

Several new codes were added for second-pass coding. Content Elicitations, Information statements, and Directions were further subdivided. In addition, we started the process of grouping utterances into higher-level categories we call Elicitation-Response sequences (ER sequences). Elicitation-Response sequences appear to be the next-level building block for classroom conversations. A more detailed definition of all of these categories will be presented later with the results, but for now it is useful to define ER sequences as a sequence of turns exchanged between the teacher and student(s) that begins with an initial elicitation and ends with a final uptake.

STATISTICAL ANALYSES

Most of the analyses presented in this preliminary report are simple comparisons of either means or distributions across the three countries. In all cases, the lesson was the unit of analysis. All analyses were done in two stages: First, means or distributions were compared across the three countries using either one-way ANOVA or Pearson Chi-Square procedures. Variables coded dichotomously were usually analyzed using ANOVA, using asymptotic approximations.

Next, if overall analyses were significant, pairwise contrasts were computed and significance determined by the Bonferroni adjustment. In all cases, the Bonferroni adjustment was made assuming three simultaneous tests (i.e., Germany vs. Japan, Germany vs. United States, and Japan vs. United States). In the case of dichotomous variables (for which the sample estimate is a proportion) and continuous variables, we computed Student's *t* 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 by consulting the Bonferroni *t* tables published by Bailey (1977).

For categorical variables, we followed the procedure suggested by Wickens (1989) and used the Bonferroni Chi-Square tables printed in that book. Throughout, a significance level criterion of .05 was used. All differences discussed met at least this level of significance, unless otherwise stated. Anytime we use terms such as "less," "more," "greater," "higher," or "lower," for example, the reader can be assured that the comparisons are statistically significant.

All tests were two-tailed. Statistical tests were conducted using unrounded estimates and standard errors, which were also computed for each estimate. Standard errors for estimates shown in figures in the report are provided in the table in appendix E. Standard errors for estimates indicated in the text but not shown in figures are reported in footnotes to the relevant text.

Weighting

All of the analyses reported here were done using data weighted with survey weights, which were calculated for the classrooms in the videotape study itself, separate from any weights calculated for the main TIMSS assessment. The weights were developed for each country, so that estimates are unbiased estimates of national means and distributions. The weight for each classroom reflects the overall probability of selection for the classroom, with appropriate adjustments for nonresponse.

The analyses also used procedures that accounted for the complex nature of the sample design within each country (with the samples being independent across countries). The jackknife procedure was used, via the WesVarPC software, to account for the fact that a stratified random sample of schools was selected, with one classroom chosen from each selected school. F-tests for the comparison of means across three countries were achieved through the use of linear regression, with dummy variables indicating country as the independent variables. Pairwise t-tests were computed using the 'FUNCTION' capability of the 'TABLE' statement. Chi-square tests were computed using the 'TABLE' statement also, using first-order Rao-Scott corrections to account for the complex sample design.

COMPARISON OF VIDEO SUBSAMPLES WITH MAIN TIMSS SAMPLES

Despite the exhaustive attempts to select the video subsample randomly from the TIMSS main study sample, it may still be asked: Are the classrooms selected for the video study representative of the larger TIMSS sample?

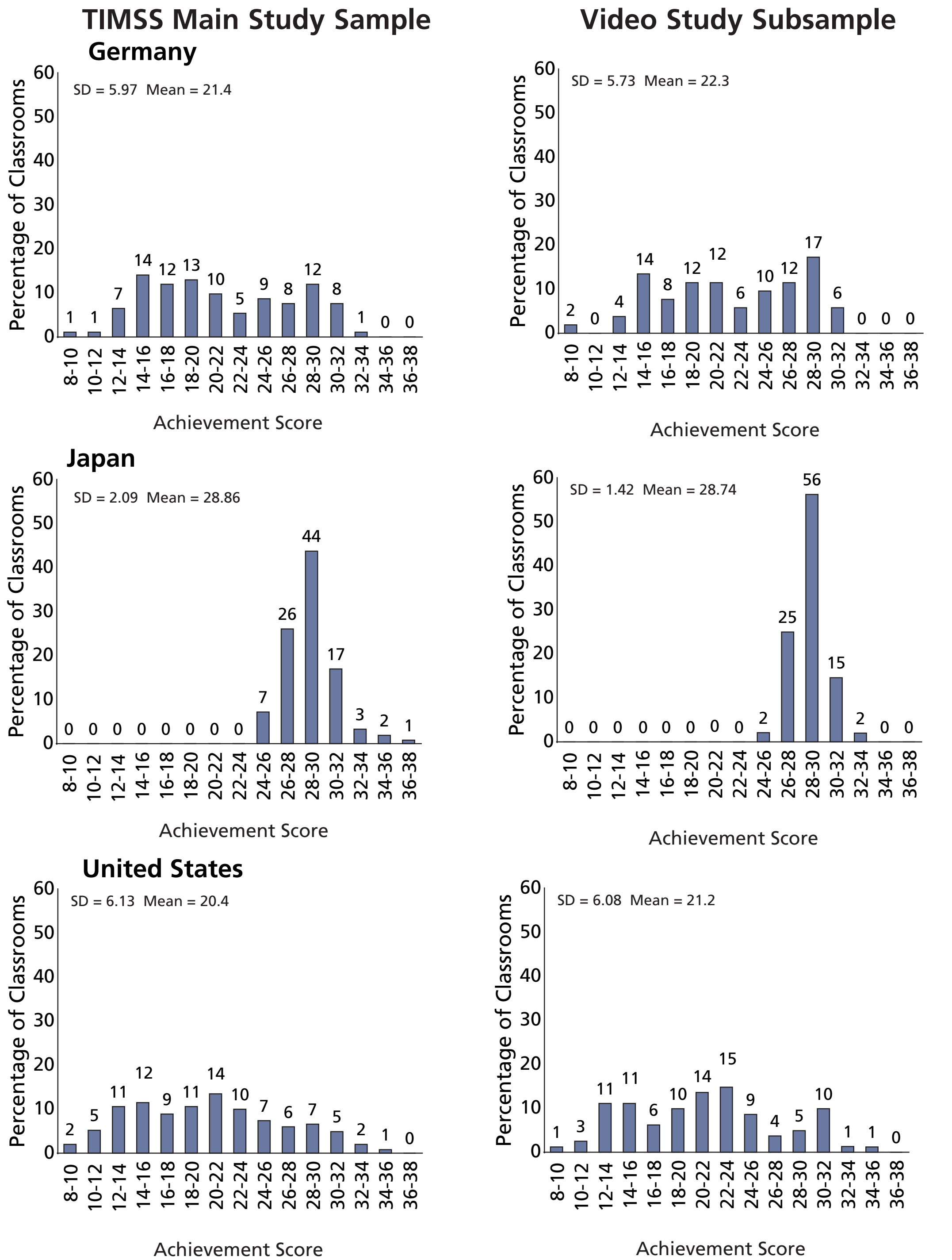
Some information relevant to this question can by found by comparing the mathematics achievement scores (i.e., performance on the TIMSS student assessments) of classrooms in the main TIMSS samples with the subsample of classrooms selected for the video study. We did not have test data for all of the classrooms included in the video study; and, data from Japan were somewhat problematic in one respect: Test data were not collected on classrooms included in the video study but on other eighth-grade classrooms in the same schools as the video classrooms. Nevertheless, we did have enough data to warrant a meaningful comparison of the two samples, and the lack of any tracking in Japan gives us some confidence that the school-level estimate of performance in Japan would be a reasonable indicator of classroom-level performance.

Distributions of mean achievement scores for classrooms in the Main TIMSS samples and in the video subsamples for each country are presented in figure 5. It is apparent in the figure that the distribution of mathematics achievement scores among the video subsamples are representative of distributions in the Main TIMSS samples.[3]

[3] These distributions are based on unweighted average achievement scores for each classroom. Our purpose is simply to compare distributions across pairs of samples within countries, not to make any inferences about true population distributions.

Figure 5

Distributions of unweighted average mathematics achievement test scores for classrooms in the Main TIMSS samples and video subsamples from each country



NOTE: SD = standard deviation.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

VALIDITY OF THE VIDEO OBSERVATIONS

As mentioned earlier, one of our concerns was that the presence of the video camera might in some way alter the nature of classroom instruction and thus threaten the validity of the study. One step we took to lessen the chances of this happening was to give all teachers a standard set of instructions in which we informed them of the goals of the study. We wanted to be certain that teachers understood that we wanted to film a typical day in their classroom, not one that was prepared especially for us.

We also attempted to assess how successful we were in sampling what typically happens in these classrooms by asking teachers, after the videotaping, to evaluate the typicality of what we would find on the videotape. We will report these results here.

Our first concern was that we might get a special lesson, one that the teacher holds in reserve for demonstration purposes. To ascertain whether or not this happened, we asked several questions on our questionnaire about how this particular lesson was chosen, and about how it related to the previous and next lessons that the teacher had taught or would teach to this same class of students. We asked, for example, whether the lesson we videotaped was a stand-alone lesson or part of a sequence of lessons. If the lesson was part of a sequence, we asked them to describe the goals and activities of the adjoining lessons so that we could judge the relationship they had to the lesson on videotape. Our reasoning was that special lessons would show up as stand-alone lessons that were unrelated to the adjoining lessons.

As it turned out, almost all of the teachers in our sample (97.8 percent in Germany, 95.7 percent in Japan, and 93.4 percent in the United States) reported that the videotaped lesson was part of a sequence of lessons designed to teach a particular topic in the mathematics curriculum. Further, they were able to give clear and reasonable descriptions of how this lesson related to the previous and next lessons in the sequence. This outcome confirmed our sense that teachers did not make drastic accommodations to prepare for our videographer.

We also asked teachers how many lessons were in the whole sequence of lessons, and where the lesson we videotaped fell in the sequence. The average sequence of lessons reported by U.S. teachers was 9.1 lessons, significantly shorter than the 13.6 and 14.4 lessons reported by German and Japanese teachers respectively.[4] However, the position of the videotaped lesson in the sequence did not differ across countries.[5]

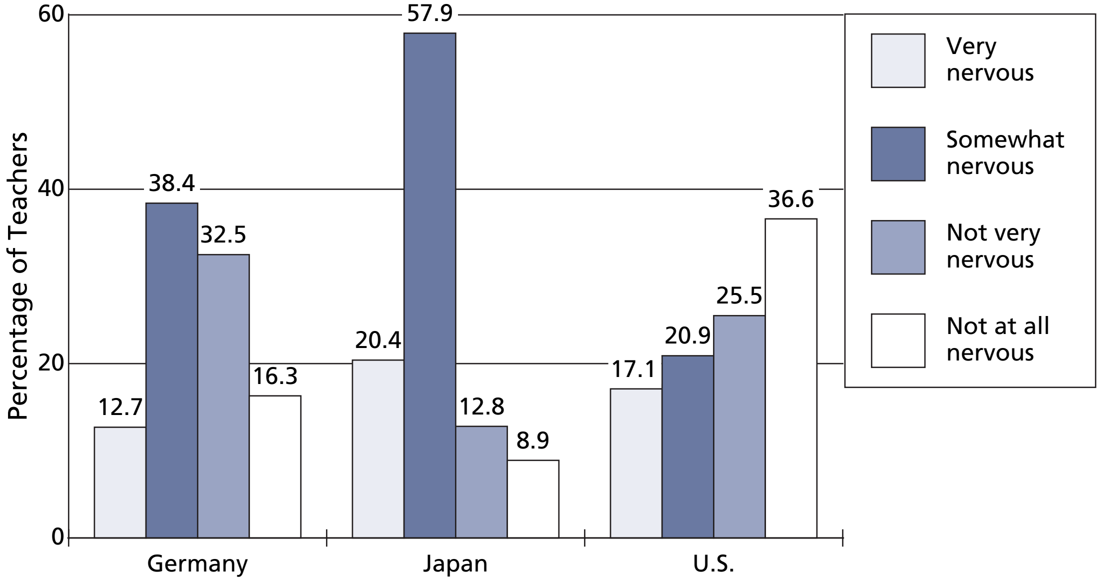
A more subtle picture of how the presence of the camera might have affected instruction emerges when we look at some of the other judgments teachers made about the lesson. First, it is interesting to see how nervous or tense the teachers felt about being videotaped. Teachers were asked to check one of four choices: Very nervous, somewhat nervous, not very nervous, and not at all nervous. Japanese teachers reported being more nervous than both German and U.S. teachers about the presence of our videographer.[6] Figure 6 shows that about three-fifths of U.S. teachers (62.1 percent), almost one-half of German teachers (48.9 percent), and about one-fifth of Japanese teachers (21.6 percent) reported being "not at all nervous" or "not very nervous."

[4] Standard errors for Germany, Japan, and the United States are 0.99, 1.26, and 0.61, respectively.

[5] Estimates for Germany, Japan, and the United States of the position in the sequence (as a ratio of the position in the sequence to the total number in the sequence) are 0.50, 0.46, and 0.53, respectively, with standard errors of 0.035, 0.036, and 0.042.

[6] Estimates for Germany, Japan, and the United States of the level of nervousness (on a four-point scale, with 1 indicating "very nervous" and 4, "not at all nervous") are 2.5, 2.1, and 2.8, respectively, with standard errors of 0.07, 0.09, and 0.12.

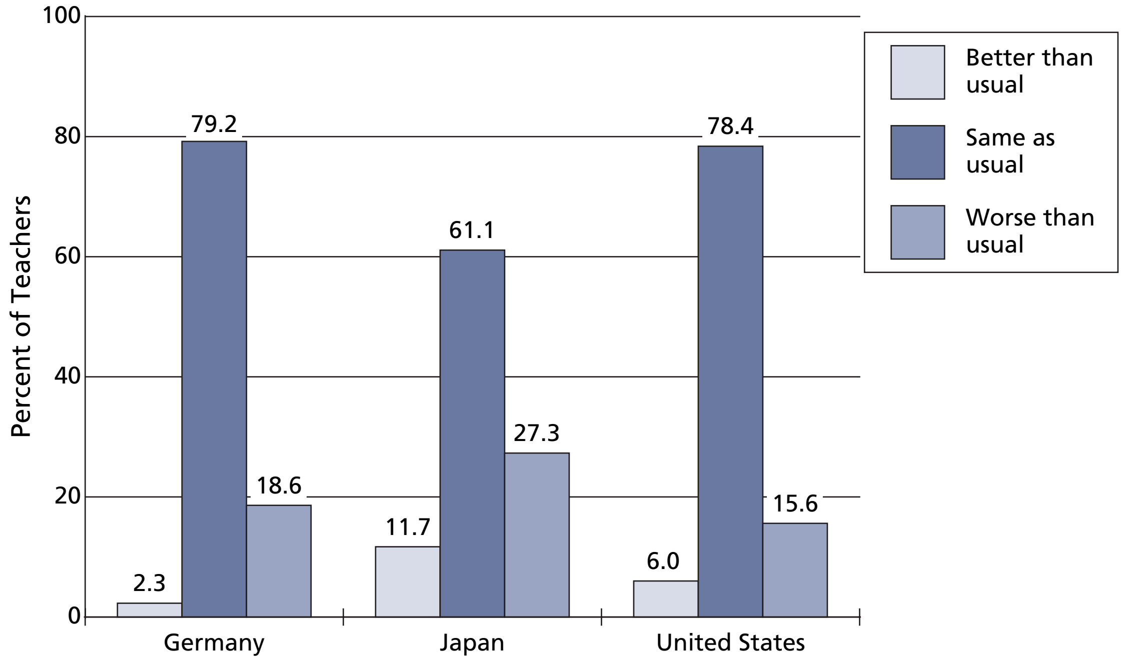
Figure 6
Teachers' reports of how nervous or tense they felt about being videotaped



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

A number of questions were designed to get teachers' evaluations of how typical the lesson on the videotape was of the lessons they normally teach to this class of students. In figure 7 we present teachers' judgments of the quality of the video lesson compared with their usual lesson. Again, we see that Japanese teachers appear to differ from their German colleagues. Twenty-seven percent of Japanese teachers and 18.6 percent of German teachers reported feeling that the lesson on tape was worse than usual. At the other end of the scale, 11.7 percent Japanese teachers and 2.3 percent of German teachers reported feeling that the videotaped lesson was better than usual.

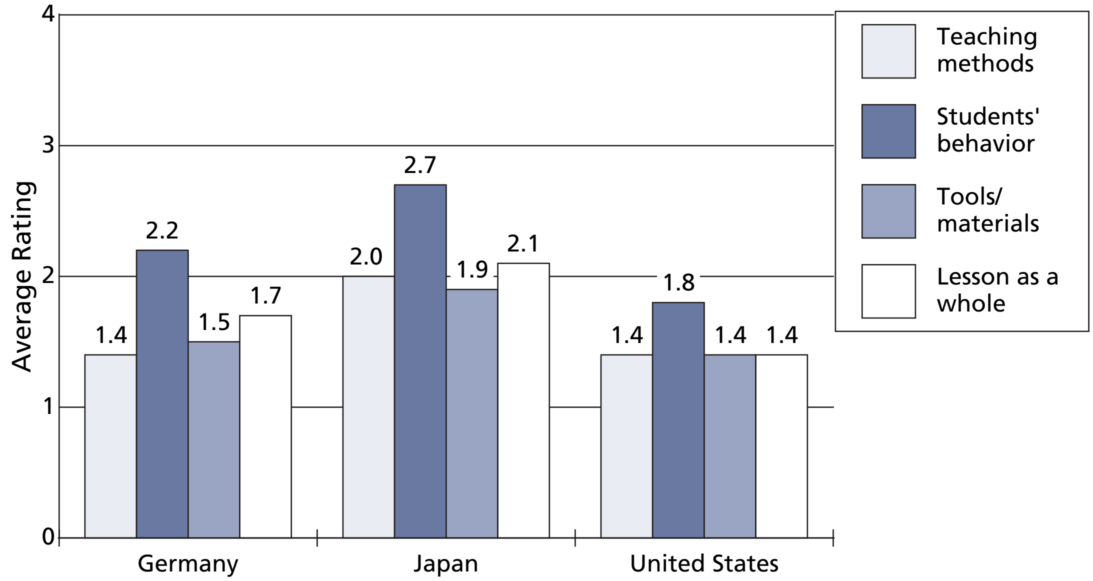
Figure 7
Teachers' ratings of the quality of the videotaped lesson compared to lessons they usually teach



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Four other questions probed teacher's judgments of how typical the videotaped lesson was of lessons they normally teach. On these questions, teachers rated the typicality of the teaching methods, the behavior of the students, the tools and materials, and the lesson as a whole. Teachers used a four-point scale, where 1 means very typical/similar to what usually happens, and 4 means completely atypical/very different from what usually happens. Responses to the four questions are summarized in figure 8. Again, the Japanese teachers, on each of the four questions, rated their lessons as less typical than did teachers in the other two countries.

Figure 8
Teachers' average ratings of the typicality of various aspects of the videotaped lesson



NOTE: 1=very typical, 4=completely atypical.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Although Japanese teachers rated the lessons as significantly less typical than did German or U.S. teachers, the overall ratings were not particularly troubling to us. In fact, when asked to rate the typicality of the lesson as a whole, most teachers in all cultures chose either "very typical" or "mostly typical." 95.6 percent of German teachers, 85.1 percent of Japanese teachers, and 97.4 percent of U.S. teachers responded in this way. In conclusion, the videotape study probably captured a fairly representative picture of what typically happens in eighth-grade mathematics classrooms in these three countries.

Chapter 3. Mathematical Content of Lessons

CONTENT: A PLACE TO BEGIN

Any observer of classroom instruction is struck by its complexity. Instruction is multidimensional. Instruction also unfolds quickly in real time, comprised as it is of an unceasing flow of events. To focus on one dimension is to lose sight of the others, which is why video is so useful in the study of classroom instruction. With video, we can make multiple passes through the lesson, focusing in each pass on a particular dimension or layer of description.

For purposes of analysis, we can describe instruction in layers. Although these layers may differ in character across cultures, the layers themselves as analytical points of reference have validity across cultures. These layers include the following:

- Setting: What is the environment, both physical and social, in which the lesson takes place?
- Content: What is the curricular content of the lesson? What does "mathematics" look like?
- Participants: We can analyze the role of the teacher and the role of the student as separable layers of classroom instruction.
- Organization: What is the social organization of the lesson (e.g., whole class, small groups, individuals) and how does the form of organization change over the course of the lesson? What is the functional organization of the lesson in terms of activities?
- Scripts and Goals: What are the cultural scripts and goals that tie the parts of the lesson together? In other words, how are all of these layers put together to make a lesson?

In practice, these different layers are woven together by the teacher who, in all three cultures, takes primary responsibility for the lesson, both in planning and execution. In planning the lesson teachers rely on physical, intellectual, and cultural resources. Physical resources include the setting and the materials available. Intellectual resources include the teacher's own mathematical knowledge and other mathematical knowledge available to the teacher. Cultural resources include the shared cultural understandings of the students in terms of goals, assumptions, routines, and roles, and a set of participant structures that teachers and students together know how to construct in the classroom. These structures include such things as classwork and seatwork.

While recognizing the multidimensionality of classroom instruction, we have chosen to begin with content. It is difficult to draw the line between what is taught and how it is taught. Still, it is useful to examine content apart from the lesson in which it is embedded. The rationale for this is simple: No matter how good the teaching is, if a lesson does not include rich mathematical content, it is unlikely that many students will construct a deep understanding of mathematics from the lesson. In this section of the report we describe the mathematical content of lessons in each country.

GENERAL DESCRIPTIONS OF CONTENT

Our first step in describing the mathematical content of each lesson was to apply the TIMSS content coding system. The complete system, which included 44 categories, is available in Robitaille, McKnight, Schmidt, Britton, Raizen, and Nicol (1993). All coding was done from the lesson tables. For each segment in the table, the coder wrote down the TIMSS code that best described the mathematical content. Each lesson was thus described using one or more content codes.

TIMSS content coding was checked by independent coders at Michigan State University (MSU)—the same coders who had done the textbook coding for the TIMSS curriculum analysis. They, like the UCLA coders, based their analysis on the video lesson tables. There was perfect agreement between coders at UCLA and at MSU.

In general, more topics were represented in the sample of U.S. videotapes (24 topics) than in the samples of German (18 topics) or Japanese (13 topics) tapes. The 44 TIMSS content coding system categories were further grouped into 10 major categories:

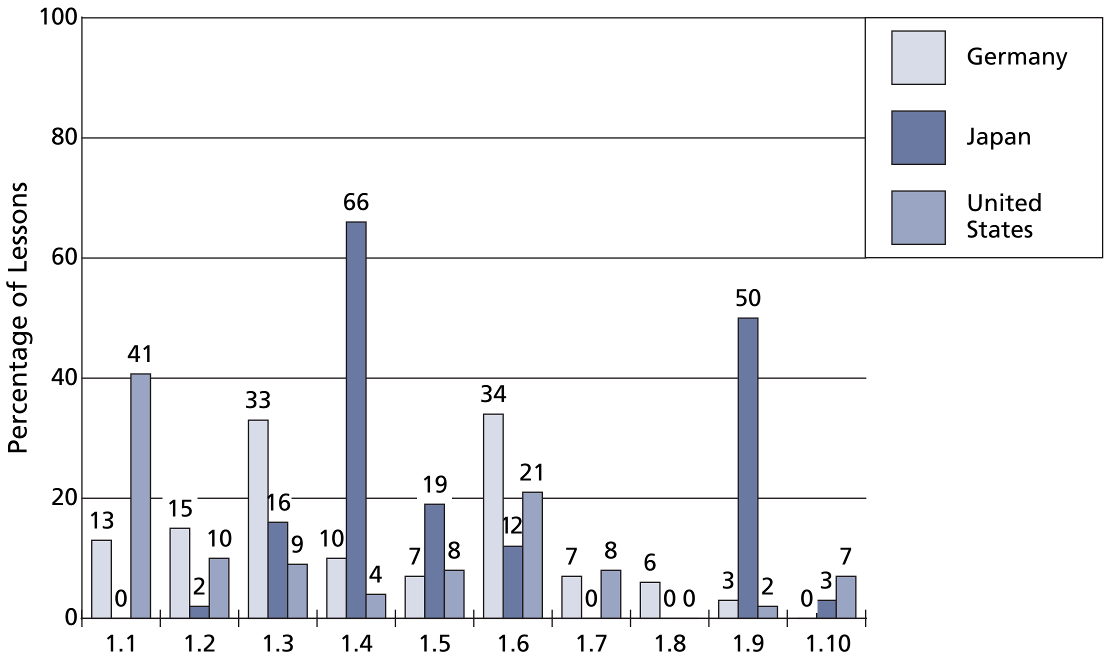
- 1.1 Numbers—including whole numbers, fractions and decimals; integers, rational, and real numbers; number theory; estimation and number sense.
- 1.2 Measurement—including units, perimeter, area, and volume.
- 1.3 Geometry: Position, Visualization, and Shape—including two dimensional and three dimensional geometry.
- 1.4 Geometry: Symmetry, Congruence, and Similarity—including transformations; congruence and similarity; and constructions using straight-edge and compass.
- 1.5 Proportionality—including proportionality concepts and problems; slope and trigonometry; and linear interpolation and extrapolation.
- 1.6 Functions, Relations, and Equations—including patterns, relations, and functions; and equations and formulas.
- 1.7 Data Representation, Probability, and Statistics—including data representation and analysis; uncertainty and probability.
- 1.8 Elementary Analysis—including infinite processes and change.
- 1.9 Validation and Structure—including validation and justification; structuring and abstracting.
- 1.10 Other Content

In figure 9, we show the (unweighted) percentage of lessons in our sample that included content belonging to each of the 10 major categories. Our purpose in presenting these data is to better describe the mathematical content of our sample of videotapes; the resulting differences across the three samples should not necessarily be generalized to the populations of eighth-grade classrooms in the three countries. It is for this reason that no statistical test was done on this variable.

There are some clear differences in the percentage of lessons in our sample that were devoted to various topics across the three countries. The most frequent topic in the U.S. sample (about 40 percent of the lessons) was Numbers, which included such topics as whole number operations, fractions, and decimals. In the German sample, the two most common topics were Geometry (Position, Visualization, and Shape), and Functions, Relations, and Equations. In the Japanese sample, the most common was Geometry

(Symmetry, Congruence, and Similarity), followed by Validation and Structure. Recall, however, that the emphasis on geometry in Japan is partly a result of bias in our sampling procedure.

Figure 9
Percentage of lessons in each country in which content belonged to each of the ten major content categories



NOTE: 1.1=Numbers; 1.2=Measurement; 1.3=Geometry (Position, Visualization, Shape); 1.4=Geometry (Symmetry, Congruence, Similarity); 1.5=Proportionality; 1.6=Functions, Relations, Equations; 1.7=Data Representation, Probability, Statistics; 1.8=Elementary Analysis; 1.9=Validation and Structure; 1.10=Other.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

HOW ADVANCED IS THE CONTENT BY INTERNATIONAL STANDARDS?

It is not possible, a priori, to say that one topic is more complex than another. However, it is possible to make an empirical judgment of how advanced a topic is based on its placement in mathematics curricula around the world. We were able to make use of the TIMSS curriculum analyses, conducted by William Schmidt and his colleagues at MSU, to make such a judgment. The TIMSS content codes for each lesson (Robitaille et al., 1993) were assigned a number indicating the modal grade level at which the majority of the 41 countries studied gave the most concentrated curricular attention to the topic. Average level for each lesson was obtained by averaging the MSU index for all topics coded for the lesson.

The average grade level of topics covered in the video sample, as indicated by the MSU index, is shown in figure 10. In terms of this index, the average grade level of topics covered in the U.S. sample was significantly different than in Germany and Japan. Based on the MSU index of international standards, the mathematical content of the U.S. lessons in the videotape study was at a seventh-grade level, whereas the German and Japanese lessons fell at the high eighth- or even ninth-grade level.

Figure 10
Average grade level of content by international standards

	Mean
Germany	8.7
Japan	9.1
United States	7.4

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

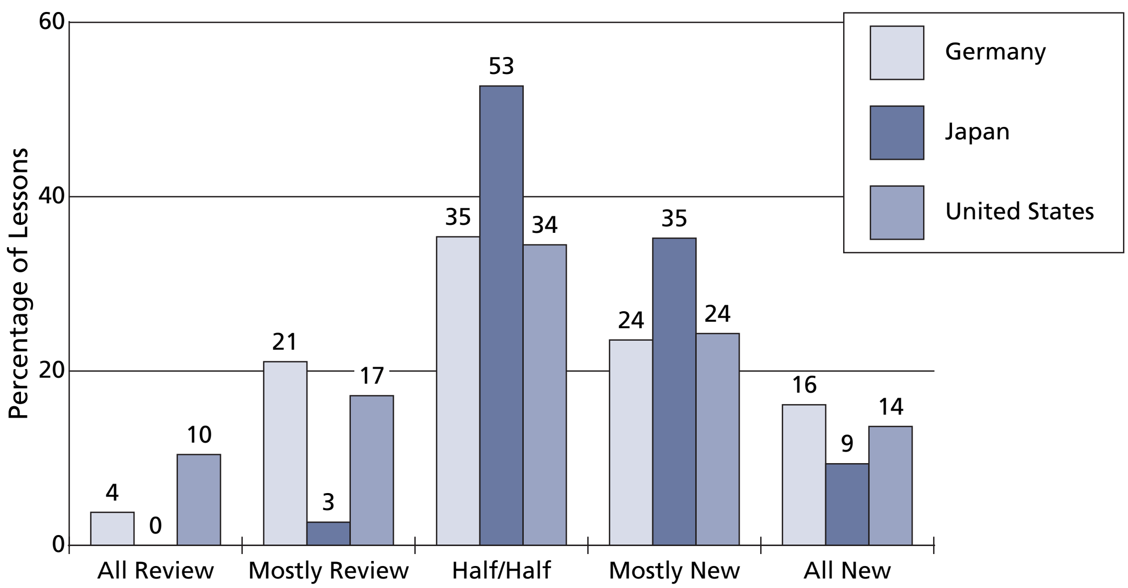
A CLOSER LOOK AT CONTENT

Teacher's Goal for the Lesson

We begin a more detailed look at content by examining teachers' responses on the teacher questionnaire.

We asked teachers whether the content of the lesson was all review, all new, or somewhere in between. Responses to this question are shown in figure 11. Analyses indicated that the distribution of responses in Japan differed significantly from that in both Germany and the United States.

Figure 11
Teachers' description of the content of the videotaped lesson on a continuum from "all review" to "all new"



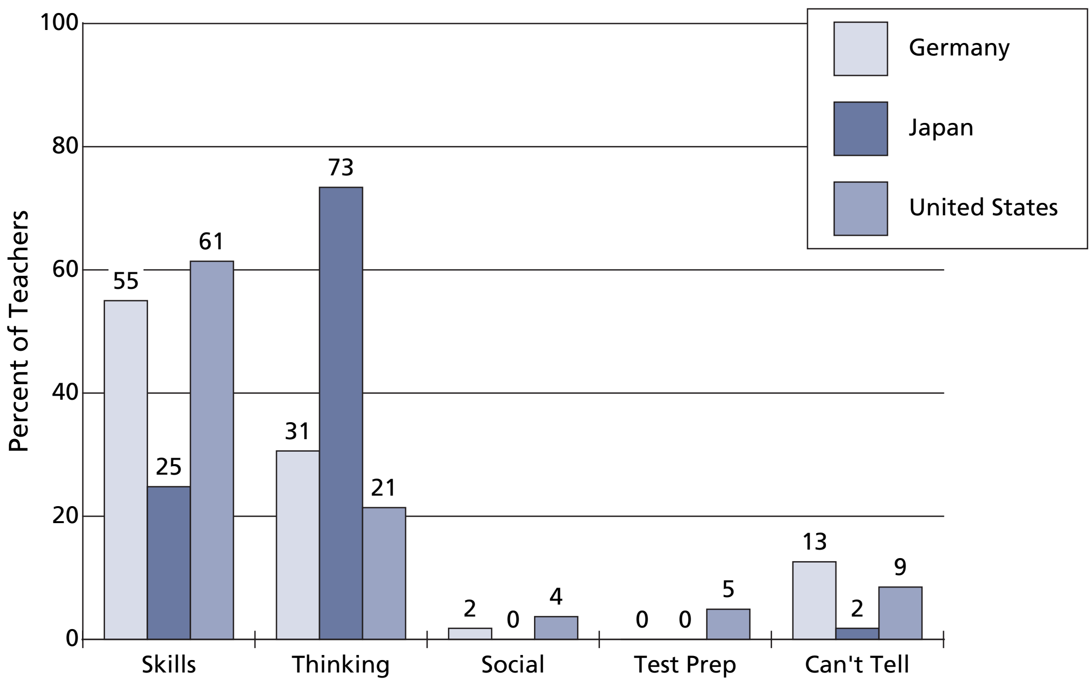
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

We also found significant differences when we asked teachers what main thing they wanted students to learn from the lesson. Responses were coded into five categories:

- *Mathematical Skills*—responses that emphasized the teaching of how to solve specific kinds of problems, use of standard formulas, etc.
- *Mathematical Thinking*—responses that emphasized students' exploration, development, and comprehension of mathematical concepts, or the discovery of multiple solutions to a problem.
- *Social/Motivational*—responses that emphasized non-mathematical goals, such as "listening to others," or the creation of interest in some aspect of mathematics.
- *Test Preparation*—responses that focused on preparing for an upcoming test.
- *Indeterminable*—responses that were not possible to categorize, usually because they were too vague or incomplete.

The teachers' responses are shown in figure 12. There was a significant difference between the reported goals of teachers in Japan and teachers in the other two countries. A majority of Japanese teachers reported that thinking was the main goal/skill to be learned from their lessons, while 55 percent of German teachers and 61 percent of U.S. teachers reported that skills were the main thing to be learned.

Figure 12
Teachers' responses, on the questionnaire, to the question,
"What was the main thing you wanted students to learn from today's lesson?"



NOTE: The numbers in this graph differ slightly from those reported in Peak (1996, page 42) where unweighted averages were mistakenly used instead of weighted averages. Specifically, the percentages of teachers who responded Skills were reported as about 60 percent for U.S. and German teachers and 27 percent for Japanese teachers. The percentages responding Thinking were reported as 24, 29, and 71, respectively. The numbers shown in this figure are the correct ones. Percentages may not sum to 100 due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Number of Topics and Topic Segments per Lesson

Having determined which topics are taught in each country, we proceeded to divide lessons into topic segments (i.e., the points in the lesson at which the topic shifts from one to another). By topic, we here refer to the TIMSS content topics, which are broadly defined. A shift in topic is a clear shift in the primary content of the lesson and is usually marked by an announcement by the teacher. For example, in lesson GR-080 (00:28:48) the teacher says: "Okay let's still start something new today. Your school exercise book please." This is a clear signal that the topic is about to change, thus marking the beginning of a new topic segment.

If a lesson has only one topic then it will, by definition, have only one topic segment. However, many lessons had more than one topic. In this case, the number of topic segments might equal the number of topics (i.e., there could be one segment for each topic) or it might exceed the number of topics if, for example, the lesson alternated from a segment of one topic, to a segment of another, to another segment of the original topic.

US-003 is an example of a lesson that contains two topics and three topic segments. The teacher begins the lesson with proportionality, assigning students to work on the following problem:

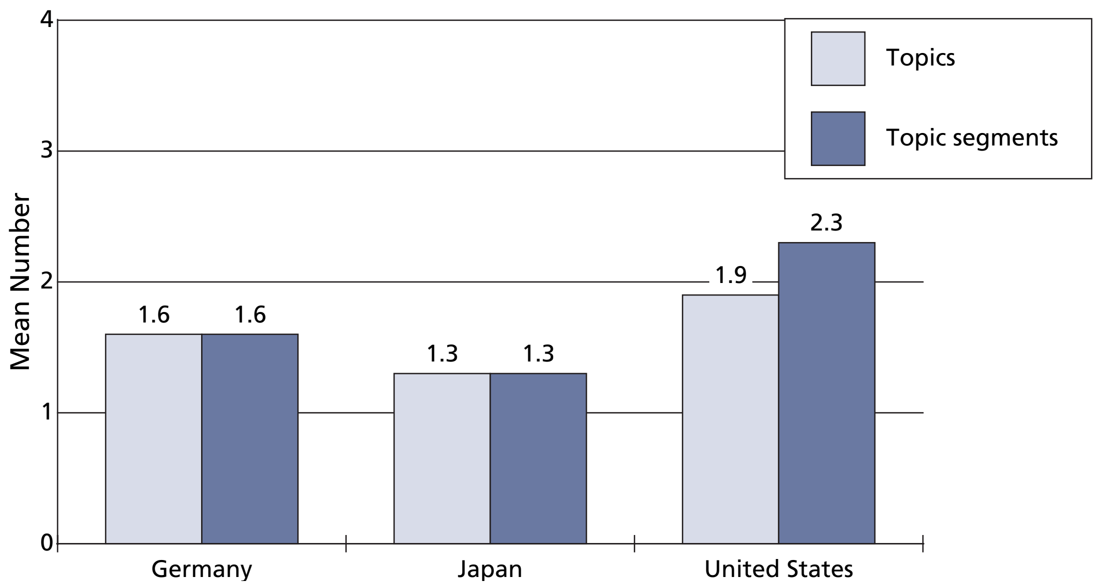
> A team must choose which of the field goal kickers to send in for a possible game winning kick. The one who is 11 for 16 or the one who is 8 out of 12.

At 22:05, the teacher shifts to focus on a new topic, inequalities, with the following statement:

> Okay? Now if you turn your books to page one ... twenty-three where that homework was let's look at the ones we had on the sideboard before we give these papers back.

Students work on discussing the inequalities they had done for homework until, at 35:12, the teacher changes the focus back to proportionality, assigning exercises from the textbook until the end of class. Thus, the first and third topic segments focused on proportionality, the second segment, on inequalities.

Figure 13
Average number of topics and topic segments per videotaped lesson in each country



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

As shown in figure 13, U.S. lessons contained significantly more topics and topic segments than did Japanese lessons. Also, German lessons contained more topics and topic segments than did Japanese lessons.

Concepts and Applications

We next examined, for each topic covered in the lesson, whether the topic included concepts, applications, or both. We defined these mutually exclusive categories broadly to catch all instances in which students might have constructed concepts or learned to apply them.

- CONCEPT was coded when the only treatment of the topic in the lesson involved the presentation of information, through either a statement or a derivation of general mathematical principles, properties, or definitions (e.g., formulas and theorems), or statement or derivation of a method for solving a class of problems. A concept can be presented through a concrete example or abstractly. It can be introduced for the first time or simply be restated.
- APPLICATION was coded when the only treatment of the topic in the lesson was as an application to the solving of a specific mathematical problem. Mathematical concepts were not explicitly stated or discussed. The emphasis was on developing skills for solving specific types of problems.
- BOTH was coded when a topic included, somewhere in the lesson, both concepts and applications. (If a mathematical concept was stated that did not directly relate to the topic, it was not counted as a concept.)

An example of a topic coded as CONCEPT can be seen in GR-074. The lesson deals with calculations involving polygons. CONCEPT was coded because solution methods were stated abstractly, as formulas, rather than using the specific numbers of a particular problem. Here is an excerpt from the lesson transcript:

00:03:45	T	Okay and now we're really getting to work here. How do we calculate this triangle's circumference?
00:03:53	T	Carolyn. ... The circumference of a triangle. How do we calculate it?
00:04:02	S	Uhm A plus B plus C.
00:04:03	T	Yes. We can do that without even looking it up right? Simply adding all three sides and we already have the sum of the borders. The surface area? That already is a little more difficult. If you don't know it anymore quickly look it up. We haven't needed that as frequently lately. Katrin you already had it?
00:04:23	S	G times H divided by two.
00:04:25	T	Right. Base times height over that base. And the whole thing over two because we only have a triangle. And since we're already at it. How does the whole thing look for a rectangle? Circumference of a rectangle? Polly?
00:04:42	S	Two A plus two B.
00:04:45	T	Let's make it U R to distinguish that we have a rectangle here. Right. We have four sides. Two times A and two times B ... added we get that. And the most difficult formula? Jim?
00:04:58	S	A times B.

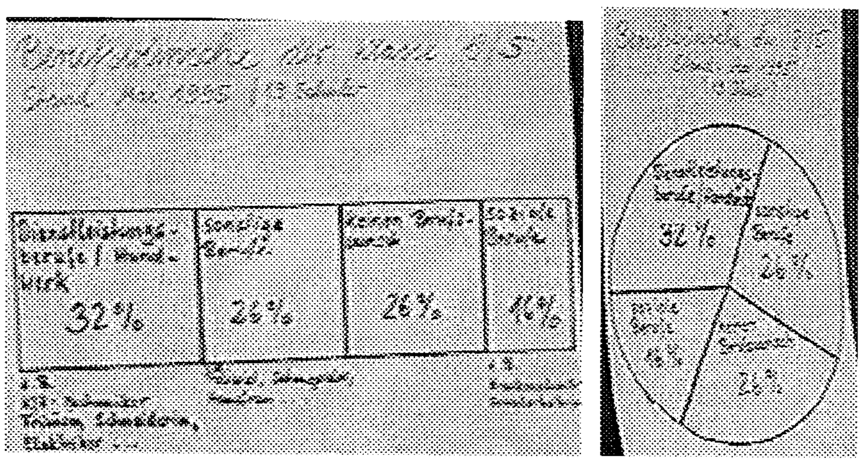
00:04:59 T Right. The area in a rectangle is calculated with A times B. You know that by heart. You have to know that.

00:05:11 T We have summarized it again then. I promise you we'll need it again.

Another example of a topic coded as CONCEPT is found in JP-040. The topic of this lesson is similarity. The teacher first presents the definition of similar figures and, together with the students, generates examples. The teacher then demonstrates that operations of multiplication and division can be used but not addition and subtraction. Finally, the students derive a triangle's similarity conditions from its congruence conditions. Students are asked to memorize the conditions for similarity.

APPLICATIONS are coded when there is no explicit mention of mathematical concepts. In GR-096, for example, the topic is representation of data. The teacher reviews different types of charts by using illustrations as examples (figure 14).

Figure 14
Pictures of the chalkboard from GR-096

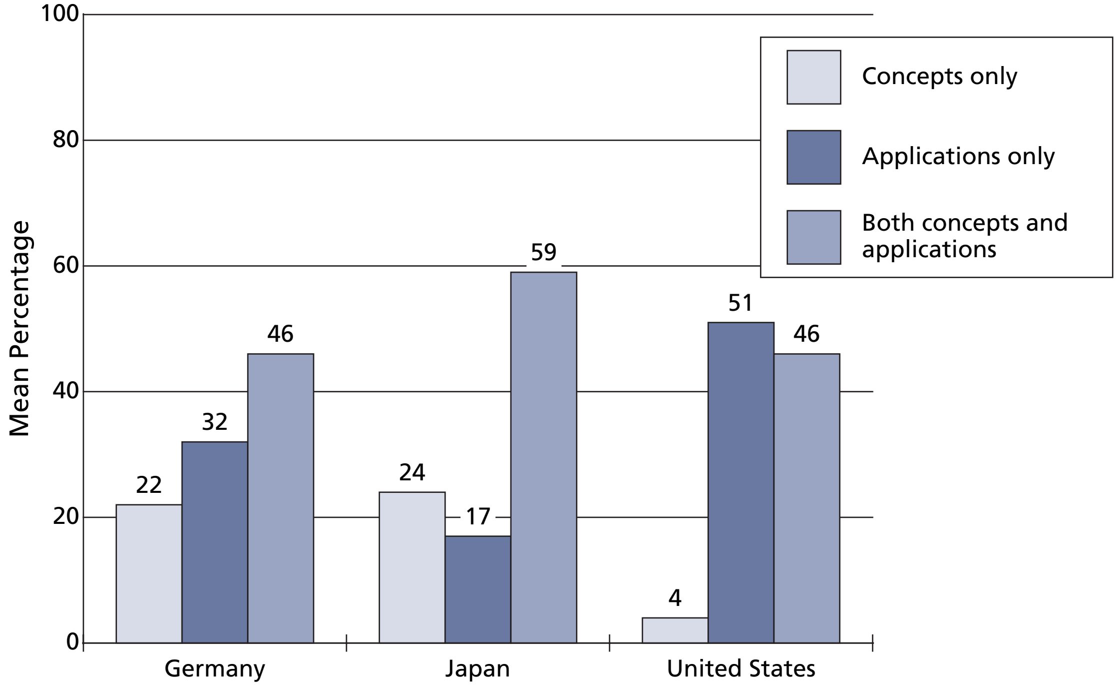


SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Teacher and students talk briefly about the names of different kinds of charts, as well as what different values in the charts represent. The application begins when the teacher hands out data in tabular form and asks students, working in groups, to create their own charts from the data.

As shown in figure 15, U.S. lessons had significantly lower percentages of topics that consisted of concepts only than did either German or Japanese lessons. And, U.S. and German lessons had a higher percentage of topics that consisted of applications only than did Japanese lessons.

Figure 15
Average percentage of topics in each lesson that include concepts, applications, or both



NOTE: Percentages may not sum to 100.0 due to rounding.

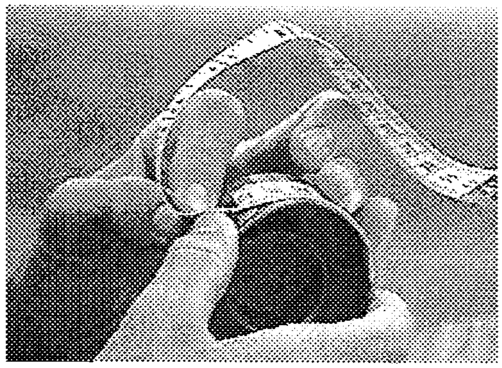
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

WERE CONCEPTS STATED OR DEVELOPED?

When concepts were included in the lesson, they could be stated or they could be developed. A concept was coded as STATED if it was simply provided by the teacher or students but not explained or derived. For example, the teacher, in the course of solving a problem at the board, might simply remind the students of the Pythagorean theorem (e.g., "The formula for finding the length of the hypotenuse of a right triangle is $a^2 + b^2 = c^2$") in order to guide the solution of the problem. The focus here is on the mathematical information itself rather than on the process of deriving it. A concept was coded as DEVELOPED when it was derived and/or explained by the teacher or the teacher and students collaboratively in order to increase students' understanding of the concept. The form of the derivation could be through proof, experimentation, or both.

US-068 provides an example of a concept being developed through experimentation. The topic of the lesson is the relationship between circumference and Pi. The teacher begins the lesson by defining terms such as circumference and diameter. Students then break into groups to work with circular objects. They measure the objects' circumferences (C) and diameters (D) with a measuring tape (figure 16). They then divide C by D and examine their answers. In a subsequent class discussion, the teacher uses the commonality across answers as a basis for defining Pi.

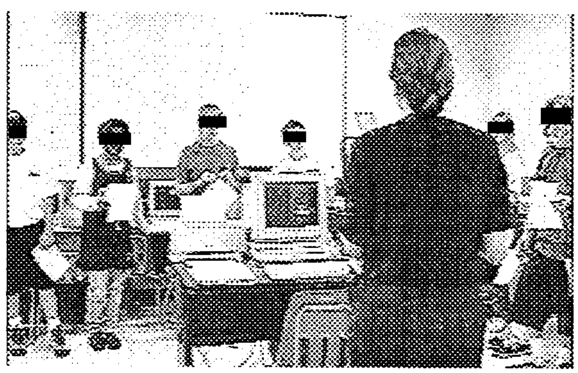
Figure 16
Materials used in US-068



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

US-061 also engages students in a concrete example. This time, however, the concept is coded as STATED. The topic is data representation. The teacher asks all students to write their shoe size on a piece of paper and asks them to stand in order according to shoe size from smallest to largest (figure 17). The teacher then defines several statistical terms (e.g., range and mean) and asks the students to calculate the statistics using their shoe sizes. This is not coded as development because the concepts are not derived, only stated and applied in the example.

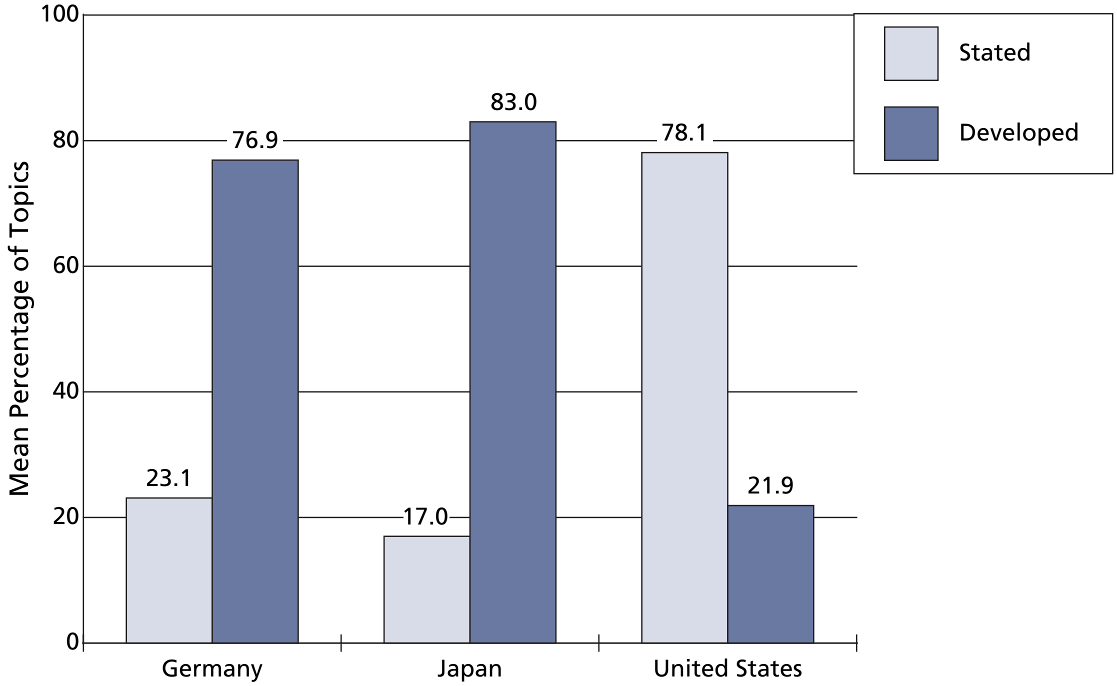
Figure 17
A view of the classroom in US-061



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

The results of coding concepts as STATED versus DEVELOPED are shown in figure 18. Concepts were significantly more likely to be STATED in the U.S. lessons than in both Germany and Japan. Conversely, concepts were more likely to be DEVELOPED in German and Japanese lessons than in the U.S. lessons.

Figure 18
Average percentage of topics in eighth-grade mathematics lessons that contained concepts that were stated or developed



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

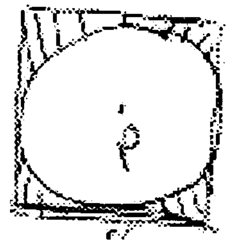
Did Applications Increase in Complexity?

Finally, in topics that contained applications, we were interested in the relationship among the application problems. Specifically, when a topic contained more than one problem, were the problems just multiple examples of the same level of complexity, or did they increase in complexity over the course of the lesson?

Change in complexity—categorized as either INCREASE or SAME/DECREASE—was coded only if there was more than one application problem within the same topic in a lesson. Increase in complexity was coded when either a procedural or conceptual difficulty was added from one application problem to another. Increasing procedural difficulties generally consisted of additional operations (i.e., calculations previously executed separately were now combined in one application). Increasing conceptual difficulties generally consisted of added mathematical information (i.e., previously learned concepts now had to be modified for a new application using more mathematical information).

US-018 provides a good example of increasing complexity across two application problems. The lesson deals with area and circumference of a circle. In the first problem students are asked to find the area of the shaded region between the circle and the square (figure 19). In order to solve this problem the students had to find the area of the circle and then subtract that from the area of the polygon.

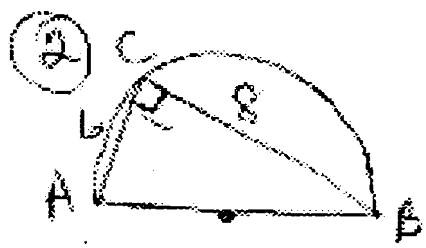
Figure 19
Drawing from chalkboard of first problem in US-018



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Students next were asked to find the area of a semicircle in which a triangle was inscribed (figure 20) and then the area of the region not covered by the triangle. This problem, although it employed some of the same concepts of the previous one, is clearly more complex. To solve it, students had to find the hypotenuse of the triangle (which is also the diameter of the circle), calculate the area of the semicircle, calculate the area of the triangle, and then subtract the area of the triangle from the area of the semicircle.

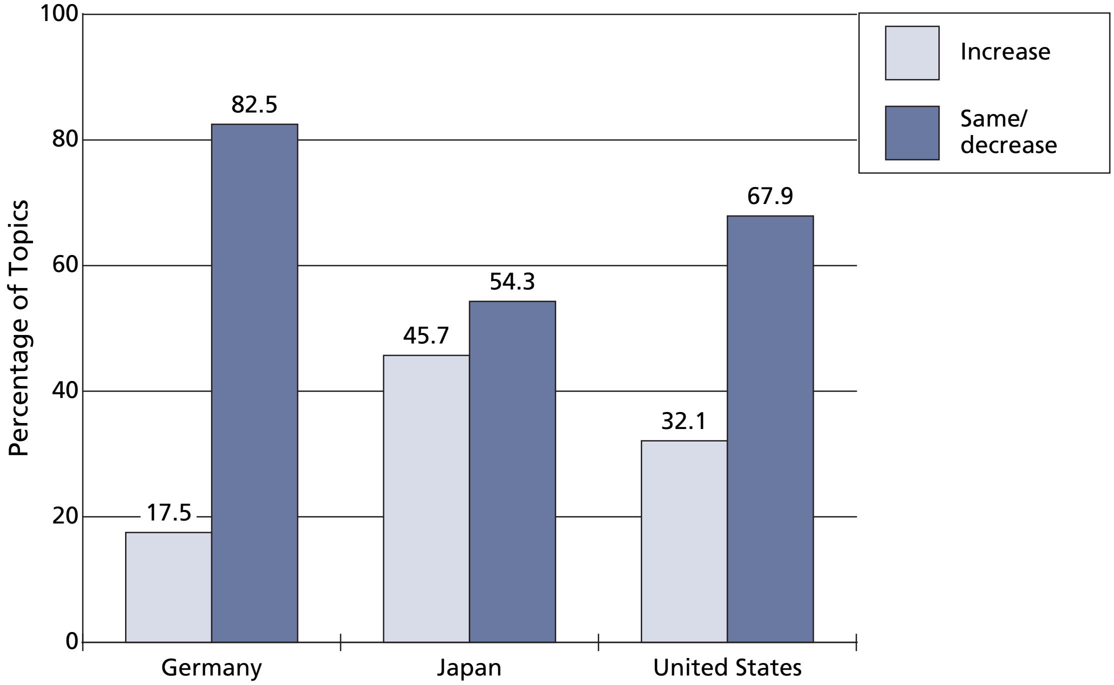
Figure 20
Drawing from chalkboard of second problem in US-018



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

The results of coding this distinction are shown in figure 21. The Japanese applications were significantly more likely to increase in complexity than those in Germany. There was no difference in this regard between the United States and the other two nations.

Figure 21
Average percentage of topics in each lesson that contained applications that increased in complexity vs. stayed the same or decreased over the course of the lesson



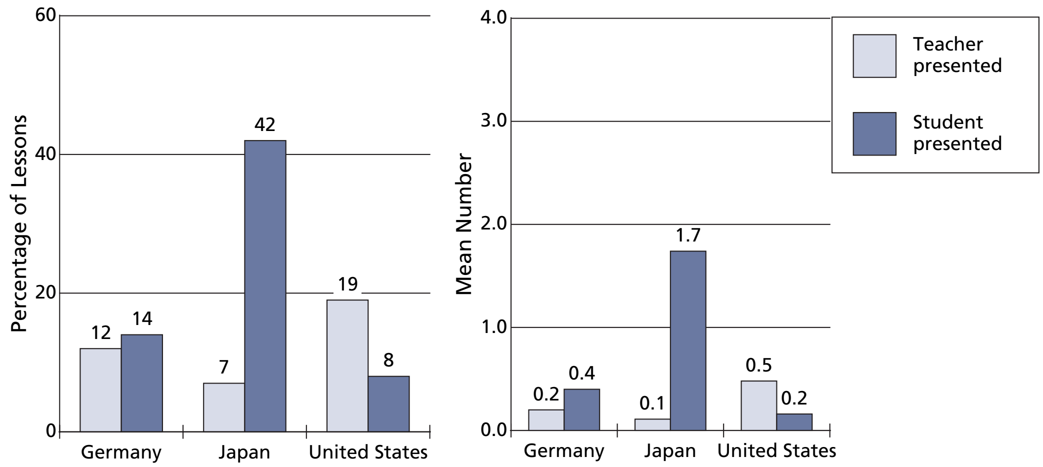
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Alternative Solution Methods

Often the teacher's goal is to teach students how to solve a specific type of mathematics problem (e.g., an algebraic equation or a geometric construction). This can be accomplished by presenting one solution method and asking students to use it on similar problems or by encouraging the development of different methods and examining their relative advantages. We coded whether an alternative solution method was presented by the teacher, or by students, during the course of each lesson.

Panel (a) of figure 22 shows the percentage of lessons that included alternative solution methods of each type; Panel (b) shows the average number of alternative solution methods of each type presented in the lessons of the three countries. U.S. lessons included significantly more teacher-presented alternative solution methods than did Japan. Japanese lessons included significantly more student-presented alternative solution methods than did either German or U.S. lessons.

Figure 22
(a) Percentage of lessons that included teacher-presented and student-presented alternative solution methods; (b) average number of teacher- and student-presented alternative solution methods presented per lesson



(a) **(b)**

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Principles, Properties, and Definitions

As described earlier, when describing the content of each lesson we not only recorded tasks and situations but also the PRINCIPLES, PROPERTIES, and DEFINITIONS (PPD) that were stated in the lesson. Although these PPDs filled a relatively small percentage of lesson time, they nevertheless could be crucial for moving along the content of the lesson.

We were not able to reliably differentiate PRINCIPLES and PROPERTIES, but we were able to differentiate between these and DEFINITIONS. To be coded as a DEFINITION, a statement of mathematical information had to include a general statement (generic term) followed by the defining characteristics or properties. For example: A ray is a straight line (generic term) that has a beginning point but no end point (defining characteristics). A statement that was not a complete definition in this sense was coded as PRINCIPLE/PROPERTY. Any mathematical information stated in the lesson that was not coded as a DEFINITION was coded as PRINCIPLE/PROPERTY.

An example of a DEFINITION is seen in the following excerpt from GR-022. The topic of this lesson is congruence of triangles. Here is a part of the lesson transcript:

00:01:37 T We are dealing with the congruence theorems. Who can say once more what congruence- or congruent triangles are? (...)
00:01:54 S //(Well). Congruent triangles are triangles uhm which one - if one puts them on top of each other uhm they do not intercept. Right?
00:02:03 T Yes. Are in exact accordance is what you (wanted) to say. Right? (...)

Another example is from the Japanese lesson JP-035, which deals with the topic of similarity of geometric figures. The teacher writes the word "similarity" on the board, then says:

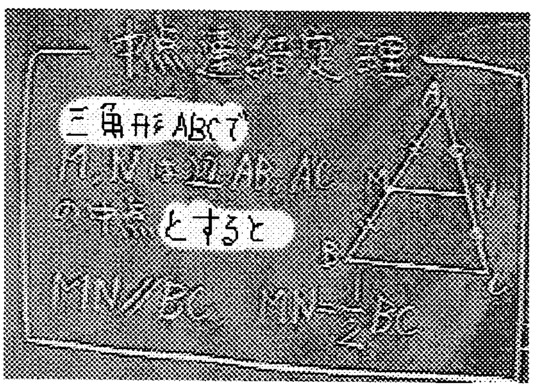
00:03:58 T We use the word similar. And you read Makoto. Okay from now I'm going to write the definition of similarity on the board ... so please prepare your notebooks.

00:04:24 B Similarity means that the figure which size is expanded or reduced is similar to the original figure.

00:07:24 T Okay. Then next I'm going to talk ... all right? What similarity means is that the figure whose size is expanded or reduced is similar to the original figure.

An example of PRINCIPLE/PROPERTY is from JP-039, which dealt with parallel lines and similarity. The teacher first reviewed four theorems on this topic that were covered in a previous lesson, then introduced the fifth theorem, namely, the midpoint connection theorem (figure 23). The theorem is clearly stated on the chalkboard. (Some of the writing on the chalkboard has been digitally enhanced to improve readability.)

Figure 23
Excerpt from chalkboard from JP-039, with English translation



Theorem of midpoint connection:

In the triangle ABC, if we make M and N the midpoints of lines AB and AC respectively,

MN//BC
MN = ½ BC

NOTE: Some of the writing on the chalkboard has been digitally enhanced to improve readability.

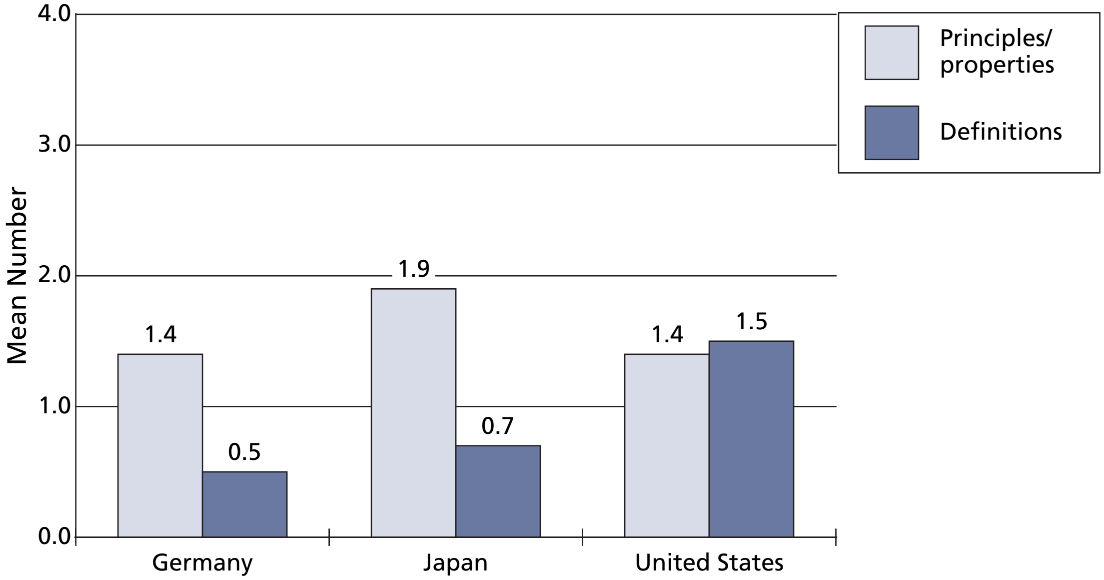
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

A few other examples of PRINCIPLES/PROPERTIES are found in the following teacher statements:

- JP-001: The sum of the interior angles of a quadrilateral is 360°.
- JP-003: Vertically opposite angles are equal.
- JP-012: The triangles between two parallel lines have the same areas.
- JP-022: In a triangle, the three median lines always intersect at one point. This is called the center of gravity of the triangle.

The average number of PRINCIPLES/PROPERTIES and DEFINITIONS in each lesson is graphed in figure 24. There was no difference across countries in the number of PRINCIPLES/PROPERTIES per lesson. U.S. lessons, however, contained more DEFINITIONS per lesson than did German lessons. The U.S. and Japan, and Germany and Japan, did not differ from each other in this regard.

Figure 24
Average number of principles/properties and definitions in each German, Japanese, and U.S. eighth-grade mathematics lesson



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Proofs

Constructing proofs is an important mathematical activity because it provides a reasoned method of verification based on the accepted assumptions and observations of the discipline. Many reform documents, for example, those by the National Council of Teachers of Mathematics (1989, 1991), recommend that students should have increasing opportunities to examine and construct mathematical proofs. We coded a lesson as including a proof if an assumption was presented, a proof executed, and the assumption confirmed as correct. The proof could be presented by the teacher, by a student, or worked out collaboratively during classwork. As long as an assumption was presented and the strategy for proving it was discussed, it was still considered a proof, even if the proof was not executed. Likewise, if a proof was started but not completed due to running out of time in the lesson, we still coded the lesson as including a proof.

Analysis revealed that a greater percentage of the Japanese lessons included proofs than either the German or U.S. lessons. Indeed, 10 percent of German lessons included proofs while 53 percent of Japanese lessons included proofs. None of the U.S. lessons included proofs.[1]

[1] Standard errors for German, Japanese, and U.S. lessons are 4.32, 3.45, and 0, respectively.

FINDINGS OF THE MATH CONTENT GROUP

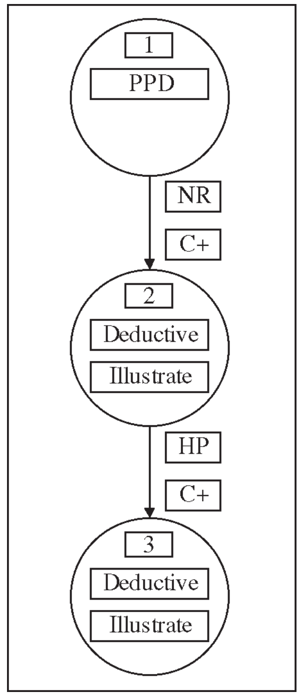
We turn now to present the findings of the independent Math Content Group. Recall that this group analyzed the content of 30 lessons from each country, 15 algebra and 15 geometry. The Math Content Group based its analyses on the detailed descriptions of mathematical content contained in the lesson tables, as previously described. To reduce the likelihood of bias, tables were disguised (e.g., references to currency and other country-specific contents were altered) so that it was not possible to tell for certain which country the lessons came from. After analyses were complete, results were then tabulated by country.

Methods of Analysis

Content descriptions in the lesson tables were subjected to a detailed series of analyses. The first step was to construct a directed graph representation of each lesson. The purpose of the directed graph was to show the content and flow of the lesson in shorthand notation so that patterns within and among lessons would become more apparent. In this graph, the content of each lesson was represented as a set of nodes (depicted by circles) and links among the nodes (depicted by arrows), depending on relationships existing among the nodes. Both nodes and links were then labeled according to the coding system developed by the group.

Definitions of each code will be presented later. For now, however, it is useful to go through one lesson in detail, showing how the Math Content Group transformed the content description in the table into a directed graph. For convenience, we will use a Japanese lesson (JP-012) to illustrate this process. The lesson table for JP-012 was presented earlier. The directed graph produced by the Math Content Group is shown in figure 25.

Figure 25
Directed graph representation of a Japanese lesson (JP-012) as constructed by the Math Content Group



NOTE: PPD = Principle/Property/Definition; NR = Necessary Result; C+ = Increase in Complexity; HP = Helpful Process.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

For purposes of content analysis, lesson JP-012 was divided into three content segments. The homework segment at the end was not counted as a separate segment. Each segment became a node on the graph. We will now review the construction of the directed graph segment by segment.

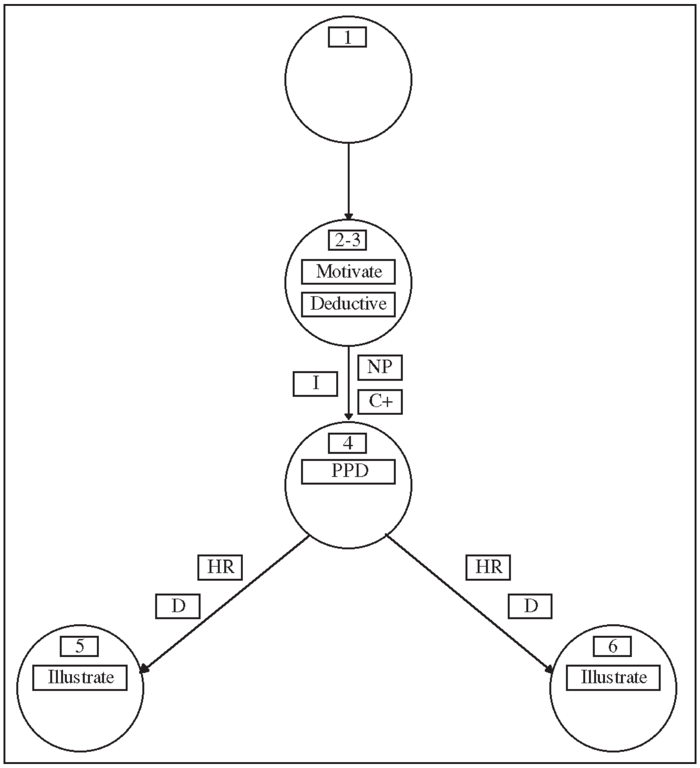
1. *First node [(00:27)-(01:26)].* This node consists of a PPD and is identified in this way by the coders. It is worth noting that the PPD is not only stated but also illustrated by several examples using a computer.
2. *Second node [(01:26)-(22:57)].* In the second node a problem is posed and then students work on the problem for several minutes. After that, students present two solutions to the problem. The solutions use the mathematical principle presented in the first node and involve explicit use of deductive reasoning in applying that principle: "Since the areas of the triangles..." (19:20). Because the earlier node has a result necessary for the content of this one, this node is directly connected to the first node by a link labeled NR (Necessary Result). Because the principle of the first node is used to solve the problem of the second node, the second node is labeled "illustrate," as the situation/task in the

second node is used to illustrate the earlier principle. The explicit reasoning mentioned earlier leads to labeling this node "deductive."

3. *Third node [(22:57)-(48:58)].* In the third node another problem is posed whose solution depends upon the mathematical principle of the first node. It is, therefore, another illustration of that principle, and the third node too is labeled "Illustrate". This problem is similar to, but more complex than, the problem of the second node. Thus, there is a link between the second and third nodes labeled "C+" to indicate a more complex conceptual setting, and a link labeled "HP" to indicate that the procedures used in the second node are helpful in solving the problem set in this one. This node was labeled "deductive" because the same reasoning had to be used as in the second node. Because the third node is indirectly linked to the first one through the second, and because the nature of the linkage between the first and the third is indicated by the two links given, there was no need to provide a direct link from the first node to the third.

The directed graphs took a number of different forms. One additional example is shown in figure 26. Our purpose in presenting this graph is simply to illustrate the general level of complexity of the graphs. We will forgo a detailed explanation of exactly how this graph was derived.

Figure 26
Additional example of directed graph produced by the Math Content Group



The first node leads into the second and third.

The second and third nodes include deductive reasoning and provide clear motivation for a PPD (principle, property, or definition) that occurs later in the lesson.

Nodes two and three were linked to node four in three ways: (1) by inductive reasoning (I); (2) a process was explained in nodes two and three that was necessary for understanding the content of node four (NP); and (3) there was an increase in complexity in tasks and situations between the nodes (C+).

The fourth node contains a PPD, as well as a result that is helpful (HR) for both nodes five and six. The content of nodes five and six are derived deductively (D) from node four. And, nodes five and six both contain illustrations of PPDs presented earlier in the lesson.

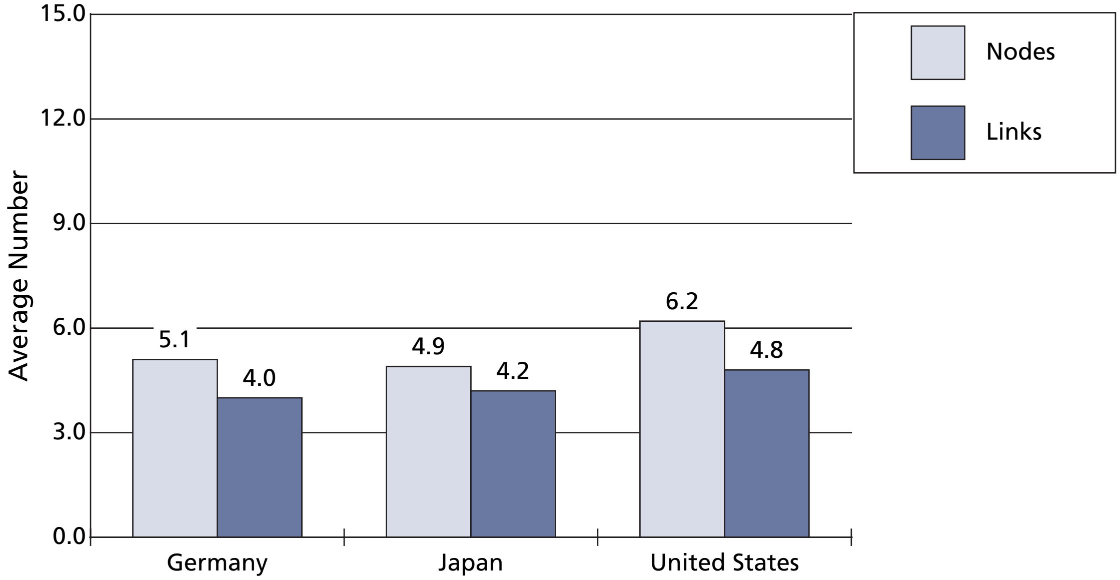
NOTE: PPD = Principle/Property/Definition; HR = Helpful Result; C+ = Increase in Complexity; NP = Necessary Process; I = Inductive Reasoning; D = Deductive Reasoning.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Analyses of the Directed Graphs

The first analysis based on the directed graphs was simply to count the number of nodes and links of each lesson. The average number of nodes and links for lessons in each country are presented in figure 27. There was no significant difference in the number of nodes or links across countries.

Figure 27
Average number of nodes and links on the directed graph representations of lessons in each country



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

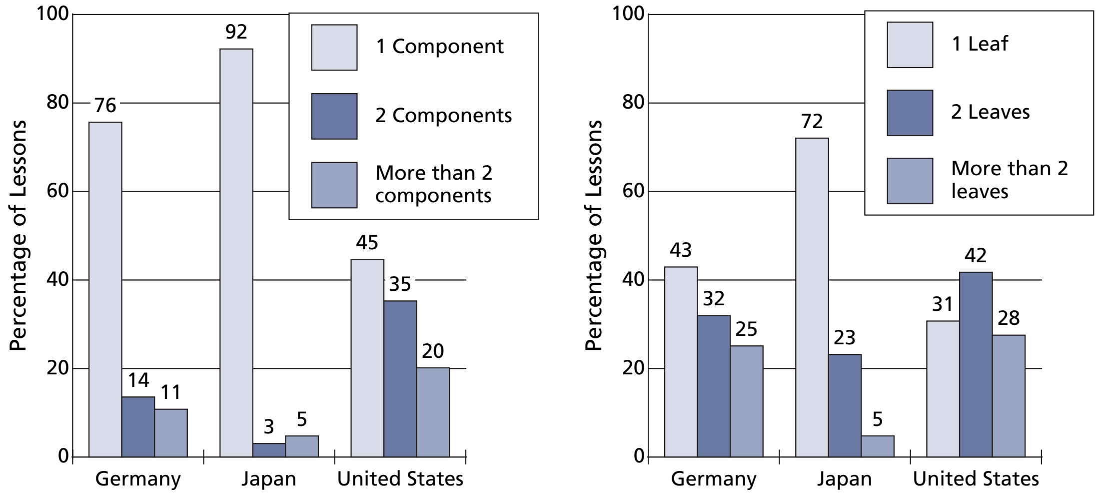
When we examine the structure of the directed graphs, however, we begin to see cross-national differences. Two indicators in particular attracted our interest:

- The number of components in each graph, defined as the number of disconnected parts of the graph. If all nodes on a graph were connected by at least one link path then the lesson was coded as having one component.
- The number of leaves in each graph, defined as the number of nodes that were touched by only a single link.

These indicators seem to measure the coherence of the lesson content because they measure the interconnectedness of different content segments.

The distribution of lessons by number of components and number of leaves is shown in figure 28. The distribution in Japan was significantly different than that in the United States. For example, Japanese and German lessons were more likely than U.S. lessons to contain only one component. Japanese lessons were also more likely than lessons in the United States to contain one leaf. This suggests that the content of lessons is more coherent in Japan than in the United States.

Figure 28
(a) Percentage of lessons that included one, two, or more than two components; (b) percentage of lessons that included one, two, or more than two leaves



NOTE: Percentages may not sum to 100 due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Further Analyses of Nodes and Links

Having examined the number of nodes and links and the structure of how nodes are connected in the lessons of each country, we turn now to examine the characterizations of nodes and links provided by the Math Content Group. A number of characteristics were coded for nodes and for links. For nodes, the following codes were applied when the group of coders judged them present:

- MOTIVATION—applied when a task or situation in the node was clearly used to motivate a PPD that occurred later in the lesson. (Marked as M on the directed graphs.)
- ILLUSTRATION—applied if the node included a task, situation, or activity that clearly illustrated a general principle that was explicitly stated earlier in the lesson. (Marked as L on the directed graphs.)
- DEDUCTIVE REASONING—applied if there was an explicit use of deductive reasoning within the node. (Marked as D on the directed graphs.)
- INDUCTIVE REASONING—applied if there was an explicit use of inductive reasoning within the node. (Marked as I on the directed graphs.)
- INCREASE IN COMPLEXITY—applied only to nodes that included more than one task or situation when there was judged to be an increase in the complexity of the tasks or situations. (Marked as C+ if increase was clearly and primarily

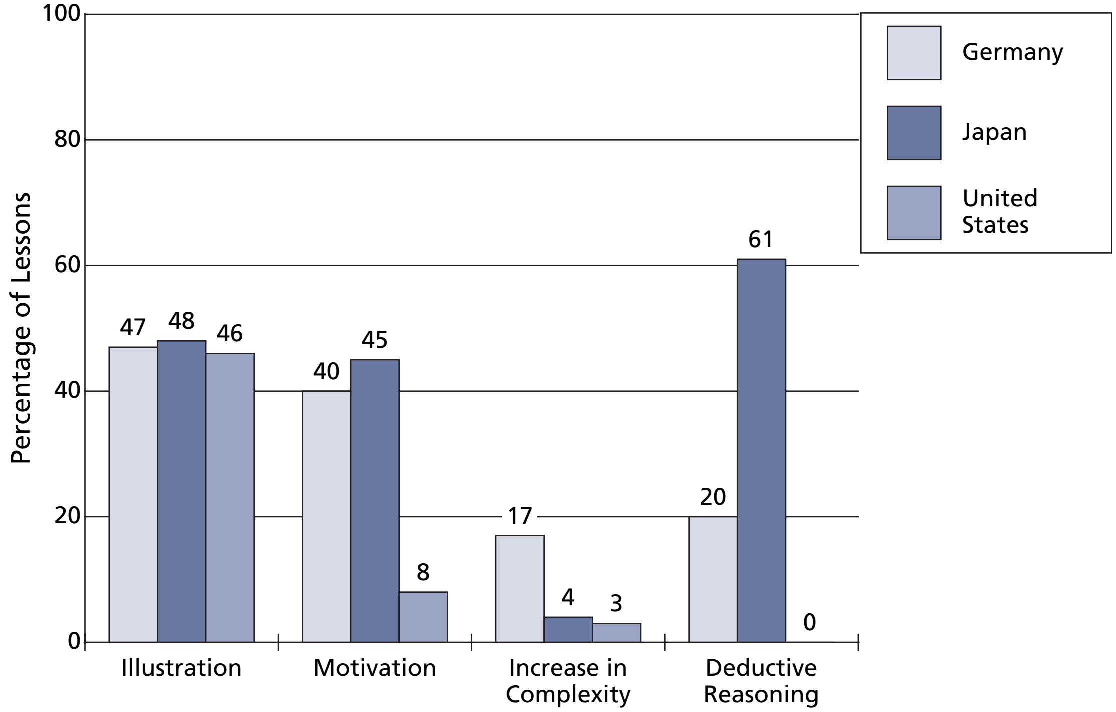
conceptual, P+ if clearly and primarily procedural. Otherwise, marked simply as +.)

For links, the following codes were applied:

- NECESSARY—applied when the content of the earlier node on a link was judged necessary in order to take up the content in the later node. The earlier node could be necessary because it provided a result that was used in the later node (marked as NR), or because it described or explained a process that was applied in the later node (marked as NP). The absence of this code might indicate a gap or discontinuity in the development of content.
- HELPFUL—applied when the content of the earlier node was clearly helpful, though not necessary, for presenting or understanding the content of the later node. Again, it could be either a result (marked as HR) or a process (marked as HP) that rendered the earlier node helpful.
- SIMILAR—applied when a process (marked as SP), result (marked as SR), or central concept of the later node was similar to a process or result in the earlier node.
- DEDUCTIVE AND INDUCTIVE REASONING—coded when either deductive (marked as D) or inductive (marked as I) reasoning was a significant component of the connection between the linked nodes.
- INCREASE IN COMPLEXITY—coded when there was an increase in complexity in tasks or situations from one node to the next. (Marked as C+ if increase was clearly and primarily conceptual, P+ if clearly and primarily procedural. Otherwise, marked simply as +.)

The percentage of lessons that included nodes coded as ILLUSTRATION, MOTIVATION, INCREASE IN COMPLEXITY, or DEDUCTIVE REASONING is shown in figure 29. (Few nodes were coded as INDUCTIVE REASONING, and, therefore, were not included in the following analysis.) There was no significant difference across countries in the percentage of lessons containing ILLUSTRATION or INCREASE IN COMPLEXITY nodes. However, Germany and Japan had significantly more lessons containing MOTIVATION nodes than did the United States. And Japan had the largest percentage of lessons containing DEDUCTIVE REASONING nodes, while the United States had the smallest percentage.

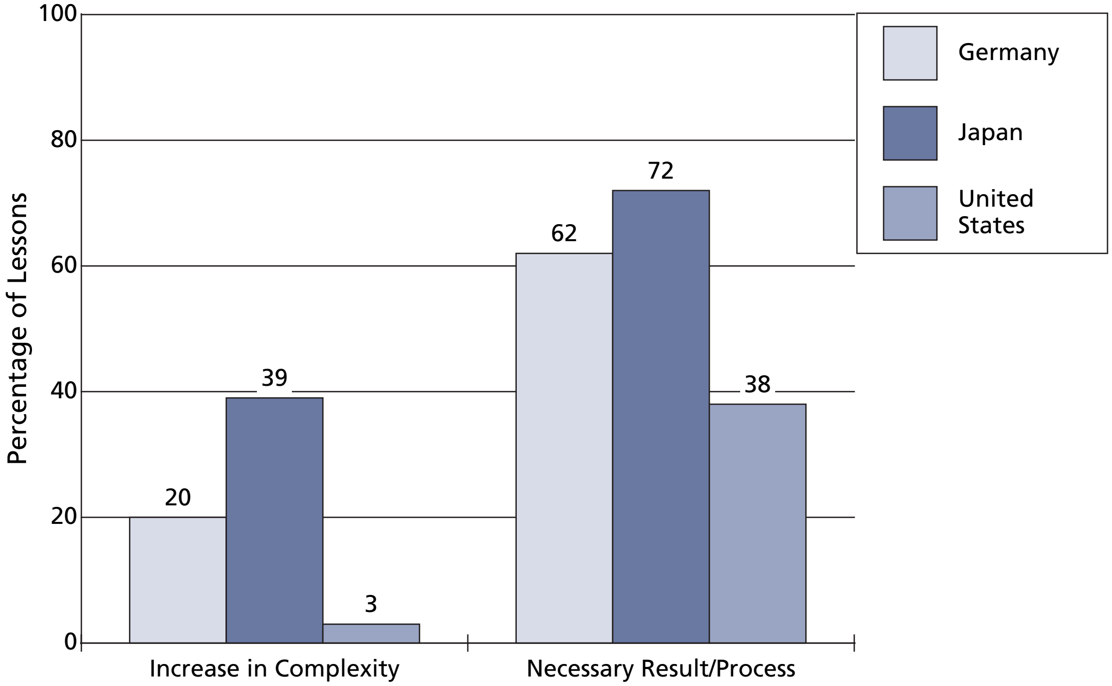
Figure 29
Percentage of lessons with nodes coded to include illustrations, motivations, increase in complexity, and deductive reasoning



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Some indication of how content was developed over the course of each lesson was gained by coding the kinds of links that connected parts of the lesson together (see figure 30). Increasing complexity between two nodes was coded more often in Japanese and German lessons than U.S. lessons. Japanese lessons contained significantly more links coded as necessary than did U.S. lessons.

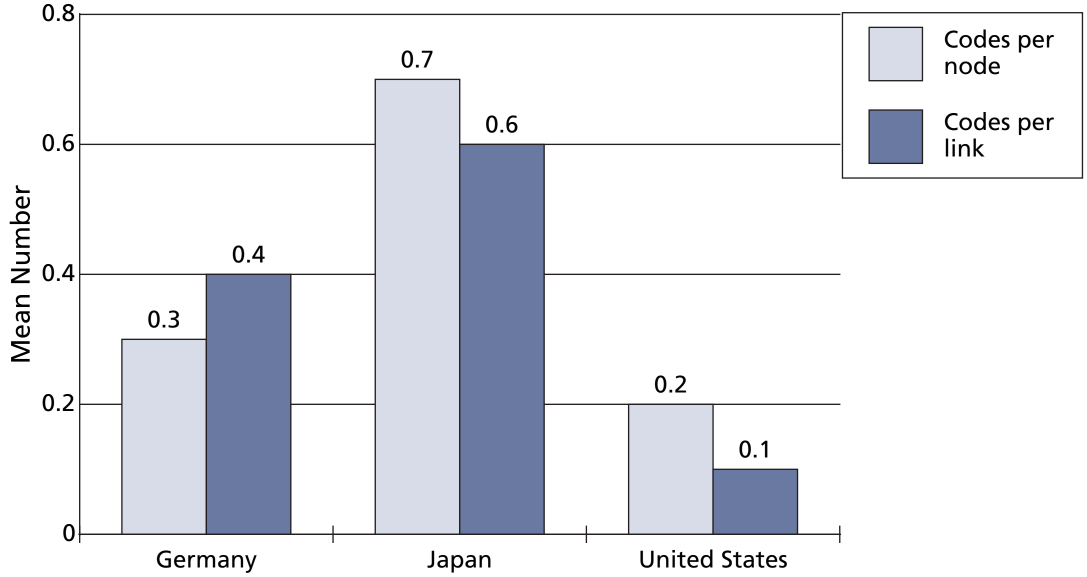
Figure 30
Percentage of lessons containing links coded as increase in complexity and necessary result/process



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

An overall summary of the Math Content Group's coding can be obtained by adding up the total number of positive characteristics coded for each lesson. Thus, for each directed graph, we simply added up the number of codes that were attached to nodes (i.e., Motivation + Illustration + Deductive Reasoning + Inductive Reasoning + Increase in Complexity), and the number of codes that were attached to links (i.e., Necessary + Helpful + Similar + Deductive/Inductive Reasoning + Increase in Complexity). Figure 31 shows the average number of codes per node and per link for lessons in the three countries. Japanese lessons contained significantly more codes per node than either German or U.S. lessons; and U.S. lessons contained significantly fewer codes per link than lessons in the other two countries.

Figure 31
Average number of codes per node and per link in German, Japanese, and U.S. lessons



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

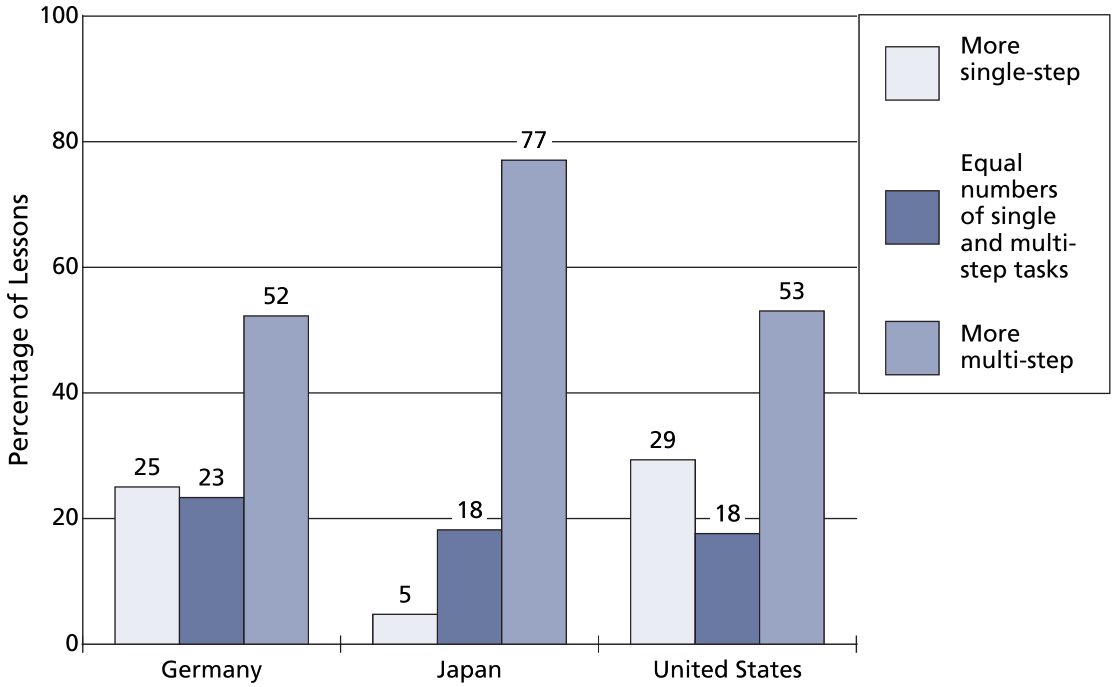
Additional Coding of Tasks

The Math Content Group developed two more codes to indicate the kinds of tasks which were engaged in during the lesson.

The first of these additional codes was Task Complexity. Each task was categorized as either single-step or multi-step. Lessons were then categorized as containing mostly single-step tasks, equal number of single- and multi-step tasks, or mostly multi-step tasks. The results are shown in figure 32. None of the pairwise differences was significant.[2]

[2] On a three-point scale where 1 indicates "More single-step" and 3 indicates "More multi-step," the averages (with standard errors) for German, Japanese, and U.S. lessons were 2.3 (0.17), 2.7 (0.11), and 2.2 (0.19), respectively.

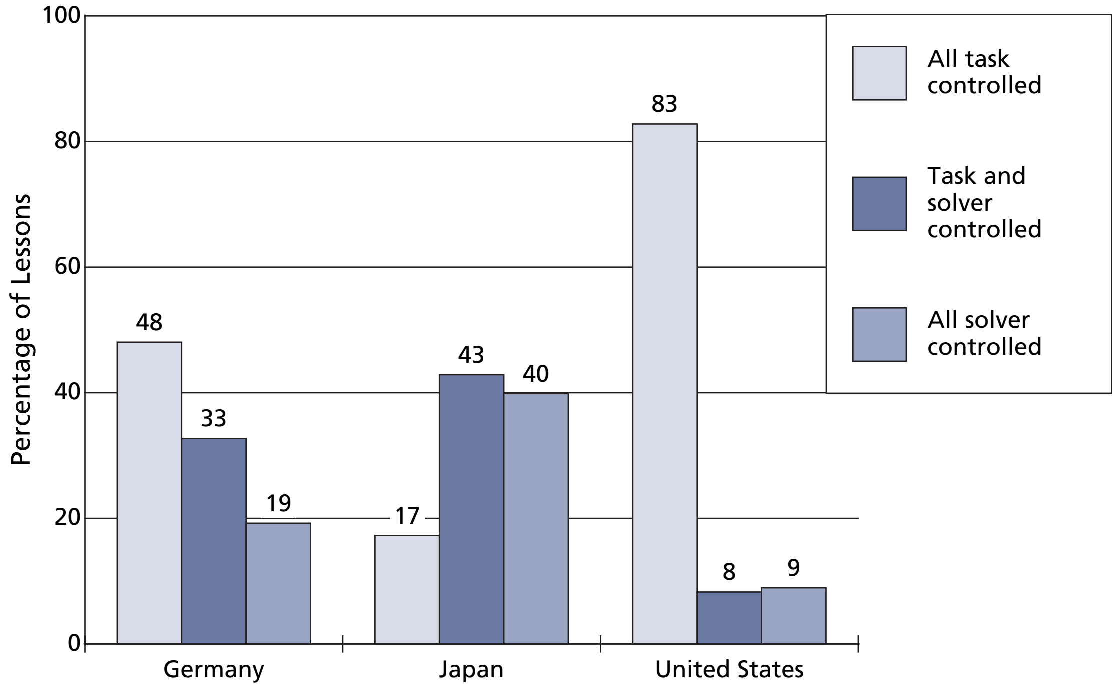
Figure 32
Percentage of lessons in each country containing mostly single-step, mostly multi-step, or equal numbers of the two types of tasks



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

The second coding of tasks was for what the Math Content Group called LOCUS OF CONTROL. What level of choice did students have in determining how to perform the task? Were they all controlled by the task, or was some of the control left up to the student? For example, if the teacher had just demonstrated how to solve a problem, then asked the students to try applying the same method to a similar problem, it was coded as TASK CONTROLLED. This is because the students were not asked to make any decisions about how to approach the problem, only to follow the exact procedure demonstrated by the teacher. On the other hand, if the teacher asked students to see if they could think of another method for solving a problem it was coded as SOLVER CONTROLLED, because the student had the freedom to decide which of several possible approaches they would take. In figure 33 we show the percentage of lessons that contained all Task Controlled tasks, all Solver Controlled, or a mixture of the two. Seventeen and 48 percent of Japanese and German lessons contained all task controlled tasks, respectively, while the share was 83 percent for U.S. lessons.

Figure 33
Percentage of lessons containing task controlled tasks, solver controlled tasks, or a combination of task and solver controlled tasks



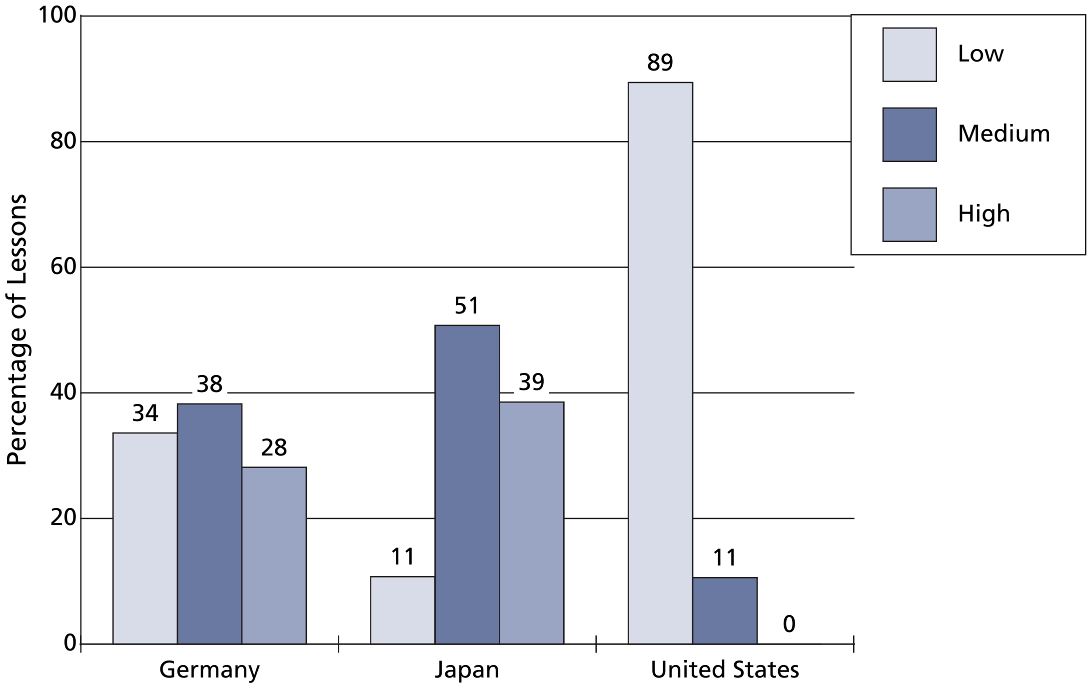
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Global Ratings of Quality

In addition to constructing the directed graph representation of each lesson and coding the nodes and links, the members of the Math Content Group also assigned a rating to each lesson that reflected their overall judgment of the quality of the mathematical content in the lesson. They rated each lesson on a three-point scale: low, medium, or high. (Where raters disagreed on this judgment, the disagreement was resolved by discussion so that, in the end, all four raters agreed.) Although the measure is subjective, there was high agreement among the independent raters. A summary of these ratings is presented in figure 34. The distributions of ratings differed significantly between the United States and the other two countries; German and Japanese lessons were rated higher in quality than U.S. lessons.[3] Twenty-eight and 39 percent of German and Japanese lessons, respectively, received the highest rating; none of the U.S. lessons did. Eighty-nine percent of U.S. lessons received the lowest rating.

[3] On a three-point scale where 1 indicates "low" and 3, "high" quality, estimates (and standard errors) for German, Japanese, and U.S. lessons were 1.9 (0.14), 2.3 (0.12), and 1.1 (0.06), respectively.

Figure 34
Percentage of lessons rated as having low, medium, and high quality of mathematical content



NOTE: The numbers in this graph differ slightly from those reported in Peak (1996, page 45) where unweighted averages were mistakenly used instead of weighted averages. Specifically, the percentages of lessons rated as low, medium, and high quality were reported as 40, 37, and 23 for Germany; 13, 57, and 30 for Japan; and 87, 13, and 0 for the United States, respectively. The numbers shown in this graph are the correct ones. Percentages may not sum to 100 due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

How well do the specific codes we have described relate to these more global judgments? Correlations were calculated between specific indicators constructed by the Math Content Group and the group's overall ratings of content quality. The single highest predictor (Pearson r = .66) was the total number of codes on both nodes and links in a lesson divided by the total number of nodes in the lesson. The average of this index for Japanese lessons was 1.17, for German lessons, .60, and for U.S. lessons, .25. The differences between all pairs of countries were statistically significant.[4]

In summary, we found a number of differences in the mathematical content of lessons across the three countries. Japanese lessons, in general, were found to be more advanced by international standards, and richer in content on several dimensions, than were U.S. lessons. German lessons tended to fall between the Japanese and U.S. lessons on some dimensions, or differed from the United States on one measure and from Japan on the next. We move, now, to consider how teachers presented content to students in the three countries.

[4] Standard errors for Germany, Japan, and the U.S. are 0.07, 0.16, and 0.05, respectively.

Chapter 4. The Organization of Instruction

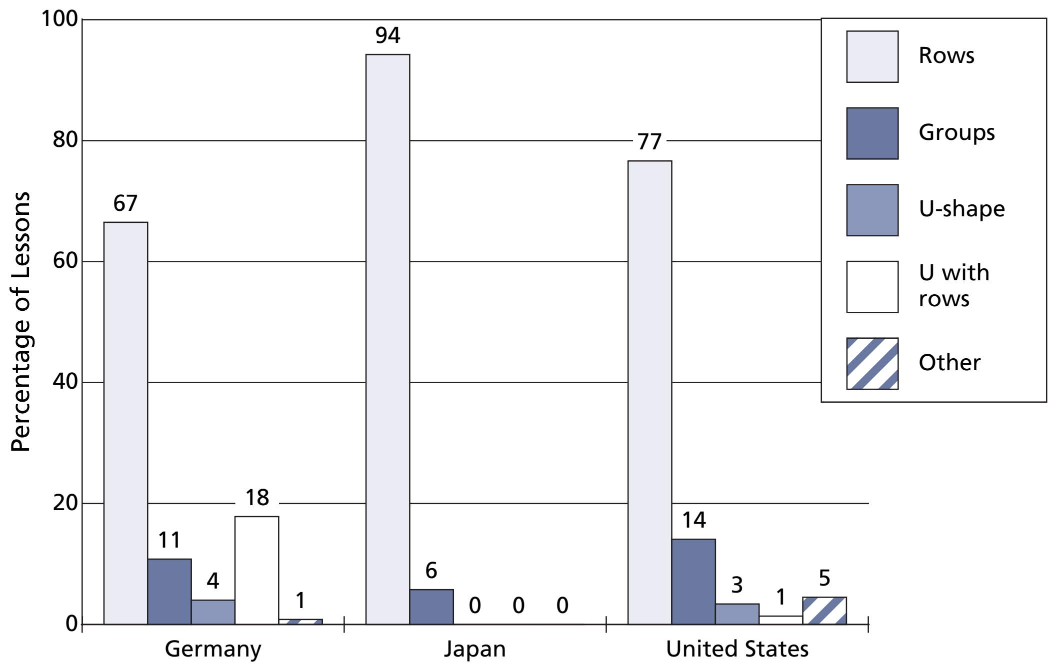
Having described the mathematical content of lessons, we turn now to examine more fully the methods by which teachers structure lessons so as to engage students with the content. It would seem that without rich mathematical content, students would be unlikely to learn mathematics. Yet even with content that is of high quality, there are many options available to teachers as they begin the practical task of constructing classroom lessons. In this section we will begin to describe the kinds of lessons teachers construct in the three countries.

CHARACTERISTICS OF THE CLASSROOM

An initial viewing of the videotapes suggests that the similarities among the eighth-grade classrooms we visited were more striking than the differences. Although we were videotaping in three highly distinct cultures, there was nevertheless considerable similarity in classroom arrangement in all countries. In all three nations, classrooms typically contained one teacher and many students. All of the classrooms we visited contained chalkboards, and all contained individual desks for each student.

The arrangement of desks in the classrooms varied somewhat by country (see figure 35). In Japan, 94 percent of lessons had desks arranged in rows facing the front; the share was 77 percent in the United States and 67 percent in Germany. A substantial number of classrooms in Germany (22 percent) had desks arranged in a U-Shape configuration, with the open end of the U facing the front. (Most often, the inside of the U contained rows of desks facing the front of the room.) Some classrooms in each country had desks arranged in groups.

Figure 35
Arrangement of desks in German, Japanese, and U.S. classrooms



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

BASIC CHARACTERISTICS OF THE LESSON

Our first step in coding the videotapes was to mark the points at which the lesson itself began and ended. Videographers made a judgment about when to turn on the camera, but this was usually before the beginning of instruction proper. Similarly, the camera kept on rolling for a few minutes after the teacher ended the lesson. We needed a consistent way of coding the beginning and ending of the lesson because many of our outcome measures (e.g., duration of time devoted to some activity) are reported as a percentage of total lesson time. Obviously, these measures will be affected by the total measured duration of the lesson.

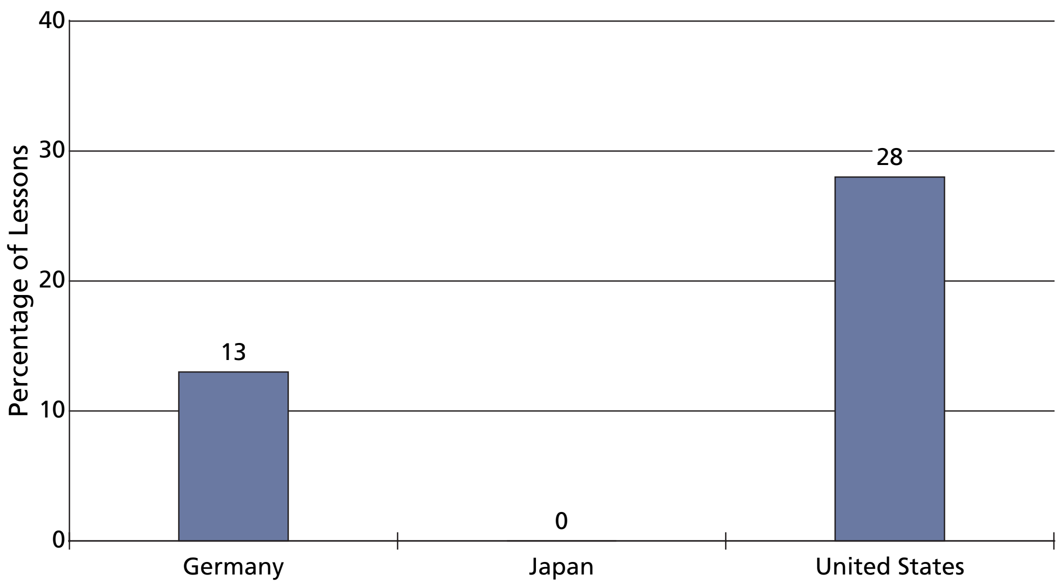
We divided each class into three segments: Pre-Lesson Activity, Lesson, and Post-Lesson Activity.

- PRE-LESSON ACTIVITY comprises all nonmathematical activity or talk that occurs before the lesson proper begins. Examples of such activity include greetings, teacher announcements about extracurricular activities, changes in school schedule, past and future exams, or housekeeping. In addition, talking about previously assigned homework in general terms, such as grades, points, etc., without an elaborated explanation would also be part of PRE-LESSON ACTIVITIES;
- LESSON is defined as the period of time that is allocated by the teacher for mathematical-instructional activities. The teacher generally indicates the beginning of the lesson with an explicit verbal marker, such as “OK, today we are going to...” Likewise, the end of the lesson is marked as well. Often there are bells that coincide with the beginning and end of the lesson; and
- POST-LESSON ACTIVITY is defined as all nonmathematical activities that occur after the end of the LESSON and before the end of CLASS. These segments include teacher announcements about homework without elaboration, directions to clean up desks and chairs, and other housekeeping activities.

The average duration of the lessons in our sample (subtracting out time devoted to pre- and post-lesson activities) was 43.2 minutes in Germany, 49.5 minutes in Japan, and 49.4 minutes in the United States. German lessons were significantly shorter than those in Japan or the United States. In the United States the shortest lesson was 32.0 minutes, the longest, 91.2 minutes. The range in Germany was 34.2 to 49.4 minutes, and in Japan, 43.4 to 54.9 minutes.[1]

[1] Standard errors of the average duration of lessons in Germany, Japan, and the United States were 0.24, 0.35, and 1.92, respectively.

Figure 36
Percentage of lessons with at least one outside interruption



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

In addition to the starting and ending points of the lesson, we also coded whenever there were interruptions from outside the classroom during the lesson. This included such events as announcements over the public address system or visitors who interrupted the lesson. These results are presented in figure 36. Such interruptions were never observed during the Japanese lessons but were relatively common during the U.S. lessons. Interruptions were significantly more likely to occur in U.S. and German lessons than in Japanese lessons.

ORGANIZATION OF THE LESSON

Classwork and Seatwork

Having marked the beginning and end of each lesson, our next step was to divide the lessons into organizational segments. Although in many respects lessons look quite different across the three cultures, teachers everywhere tend to divide their lessons into periods of classwork and periods of seatwork, and it is not difficult to reliably code the beginnings and ends of these segments. We identified three distinct kinds of segments:

- CLASSWORK segments (coded as CW) are defined as those times when the teacher is working with all (or most) of the students in a whole-class situation; the type of talk is predominantly public, that is, the audience is the whole class. Possible activities during CLASSWORK include the teacher and students engaged in learning a new concept, reviewing a previously learned concept, solving mathematical problems together, or sharing solutions to problems they are solving;

- SEATWORK (coded as SW) is defined as a period of time during the lesson when students work independently on assigned tasks, either alone or in small groups (two students or more). The type of talk is predominantly private, but there are instances of public talk when the teacher makes announcements to the whole class. The beginning of SEATWORK is usually marked by one or more of the following: (1) The teacher announces that students should begin their independent work; (2) There is a period of silence after the teacher provides necessary information to students, or (3) students actually start working; and
- CLASSWORK/SEATWORK COMBINATION segments are the rare cases in which the teacher assigns some students to work independently on a particular task and situation, while the rest of the class works with the teacher. In such cases, the Organization of Interaction is coded as CW/SW.

Organizational segments were coded exhaustively, meaning that the end of one segment was the beginning of the next. Seatwork segments were further characterized as being individual, group, or both. During individual seatwork segments students worked independently, by themselves; during group seatwork segments they worked in groups.

Results indicate that Japanese lessons contain considerably more organizational segments on average than do German or U.S. lessons (figure 37).

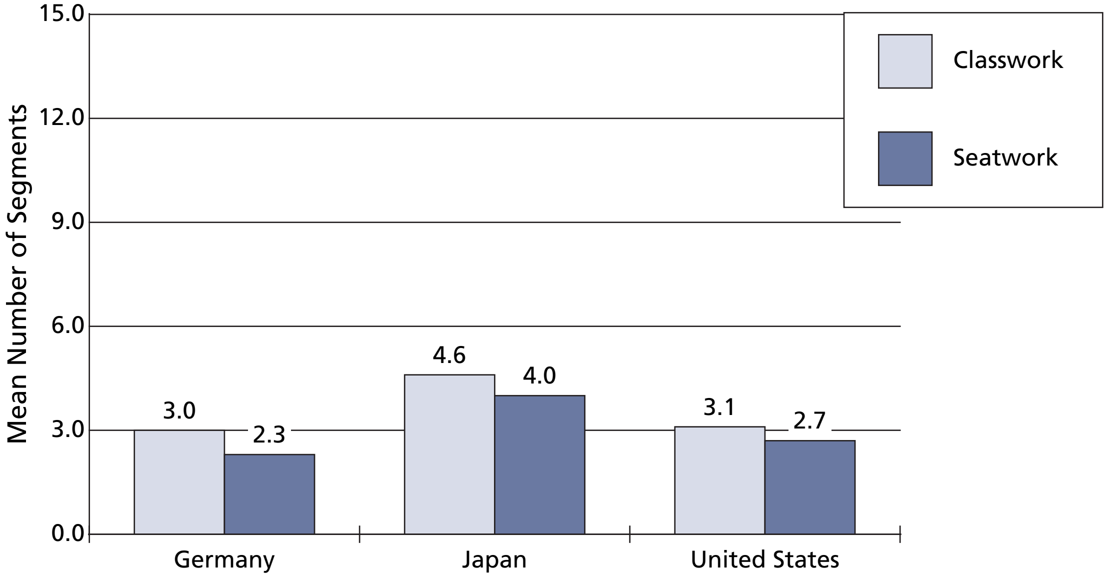
Figure 37
Average number of organizational segments in German, Japanese, and U.S. lessons

	Mean
Germany	5.5
Japan	8.7
United States	5.9

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

As it turned out, very few of the segments in any country were coded as classwork/seatwork combinations, and these segments accounted for less than 1.5 percent of lesson time, on average. The bulk of the segments were either classwork or seatwork. Figure 38 shows the average number of classwork and seatwork segments per lesson in each country. Again, Japanese lessons contain the highest average number of segments of both types. There was no significant difference between U.S. and German lessons in this regard.

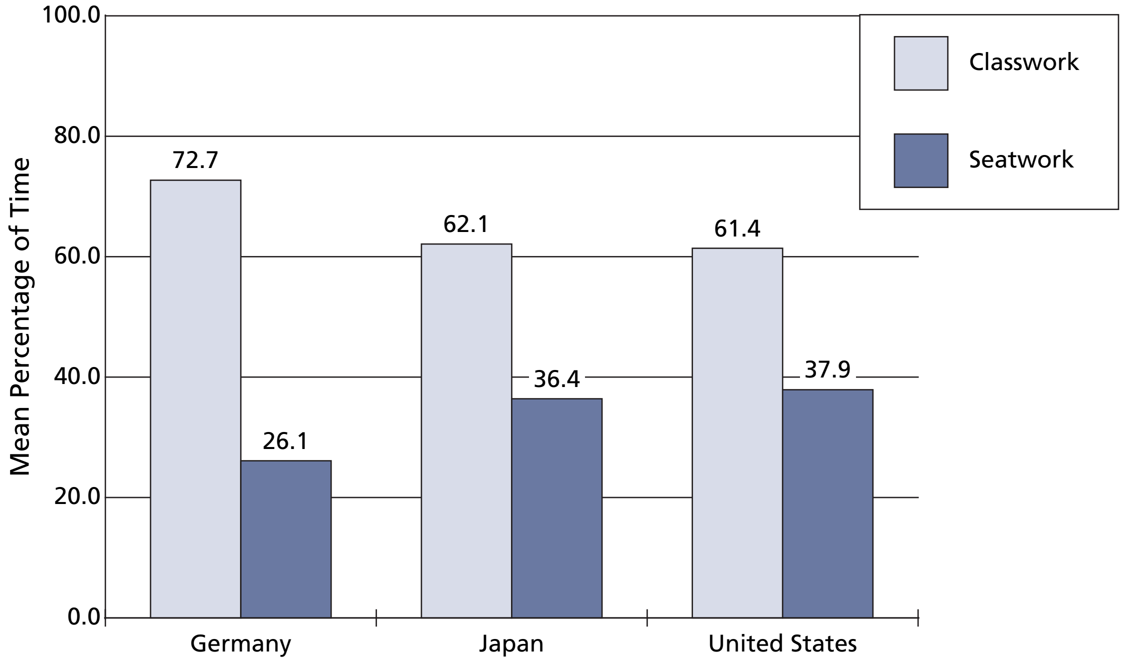
Figure 38
Average number of classwork and seatwork segments per lesson in each country



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

The picture changes somewhat if we look at the percentage of time during the lesson that is spent in classwork and seatwork (figure 39). Here, Japan and the United States look quite similar, and Germany looks different. Japanese and U.S. teachers spend almost identical percentages of time in classwork (approximately 60 percent) and in seatwork (approximately 40 percent). German teachers spend a higher percentage of their lesson time in classwork (about 70 percent) than do teachers in the other two countries. Conversely, Japanese and U.S. teachers spend a higher percentage of their lesson time in seatwork than do German teachers.

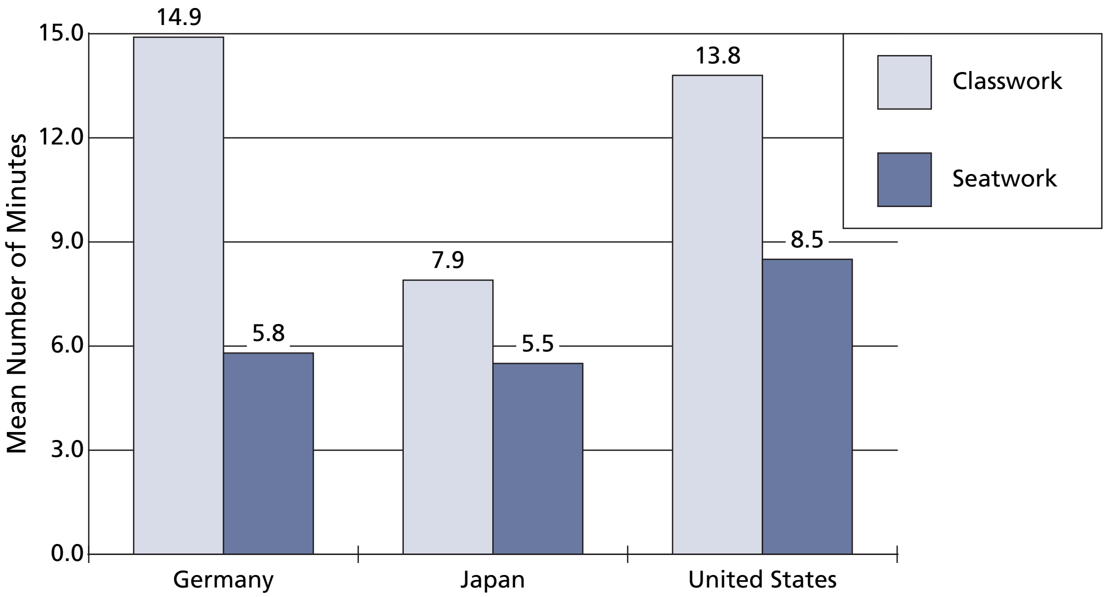
Figure 39
Average percentage of time during the lesson spent in classwork and seatwork in each country



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

While Japanese lessons have more organizational segments, the length of each segment is shorter in Japan than it is in Germany or the United States. Figure 40 shows the average duration in minutes of CLASSWORK and SEATWORK segments in each country. The average duration of SEATWORK segments in U.S. classrooms was significantly longer than in either German or Japanese classrooms. The average duration of CLASSWORK segments were significantly shorter in Japan than in either Germany or the United States.

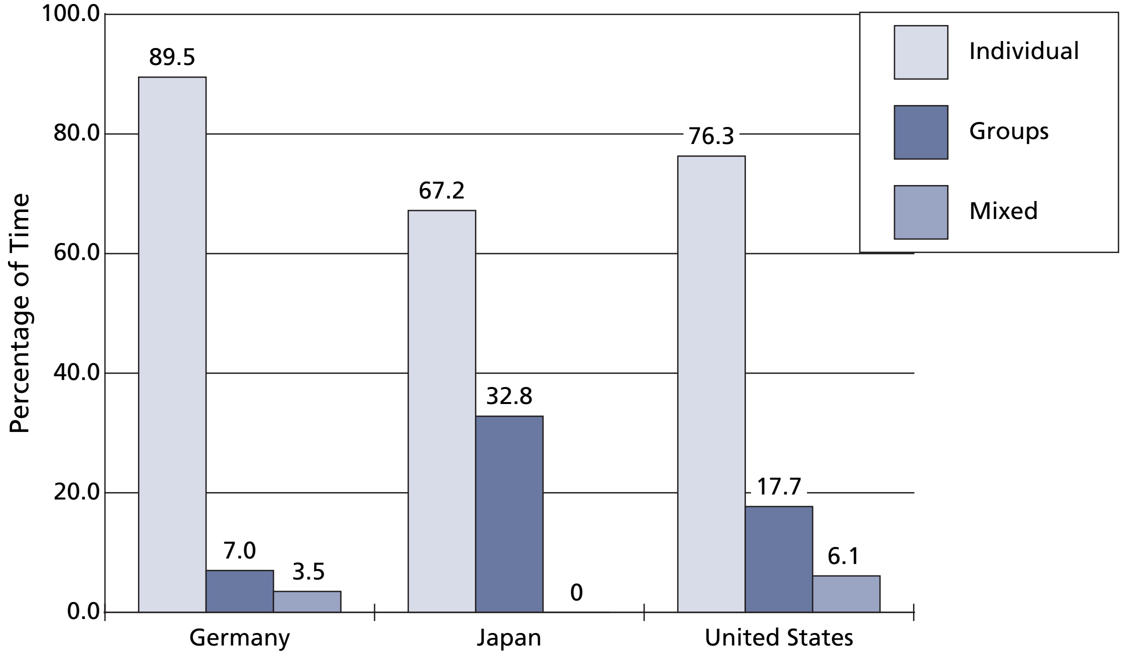
Figure 40
Mean duration of classwork and seatwork segments in each country



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Moreover, analysis of the data shows that students in German classrooms spent a higher percentage of time working individually during SEATWORK than did students in Japanese classrooms (figure 41). Students in Japan spent a higher percentage of time working in groups during SEATWORK than students in Germany or the United States.

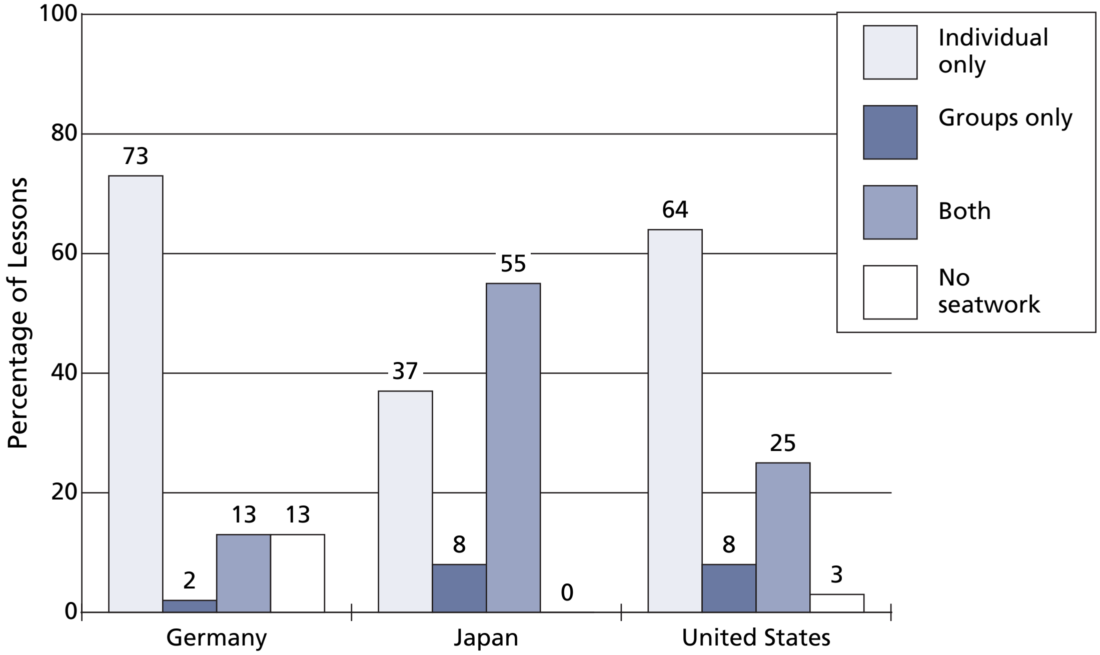
Figure 41
Percentage of seatwork time spent working individually, in groups, or in a mixture of individuals and groups



NOTE: Percentages may not sum to 100.0 due to rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

We also looked at patterns of seatwork within the lesson. In figure 42 we show the percentage of lessons in each country that include individual seatwork, group seatwork, both kinds of seatwork, or no seatwork. Although there was no significant difference across countries in the percentage of lessons containing only groupwork, a greater percentage of Japanese lessons contained both group and individual seatwork than did the German or U.S. lessons. German classrooms were more likely to have no seatwork than the Japanese or U.S. classrooms. Both German and U.S. classrooms contained a higher percentage of individual-only seatwork than Japanese classrooms.

Figure 42
Percentage of lessons in each country in which seatwork of various kinds occurred



NOTE: Percentages may not sum to 100 due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

To summarize, Japanese teachers tend to change more frequently from classwork to seatwork and back again within the lesson than do German and U.S. teachers. And, they tend to alternate between seatwork segments where students work individually and those where students work in small groups. German teachers concentrate on classwork, and provide less variation in organization of interaction over the course of a lesson than do either Japanese or U.S. teachers. U.S. teachers resemble Japanese teachers in terms of the time they devote to seatwork, but look more like German teachers in terms of the less frequent change between classwork and seatwork.

Activity Segments

Having coded lessons into segments of classwork and seatwork, we now proceed to the next layer. Classwork and seatwork, after all, represent only the most superficial view of what happens in a mathematics lesson. What goes on during these segments, and what goals are teachers trying to achieve? Our next step was to further divide the lesson into activity segments.

What we call activity segments are segments of the lesson that serve some specific pedagogical function. Examples of such functions would be setting up for seatwork (i.e., getting students ready to work on their own), working on tasks, or sharing the results of seatwork. These kinds of activities appear in all cultures and can be defined in a cross-culturally valid way. They also are categories that map well onto teachers' views of how lessons are planned and implemented. In fact, teachers generally mark the transitions between these activities with explicit words and actions. For example, a teacher might say, "Everyone get in your groups and do the problem I've written on the board." This marks a clear shift in activity, as well as a shift from classwork to seatwork.

We defined four major categories of activities: SETTING UP, WORKING ON, SHARING, and TEACHER TALK/DEMONSTRATION. The goal of SETTING UP segments is to prepare students for a subsequent seatwork segment. SETTING UP situations occur when the teacher assigns the task(s) and/or situation(s) for students to work on independently during seatwork. We identified two subtypes of SETTING UP segments:

- SETTING UP: MATHEMATICAL was coded when the teacher presented task(s) and/or situation(s) to the students with explanations or discussion; and
- SETTING UP: PHYSICAL/DIRECTIONAL was coded when the teacher presented students with task/situations without additional explanations or discussions. These segments usually included physical activities, such as moving into groups, passing out handouts, writing down task(s) and/or situation(s), and/or directions.

WORKING ON segments were the most common. Although WORKING ON was most commonly coded during periods of SEATWORK, it also could be coded during CLASSWORK. During CLASSWORK segments, WORKING ON occurred whenever the teacher and the students worked collaboratively on task(s) and/or situation(s), or derived/learned principles, properties, or definitions (PPDs). When the shift between SETTING UP and WORKING ON was not clearly identified, SETTING UP was included in the WORKING ON segment. We coded four types of WORKING ON segments:

- WORKING ON TASK/SITUATION was coded whenever the teacher and/or students worked on tasks and situations not included in the following three categories;
- WORKING ON HOMEWORK. This included segments in which homework was assigned but not necessarily started;
- WORKING ON TEST; and
- WORKING ON MULTIPLE ACTIVITIES was coded when the students were engaged in two or more assignments, such as checking homework answers and starting on a worksheet.

We coded SHARING segments when the activity focused on presenting, discussing, and reflecting on previously completed tasks and situations. The results might be shared in the form of teacher presentation, student presentation, interactive discussion, or visual representations. The segment began when the teacher expressed the intention of sharing the produced results, and it ended when there was a shift in activity that normally occurred in conjunction with a shift in content.

There were three kinds of SHARING segments, depending on what was being shared:

- SHARING TASK/SITUATION;
- SHARING HOMEWORK; and
- SHARING TEST.

TEACHER TALK/DEMONSTRATION was coded when the teacher talked about concepts, ideas, solution strategies or methods, lesson goals, or demonstrated solution steps. This segment was limited to situations where the teacher was transmitting lesson-relevant information to the students, and the students' role was to listen. Students could ask questions of the teacher, and the teacher could ask questions if his/her primary purpose was to maintain the students' attention. For example, if the teacher lectured but paused every 5 minutes to say, "OK, got it?" this would still be coded as TEACHER TALK/DEMONSTRATION, provided the responses were limited to a simple nodding of heads or murmuring of "Uh huhs." If the teacher elicited responses from the students the segment was coded as WORKING ON TASK/SITUATION. Instances where the teacher sets up a task and situation or comments on students' solution methods in the form of a presentation do not fall into this category but into WORKING ON or SHARING.

Finally, OTHER was coded if the content of the segment was not related to the mathematical content of the lesson. These segments may include instances when the teacher checks off that students completed their homework, small talk, housekeeping, or discipline. If a segment contained two or more types of activity occurring simultaneously, we coded it as MIXED.

The resulting 12 categories of activity segments are listed in figure 43.

Figure 43
Overview of categories for coding lesson activity segments

I. Setting Up
- A. Setting Up: Mathematical
- B. Setting Up: Physical/Directional

II. Working On
- A. Working On Task/Situation
- B. Working On Homework
- C. Working On Test
- D. Working On Multiple Activities

III. Sharing
- A. Sharing Task/Situation
- B. Sharing Homework
- C. Sharing Test

IV. Teacher Talk/Demonstration

V. Other

VI. Mixed

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

As can be seen in figure 44, Japanese lessons contained significantly more activity segments on average than lessons in Germany. Thus, the more frequent changing from classwork to seatwork in Japan is accompanied by a more frequent changing from one activity to another.

Figure 44
Mean number of activity segments in German, Japanese, and U.S. lessons

	Mean
Germany	8.8
Japan	12.5
United States	9.7

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

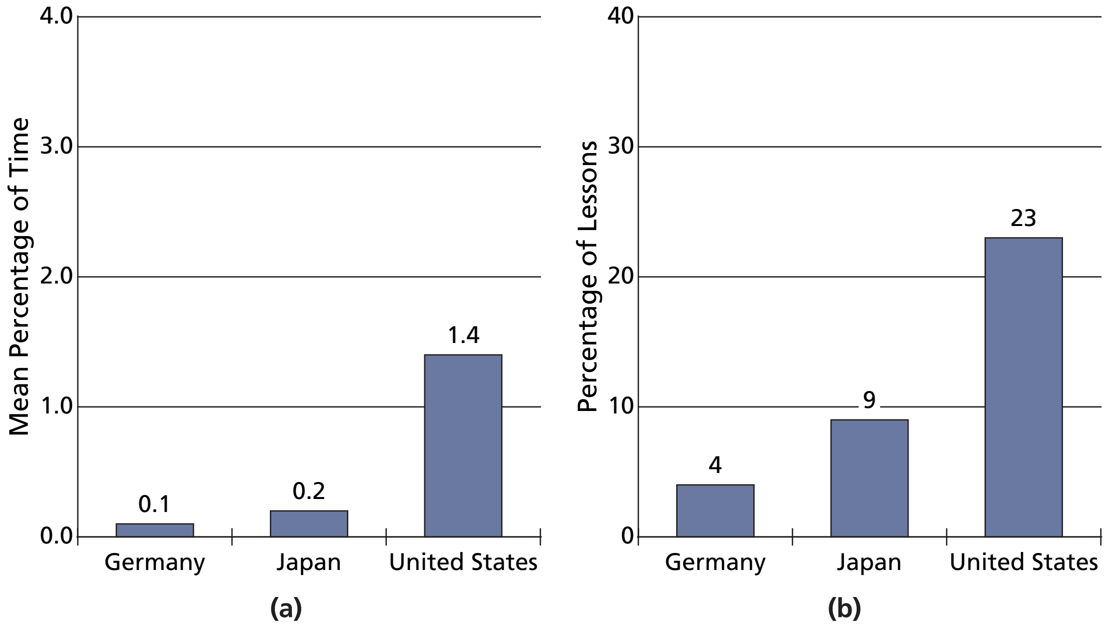
We turn now to a discussion of the time devoted to different activities within each culture. Three of the codes—WORKING ON TEST, WORKING ON MULTIPLE ACTIVITIES, and MIXED—each accounted for less than one-half percent of lesson time. Thus, we will not present any further data concerning these codes. On the other hand, some interesting differences emerged when we examined the distribution of the various activity types across cultures. We will present some of these differences next.

Time Spent in Other Activity

We start by looking at the time devoted to OTHER activities during the mathematics lesson. OTHER was coded whenever the activity was unrelated to mathematics or to the current lesson. For example, a teacher might pause during the lesson to discuss a recent field trip or sporting event of interest to the students.

This kind of unrelated activity was more common in the United States than in either of the other two countries. In figure 45 we show two different measures of OTHER: as a percentage of total time (panel a) and as a percentage of lessons in which OTHER was coded (panel b). Using percentage of total time as the measure, U.S. lessons devoted significantly more time to unrelated activities during the lesson than did either German or Japanese teachers. Similarly, a significantly higher percentage of U.S. lessons contained such activities than did German lessons. Although the total percentage of time devoted to unrelated activities during the lesson is small in all countries, even a brief diversion of this sort may break the flow of the lesson. Thus, it may be important that nearly 25 percent of the U.S. lessons included this kind of activity.

Figure 45
Time devoted to unrelated activities during the mathematics lesson: (a) as a percentage of total lesson time and (b) as percentage of lessons in which any activity is coded as "other"



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

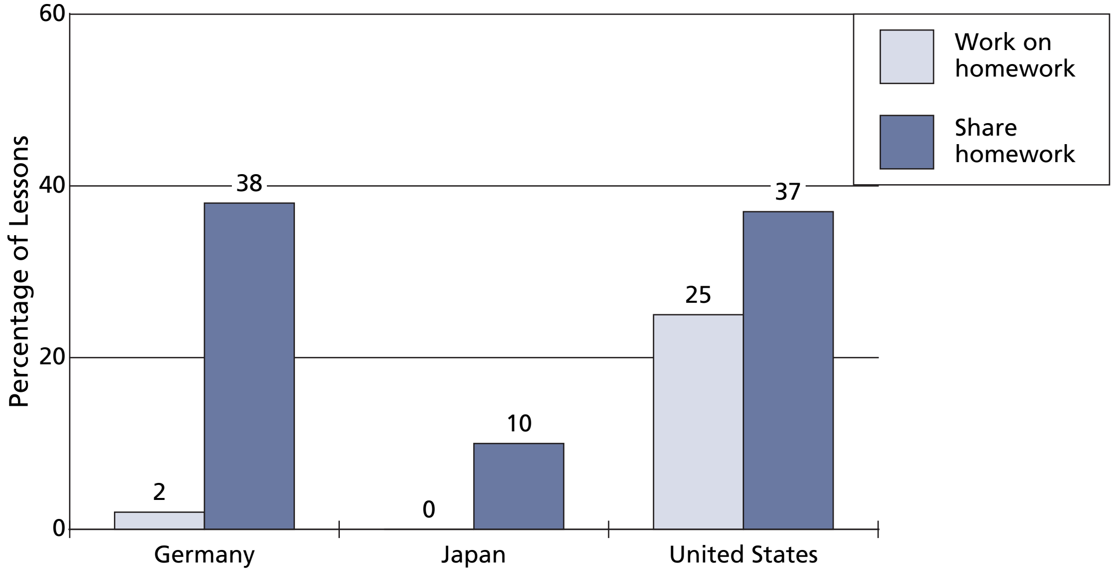
Homework During the Lesson

There is a great deal of interest in the United States in the topic of homework: Many Americans believe that more homework will translate into higher achievement for students. In fact, our observations confirm that homework plays a more prominent role in eighth-grade mathematics classrooms in the United States than it does in Japan, and somewhat more prominent than in Germany.

In figure 46 we show the percentage of lessons in which students actually work on or share homework. Japanese students were never observed to work on the next day's homework during class, and were relatively rarely observed to share homework results. Both German and U.S. students share homework in class frequently, but only U.S. students spend significant amounts of time in class actually working on the next day's homework. When we look at total percentage of time during the lesson devoted to assigning, working on, or sharing homework, we get a similar result: Two percent of lesson time in Japan involves homework in any way which is significantly less than the 8 percent of lesson time in Germany and 11 percent in the United States.[2]

[2] Standard errors for Germany, Japan, and the United States were 1.75, 0.46, and 1.89, respectively.

Figure 46
Percentage of lessons in which class works on and shares homework (not including assigning homework)



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

In our teacher questionnaire we asked whether or not they had previously assigned homework that was due on the day of the videotaped lesson. The countries differed in the proportion that had done so. In 55 percent of both the U.S. lessons and the German lessons, the teachers said that they had assigned such homework, compared to 14 percent of Japanese teachers who responded in this way.[3]

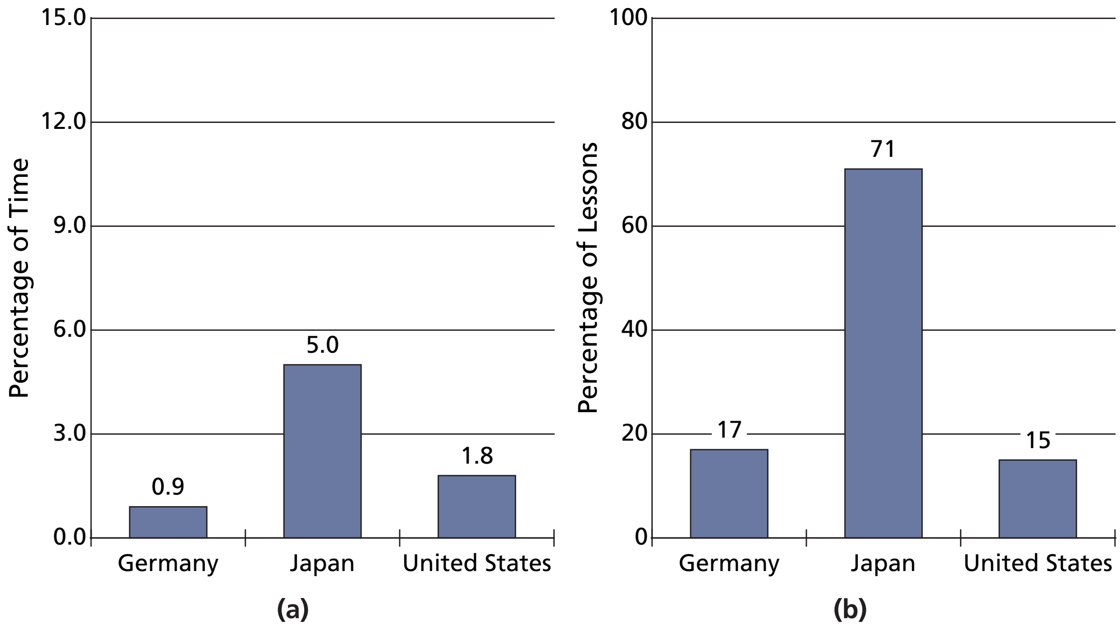
Teacher Talk/Demonstration

TEACHER TALK/DEMONSTRATION was coded when teachers engaged in lecturing. Here teachers would simply talk, with or without objects, presenting lesson-relevant information to students.

As can be seen in figure 47, there was more TEACHER TALK/DEMONSTRATION in Japan than in the other two countries. As with OTHER, the overall percentage of time devoted to this activity was not large (see panel a), but more Japanese lessons included at least some TEACHER TALK/DEMONSTRATION than did those in either Germany or the United States (panel b).

[3] Standard errors for Germany, Japan, and the United States were 4.86, 5.46, and 7.12, respectively.

Figure 47
Emphasis on teacher talk/demonstration as indicated by (a) percentage of lesson time, and (b) percentage of lessons in which such segments occur



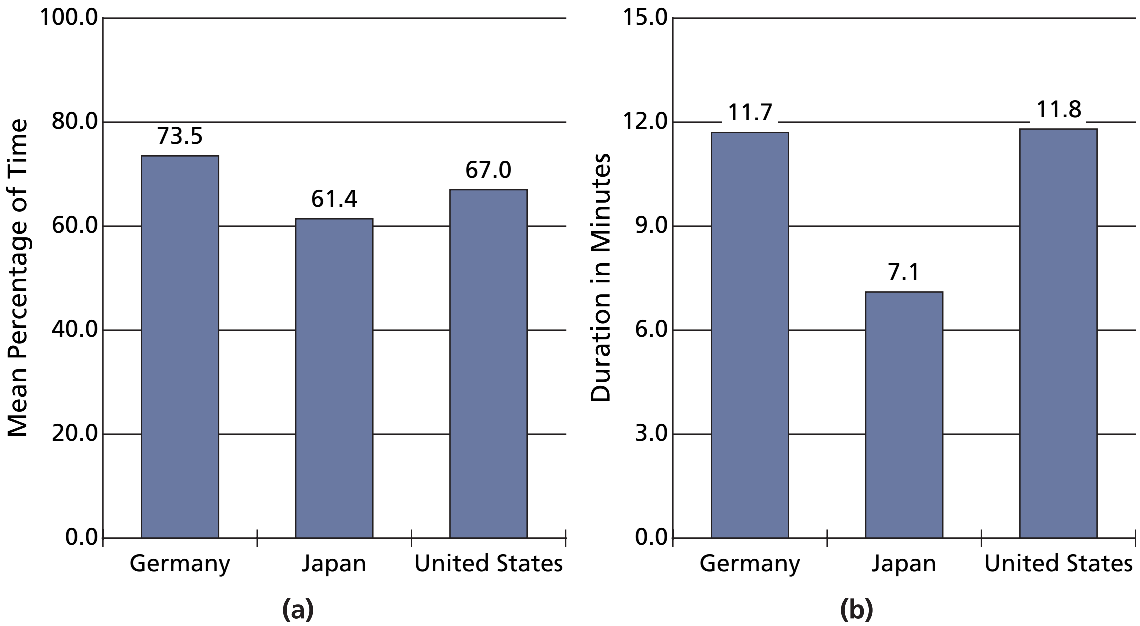
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Working On Tasks and Situations

Much of the instructional time in all countries occurs during three types of segments: SETTING UP tasks and situations, WORKING ON tasks and situations, and SHARING of tasks and situations. SETTING UP and SHARING occur during classwork; WORKING ON tasks and situations can occur during either classwork or seatwork.

By far the most frequent and time-consuming type of activity segment in all countries is WORKING ON TASK/SITUATION. As can be seen in panel (a) of figure 48, classrooms in all countries spent 60 percent or more of lesson time in WORKING ON TASK/SITUATION segments. German classrooms, however, spent more time this way than Japan. Consistent with our findings on classwork and seatwork, Japanese WORKING ON TASK/SITUATION segments were significantly shorter in duration on average than were such segments in Germany and the United States (panel b).

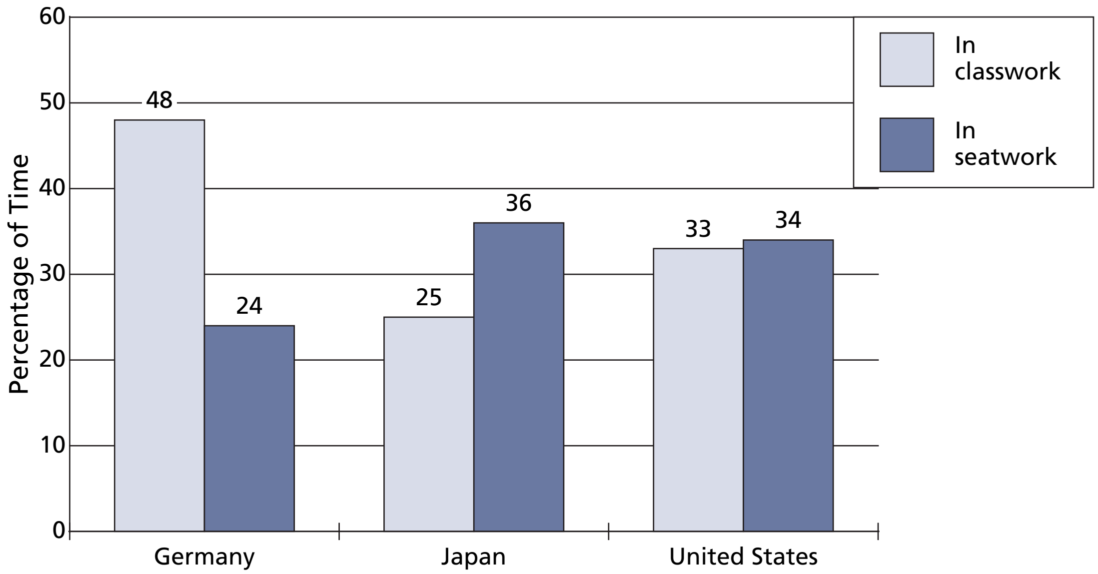
Figure 48
(a) Percentage of total lesson time spent in and (b) average duration of working on task/situation segments



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

A more detailed picture emerges when we break down WORKING ON TASK/SITUATION according to whether it occurs during classwork or during seatwork. As seen in figure 49, German classes spend more time WORKING ON tasks and situations during classwork than during seatwork; Japanese classes spend more time WORKING ON tasks and situations during seatwork than during classwork. U.S. classes spend roughly equal time WORKING ON tasks and situations during classwork as they do during seatwork. Moreover, German classes spend more time WORKING ON tasks and situations during classwork than either their Japanese or U.S. counterparts.

Figure 49
Average percentage of lesson time spent in (a) working on task/situation during classwork, and (b) working on task/situation during seatwork

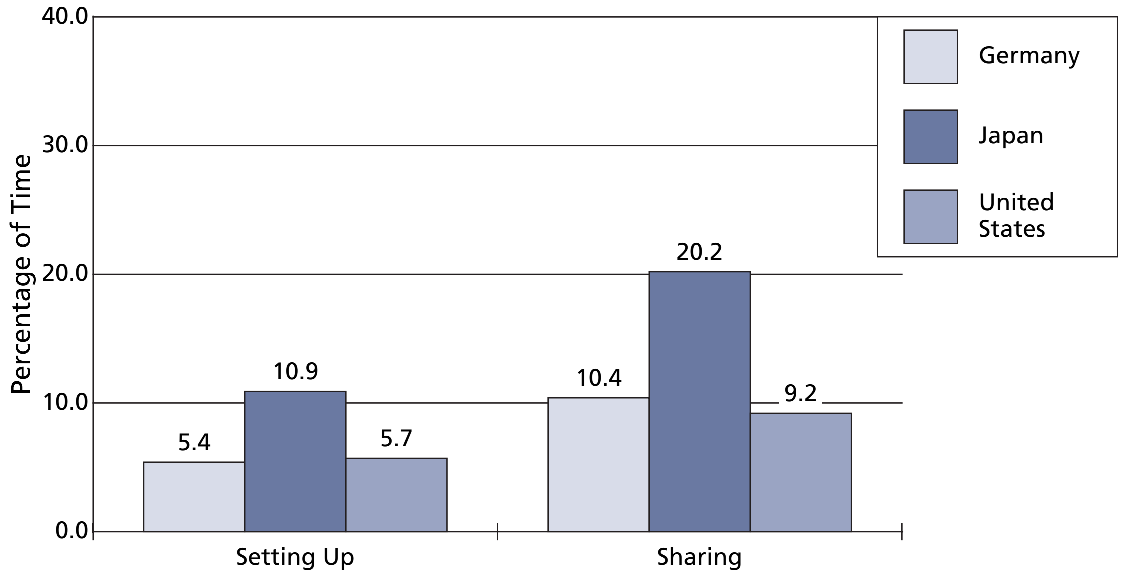


SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Setting Up and Sharing Tasks and Situations

Finally, although Japanese classes spend relatively less time WORKING ON tasks and situations during classwork, they spend more time SETTING UP and SHARING tasks and situations than do both German and U.S. classes, as seen in figure 50.

Figure 50
Average percentage of total lesson time spent in setting up and sharing task/situation



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

We have examined the content of the lessons and the ways that teachers in the three countries organize their lessons. We now present some preliminary analyses of the processes that go on as lessons unfold.

Chapter 5. Processes of Instruction

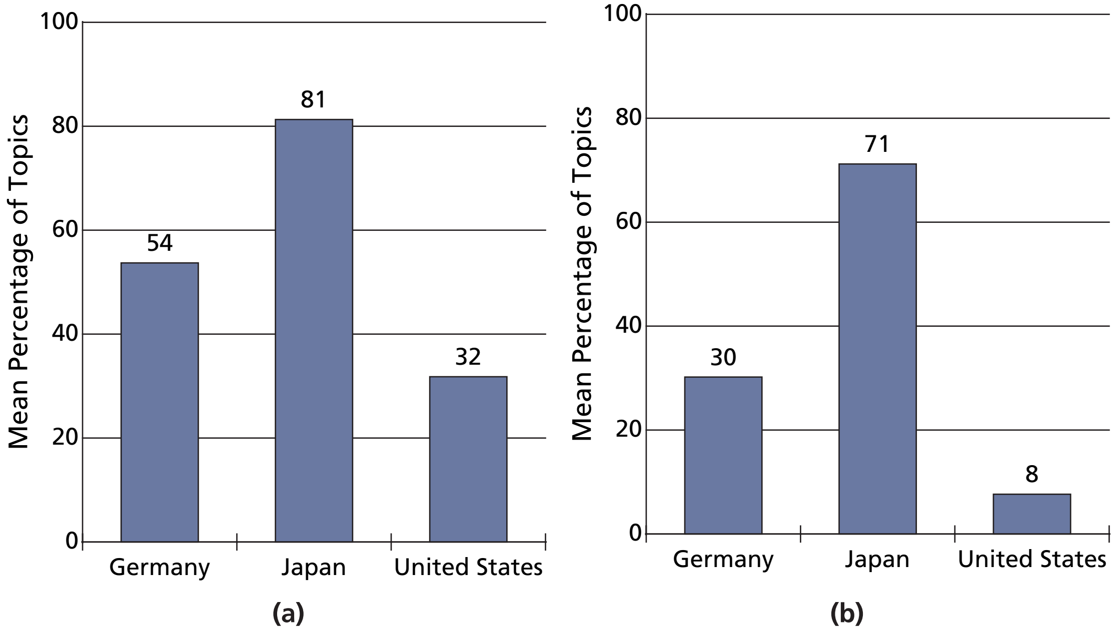
Both the content and organization of lessons are generally planned in advance; they represent conscious decisions on the teacher's part. But not all that happens in classrooms is planned. Some processes only become evident as instruction unfolds and sometimes only through detailed analyses. In this chapter we present some additional analyses that describe, more fully, the nature of instruction in these three countries. Again, we remind the reader that these analyses are preliminary; much remains to be done.

DEVELOPING CONCEPTS AND METHODS

Earlier we distinguished two ways of including concepts in a lesson: A concept might simply be stated by the teacher or students, but not explained or derived; or, it might be developed (i.e., derived and/or explained) by the teacher or the teacher and students collaboratively in order to increase students' understanding of the concept. Our analyses indicated that development of concepts occurs more often in Germany and Japan than in the United States. Further analysis reveals, however, that there are some significant differences in the way in which concepts are developed in Germany and Japan. Development happens primarily during classwork in Germany, with most of the work being done by the teacher. In Japan, on the other hand, seatwork segments play a more critical role in the development of mathematical concepts, consistent with a strategy of giving students themselves more responsibility for the process.

Evidence for this conclusion is presented in figure 51. Recall that concepts were coded as "developed" if they were derived or explained by the teacher and/or the students in order to increase students' understanding. For each topic within each lesson, we first determined whether or not development of concepts was included. If it was, we next coded whether or not there were any seatwork segments within the topic/lesson. In panel (a) we show the average percentage of these topics within each lesson that included any seatwork at all. By this loose definition, the Japanese development segments included significantly more seatwork than did the German segments. Of course, the seatwork may not have been the part of the segment in which the development actually occurred. If we tighten the definition, as we have done in panel (b), we get a similar result. Here, we show only those topics for which development actually occurred during seatwork. The percentage in Japan was significantly higher than in either of the other two countries.

Figure 51
Average percentage of topics including development that (a) include at least some seatwork and (b) include actual development of concepts during a seatwork segment



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

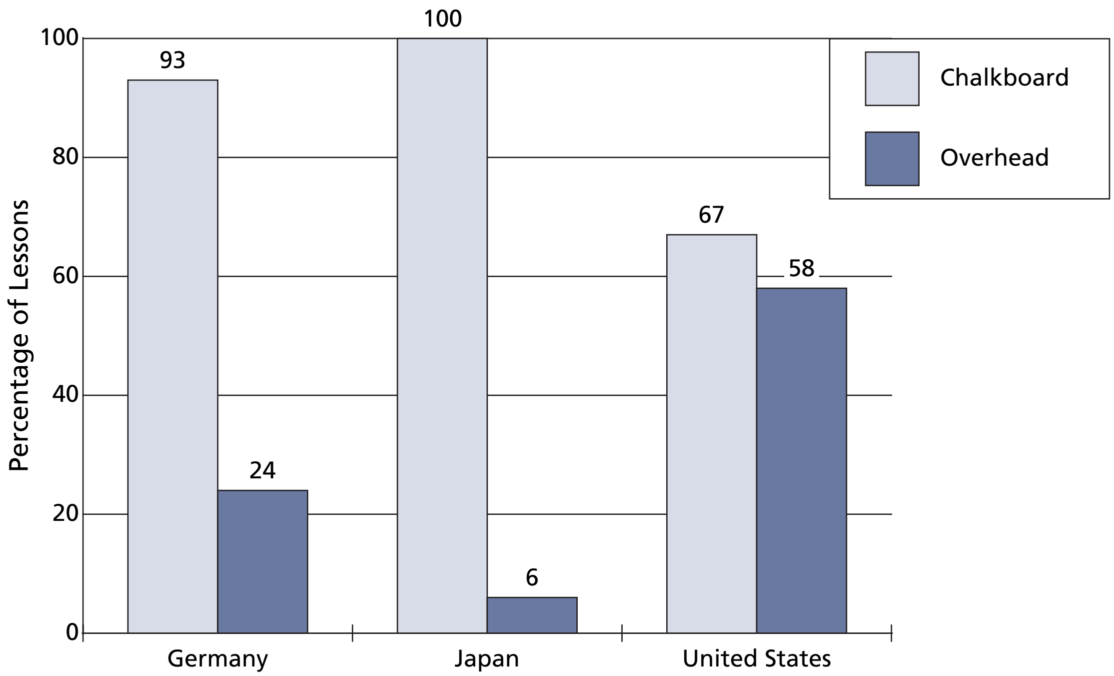
THE USE OF INSTRUCTIONAL MATERIALS

We observed a wide variety of tools and materials being used in our samples of eighth-grade mathematics classrooms. The most commonly used instructional tools were the chalkboard and the overhead projector. Indeed, four lessons in Germany, none in Japan, and three in the United States used neither of these tools.[1]

The percentage of lessons in each country in which the chalkboard and overhead projector were used is displayed in figure 52. (In some lessons, teachers used both.) The German and Japanese teachers used the chalkboard significantly more often than teachers in the United States. In contrast, U.S. teachers used the overhead projector more often than teachers in Japan or Germany, and German teachers used the overhead more than teachers in Japan.

[1] These lessons were GR-16, GR-21, GR-34, GR-84, US-9, US-16, and US-42.

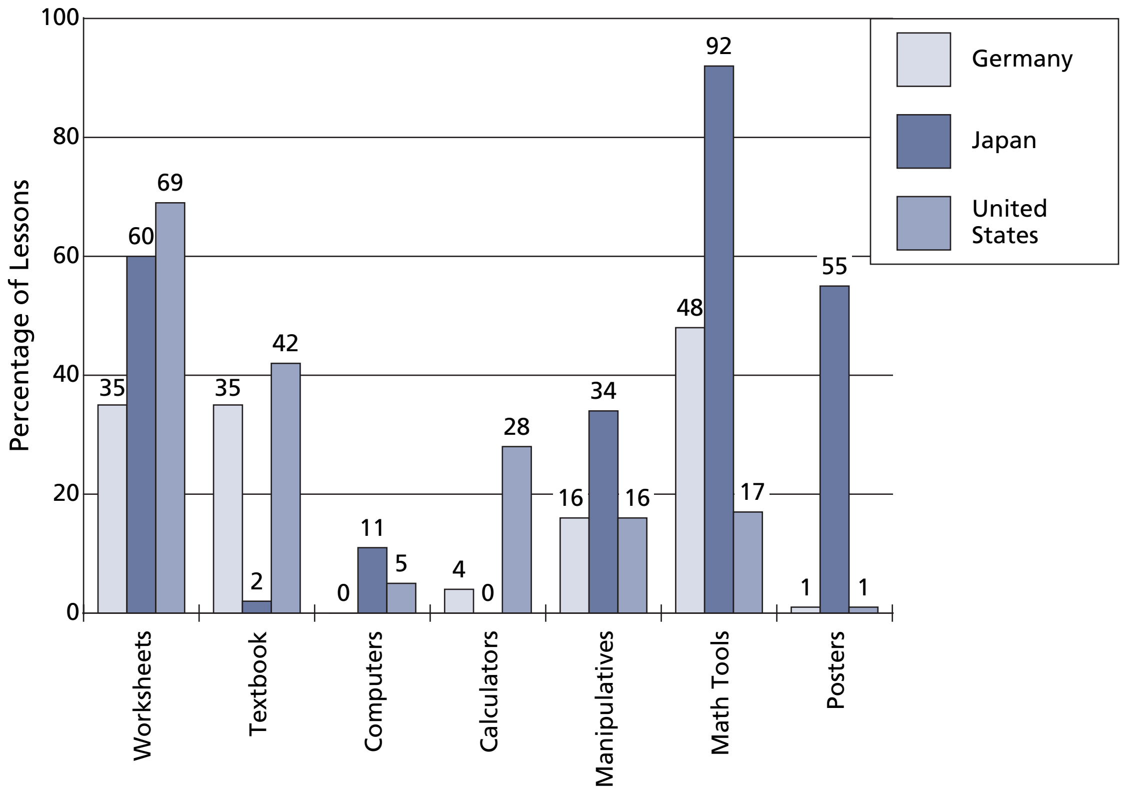
Figure 52
Percentage of lessons in which chalkboard and overhead projector are used



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

The percentage of lessons in which other kinds of materials were used is shown in figure 53. Most of the categories used are self-explanatory. The term “Manipulatives” refers to any concrete materials used to represent quantitative situations, such as paper circles, plastic triangles, unit blocks, or geoboards. Posters, used mostly in Japan, refer to prepared paper materials that are brought out and attached to the board during the lesson. Mathematics tools include objects specifically designed for use in solving mathematical problems. Examples of this category include rulers and graph paper.

Figure 53
Percentage of lessons in which various instructional materials were used



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

A number of cross-national differences emerged. Worksheets were significantly more common in Japan and the United States than in Germany. Textbooks, on the other hand, were seldom used in Japan but were rather common in Germany and the United States. Calculators were used primarily in the United States and rarely or never in the other two countries. Japanese teachers used significantly more mathematical tools and posters than did German and U.S. teachers.

A summary analysis in which we simply added up the total number of different materials categories represented in each lesson revealed that Japanese lessons had the most (average of 3.7) followed by the United States (3.1) and then Germany (2.6).[2]

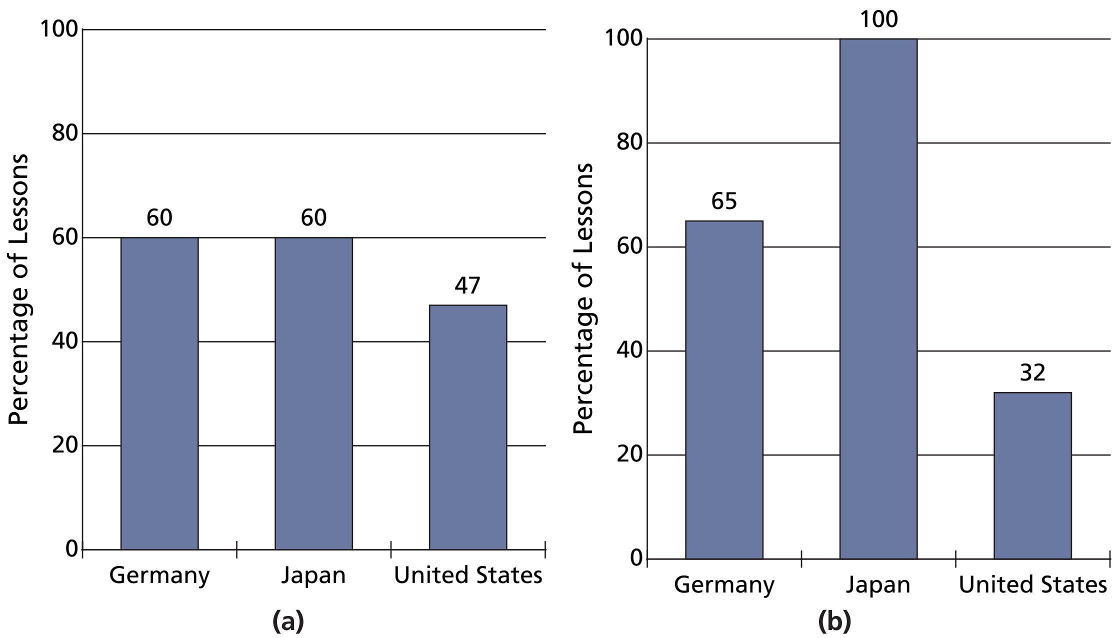
Use of the Chalkboard

We discovered some differences in the way that chalkboards and overhead projectors are used in the three countries. One of these concerned the frequency with which students, as opposed to the teacher only, come to the front and use the chalkboard or overhead projector. In figure 54 we show (a) of the lessons in which the chalkboard is used at all, the percentage in which students actually use the chalkboard; and (b) of the lessons in which the overhead projector is used at all, the percentage in which students use it. The cross-country differences were not significant in the case of use of the chalkboard

[2] Standard errors for Germany, Japan, and the United States were 0.10, 0.09, and 0.12, respectively.

by students. Japanese students, on the other hand, did use the overhead projector significantly more often than did German students by this measure, and German students did so significantly more often than U.S. students.

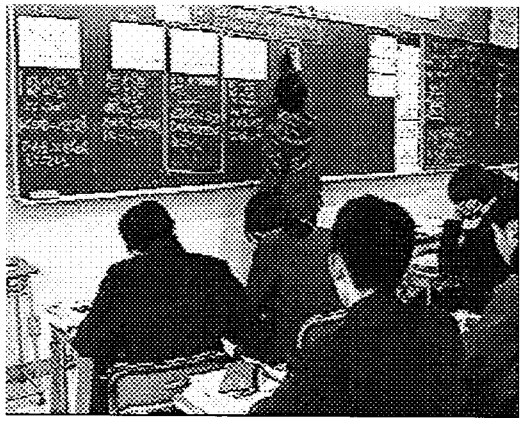
Figure 54
Percentage of lessons including (a) chalkboard or (b) overhead projector in which students come to the front and use it



NOTE: The overhead projector was used in only three Japanese lessons.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

The second discovery was in the way the chalkboard is used. In Japan, the chalkboard is used in a highly structured way: Teachers appear to begin the lesson with a plan for what the chalkboard will look like at the end of the lesson, and by the end of the lesson we see a structured record or residue of the mathematics covered during the lesson (see figure 55). In the United States, in contrast, the use of the chalkboard appears more haphazard. Teachers write wherever there is free space and erase frequently to make room for what they want to put up next.

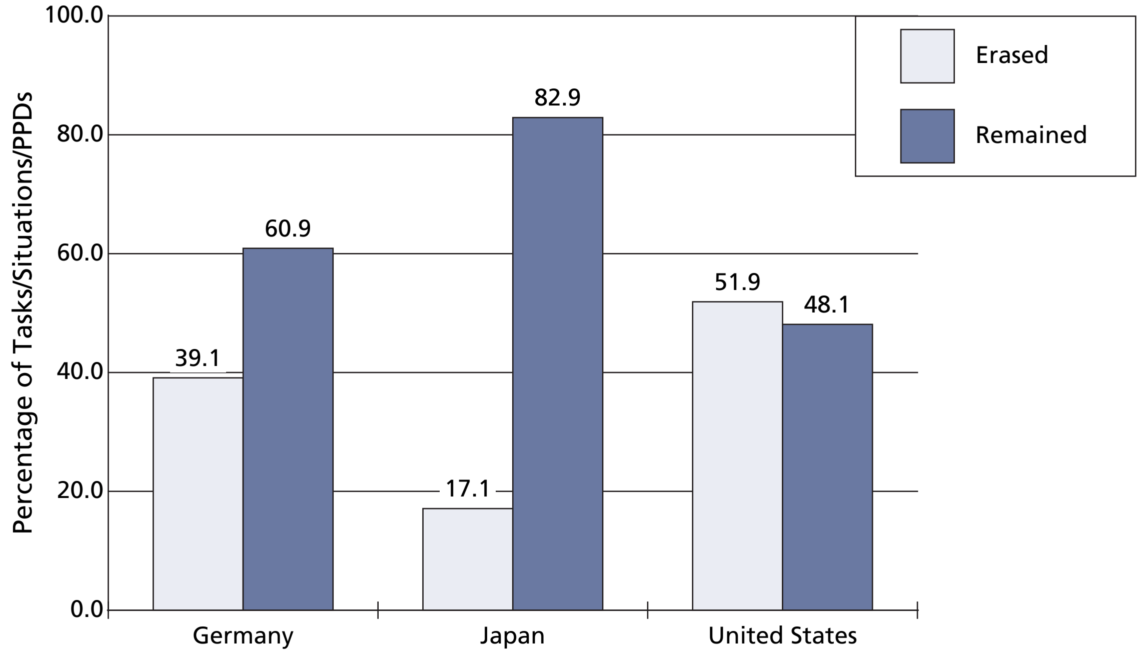
Figure 55
Example of chalkboard use from a Japanese lesson



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Objective support for this impression comes from an analysis of erasures during the lesson. We performed this analysis on the subset of 90 lessons used by the Math Content Group. We counted all of the tasks and situations that were represented on the chalkboard during the lesson, then looked, at the end of the lesson, to see what percentage remained. The results are shown in figure 56.

Figure 56
Percentage of tasks, situations, and PPDs (principles/properties/definitions) written on the chalkboard that were erased or remained on the chalkboard at the end of the lesson



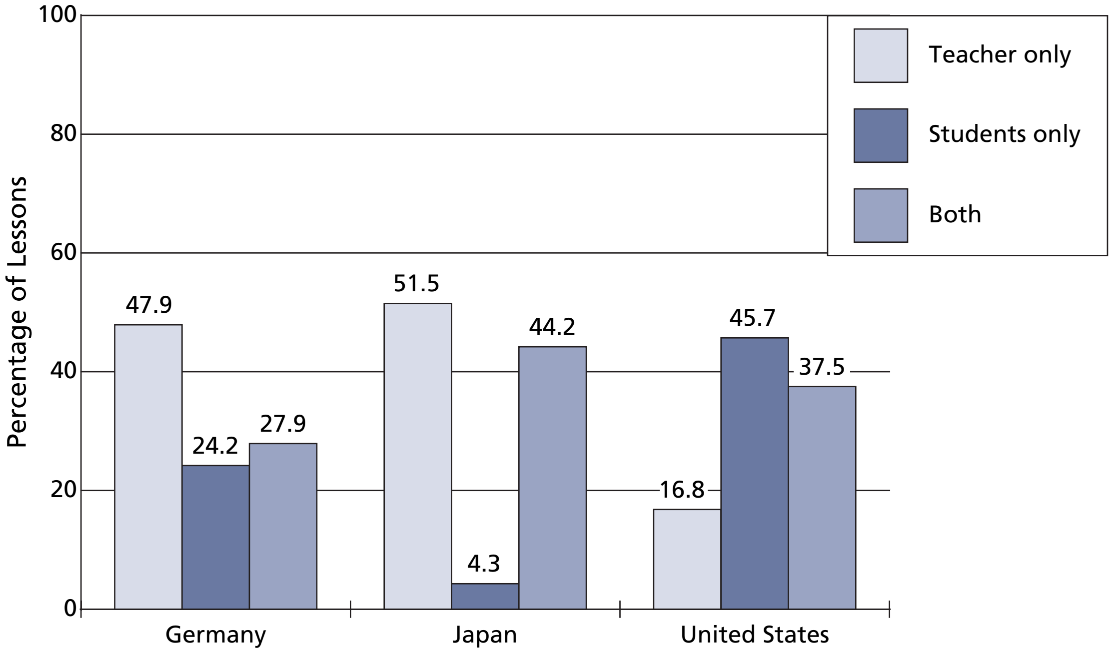
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Analyses revealed that Japanese teachers left more information remaining on the board at the end of the lesson than did either German or U.S. teachers. It is interesting to consider the potential effect this practice might have on student comprehension of the lesson. If information is erased, it is no longer available to the student who may need more time to process it. Having the information available throughout the lesson, in an organized fashion, may provide a crucial resource to the student. Alternatively, students may absorb material on a chalkboard more completely if there is less information on it at a given time.

Use of Manipulatives

Although the Japanese teachers in our sample used manipulatives more frequently than teachers in the other countries, teachers in all countries did use them to some degree. But they were not always used in the same way across countries. One aspect in which we were interested was who used the manipulatives. In figure 57 we show the percentage of lessons in which manipulatives were used by the teacher only, the students only, or both teachers and students. The only significant difference was in the percentage of manipulatives used by the students only: Japanese lessons were significantly less likely to include manipulatives used by students only than were the U.S. lessons.

Figure 57
Average percentage of lessons where manipulatives were used in which the manipulatives were used by teacher, students, or both



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

PROCESSES DURING SEATWORK

It is difficult to infer what students are doing during classwork. They may appear to be listening to the teacher, but beyond that we have little information about the kind of thinking in which they are engaged. Seatwork is somewhat different in this respect: Students are generally given an explicit task to work on, and the task usually leads to some visible product. We wanted to describe the number of tasks and situations, as well as the kinds of tasks, students were assigned to work on during seatwork.

Tasks and Situations During Seatwork

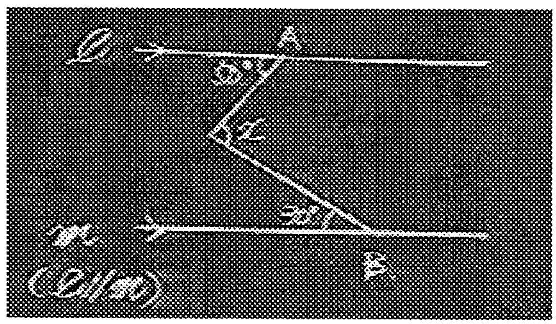
It was possible to reliably identify tasks and situations engaged in seatwork. By examining the tasks and situations students were assigned to work on during each seatwork segment, we identified four distinct patterns:

- ONE TASK/ONE SITUATION—this would typically occur when the teacher had students do one example, then come back as a class to discuss it;
- MULTIPLE TASKS/ONE SITUATION—this would typically happen when students were given a single mathematical situation and asked to explore it from a variety of perspectives, performing multiple tasks;

- ONE TASK/MULTIPLE SITUATIONS—this typically happens when students, having just been taught how to perform a specific task, are asked to practice it in a number of exercises; and
- MULTIPLE TASKS/MULTIPLE SITUATIONS—this is a typical worksheet or problem set from the textbook in which students are asked to do a variety of exercises.

An example of ONE TASK/ONE SITUATION comes from lesson JP-007, which dealt with angles between parallel lines (figure 58). Presented with the diagram on the chalkboard, students were asked to find the angle (X) in the bend using any of the three methods that they had learned previously. After completing this task they reconvened to discuss their answers.

Figure 58
Excerpt from chalkboard of JP-007



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

An example of MULTIPLE TASKS/ONE SITUATION can be seen in US-012. Students are presented with a single situation, the equation $x^2 + 14x - 43 = 0$. They are told to solve the equation twice, the first time by completing the square, the second, by using the quadratic equation. In this example, each solution would be coded as a separate task.

ONE TASK/MULTIPLE SITUATIONS is exemplified in GR-103. Students are told to turn to page 95 of their textbooks and do exercises 12a, 12b, and 12c (figure 59). In each case the task is the same, namely, to solve the systems of linear equations.

Figure 59
Excerpt from textbook page used in GR-103

12. a) $\frac{3x}{4} + \frac{7}{12} = 2 - \frac{2y}{9}$ b) $\frac{x}{3} + 2 = \frac{y}{2} + \frac{5}{6}$ c) $\frac{2y-5}{9} = \frac{5}{6}(x-1) - 5y$

$\frac{2y}{5} + \frac{3}{10} = 1 + \frac{x}{2}$ $\frac{y}{4} + 1 = \frac{3}{10} - \frac{3x}{5}$ $\frac{3x+1}{12} = \frac{8}{3}(y-2) + \frac{33x}{2}$

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Finally, an example of MULTIPLE TASKS/MULTIPLE SITUATIONS can be seen in US-016. The teacher hands out a worksheet with problems and asks students to work on them during seatwork (figure 60):

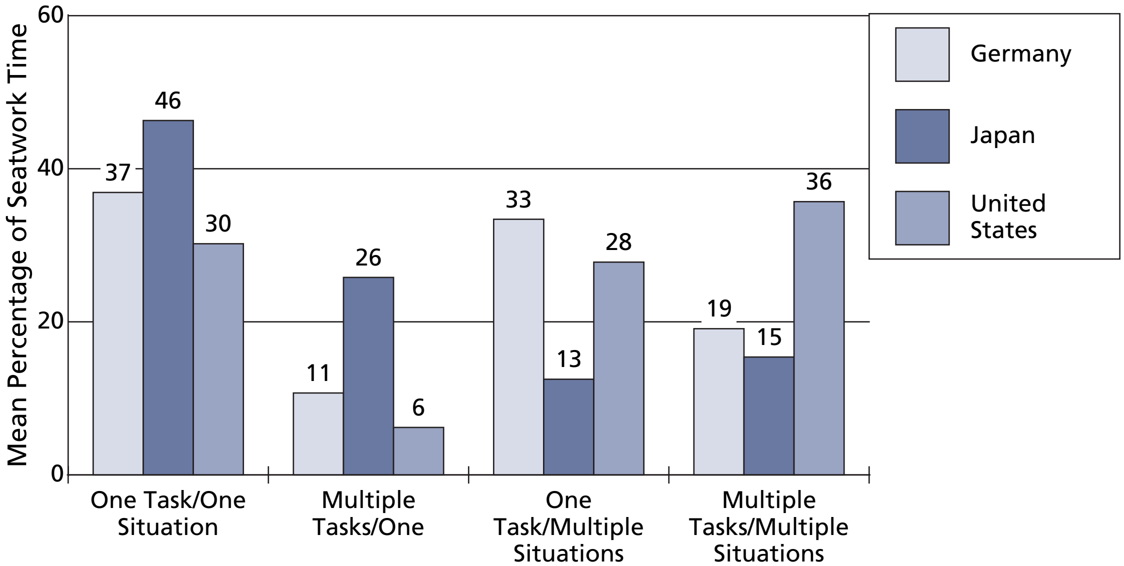
Figure 60
Problems from worksheet used in US-016

Task 1/Situation 1	The Greek alphabet has 24 letters. How many 3-letter monograms are possible? (3-letter Greek monograms often are used to name fraternities and sororities).
Task 2/Situation 2	Bill is a streak hitter in baseball. He gets hits 25 percent of the time he is at bat. But when he gets a hit his first time up, the probability he will get a hit the next time up is 32%. What is the probability Bill will get hits twice in a row at the beginning of the game?
Task 3/Situation 3	The estimated probability of being able to roll your tongue is 1/8. The estimated probability of having attached earlobes is 1/16. What is the probability of a person being able to roll his/her tongue and having attached earlobes?
Task 4/ Situation 4-1 and 4-2	The digits 0, 1, and 8 read the same right side up or upside down. a. How many different two-digit numbers read the same either way? (A number may not begin with zero). b. How many different three-digit numbers read the same either way? (A number may not begin with zero).

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

The average percentage of seatwork time spent working in these four patterns is shown in figure 61. U.S. students spent significantly more time working on MULTIPLE TASKS/MULTIPLE SITUATIONS than did Japanese students. Japanese students spent more time working on MULTIPLE TASKS/ONE SITUATION than did German or U.S. students. German students spent more time than Japanese students working on ONE TASK/MULTIPLE SITUATIONS.

Figure 61
Average percentage of time in seatwork/working on task/situation segments spent working on four different patterns of tasks and situations in each country



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

These patterns may affect students' experience during seatwork. For example, in Japan, where students generally work on only one situation during a seatwork segment (72 percent of the time), the students may experience the work as more unified or coherent than do U.S. students, who tend to work on multiple situations (64 percent of the time). Alternatively, students may develop a more coherent sense of a concept when presented with a variety of tasks and situations to approach.

Performance Expectations

What kinds of tasks were students working on during seatwork? We coded tasks into three mutually exclusive categories:

- PRACTICE ROUTINE PROCEDURES;
- INVENT NEW SOLUTIONS/THINK; and
- APPLY CONCEPTS IN NEW SITUATIONS.

PRACTICE ROUTINE PROCEDURES was coded to describe tasks in which students were asked to apply known solution methods or procedures to the solution of routine problems. Generally, the function of these seatwork segments was to practice previously learned information. For example, in GR-033 the class first goes over the solutions to two linear equations that had been assigned for homework. After sharing the solution to two equations that were homework, teacher and students together solve one more equation as a practice example : $x + 2(x - 3) = 5x - 4(2x - 9)$. Then the teacher assigns two more equations for seatwork:

(1) $60 - 8(6 - 2x) = 44$

(2) $8x + 12 + (5x - 8) = 10x - (3 - 2x - 8)$

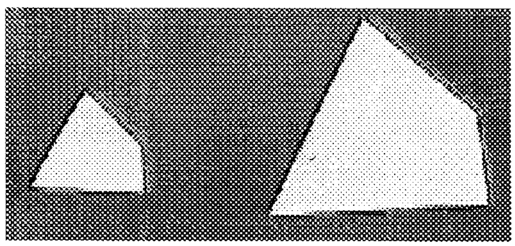
These seatwork tasks were coded as Practice Routine Procedures.

INVENT NEW SOLUTIONS/THINK was coded to describe tasks in which students had to create or invent solution methods, proofs, or procedures on their own, or in which the main task was to think or reason. The expectation, in these cases, was that different students would come up with different solution methods.

An example of this category can be seen in JP-034 (figure 62). The topic of the lesson is similarity of two-dimensional figures. In a preliminary discussion, students are asked to think of objects that have the same shape but different sizes. After a number of objects are listed, the teacher presents quadrilateral ABCD on the board (on the left), together with a similar quadrilateral that is expanded twofold (on the right). She asks students, during seatwork, to think of as many methods as they can to expand the figure on the left into the one on the right.

Figure 62

Excerpt from chalkboard in JP-034



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

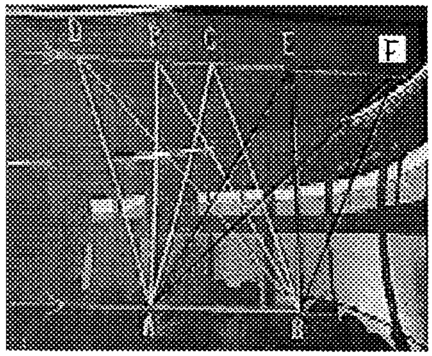
Only one seatwork segment in the U.S. data was coded as INVENT NEW SOLUTIONS/THINK. The lesson, US-033, dealt with common fractions. The teacher gives definitions of equivalent and improper fractions, then illustrates each definition with an example. At this point the teacher tells students to come up with their own definition for "proper fraction":

00:13:39	T	What about a proper fraction?
00:13:52	T	Okay you guys should be thinking about what a proper fraction is while I try to take care of this.
00:14:04	T	You can go ahead and chat with your study buddy and figure out, come up with, good definition for proper fraction.

It is important to note that performance expectations cannot be coded simply by analyzing the problem. It is also necessary to see what students do, both while solving the problem and afterwards. When seatwork is followed by students sharing alternative solution methods, this generally indicates that students were to invent their own solutions to the problem.

We have labeled the third category of performance expectations as APPLY CONCEPTS IN NEW SITUATIONS because most of the tasks coded into this category involved transferring a known concept or procedure into a new situation. In point of fact, however, we coded this category whenever the seatwork task did not fall into one of the other two categories. An example of this code can be seen in JP-012, which dealt with geometric transformations (figure 63). First, the teacher uses the display to review with students the fact that any two triangles between the same two parallel lines will have the same area.

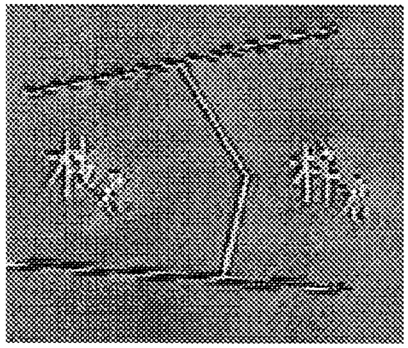
Figure 63
Excerpt from computer monitor used in JP-012



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Then the teacher assigns students to work on the following problem: The border between Eda's land and Azusa's land is bent (figure 64). How can we straighten the border without changing the area of either person's land? We coded this as APPLY CONCEPTS IN NEW SITUATIONS because the teacher suggested which concept students should apply to the solution of the problem.

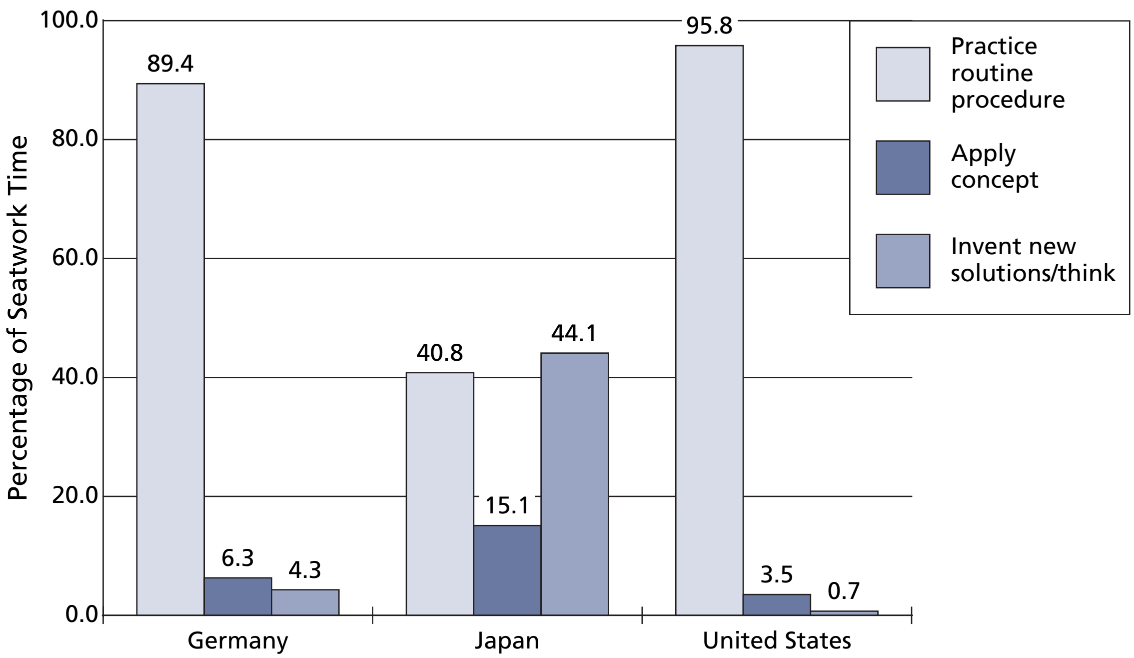
Figure 64
Excerpt from chalkboard in JP-012



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

The average percentage of seatwork time spent in each of the three kinds of tasks is shown in figure 65.

Figure 65
Average percentage of seatwork time spent in three kinds of tasks



NOTE: Percentages may not sum to 100.0 due to rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Japan differed significantly from the other two countries, spending less time on practice of routine procedures during seatwork and more time inventing new solutions or thinking about mathematical problems.

CLASSROOM DISCOURSE

In this section we will present the coding categories and results from the discourse coding that has been completed thus far. We stress that these analyses of discourse are quite preliminary. However, they do provide a foundation on which to build subsequent analyses.

We report two sets of analyses. The first set (First-Pass Coding) utilized the full sample of lessons but sampled 30 utterances to represent each lesson. (See section "Coding of Discourse" for more details on how utterances were sampled.) The second set of analyses (Second-Pass Coding) utilized 30 lessons in each country (the same 90-lesson subsample used by the Math Content Group), but analyzed all of the utterances in these lessons.

First-Pass Coding: Categorizing Utterances

The unit of analysis for first-pass discourse coding was the utterance. An utterance was defined as a sentence or phrase that serves a single goal or function.

The first step was to categorize each utterance during public discourse into 1 of 12 mutually exclusive categories. Six of the categories were for teacher utterances, 5 for student utterances, and 1 (Other) for both teacher and student utterances. These categories are briefly described in figure 66.

Figure 66
Categories used for first-pass coding of utterances during public discourse

Category		Description
Elicitation	E	A teacher utterance intended to elicit an immediate communicative response from student(s), including both verbal and non-verbal responses.
Information	I	A teacher utterance intended to provide information to the student(s). Does not require communicative or physical response from students.
Direction	D	A teacher utterance intended to cause students to perform some physical or mental activity. When the utterance is intended for future activities, it is coded as Information even if the linguistic form of the utterance is a directive.
Uptake	U	A teacher utterance made in response to student verbal or physical responses. It may be evaluative comments such as "Correct," "Good," or "No," repetition of student response, or reformulation of student response. Uptake is intended only for the respondent, and when it is clear that the utterance is intended for the entire class, it is coded as Information instead of Uptake.
Response	R	A student utterance made in response to an elicitation or direction.
Student Elicitation	SE	A student utterance intended to elicit an immediate communicative response from the teacher or from other students.
Student Information	SI	A student utterance not intended to elicit any immediate response from teacher or from other students.
Student Direction	SD	A student utterance intended to cause the teacher or other students to perform immediately some physical/mental activity.
Student Uptake	SU	A student utterance intended to acknowledge or evaluate another student's response.
Teacher Response	TR	A teacher utterance made in response to a student elicitation.
Provide Answer	PA	A teacher utterance intended to provide the answer to the teacher's own elicitation.
Other	O	An utterance that does not fit into any of the above categories or that is not intelligible.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Elicitations (E) were further subdivided into five mutually exclusive categories, presented in figure 67.

Figure 67
Subcategories of elicitations

Category		Description
Content Elicitation	EC	An elicitation that requests information directly concerned with mathematics, mathematical operations, or the lesson itself. Such elicitations may request the student to supply a quantity, identify a geometric shape, explain a mathematical procedure, define some mathematical term, or evaluate a mathematical answer, among other things.
Meta-cognitive Elicitation	EM	An elicitation designed to determine a student's current state of mind or level of understanding. These types of elicitations are often used to assess student progress as well as student understanding.
Interactional Elicitation	EI	An elicitation that requests a student to modify his/her behavior, to acknowledge his/her participation in some current activity, to recall specific classroom procedures or rules, or to gain students' attention.
Evaluation Elicitation	EE	An elicitation that requests a student or students to evaluate another student's answer, response, etc. Generally, the evaluation of responses is a role taken by the teacher, but on occasion, the teacher may turn that role over to a student or students.
Other Elicitation	EO	An elicitation that does not fit into any of the above categories, including all forms of conversational repair. When an elicitation occurs in the middle of a student's long response, it could be coded as [EO] when it is obvious that the teacher does not intend to terminate the response but to clarify a part of response.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Finally, Content Elicitations (EC) were further subdivided into three mutually exclusive categories, as outlined in figure 68.

Figure 68
Subcategories of content elicitations

Category		Description
Yes/No	YN	Any content elicitation that requests a simple yes or no response from student(s).
Name/State	NS	Any content elicitation that requests a relatively short response, such as vocabulary, numbers, formulas, a single rule, an answer to some mathematical operation, etc. Also, an elicitation that requests a student to read a response (from a notebook, book of formulae, etc.) or that requests a student to choose among alternatives.
Describe/Explain	DE	Any elicitation that requests description of a mathematical object (rather than its label), explanation of a generated solution method (rather than an answer), or a reason why something is true or not true.

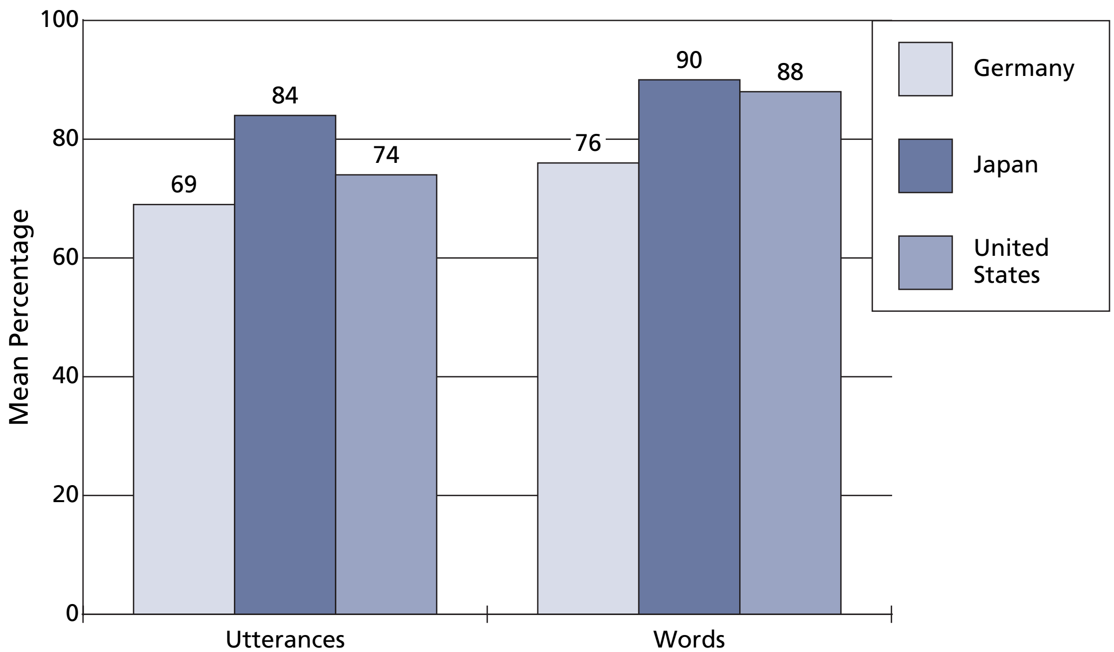
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

First-Pass Coding: Results of the Sampling Study

In the sampling study, recall, the total number of utterances coded was held constant at 30 per lesson. However, the number of these produced by teachers versus students could vary. In all three countries, teachers talked more than students, whether measured in terms of utterances or words. In figure 69, we show the average percentage of coded utterances made by the teacher and the average percentage of the total words spoken by the teacher in the 30-utterance corpus. When comparing utterances of teachers relative to those of students, German teachers talked less than U.S. teachers, who talked less than Japanese teachers. When we look at words, German teachers still talked significantly less than teachers in the other two countries, but U.S. and Japanese teachers are not distinguishable in this regard.

Figure 69

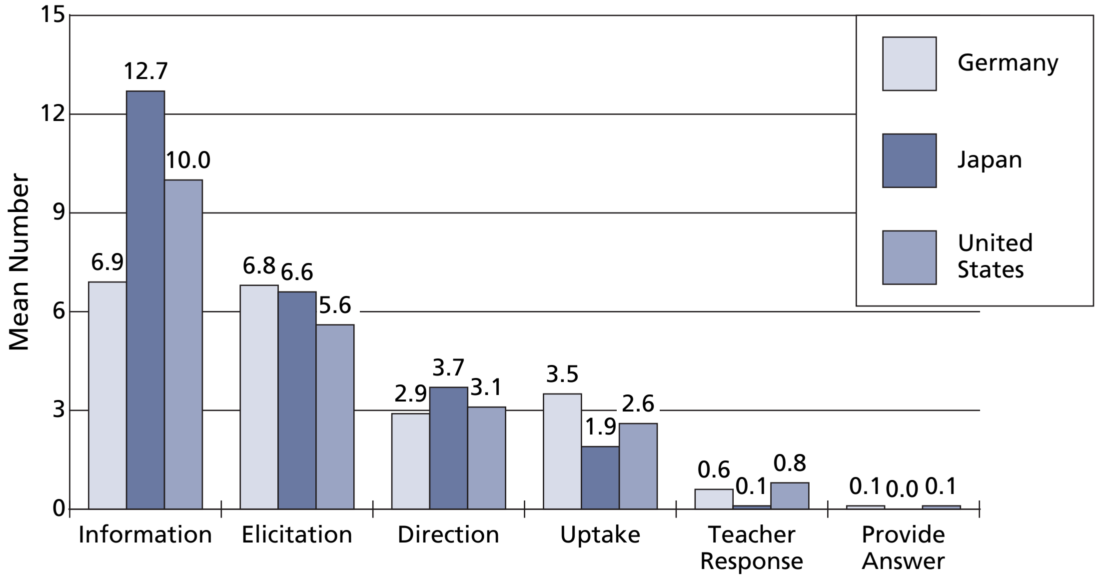
Average percentage of utterances and words spoken by teachers in each country



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

We look next at the kinds of utterances produced by teachers in the three countries. In figure 70 we show the average number of utterances (out of 30) made by teachers broken down by category. Japanese and U.S. teachers produced significantly more Information utterances than did German teachers. U.S. teachers produced significantly fewer Elicitations than German teachers, and German teachers produced significantly more Uptakes than both Japanese and U.S. teachers. German and U.S. teachers produced more Teacher Responses to student elicitations than did teachers in Japan.

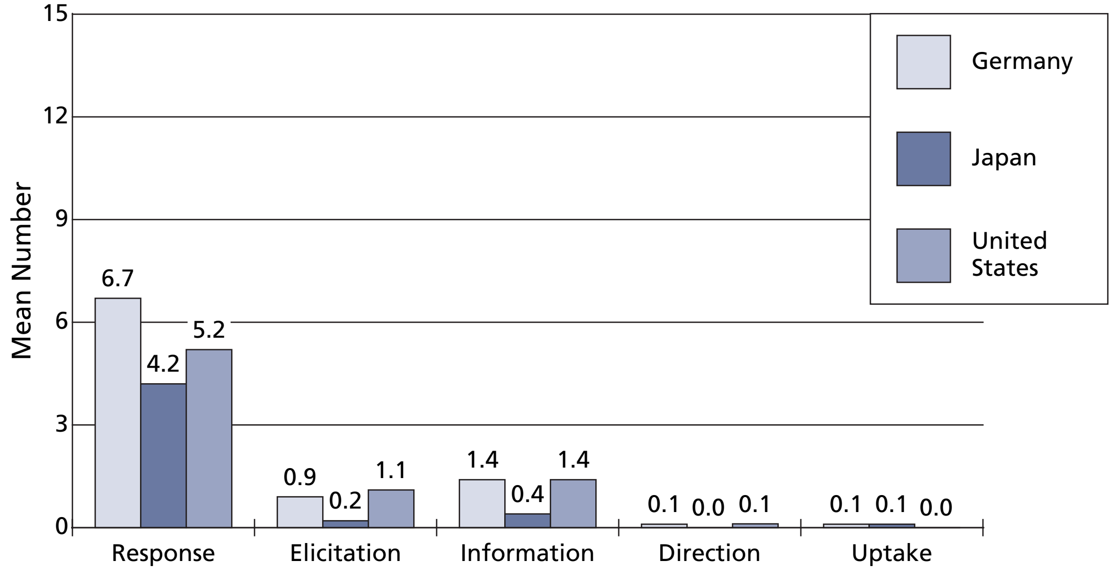
Figure 70
Average number of utterances (out of 30 sampled per lesson) coded into each of six teacher utterance categories



NOTE: Numbers less than 0.05 are rounded to 0.0.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

The distribution of student utterances is shown figure 71. The German lessons contained significantly more Student Responses than did lessons in the other two countries. German and U.S. lessons contained significantly more Student Elicitations and Student Information utterances than did Japanese lessons.

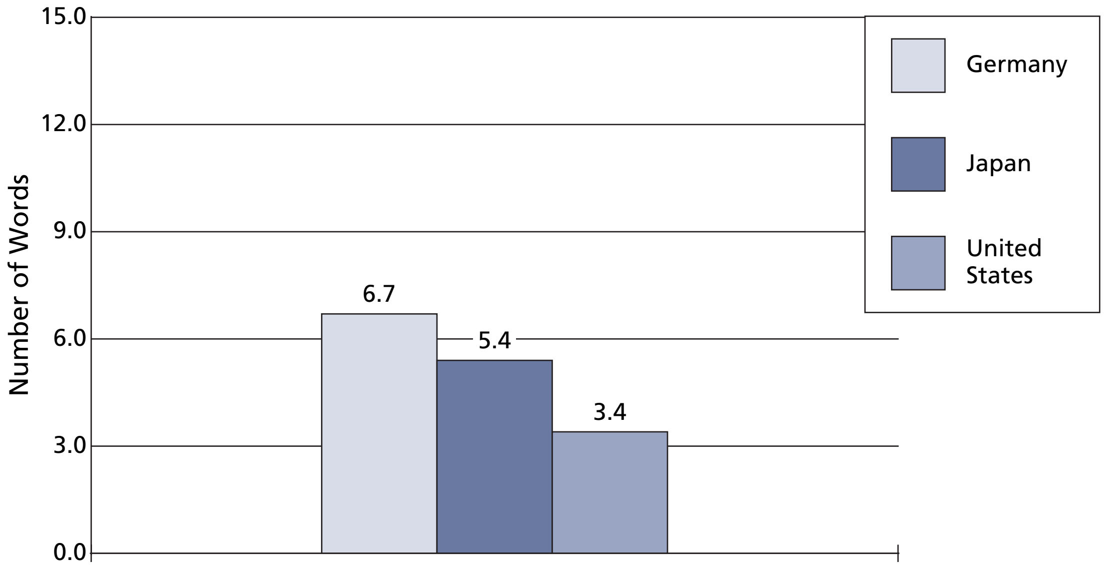
Figure 71
Average number of utterances (out of 30 sampled) coded into each of five student utterance categories



NOTE: All values less than 0.05 are rounded to 0.0.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Not only did the German students respond more frequently than Japanese and U.S. students but they also spoke at greater length during each response, as indexed by the number of words in the response. As depicted in figure 72, U.S. student responses are significantly shorter than responses produced by German students.

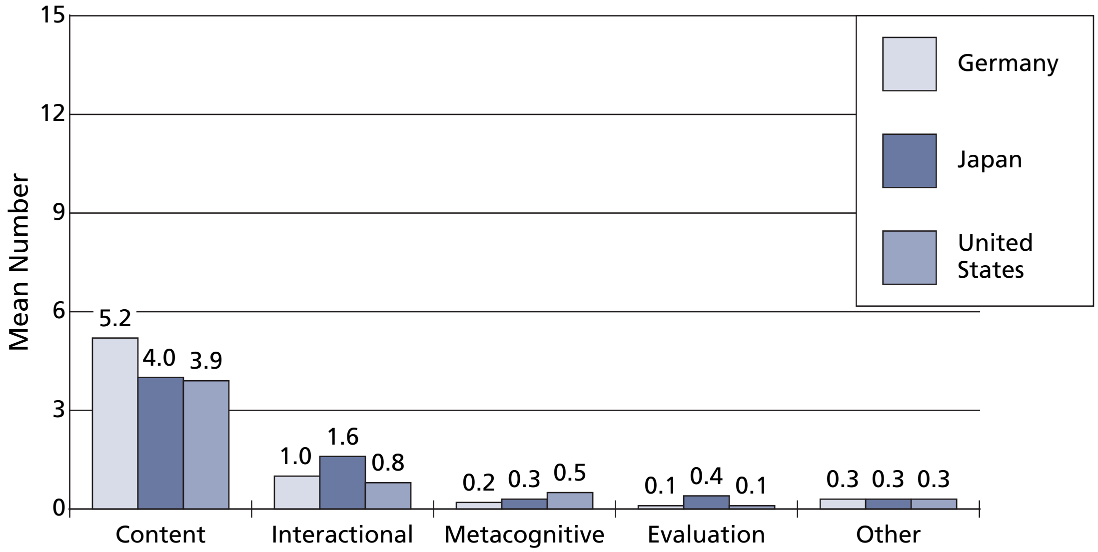
Figure 72
Average length of student responses as measured by number of words



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Let us next examine the types of Elicitations teachers produced in the three countries (figure 73). German teachers produced more Content Elicitations than did Japanese and U.S. teachers. Japanese teachers produced more Interactional Elicitations than did U.S. teachers. U.S. teachers produced more Metacognitive Elicitations than did the German teachers.

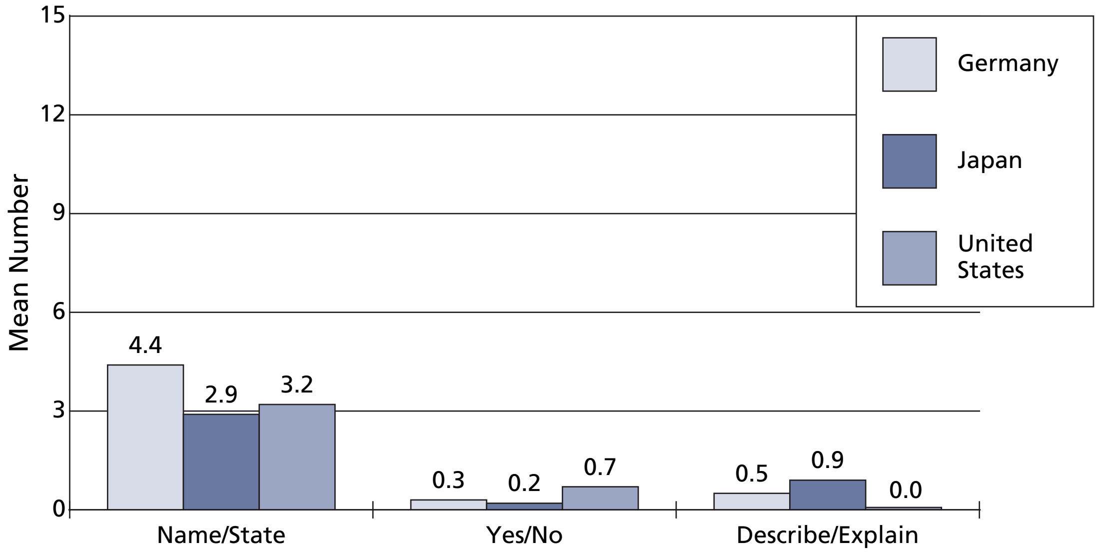
Figure 73
Average number of utterances (out of 30 sampled per lesson) coded into each of five categories of teacher elicitations



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

We were most interested in the Content Elicitations because these elicitations generate much of the mathematical content that is discussed in the lesson. In the next figure (figure 74) we show the average number of Content Elicitations (out of the 30 utterances sampled per lesson) that were coded as Name/State, Yes/No, and Describe/Explain. German teachers asked significantly more Name/State questions than did either Japanese or U.S. teachers; U.S. teachers asked significantly more Yes/No questions than did Japanese teachers; and, German and Japanese teachers asked significantly more Describe/Explain questions than did U.S. teachers.

Figure 74
Average number of utterances (out of 30 sampled) coded into each of three categories of content elicitations



NOTE: Values less than 0.05 have been rounded to 0.0.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Second-Pass Coding Categories

Whereas the first-pass coding of discourse was based on a sample of 30 utterances from each lesson in the full data set, second-pass coding was based on all the utterances in each lesson, but only for the subsample of 90 lessons used by the Math Content Group. (These 90 lessons also were coded in the first-pass coding.) Several new codes were added for second-pass coding. Content Elicitations, Information statements, and Directions were further subdivided. In addition, we started the process of grouping utterances into higher-level categories we call Elicitation-Response Sequences. We will briefly describe each of these new codes.

1. Content Elicitations. In our reading of transcripts we discovered that Content Elicitations, regardless of type, can be further divided into two categories according to their function:

Elicitation of Factual Information [Ef] is defined as any elicitation that requests a piece of mathematical information in pursuit of a correct answer. The purpose of the elicitation is for the teacher to assess whether the students know the answer or whether they are able to produce the answer. The teacher is not interested in finding out a particular student's thinking, and the response could be given by any student, even by the teacher.

Elicitation of Individual Ideas [Ei] is defined as any elicitation that requests a student to report on their individual opinions, ideas, or thinking processes. There may be a mathematically correct answer to the elicitation, but the purpose of the inquiry is for the teacher to find out what individual students have in mind. Control of the response rests more with the student than with the question itself. There is no specific response that the teacher is pursuing, and therefore it is less likely that the response is evaluated by the teacher as right or wrong.

Several examples from the data will help to illustrate this distinction. In the first two examples content elicitations were coded as Factual Information. (Letters in brackets indicate both first- and second-pass discourse codes.)

(Example 1)

[E][EC][NS][Ef]	T	*Will, what did you get for that one?*
[R]	S	Negative nine.
[EO]	T	Negative nine?
[R]	S	I mean nine.
[U]	T	Nine. All right. Good. Nine.

(Example 2)

[I]	T	All right how could you if you were...explaining to a friend you were sitting together doing homework one night and this person kept putting down N squared...and for twice...and it should be of course two N. How could you explain to that friend what they were doing wrong.
[E][EC][DE][Ef]	T	*Why...twice...is two N and not N squared? What's the difference between...in other words two N and N squared?*
[R]	S	Um you're just multiplying the number by two and you're not squaring it.
[U]	T	Okay when we have twice a number...we're just multiplying the number by two.

It is often necessary to see what kind of uptake follows the response in order to code Factual Information versus Individual Idea. When the teacher does not provide such feedback, a coder must determine whether the expected response is something that is an objective mathematical fact or something over which the respondent has ownership, so that there is no "correct" answer to the elicitation. The most common circumstances in which Individual Idea elicitations occur is when the teacher collects a variety of responses from different students without providing evaluative feedback. Below are two examples.

(Example 3)

[E][EC][NS][Ei]	T	Who else didn't get..um..*all right Carla. What did you get?*
[R]	S	I added.
[U]	T	Added.

(Example 4)

[E][EC][DE][Ei]	T	*What do you think about when you look at bread.*

2. Further categorization of Information and Direction utterances. Information and Direction utterances were each further categorized into one of four mutually exclusive categories: Content related, Managerial, Disciplinary, and Other. Brief definitions of the four categories for Information and Directions are presented in figure 75.

Figure 75
Four subcategories of information and direction utterances

Category		Description
Information/Content	IC	Information provided by the teacher that refers to mathematical concepts and/or procedures. Any information that is necessary for the students to understand the mathematical topics, tasks, and situations are coded as Content. The statement of a lesson goal and the summary of the lesson content are also coded as [IC] because they provide a cognitive structure to the students. The information must include mathematical concepts, quantities, relationships, procedures, or reasoning to be coded as [IC].
Information/Managerial	IM	Information regarding activities, experiences, or interactions rather than mathematical content. For example, when the teacher says, "Okay we start with a few calculations in our head," the students only know the upcoming activity but not the concrete mathematical operation, and therefore, this is coded as [IM]. Also, the statement such as, "Now it will get a little bit harder," is coded as [IM] because it describes the activity rather than mathematical content.
Information/Discipline	ID	Information intended by the teacher to discipline students. Information/ Other IO Information that is irrelevant to the mathematical content of the lesson or managerial/interactional aspect of content related activities is coded as [IO]. For example, when the teacher talks about World War II or about pizza, it is coded as [IO] because it does not contain any mathematical content nor is it related to mathematical activities, even though the teacher may plan to connect this information to the mathematical topic later on.
Direction/Content	DC	A direction that assigns a mathematical task to students to be carried out immediately. For example, "Okay for number two find out the value of Y" or "Write that as a unit rate" are coded as [DC].
Direction/Managerial	DM	A direction that solicits or prohibits students' physical activities except for mathematical tasks. Examples include "All right get started," "Open your books to page fourteen," or "Leave some space between that."
Direction/Discipline	DD	A direction that prohibits students' problem behavior or solicits some behavior that is appropriate in a classroom. Direction/Other DO A direction that solicits activities other than the above two types is coded as [DO].

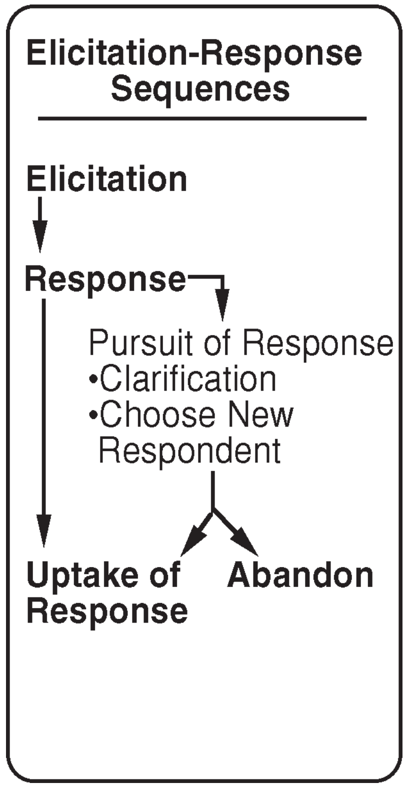
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

3. Coding of Elicitation-Response Sequences. Our next step in discourse coding was to define higher-level units we called Elicitation-Response Sequences. Discourse is organized and cannot be understood simply by characterizing the utterances. Some utterances are more important than others. Defining ER sequences was our first step in coding the organization of discourse.

Elicitation-Response Sequence [ER] was defined as a sequence of turns exchanged between the teacher and student(s) that begins with an initial elicitation and usually ends with a final uptake. The ER sequence is a cohesive unit of conversational exchange. ER sequences may consist of a single elicitation, response, and uptake, or they may consist of several of these three utterance types. They may also consist of a single Elicitation without a Response or Uptake, or of single Elicitation and Response without an Uptake.

A new ER sequence begins when there is a new Elicitation. A new Elicitation is one that requests new information. Repetitions, redirections of the initial elicitation to other students, or clarifications are not considered new elicitations. A diagram representation of the ER sequence is shown in figure 76.

Figure 76
The elicitation-response sequence



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Following are some examples.

[Example 1]

[E][EC][NS][Ef]	T	How much is in this in the beginning?
[E][EC][NS][Ef]	T	All together. How much is it?
[R]	Ss	One hundred eighty yen.
[U]	T	Yes.

[Example 2]

[E][EC][DE][Ef]	T	Okay. Someone tell me why she can't buy this dress for the prom?
[R]	S	Cause the other girl bought it.
[R]	S	The prom's already over.

[E][EO]	T	Huh?
[R]	S	It's over.
[R]	S	Nah the other girl bought it.
[U]	T	No she- no no.

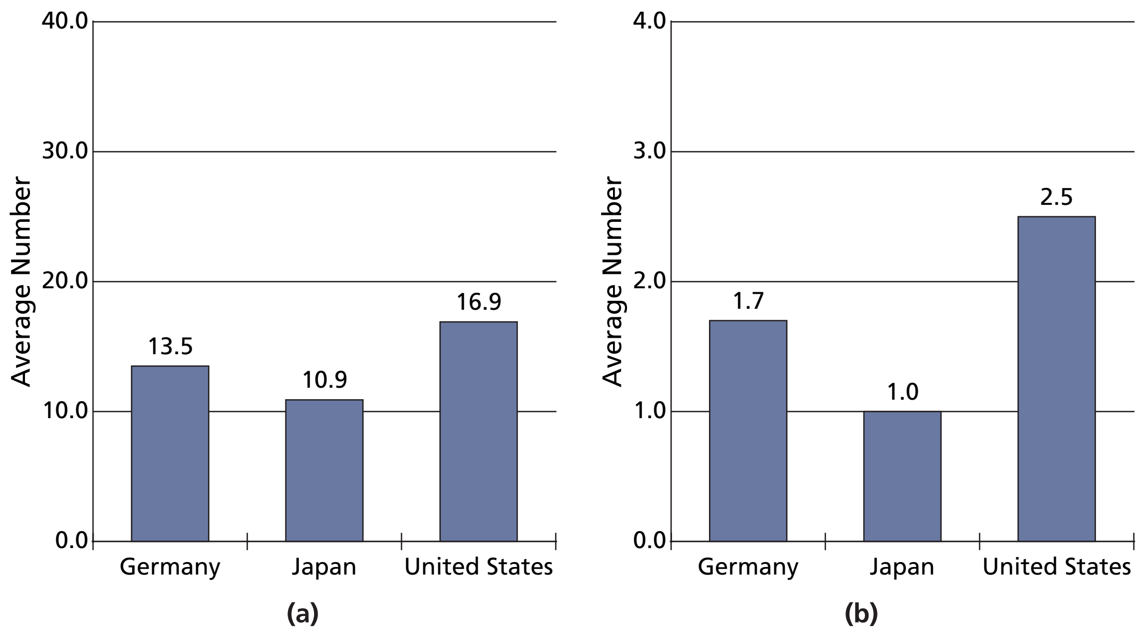
Results of Second-Pass Coding

We have not done many analyses of the second-pass coding. However, we can present a few results at this time.

In the sampling study, because we looked at only 30 utterances in each lesson, we were unable to say anything about the rate of talk in the classroom. In second-pass coding, even though the number of lessons analyzed is smaller (30 in each country), all utterances in each lesson were included in the analysis, giving us a more detailed sense of how talk occurs as the lesson unfolds. In all, for the 90 lessons, we entered more than 42,000 discourse codes.

The first analysis of interest concerns the rate of talk in the classrooms of each country. In figure 77 we show (panel a) the average number of discourse codes in each lesson (excluding Elicitation-Response Sequences) divided by the number of minutes in Classwork. U.S. classrooms had a higher rate than Japanese classrooms. In panel (b) of the figure we show the average number of Elicitation-Response Sequences divided by the number of minutes of Classwork. The rate in the United States is highest, and in Japan, lowest, among the three countries.

Figure 77
(a) Average number of discourse codes per minute of classwork in the three countries; (b) average number of elicitation-response sequences per minute of classwork in the three countries

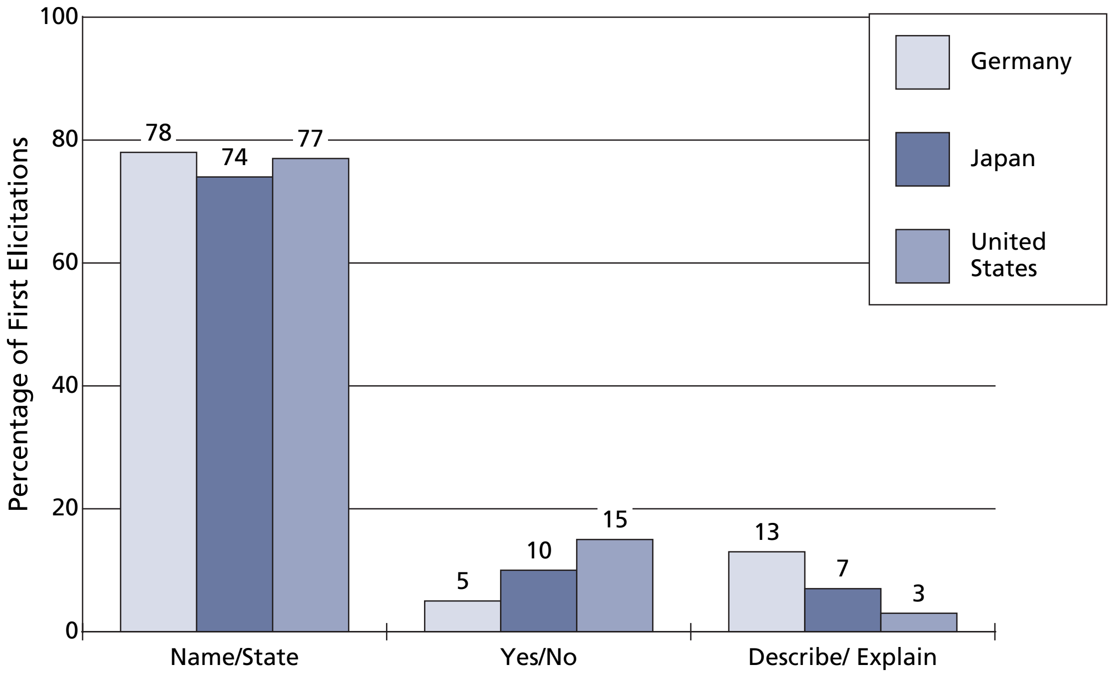


SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Not only was the rate of talk slower in Japanese classrooms but also the length of ER sequences was greater than in the United States. The average number of discourse codes per ER sequence was 9.6 in Germany, 12.1 in Japan, and 7.2 in the United States. This means that Japanese teachers stuck with a question longer than did U.S. teachers before moving to the next question.[3]

Whereas in the sampling study we counted utterances regardless of their importance in the lesson discourse, in this analysis we were able to take into account the fact that not all utterances are equally important. Specifically, we assumed that the first elicitation in an Elicitation-Response Sequence will be more significant than the follow-up elicitations. The following two figures show the average percentage of initiating elicitations of ER sequences that were of various types. In figure 78 we look at First Elicitation/Content Elicitations that were judged to be eliciting a fact or correct answer. The incidence of Name/State elicitations did not differ across countries. Lessons in the United States were more likely to contain Yes/No elicitations than those in Germany; German and Japanese lessons were more likely to contain Describe/Explain elicitations than U.S. lessons.

Figure 78
Average percentage of initiating elicitations of elicitation-response sequences in each country: Content-related elicitations seeking facts

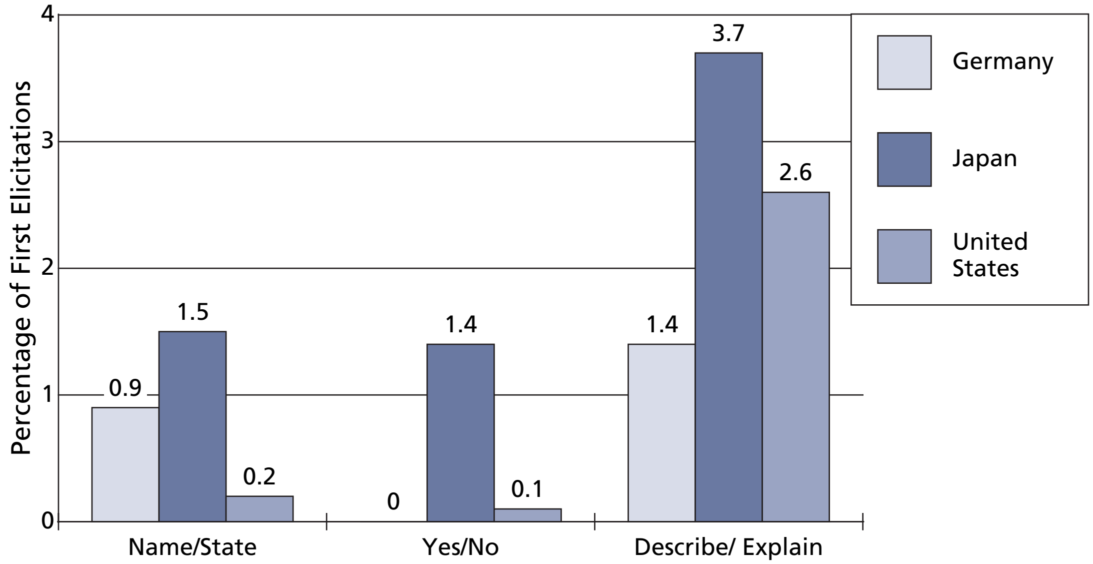


SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

In figure 79 we show the same graph but for Content Elicitations judged to elicit individual student ideas. Analyses revealed no differences between the three countries.

[3] Standard errors for Germany, Japan, and the United States were 1.25, 0.62, and 0.50, respectively.

Figure 79
Average percentage of initiating elicitations of elicitation-response sequences in each country: Content-related elicitations seeking individual ideas



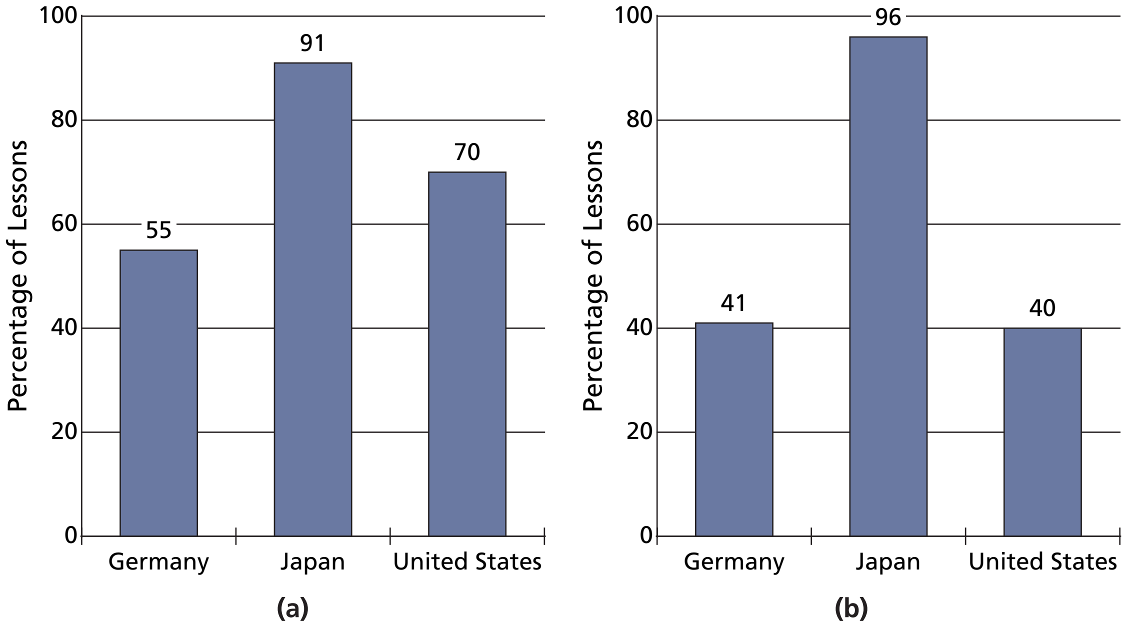
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Explicit Linking and the Coherence of the Lesson

Language can serve many functions in a mathematics lesson. One of these is to explicitly link together ideas and experiences that the teacher wants students to understand in relation to each other. Using the subsample of 90 lessons coded by the Math Content Group, we coded two kinds of linking: Linking across lessons and linking within a single lesson. We defined linking as an explicit verbal reference by the teacher to ideas or events from another lesson or part of the lesson. The reference had to be concrete (i.e., referring to a particular time, not to some general idea). And, the reference had to be related to the current activity.

The results of this coding are displayed in figure 80. The highest incidence of both kinds of linking—across lessons and within lessons—was found in Japan. Indeed, teachers of Japanese lessons linked across lessons significantly more than did teachers of German lessons, and linked within lessons significantly more than teachers of both German and U.S. lessons.

Figure 80
Percentage of lessons that include explicit linking by the teacher (a) to ideas or events in a different lesson, and (b) to ideas or events in the current lesson



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Our observations of the videotapes reveal that the Japanese teachers appear to use linking in a systematic, almost routinized, way. They tend to start the lesson by recalling or reviewing what was done in the previous lesson. JP-036, for example, opened with the teacher saying, "Then, um yesterday the last part on ratio we did two characteristics, but characteristics that can be changed into a multiplied form, right? Please summarize what we practiced." The same lesson closed with the teacher saying, "Well we could not do concrete practice so I would like to do it using this in the next class period. I also want to review and go over number five in the next class period." Twenty-six of the 30 Japanese lessons included linking to a past lesson, and 19 included linking to a future lesson.

Twenty-nine of the 30 Japanese lessons included linking to different parts of the same lesson. In one example, the teacher referred back to a statement made by a student several minutes earlier: "Just now you've heard the opinion of your friend and um while using that as a reference you can continue on your computer or um you can continue on your computers by reading ahead yourselves or you can prove it while getting hints from the computers. Okay? Or you can prove it by um expressing your opinions um with your friends." These kinds of statements were far less common in German and U.S. lessons.

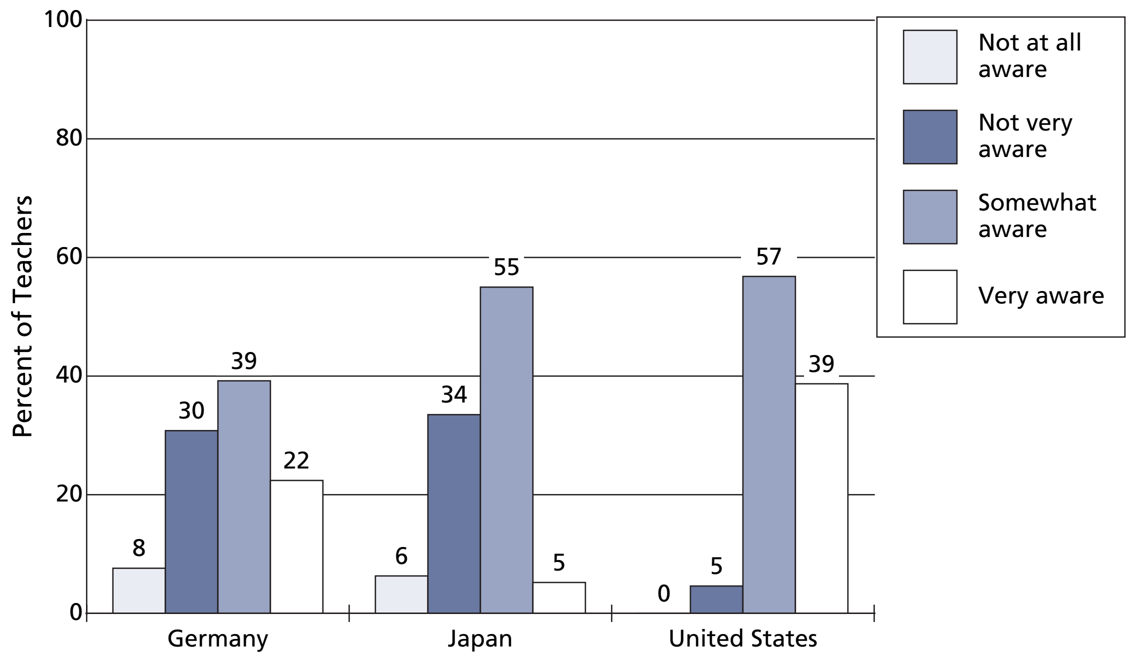
Chapter 6. Teachers and Reform

One goal of this study was to determine the extent to which U.S. teachers have been influenced by current ideas about the teaching of mathematics. Reform documents—most notably the NCTM *Professional Standards for Teaching Mathematics* (1991)—provide guidance on how to teach mathematics in the classroom, or at least on what features of instruction ought to be evident in the mathematics classroom. Many of our codes were inspired by these reform ideas. In this section we will examine how the teachers in our sample think of themselves in relation to current reform ideas, both in general and in relation to the lesson we videotaped. Analyses presented in this section are based on answers given in the Videotape Classroom Study teacher questionnaire.

GENERAL EVALUATIONS

Although the question clearly has a different meaning across cultures, where ideas about education are communicated to teachers in quite different ways, it nevertheless is interesting to see how teachers responded to the question, "How aware do you feel you are of current ideas about the teaching and learning of mathematics?" The teachers could answer "Very Aware," "Somewhat Aware," "Not Very Aware," or "Not at All Aware." The results are presented in figure 81. The distributions of responses to this question differed significantly across countries. Thirty-nine percent of teachers of U.S. lessons report being "Very aware" of current ideas; 5 percent of those teaching Japanese lessons indicated this level of awareness.

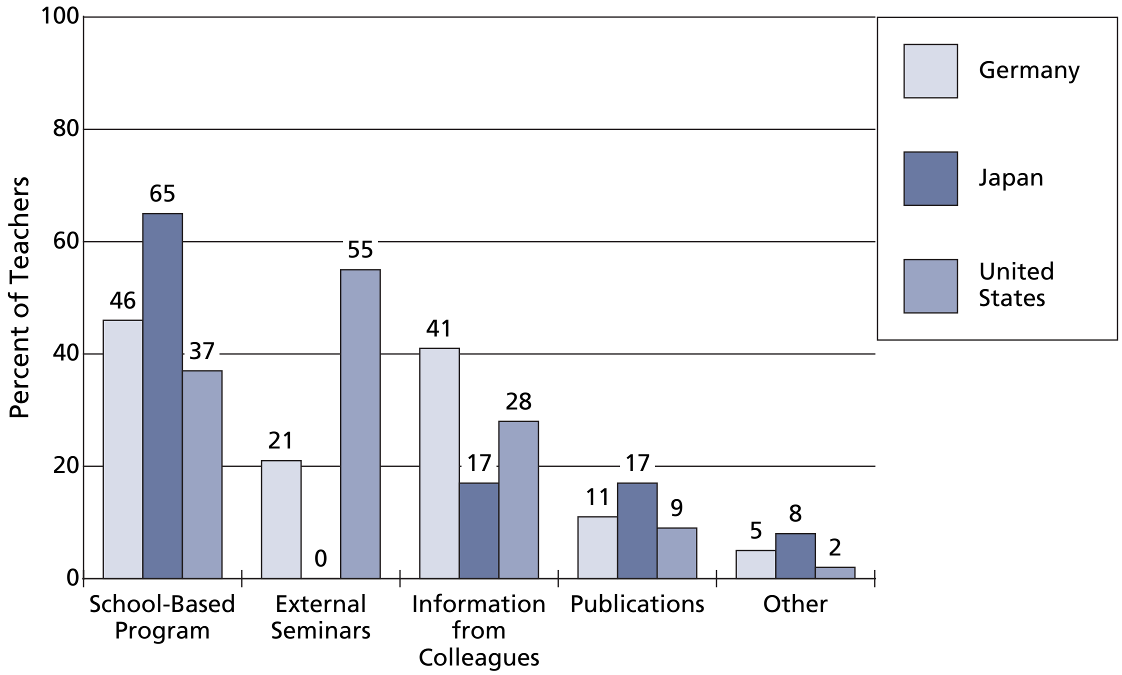
Figure 81
Teachers' ratings of how aware they are of current ideas about the teaching and learning of mathematics



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

We asked teachers how they usually hear about current ideas about the teaching and learning of mathematics. Responses were open-ended; we coded them into five categories: School-Based Programs, Information from Colleagues, External Seminars, Publications, and Other. Percentages of teachers in each country who included responses in each category are shown in figure 82. Significantly more Japanese than U.S. teachers mentioned school-based programs. Significantly more U.S. teachers than German teachers, and more German teachers than Japanese teachers, mentioned attending external seminars or workshops. More German teachers than Japanese teachers mentioned their colleagues as a source of knowledge.

Figure 82
Teachers' responses when asked where they get information regarding current ideas about the teaching and learning of mathematics



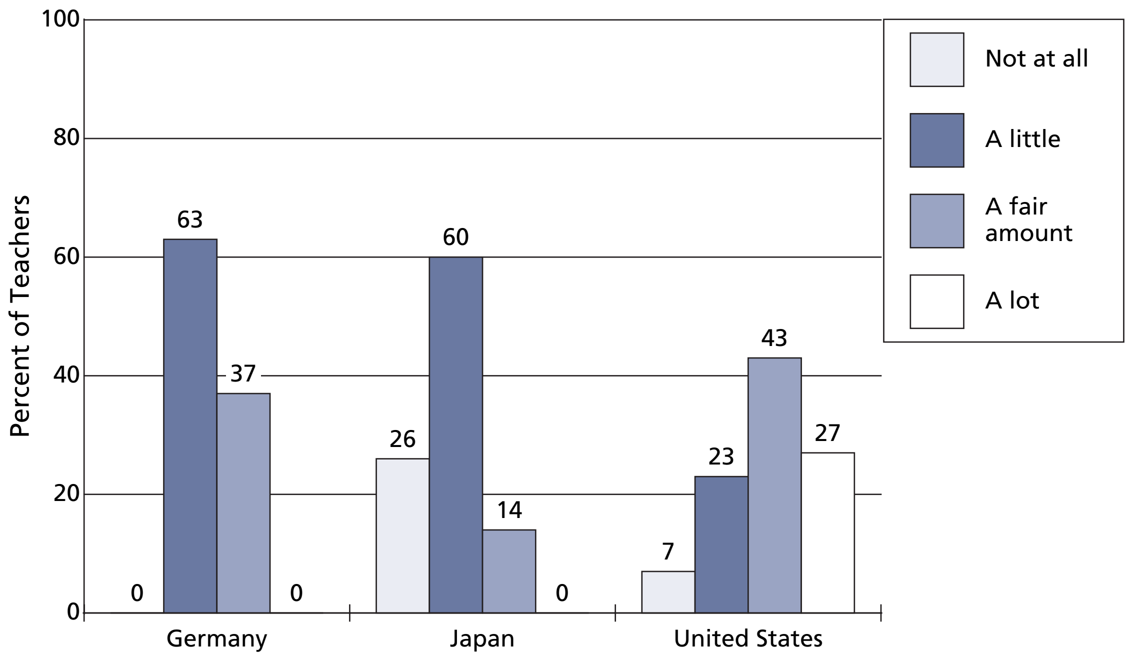
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

U.S. teachers were asked what written documents or materials they had read to stay informed about current ideas. Dozens of documents were mentioned. Thirty-three percent mentioned the NCTM standards by name (either the *Curriculum and Evaluation Standards for School Mathematics* or the *Professional Standards for Teaching Mathematics*). Forty percent mentioned some other NCTM publication by name. And 59 percent mentioned either the Standards or some other NCTM publication.

EVALUATIONS OF THE VIDEOTAPED LESSONS IN TERMS OF CURRENT IDEAS

Claiming to be aware of current ideas is one thing; implementing those ideas in the classroom is another thing entirely. Still, a large percentage of U.S. teachers reported that the lesson we videotaped was in accord with current ideas about teaching and learning mathematics. Bearing in mind that "current ideas" may differ between Germany, Japan, and the United States, we nevertheless asked teachers to specifically evaluate their own videotaped lesson in terms of current ideas. Teachers could say that it was "not at all" in accord with current ideas, "a little" in accord, "a fair amount," or "a lot." Twenty-seven percent of the U.S. teachers responded "a lot," and 70 percent responded either "a lot" or "a fair amount." None of the German or Japanese teachers responded "a lot," and 37 percent (German) and 14 percent (Japanese) responded "a fair amount" (figure 83).

Figure 83
Teachers' perceptions regarding the extent to which the videotaped lesson was in accord with current ideas about the teaching and learning of mathematics



NOTE: Due to a clerical error in which an unweighted average was mistakenly substituted for a weighted average, the numbers in this graph differ slightly from those reported in Peak (1996: page 46). Specifically, the total of "a fair amount" plus "a lot" for the U.S. teachers was reported as "almost 75 percent." The correct total should be 70 percent, as shown in this graph.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Teachers who said that the videotaped lesson was either "a lot" or "a fair amount" in accord with current ideas about the teaching and learning of mathematics were further asked to justify their responses by citing specific aspects of the lesson that exemplified these ideas. This gave us the opportunity to see which aspects of the lesson teachers focused on and to see what, in the video, could be found to connect with their descriptions.

We analyzed these responses only for the U.S. teachers. Although the range and variety of responses to this question were great, the vast majority of teachers' responses fell into three categories:

- *Real-World/Hands-On.* Thirty-eight percent of the U.S. teachers gave answers that (1) focused on the application of math to daily life, or (2) involved the use of physical or manipulative representations of mathematical concepts. For example: "The four problems dealt with temperature in Anchorage, Alaska. This gave me a chance to relate mathematics to everyday life."
- *Cooperative Learning.* Thirty-one percent of the U.S. teachers mentioned cooperative learning in their answer. One teacher, for example, mentioned her practice of having "study buddies" where students pair up to work together; other teachers pointed to their use of peer tutoring, having students explain answers to each other.
- *Focus on Thinking.* Finally, 19 percent of the U.S. teachers mentioned a focus on thinking, specifically conceptual understanding, a focus on process over product, or a focus on problem solving. Some of these teachers specifically contrasted this focus with one that emphasizes computational skills.

In general, we can see what teachers are talking about when we review their videotapes: All teachers who pointed to real-world applications did include such applications in their lessons, and the same was true for cooperative learning. Whether or not these features led to lessons that were, in fact, more in line with those envisioned by reformers is a question we shall return to.

U.S. REFORM IN CROSS-CULTURAL PERSPECTIVE

Although it is unclear exactly what is meant by "reform," or even "current ideas" in the context of Germany and Japan, it is quite clear in the United States. Thanks to the influence of the National Council of Teachers of Mathematics, we have some fairly well-specified ideas about what the mathematics classroom should look like, and many teachers claim to be familiar with these ideas. Furthermore, the majority of teachers in our video sample believed that we would find evidence of these current ideas in the lessons we had videotaped. Is there any evidence, in our data, that U.S. teachers are, in fact, implementing these ideas in their classrooms?

Although most of the current ideas stated in such documents as the NCTM *Curriculum and Evaluation Standards for School Mathematics* (1989) and the NCTM *Professional Standards for Teaching Mathematics* (1991) are not operationalized to the extent that they could be directly coded, it is possible to use some of the indicators we have developed in conjunction with these current ideas. When we view our data in this way, we come to this conclusion: Japanese classrooms, on average, appear to more closely exemplify current ideas advanced by U.S. reformers than do classrooms in the United States and Germany. Because the reform ideas considered here emerged from the United States, we limit our discussion to a consideration of the contrasts between the United States and Japan.

Let us take a couple of examples. In both of the NCTM documents just mentioned, problem solving is proposed as the central focus of curriculum, teaching, and learning. The *Curriculum and Evaluation Standards for School Mathematics* states that, "In grades 5-8, the mathematics curriculum should include numerous and varied experiences with problem solving as a method of inquiry and application so that students can use problem solving approaches to *[among other things]* investigate and understand mathematical content; ...develop and apply a variety of strategies to solve problems, with emphasis on multistep and nonroutine problems..." (NCTM 1989, page 75. Italics added.). Similarly, the NCTM *Professional*

Standards for Teaching Mathematics proposes the posing of "worthwhile mathematical tasks" as the first standard for teaching. Worthwhile mathematical tasks are further defined as tasks based on "sound and significant mathematics" that "engage students' intellect; develop students' mathematical understandings and skills; stimulate students to make connections and develop a coherent framework for mathematical ideas; call for problem formulation, problem solving, and mathematical reasoning" (NCTM 1991, page 25).

Several indicators in our study point to the greater consistency of Japanese lessons in terms of these criteria. Content analyses showed that Japanese lessons included more advanced levels of mathematics, and that the mathematics was presented in a more coherent way than in U.S. lessons (see, for example, the analyses by the Math Content Group). Japanese lessons included more emphasis on concepts than U.S. lessons, and were more likely to develop instead of merely state the concepts. Japanese teachers also were more likely than U.S. teachers to make explicit the connections within a lesson. These facts would appear to give Japanese students an advantage in the quest to "make connections and develop a framework for mathematical ideas." Finally, our analyses of performance expectations of tasks posed during seatwork showed that Japanese students, more than U.S. students, were engaged in genuine problem solving during the lesson, rather than simply the application and practice of routine problem-solving skills.

Let us take another example from the reform documents. Both of the NCTM documents we are discussing place communication and discourse at the center of their proposed reforms. The *Curriculum and Evaluation Standards for School Mathematics* states that the study of mathematics should include opportunities to communicate so that students can, for example, "reflect on and clarify their own thinking about mathematical ideas and situations" (NCTM 1989, page 78). The *Professional Standards for Teaching Mathematics* (1991) devotes three of its six teaching standards to discourse. Teachers, according to the document, should "orchestrate discourse by posing questions and tasks that elicit, engage, and challenge each student's thinking" (page 35). Students should "listen to, respond to, and question the teacher and one another," and "make conjectures and present solutions" (page 45).

Although we have only completed a rudimentary analysis of classroom discourse, we already can find some evidence that Japanese teachers, more than U.S. teachers, orchestrate the kind of discourse called for in these reform documents. For example, we find Japanese teachers asking more describe/explain questions, and fewer yes/no questions, than U.S. teachers. Also relevant is the analysis of student-generated solution methods, which occurs more frequently in Japan than in the United States. The reason for this pattern is clear: Japanese teachers often have students struggle with a problem for which they have not yet been taught a solution, then present the solutions they generated to their classmates. Presentation and discussion of alternative solution methods may provide a natural opportunity for engaging in the kind of mathematical discourse reformers are seeking to foster.

Of course Japanese teachers may not teach the way they do because they are following the recommendations of U.S. reformers. And it is also worth pointing out that there are some respects in which they do not appear to teach in accordance to the proposals of U.S. reformers. For example, Japanese teachers engaged in far more direct lecturing/demonstration than U.S. teachers—a practice frowned on by reformers. And, contrary to specific recommendations made in the NCTM *Professional Standards for Teaching Mathematics,* Japanese teachers never were observed using calculators in the classroom.

REFORM IN THE U.S. CLASSROOM: OBSERVATIONAL INDICATORS

We have seen that on many of the variables coded, Japanese teachers teach more in accord with current U.S. reform ideas than do U.S. teachers. But what about differences among U.S. teachers? Do some teachers show more evidence of reform than others? We have already shown that the majority of the U.S. teachers in our sample felt that their lesson was in accord with current ideas. Of course, there were some U.S. teachers who did not feel this way. Could we find some differences among teachers who did and did not report implementation of current ideas in their classrooms?

For purposes of analysis we defined two groups of U.S. teachers: One group (N=22) responded "a lot" when asked the degree to which the videotaped lesson was in accord with current ideas (i.e., those responding "a fair amount" were excluded from this analysis); the other group (N=19) answered either "a little" or "not at all." We can call the first group Reformers, the second, Non-Reformers. We compared the classrooms of these two groups of teachers on all of the variables discussed earlier in this report.

Overall, the analyses revealed very few significant differences between the two groups of teachers. Although this lack of differences may be due, in part, to the lack of statistical power given the small size of our sample, we do not believe this is the primary reason. In our own viewing of the tapes, we did not see a strong distinction between these groups.

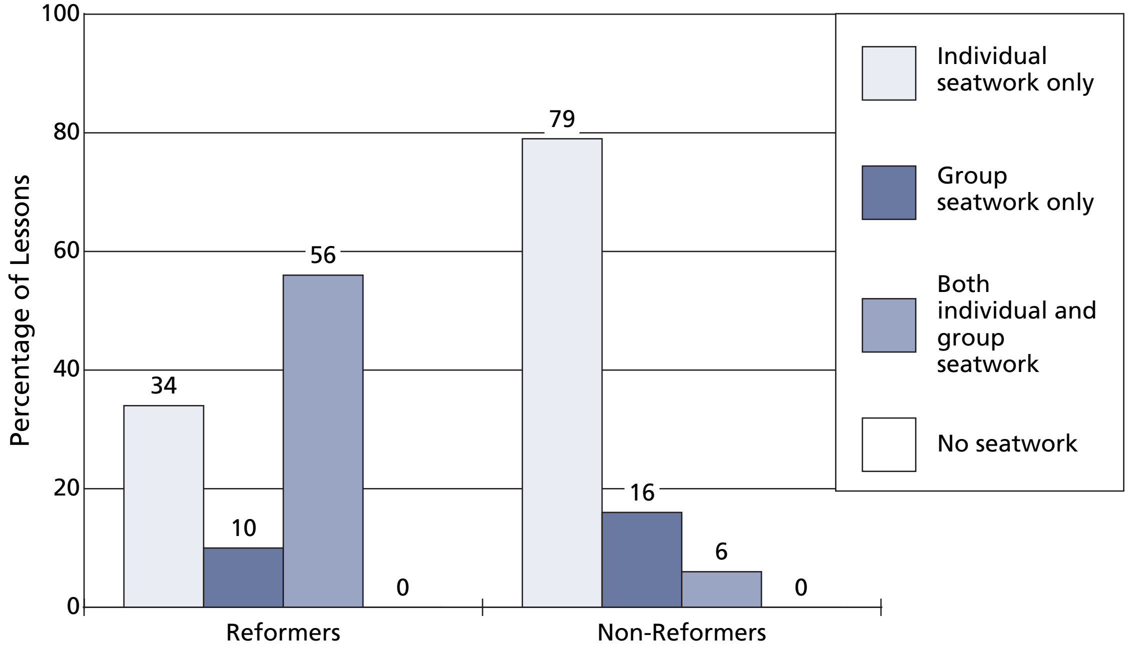
Statistically significant differences did emerge, however, in organization of the lesson and materials.

Organization of the Lesson

As a group, Reformers spent a higher percentage of their time in seatwork than did Non-Reformers. Reformers spent 43.1 percent of lesson time in seatwork, compared with 28.6 percent for Non-Reformers. (The average for all U.S. teachers was 37.3 percent.)[1]

[1] Standard errors for Reformers and Non-Reformers were 3.99 and 5.26, respectively.

Figure 84
Percentage of lessons among Reformers and Non-Reformers in the United States in which seatwork of various kinds occurred



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

The two groups also differed in the kind of seatwork in the lesson: individual, group, or a combination of the two. In figure 84 we show the percentage of lessons in the two groups that included individual seatwork only, group seatwork only, both kinds of seatwork, or no seatwork. We can see that the Reformers were far more likely to include both individual and group seatwork in their lessons than were the Non-Reformers, while they were less likely to use individual seatwork only.

Instructional Materials

Finally, we found one significant difference in the tendency of U.S. Reformers and Non-Reformers to use certain types of materials in the classroom. Specifically, Reformers were less likely than Non-Reformers to use textbooks in the lesson: 18 percent of lessons for Reformers, and 63 percent for Non-Reformers.[2]

REFORM IN THE CLASSROOM: QUALITATIVE ANALYSES

Other than these areas of differences, we have little quantitative evidence that reform teachers in the United States differ much from those who claim not to be reformers. Most of the comparisons were not significant. But it is useful to look more qualitatively at the lessons taught by Reformers and Non-Reformers. Our conclusion is interesting: It is true that teachers who cite features of instruction, such as the use of real-world problems or cooperative learning, do implement such features in the lessons we videotaped. However, these features alone do not necessarily indicate the quality of instruction as intended by the NCTM standards, and in fact may only bear a superficial relationship to the quality of instruc-

[2] Standard errors for Reformers and Non-Reformers were 8.56 and 8.44, respectively.

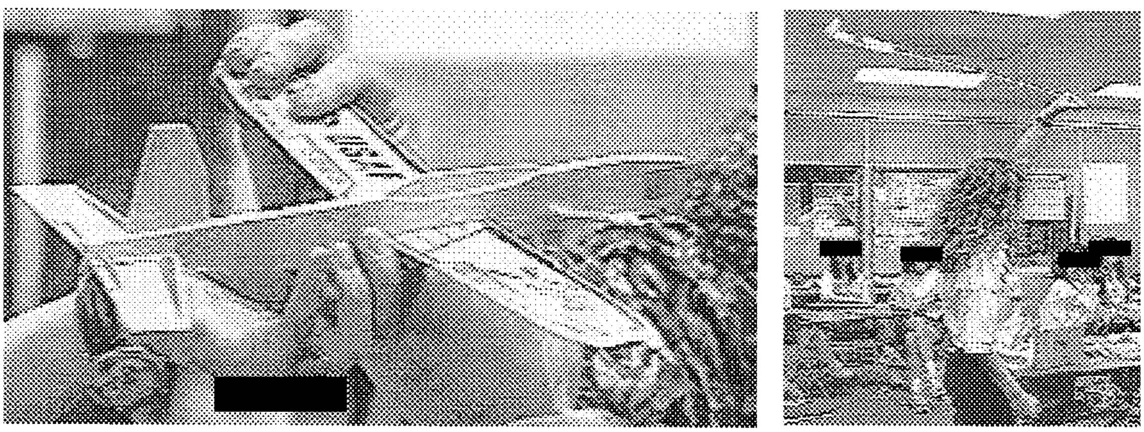
tion. Similarly, a high quality lesson could be constructed that did not contain these features. Quality of mathematical activity depends on how features are implemented. Let us explore a few examples.

Example 1: Airplane on a String (US-060)

US-060 was taught by a teacher who judged it to be a good example of current ideas about the teaching and learning of mathematics. The lesson included a real-world problem situation and a period of cooperative group work. It also included a writing assignment in which students were asked to reflect on what they learned in the lesson. We agreed with the teacher's assessment of the lesson, and we will try to explain why.

The lesson started with the teacher asking for a volunteer to come to the front of the room and swing a model airplane on the end of a string around her head in a circular motion (figure 85). Everyone appeared attentive, and enthusiasm was high as students wondered what the airplane on the string would have to do with mathematics.

Figure 85
Frames from the video of US-060



SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

The teacher then posed a question: How fast is the plane going? Many students gave answers, which the teacher wrote on the overhead projector: "About thirty miles per hour." "Fifteen miles per hour." "Two-hundred thirty." And so on. Why are the answers in miles per hour, the teacher wondered aloud. What would be a rate we could measure? Revolutions per minute? Per second? If we measured such things, could we convert them to miles per hour? Discussion of these questions took up the first 10 minutes of the lesson.

After the teacher could see that most students in the class understood the problem and had thought through some of the issues involved in solving the problem, she had them work in groups for the next 30 minutes. Each group had its own airplane and string, and each worked to measure the speed of the plane. The general strategy adopted by all groups was the same: Measure the radius of the circle from the plane to the point around which the plane was swung; then, calculate the circumference in inches, count the number of times the plane traversed the circumference in a set amount of time, and get an answer in inches per second or per minute; finally, convert the answer into miles per hour. Students appeared to be very involved in the activity.

As students worked, the teacher circulated and posed a new question to each group: If a bird were sitting on the string, halfway between the plane and the center of the circle, would it be traveling the same speed as the plane, faster, or slower? Groups actively began to discuss this possibility, with differing opinions offered and justified. One student noticed that her hand was not completely still as she twirled the plane, and that this must have affected the radius of the circle. The teacher asked the group to consider how they might need to adjust their method to take this problem into account.

Thirty-nine minutes into the lesson students convened again as a class. Over the next 12 minutes all groups presented their answers, then two groups presented their solution methods to the class. Someone brought up the bird: how fast would the bird be going? The teacher told them that they would discuss this tomorrow, then handed out a writing assignment for homework. The writing assignment asked students to describe the problem they had worked on, then summarize the approach their group took to solving the problem. It also asked them to write about the role they played, specifically, in the group's work.

We thought this was in line with NCTM standards for several reasons. Students were engaged in a rich mathematical problem that appeared to be perceived as a real problem by most of the students in the class. The problem was closely tied to mathematical concepts—circumference and radius of a circle, and rate. The task encouraged students to make connections among these concepts and between these concepts and a real-world domain. Students were encouraged to come up with their own ways of solving the problem, and much of the lesson focused on discussing the validity of the methods they devised. Finally, students were left to reflect on their activity and to ponder a new dimension, the addition of the bird to the string. There was a clear sense of what the next step would be as the class pursued the topic further.

Example 2: The Game of Pig (US-071)

US-071 was also judged by the teacher to exemplify current ideas about the teaching and learning of mathematics. On a superficial level, the lesson had much in common with the one we just described: It included a hands-on learning experience, working in groups, and a writing assignment. However, we judged this to be less in line with NCTM standards than the previous lesson. Let us try to explore the reason for our judgment.

The lesson started by asking students to recall the game of Pig, which they had played previously. Take 5 minutes, the teacher told the students, and write down everything you remember about the rules of the game.

The game of Pig is played with dice. Students work in groups. One student rolls a pair of dice, and all students who are playing receive the number of points that is the product of the two numbers rolled. The process is repeated, and the number of points on each turn is added to each player's total. However, if a one is rolled, players receive no points for that turn; and if two ones are rolled, players lose all of their points and have to start over. Players may elect to stop playing at any time, in which case they are left with whatever number of points they have accumulated up to that point.

After students completed the 5-minute writing assignment, the teacher went over the rules of the game. She then told the students that today they would play the game twice, first with 6-sided dice, then with 10-sided dice. "We will try to see if you can get a higher score playing with 10-sided dice than with 6," the teacher remarked. She told students to prepare a score sheet, then handed out the 6-sided dice to each group of students and the play began. Halfway through the period the teacher handed out the 10-sided dice. At the end of the lesson, students were given a 10-minute writing assignment:[3]

0:38:51	T	Thank you. Hands down. Okay. Um looks like just about everybody is finished with their five rounds. Um- you need to come up with a grand total ... for- what was your total score for six sided dice- what was your total score for ten sided dice.
0:39:18	S	Awwww.
0:39:44	T	Okay. Back of your score sheet.
0:40:16	T	Okay. Um we are going to do compare and contrast. We've done a lot of this. Should be relatively easy for you. You are going to start with the Venn diagram //... to summarize your ideas. Then you're going to put it into a paragraph. Okay.
0:40:29	S	// ( )
0:40:54	T	So ... the things that you found out about playing Pig with six sided dice- um gentlemen ... I'm not going to repeat myself- go in the first circle ... um the things that are true of ten sided dice ... Pig ... go in second circle. The things that are 0:41:46 T Okay. So these are the main ideas I want you to include. ... Okay the rules - Yes ma'am. - Probability ... and your total scores.

This teacher reported, on her questionnaire, that this lesson dealt with probability and uncertainty. When asked what she wanted students to learn from the lesson she wrote: "How does theoretical probability compare with actual experimental data?" And indeed, it would seem possible to develop a lesson on probability that involved comparing 6-sided with 10-sided dice. However, when we watch the lesson and read the transcript in detail, we find no evidence that this lesson involved probability. Virtually all of the mathematical talk that went on in the lesson concerned multiplication of single-digit numbers and addition (i.e., those operations required to keep score). In our judgment, that was the mathematics that students were getting out of this lesson.

Twice the teacher brought up probability for discussion, but in each case, she failed to pursue the discussion for more than a few seconds. The first such instance occurred near the beginning of the lesson when the teacher was explaining the rules of the game. She was discussing what happens when double-ones are rolled on the dice:

0:07:07	T	If the one comes // up you lose all your points for that round ... // if you're still playing. But if you've already circled your total and decided you want to stay there ... you get to keep it. Okay? How about if I roll again and I get double ones?
0:07:08	S	// Oh.
0:07:10	S	Oh.
0:07:23	S	You lose it. You lose all your cash.
0:07:26	T	Okay. But what if you're not playing right then?

[3] A double slash mark (//) on two succeeding turns in a transcript indicates overlap in speech. Thus, the // in the beginning of the turn at 40:29 indicates that this utterance (which was inaudible, as indicated by the blank space between the parentheses) started right after the teacher said "Venn diagram" in the preceding turn. An explanation of this and other transcription conventions is included as appendix F.

0:07:30	Ss	( )
0:07:31	T	Then you're safe. Okay. So. Um ... what did we decide the probability was that we could come up with a one on one die?
0:07:42	Ss	One out of six.
0:07:43	T	One out of six. Okay. And what was the probability that you'd come up with double ones?
0:07:49	S	Um two out of six- uh two- two out of twelve.
0:07:51	T	Okay. Remember. What are all the different combinations? How many different combinations are there?
0:07:56	S	Three.
0:07:57	Ss	// Thirty six.
0:07:57	S	// I mean thirty six.
0:07:57	T	// Thirty six. So there's only one way to get double ones right? So what's the probability of getting double ones?
0:08:04	Ss	One out of thirty six.
0:08:05	T	There you go.
0:08:06	S	One out of thirty six?
0:08:07	T	Okay. So that's the game basically.

There was no attempt on the teacher's part to explain how knowing the probability of a dice throw might influence the playing of the game. No additional problem was posed for which students could use probability, and there was no discussion of probability throughout the rest of the lesson.

The other point at which the teacher asked a question that could have led to a discussion of probability happened 34 minutes into the lesson. The teacher suggested that students who kept ending up with a score of zero might want to reconsider their strategy. However, the point was dropped after this isolated comment and did not lead into any discussion of probability, game strategy, or the relation between the two. It is interesting that the teacher raised the question of strategy. One could, from there, potentially get into a discussion of probability theory. However, there really is no clear strategy for this game, which makes it harder for the teacher to get a discussion going about strategy.

At the end of the questionnaire we asked the teacher which part of the lesson exemplifies current ideas about the teaching and learning of mathematics and why. This teacher wrote:

> Students were involved in a hands-on, interactive activity. They were allowed to come to their own conclusions about their experience. They were required to communicate their experience to others, both verbally and in writing.

It is clear to us that the features this teacher uses to define high quality instruction can occur in the absence of deep mathematical engagement on the part of the students.

Example 3: A Non-Reformer (US-062)

Let us briefly present one more example, this from a teacher who, although she claimed to be "very aware" of current ideas about the teaching and learning of mathematics, judged her videotaped lesson to be "only a little" in accord with these current ideas. The lesson dealt with factoring of polynomials, and discussed factoring in the context of slope and the solving of simultaneous equations. This teacher stated the goal of the lesson in terms of student understanding: "I wanted the students to arrive at an understanding of what factoring is, and to be able to use the language." In fact, this lesson showed a great emphasis on student thinking and understanding and appeared to us to be in line with NCTM standards. It is an interesting example to consider because the teacher did not see it as in accord with current ideas.

At the beginning of the lesson students were given back a test they had taken the previous day that dealt with factoring of polynomials. She asked students to go over the tests in their groups, for 5 minutes, discuss the questions they got wrong, and then decide on one item, presumably the most problematic one, to present to the class. As the students deliberated, the teacher went from group to group answering questions and facilitating discussion.

Eight minutes into the lesson, the class reconvened. Each of the five groups sent a representative to the front, one by one, to present and discuss a problem from the test with the class. Let us look at an excerpt from the transcript to get a sense of what the class discussion was like. Here, for example, the student, at the front of the room is presenting problem number 16: "Factor completely, by first factoring out the greatest common factor and then factoring the resulting polynomial: $8x^2 + 8$." We present a somewhat lengthy excerpt.

00:14:17	T	Question?
00:14:18	S	That's the answer that's the//( )
00:14:19	T	//Wait let her- she's she's she'll tell.
00:14:22	S	Then it says you have to factor the resulting polynomial and so you have to figure out if you can factor this right there
00:14:29	S	And you can't.
00:14:31	T	Explain why though or how-have them explain why.
00:14:35	S	Does anybody know why you can't?
00:14:36	S	Because
00:14:38	S	//Cause the
00:14:38	S	//There's no similar
00:14:38	S	There's no common factor.
00:14:40	S	Because( ) there can't be any- you have one negative one positive and the negative one minus one ( )
00:14:47	T	Let's have someone else too have a chance.
00:14:51	S	Um pretty much what Jared said. You can't um cause you need a negative or a positive to . . . the middle term and
00:14:59	T	Okay. You know what I love what's happening.
00:15:02	T	I like your explanation. I think though, would you help by summarizing- He said if it were a binomial I believe.
00:15:11	T	Right?
00:15:12	S	Uh well I'm ( ) (polynomials)
00:15:16	T	Okay.
00:15:17	S	Umm
00:15:18	T	Say again what you said and then I think we'll have

00:15:20	S	You've got to cancel out the middle term so you only have uh the one and not....
00:15:28	T	//Maybe-would you just write the X squared plus one to the side just because that's what we're really hung up on.
00:15:33	S	See if you want to get rid of the middle term
00:15:35	T	And what does he mean by that you want to get rid of the middle term? He means that this has an X squared and has no?
00:15:43	Ss	Middle term.
00:15:45	T	No middle X term and a one. Continue.
00:15:48	S	Umm. So you have to have a negative one and a positive one and if you multiply a negative ( ) you when you multiply a negative and a positive you get a negative.
00:15:58	T	You get a negative. So the only- woah- I'm not going to summarize that. I want someone else to summarize the situation.
00:16:07	T	Someone else. Go for it.
00:16:09	S	If it was if it was an X if you wanted it to be X plus one so if you had it both positives then it would be a middle term which would be two 'cuz X //plus one
00:16:19	T	//Lovely.
00:16:20	T	Could you copy down what she's saying please. She's saying for example if it were X plus one times X plus one.
00:16:27	S	Like X plus one squared it would be you get X squared plus two X plus one.
00:16:34	T	Can everybody// see that?

This excerpt is quite typical of the whole lesson. On the tape we see a class struggling to attach words to their understandings and their solution methods, and the teacher constantly acting to facilitate the exchange. The last ten minutes of the lesson were spent on a new problem that involved finding the slope of the line crossing through two points. The slope is negative, which some of the students find confusing. Again, there is a lively discussion as the teacher tries to mediate between two different solution methods, only one of which ends up being correct.

This teacher explained that the lesson was not in accord with current ideas because "current mathematical thinking is that factoring is not an important concept." She then added the comment: "I disagree."

Chapter 7. Discussion and Conclusions

We have taken numerous measurements and found numerous differences among German, Japanese, and U.S. lessons. But what do they all mean? Which of these indicators are important, and which not? How can we put them all together to describe a German math lesson, a Japanese math lesson, and a U.S. math lesson?

The real answer is that we cannot fully put them together, at least not yet. We are at the beginning stages of this inquiry. Although we are beginning to understand the characteristics of the indicators themselves, we still have much to do to understand how these indicators relate to the underlying models of instruction that govern their performance. Nevertheless, we can give our impressions, our speculations, about the models of instruction that produce the findings presented in this report. What follows, therefore, should be taken as preliminary and speculative.

TYPICAL LESSONS: GERMANY, JAPAN, AND THE UNITED STATES

Throughout this report we have used "lesson" as the unit, both of sampling and analysis. Although we believe there is validity in this approach, it also is important to consider that lessons, for teachers in all three cultures, are related to each other in sequences that form units of instruction across the school year. In our attempts to describe lessons, we necessarily must take this fact into account. Not all the parts of a typical lesson in a given culture, for example, will appear at all points in the course of a unit of instruction. Early lessons will include more development, later ones more practice. Because of our methodology, we cannot take relations among lessons into account.

With this limitation in mind, we nevertheless will proceed to construct, based on the data, typical lessons in Germany, Japan, and the United States. By combining a number of the indicators we have reported and by using the country comparisons as foils against which we can get sharper images of instructional patterns, we can begin to piece together descriptions of how "typical" lessons look in each country. The data presented in this report allow us to describe some characteristics of typical lessons.

For ease of reading, we present the typical lessons in as simple and straightforward way as possible. This means that we have overgeneralized and have omitted qualifiers and caveats that indicate the tentativeness with which these stories should be interpreted. Nevertheless, the primary features of the stories are supported by the coded data presented earlier.

Germany

The mathematics teacher sets the goal for the lesson as the acquisition of a skill or procedure for solving a mathematical problem (figure 12). It is likely that the particular skill or procedure the teacher intends to teach is relatively challenging mathematically (figure 10). She probably intends for the students to understand the rationale for the procedure, why it works (figure 18).

To achieve these goals, the teacher organizes the lesson so that most of the mathematical work during the lessons is done as a whole class (figure 39). The teacher does not lecture much to the students (figure 47); instead, she guides students through the development of the procedure by asking students to orally fill in the relevant information (figures 71-74, 78). This is done by presenting the students with a task such as finding the solution set to two simultaneous linear equations in two unknowns. If the problem is a relatively new one, the teacher generally works the problem at the board, eliciting ideas and procedures from the class as work on the problem progresses. If the problem is one they have already been introduced to, a student might be called to the chalkboard to work the problem. The problem might be slightly different than problems students have worked before but the method to solve the problem has been introduced previously and applied in related situations. The class is expected to monitor the student's work, to catch errors that are made, and to help the student when he or she gets stuck. The teacher keeps the student and class moving forward by asking questions about next steps and about why such steps are appropriate.

After two or three similar problems have been worked in this way, the teacher summarizes the activity by pointing to the principle or property (figure 24) that guides the deployment of the procedure in these new situations. For the remaining minutes of the class period, she assigns several problems in which students practice the procedure in similar situations (figure 65).

Japan

The goal the teacher sets for the lesson is for students to develop mathematical thinking (figure 12) rather than to acquire a particular mathematical procedure as in other countries. Planning for the lesson involves selecting challenging mathematical problems (figure 10) that might involve the development of several student-presented methods of solution (figure 22) or the development of a mathematical proof (see chapter 3 section "Proofs").

The lesson begins with the teacher posing the selected problem. The students are then asked to work on the problem at their seats (figures 39, 41) in order to generate a solution method for the problem (figures 33, 51), sometimes individually and sometimes in groups. During the seatwork period, the teacher circulates around the class, noting the different methods that students are constructing. She then reconvenes the class and asks particular students to come to the front and share their methods (figures 22, 54). Occasionally, the teacher provides a brief lecture (figure 47), pointing out a property or principle inherent in the method (figure 24) or explaining the advantage of particular methods that have been shared. The cycle of the teacher presenting a problem, students working on the problem at their seats, and students sharing their solutions with the class, is repeated several times during lesson (figure 38).

Much mathematical work in the lesson gets done during seatwork (figures 49, 51, 65). But students are given support and direction through the class discussion of the problem when it is posed (figure 50), through the summary explanations by the teacher (figure 47) after methods have been presented, through comments by the teacher that connect the current task with what students have studied in previous lessons or earlier in the same lesson (figure 80), and through the availability of a variety of mathematical materials and tools (figure 53).

United States

The mathematics teacher sets the goal for the lesson as the acquisition of a skill or procedure for solving a mathematical problem (figure 12). The goal probably specifies a particular procedure that all students are to acquire rather than generating alternative solution methods (figure 22) and the emphasis will be placed on acquiring the mechanics of the procedure rather than learning why the procedure works (figures 18, 65). Compared to Germany and Japan, the procedure to be learned is at a relatively simple mathematical level (figure 10).

The lesson typically includes the teacher asking students about their homework (figure 46). The teacher works one or two problems with which students have had difficulty and the homework is collected. The teacher then presents one or more definitions or properties or principles (figure 24), often in the form of procedural rules, that will guide the students' application of the procedures for the next set of problems (figure 15). This might be in the form of a demonstration of a new procedure or a reminder of how a procedure is used in the situations presented in this lesson. Several examples are then worked together as a class (figures 38, 39), with the teacher using the chalkboard or the overhead projector (figures 52, 54). The problems are probably drawn from the textbook or a teacher-made worksheet (figure 53), and few materials or tools are used other than paper and pencil. The teacher guides the students through the procedure by asking short-answer questions of the students (figure 72), such as what is the partial result for this step in the procedure, or what operation is called for in the next step.

The teacher assigns a number of problems of a similar kind (figures 33, 46) as homework. The students work on the problems until the end of the period. The problems function as practice exercises for the procedures demonstrated earlier (figures 51, 65).

COMPARING LESSON SCRIPTS

Comparing the lesson patterns or scripts across countries leads to some interesting observations. Both German and U.S. lessons follow what might be called an acquisition/application script (Hiebert et al., 1996). During the acquisition phase, students are expected to learn how to solve particular types of problems, often through watching demonstrations by the teacher or their peers. During the application phase, students are expected to practice what they have learned.

The acquisition part of the lesson occurs during classwork, the application part during seatwork. The role of the teacher during classwork is to lead students through an example and show them how to do it. The role of students is to pay attention, follow each step, respond to teachers' questions about next steps, and ask questions if they do not understand. In Germany, students often work demonstration examples at the chalkboard, as well, and serve as scribes for the class suggestions for how to work the problems. During seatwork, the role of teachers is to monitor students' work, giving help to students who are stuck or are making mistakes. The role of students is to practice procedures on similar problems.

Lessons in Japan look quite different. They follow what might be called a "problematizing" script (Hiebert et al., 1996). In this script, problem solving becomes the context in which competencies are simultaneously developed and utilized. This means that problems play a very different role in the lesson. Rather than problem solutions serving as the goal of the lesson, they are the means through which students are to understand properties or principles of mathematics. This difference in how problems are viewed leads to a difference in the way lessons are structured.

Instead of beginning with teacher-directed classwork followed by seatwork, the order is reversed. Students try to solve problems on their own first, then comes a period of teacher-directed discussion. In fact, there are often several of these cycles in one lesson. In the opening classwork segment, the teacher's role is to pose a task that students find problematic. In the following seatwork segment, the students' role is to develop a method for solving the problem. It is expected that students will struggle because they have not already acquired a procedure to solve the problem. The time spent struggling on one's own to work out a solution is considered an important part of the lesson. To complete the cycle, the class reconvenes to share the methods that have been developed.

In this script, the students are given more responsibility for doing the mathematical work. The teacher takes an active role in posing problems and helping students examine the advantages of different solution methods, but students are expected to struggle with mathematical problems and invent their own methods. By doing this, students may be engaged in different kinds of mathematical thinking than students in Germany or the United States. This means that the differences between lessons may be more than the quantitative differences shown in the figures; there may be qualitative differences in the kinds of mathematics in which students are engaged.

Although lessons in Germany and the United States seemed to follow a similar general script, there were some differences between them. As measured by international standards, the mathematical content of German lessons was significantly higher than that of U.S. lessons (figure 10). The content of German lessons was also structured in a more coherent way and focused to a greater extent on the development of concepts (rather than just the application of procedures; figure 18).

In fact, in the development of concepts, German lessons more closely resembled Japanese lessons than U.S. lessons. In Germany and Japan, concepts were usually developed over the course of the lesson. Formulas, for example, were not just stated. They might be derived, or, at least, the rationale for them would be developed. In the United States, concepts were usually just stated. A formula would be presented without explaining why it worked.

The differences in cultural scripts that we have outlined can be summarized as follows (figure 86):

Figure 86
Comparison of steps typical of eighth-grade mathematics lessons in Japan, Germany, and the United States

The emphasis on understanding is evident in the steps typical of Japanese eighth-grade mathematics lessons:

- Teacher poses a complex, thought-provoking problem.
- Students struggle with the problem.
- Various students present ideas or solutions to the class.
- The teacher summarizes the class' conclusions.
- Students practice similar problems.

In contrast, the emphasis on skill acquisition is evident in the steps common to most U.S. and German mathematics lessons:

- Teacher instructs students in a concept or skill.
- Teacher solves example problems with the class.
- Students practice on their own while the teacher assists individual students.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

U.S. LESSONS RECONSIDERED

What general principles might underlie the decisions to shape lessons into the script often followed in the United States? We have no country-wide policy that guides the formation of mathematical lessons, but many lessons showed a very similar pattern. What principles might be guiding the development of such lessons? Apparently, they are principles that many teachers share.

We can suggest three principles as possible candidates, one about students' learning, one about the subject, and one about teaching. With regard to students' learning, many teachers may believe that the best way to learn something is to acquire it through a clear, orderly, incremental process. If teachers adhere to this principle, at least implicitly, we would expect them to design lessons that remove obstacles and minimize confusion. Procedures for solving problems would be clearly demonstrated so students would not flounder or struggle. The ambiguity of situations would be removed by making sure that students had learned all of the skills needed before new tasks were presented and by providing immediate feedback on the correctness of solutions. To reduce confusion even further, tasks would be broken into smaller subtasks and each would be mastered in turn. A single, clearly demonstrated procedure would be viewed as the best route to solving a problem; alternative methods would be viewed with some skepticism because they might introduce confusion. These characteristics of classroom lessons are similar to those we often see in the United States.

With regard to beliefs about the subject, many teachers in the United States may believe that mathematics is useful, in the end, as a set of skills. If teachers held to this principle, we would expect that the goals set for most lessons would be the acquisition of skills. Lessons would revolve around the demonstration and practice of skills. Problems would be viewed as an opportunity to apply or practice skills rather than to explore properties or principles of mathematics. A good deal of time would be spent during each lesson practicing skills and homework would require further practice. The development of the rationale or conceptual underpinnings for procedures would be viewed as optional. Again, these characteristics are common in U.S. lessons (see, e.g., figure 12).

Finally, U.S. teachers seem to believe that instruction is a collection of a variety of features rather than a tightly connected system. The collection-of-features view allows one to think about changing instruction by adding new features to an existing routine or substituting one feature for another. This makes it possible to retain the same goals and general scripts for lessons while adding new activities or forms, such as cooperative groups or concrete materials or calculators. This may be the trend.

We compare these three principles with those that are explicit and implicit in the recent reform documents (NCTM, 1989, 1991). Each of the principles is at odds with those that underlie the reform recommendations. In reform documents learning is viewed as a constructive process that involves personal struggle and discovery. Mathematics is seen as much more than a set of skills. Teaching is described as an integrated set of characteristics, including tasks, discourse, and particular roles for teachers and students. Based on this analysis, the differences between current instruction and reform recommendations may lie not just in classroom practices but in the deeper principles that support them.

THE STUDY OF TEACHING: SOME FINAL THOUGHTS

The significance of the results we have presented should be found, not only in the specific findings, but in the fact that collecting this kind of information about teaching was possible. This is the first attempt to collect videotapes of classroom instruction from a national sample of teachers. The potential of these data to yield information about what goes on inside classrooms nationwide is great. The potential is matched, however, by the considerable challenges posed by the collection, coding, and analysis of such a large quantity of video data. This study represents only the first step in learning how to effectively use such data.

Videotapes of classroom instruction provide the kind of detailed permanent real-time records of teaching that enable coding a variety of characteristics reliably and detecting patterns within and among lessons. Indeed, taping in many classrooms allowed us to identify different national scripts or patterns of instruction across lessons. Collecting, coding, and analyzing data from mathematics classrooms in three nations revealed the kinds of contrasts between U.S., German, and Japanese scripts that otherwise might have remained hidden with research based on U.S. data solely. Thus, in the end, the videotapes provided the kind of information that described the nature of current practice and, hopefully, will encourage and enable informed discussions of what practice should look like.

References

Bailey, B.J.R. (1977). Tables of the Bonferroni t statistic. *Journal of the American Statistical Association, 72,* 469-478.

Bellack, A.A. (1966). *The language of the classroom.* New York: Teachers College Press.

Cazden, C.B. (1988). *Classroom discourse: The language of teaching and learning.* Portsmouth, NH: Heinemann.

Cronin, J.M. (1985). *Issues in national educational data collection.* Unpublished paper prepared for NCES Redesign Project. Washington, DC: The Office of Educational Research and Improvement, U.S. Department of Education.

Fisher, R.A. (1951). *The design of experiments.* Edinburgh: Oliver & Boyd.

Foy, P., Rust, K., and Schleicher, A. (1996). Sample design. Chapter 4 in M.O. Martin and D.L. Kelly (Eds.), *Third International Mathematics and Science Study (TIMSS) technical report, Volume I: Design and development.* Chestnut Hill, MA: Boston College.

Hall, G., Jaeger, R.M., Kearney, C.P., and Wiley, D.E. (1985). *Alternatives for a national data system on elementary and secondary education.* Evanston, IL: Northwestern University.

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K., Human, P., Murray, H., Olivier, A., and Wearne, D. (1996). *Problem solving as a basis for reform in curriculum and instruction: The case of mathematics.* Educational Researcher, 25, (4), 12-21.

Hiebert, J., and Wearne, D. (1993). *Instructional tasks, classroom discourse, and students' learning in second-grade arithmetic.* American Educational Research Journal, 30, 393-425.

Husen, T. (1967). *International study of achievement in mathematics.* New York: Wiley.

Lampert, M. (1991). Connecting mathematical teaching and learning. In E. Fennema, T.P. Carpenter, and S.J. Lamon (Eds.), *Integrating research on teaching and learning mathematics* (pp. 121-152). Albany: State University of New York Press.

McKnight, C.C., Crosswhite, F.J., Dossey, J.A., Kifer, E., Swafford, J.O., Travers, K.J., and Cooney, T.J. (1987). *The underachieving curriculum: Assessing U.S. school mathematics from an international perspective.* Champaign, IL: Stipes.

Mehan, H. (1979). *Learning lessons: Social organization in the classroom.* Cambridge, MA: Harvard University Press.

References—Continued

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics.* Reston, VA: National Council of Teachers of Mathematics.

Nisbett, R.E., and Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice-Hall.

Peak, L. (1996). P*ursuing excellence: Initial findings from the Third International Mathematics and Science Study, NCES 97-198.* Washington, DC: U.S. Government Printing Office.

Peterson, P.L. (1985). *The Elementary/Secondary Redesign Project: Assessing the condition of education reform.* Unpublished paper prepared for NCES Redesign Project. Washington, DC: The Office of Educational Research and Improvement, U.S. Department of Education.

Robitaille, D.F., McKnight, C., Schmidt, W.H., Britton, E., Raizen, S., and Nicol, C. (1993). *Curriculum frameworks for mathematics and science.* TIMSS Monograph No.1. Vancouver: Pacific Educational Press.

Stigler, J. W. and Fernandez, C. (1995). *TIMSS videotape classroom study field test report.* Unpublished manuscript. UCLA.

Wickens, T.D. (1989). *Multiway contingency tables analysis for the social sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Appendix A. Information Given to U.S. Teachers Prior to Videotaping

Third International Math and Science Study (TIMSS)

Videotape Classroom Study

Information for Teachers

The Third International Math and Science Study is a cross-national study of students and teachers in more than 40 countries. You and your students have been asked to participate in TIMSS and in the Videotape Classroom Study, a new and experimental part of TIMSS. What are the goals of the videotape study, and what, exactly, do we want you to do?

Goals of the Study

Most international studies of educational systems have focused on achievement, that is, measuring the outputs of the system in terms of what students learn. Although the TIMSS project includes tests of what students have learned, great emphasis is also being placed on gathering information that will help to explain cross-national differences in achievement. The TIMSS Videotape Classroom Study focuses on mathematics instruction; our primary goal in this study is to characterize how teachers teach mathematics in the classrooms of different countries.

In this study we are videotaping 100 eighth-grade mathematics teachers in each of three countries: Japan, Germany, and the United States. The teachers are selected at random and are a representative sample of teachers in each country. This will be the first study ever to gather videotaped records of what actually happens in nationally representative samples of classrooms.

What We are Asking You to Do

We need your help in order to get an accurate picture of what happens in American classrooms. However, participating in the videotape study should require very little additional time commitment on your part. Aside from filling out a brief questionnaire, you will be asked to do nothing that you would not normally do in your classroom.

Our goal is to see what typically happens in American mathematics classrooms, so we really want to see exactly what you would have done had we not been videotaping. Although you will be contacted ahead of time, and you will know the exact date and time that your classroom will be videotaped, we ask that you not make any special preparations for this class. So please, do not make special materials, or plan special lessons, that would not typify what normally occurs in your classroom. Also, please do not prepare your students in any special way for this class. Do not, for example, practice the lesson ahead of time with your students.

Any kind of lesson—whether introducing new material or reviewing old—is appropriate; do not attempt to plan any particular kind of lesson for the day of the taping. The only thing we do not wish to videotape is a test that takes the entire class period. Other than that, any mathematics lesson is fine.

Appendix A. Information Given to U.S. Teachers Prior to Videotaping—Continued

Confidentiality

The tapes we collect will be used for research purposes only. Access will be restricted to researchers analyzing the tapes; no one else will be allowed to view the tapes. The results of this study will be reported only as averages across a large number of classrooms, never as information about a single classroom. The identities of the teachers and the schools will be kept in locked storage; even persons hired to code and analyze the tapes will not have access to this information.

Payment

This is an unusual study, and many teachers feel anxious about allowing a video camera into their classrooms. On the other hand, because we are selecting our sample randomly, success of the study depends on full participation by all of the teachers selected; we need your help. In appreciation of your participation we are offering to give $300 to your school when you have completed the videotaping and returned the questionnaire. Use of these funds is at the discretion of your principal, in consultation with you. We also will be happy to send you a copy of the videotape we make in your classroom.

Questions?

If you have any questions about the TIMSS Videotape Classroom Study please call the Director of the study, Dr. Jim Stigler at the UCLA Department of Psychology. You may call collect: (310) 206-9494. If there is no answer leave a message and we will return your call.

Appendix B: Response Rates

In general, a response rate reflects the proportion of total sampled eligible cases from which data were obtained. In the TIMSS Videotape Classroom Study, the response rate indicates the percentage of sampled schools for which videotapes were completed. In each country, response rates can be computed both before and after replacement. The response rate before replacement identifies the proportion of originally sampled schools that participated; the response rate after replacement gives the percentage of all schools sampled (original and replacement schools) that participated.

In addition, response rates can be either unweighted or weighted. Unweighted response rates, computed using the actual numbers of schools, reflect the success of the operational aspects of the study (getting schools to participate). Response 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. Table B1 provides unweighted and weighted response rates for each of the three countries before and after replacement.

Table B1
Unweighted and weighted school response rates before and after replacement for Videotape Classroom Study,1994-95, by country

Response rate type	Germany	Japan	United States
Unweighted			
Before replacement	85.0	96.0	63.3
After replacement	87.0	96.2	66.4
Weighted			
Before replacement	86.0	95.8	65.3
After replacement	87.7	95.9	68.5

NOTE: In Japan and the United States, the final video sampling weights included nonresponse adjustments. This adjustment was factored out in the calculation of the weighted response rates for Japan, and for the weighted U.S. response rates, base weights (which did not include the nonresponse adjustment) were used.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

In Germany, 85 of the 100 originally sampled main TIMSS schools for the videotape study agreed to participate. For the other 15 schools, one of the main TIMSS replacement schools for the school which refused the videotaping was contacted; all 15 of the replacement schools cooperated with the study.

In Japan, 50 schools were sampled; 48 of these agreed to the videotaping. One of the schools which refused was replaced with a school matched to it from the same stratum. Another replacement school was identified, but it could not be matched to the other refusal school and was weighted to be self-representing.

Sixty-nine of the originally sampled 109 schools in the United States agreed to participate in the videotaping. As described in chapter 2, the paired schools for 13 of the refusals were contacted, and 12 of these replacements agreed to participate.

Table B1 shows that response rates in Japan were very high, and response rates in Germany were also good. The response rates in the United States were somewhat lower. Many of the refusals were due to a high refusal rate for participation in the main TIMSS study: The Population 2 weighted response rate before replacement for the main TIMSS pencil-and-paper assessment was 77 percent (Martin and Mullis, 1996, Table 2.7), and roughly three-quarters of the sampled video schools which did not participate had refused to participate in TIMSS altogether. Refusal to participate in TIMSS overall also meant that the paired schools for many of the refusal schools were not contacted as potential replacements. Due to a somewhat lower response rate for the United States, caution should be used in generalizing the sample results to the population from which it was drawn.

Appendix C. Personnel

Members of the Math Content Group

Alfred Manaster received a Ph.D. in mathematics from Cornell University in 1965. His research specialty was mathematical logic. He was a mathematics instructor at the Massachusetts Institute of Technology from 1965 to 1967. He has been a member of the faculty of the Mathematics Department at the University of California since 1967. Professor Manaster joined the Mathematics Diagnostic Testing Project (MDTP) in 1980 and has served as one of its directors since 1985. He is one of the three co-directors of the University of California, San Diego, Doctoral Program in Mathematics and Science Education, which was established in 1993 and is a joint program with San Diego State University.

Phillip Emig was awarded a Ph.D. in mathematics in 1962 by the University of California, Los Angeles (UCLA). His dissertation was on a topic from the theory of Riemann surfaces. Until 1963 he remained as a research associate in the Department of Atmospheric Sciences at UCLA, after which he spent a post-doctoral year as an Alexander von Humboldt Scholar at the Friedrich-Wilhelm University in Bonn, Germany. In 1964 he came to the California State University, Northridge, where he is currently Professor of Mathematics. For 12 of his years at Northridge, Professor Emig served as department chair. During the past 10 years, he has been Faculty Consultant in Mathematics to the Office of the Chancellor of the 21-campus California State University system and a member of the Mathematics Diagnostic Testing Project work group.

Wallace A. Etterbeek received a Ph.D. in mathematics from University of California, Davis, in 1969. His research specialty was algebra. Since 1968 he has concentrated on teaching as a member of the Sacramento State University Mathematics Department. He has also usually taught a course in the San Juan Unified School District since 1972, teaching Title I students in grades 4-6 for about 4 years and then teaching gifted and talented students in those grades and in high school. Professor Etterbeek was a founding member of the MDTP workgroup in 1978. He has served as MDTP's statistician since 1981. He was a member of the first California State University Entry-Level Mathematics (ELM) test development committee.

Barbara Griggs Wells received a B.S. in mathematics from Howard University and has spent the last 30 years teaching mathematics in the District of Columbia and California public schools—evenly divided between junior and senior high school settings. In 1990-91 the UCLA Mathematics Department awarded her the Visiting High School Lecturer position. She recently received a Ph.D. in education from UCLA with a specialty in administration, curriculum and teaching studies emphasizing mathematics. Her active participation in the MDTP began in 1992, including service as a site director and school liaison coordinator. She is presently a member of the clinical faculty of the Graduate School of Education and Information Studies at UCLA where she coordinates the preservice teacher education of secondary mathematics teachers.

Appendix C. Personnel—Continued

List of Participants/Consultant

Collaborators

Jürgen Baumert
Rainer Lehmann
Toshio Sawada

Primary Consultant

James Hiebert

Coordination in Germany

Manfred Lehrke
Jan Poerschke

Technical Support

Aviram Carmi
Alyne Delaney
Tina Gill
Readius Hinton
Gundula Huber
Shirley Huang
Bennie Terry

Chief Videographer

Scott Rankin

Videographers

Ron Kelley
Andrea Lindenthal
Tadayuki Miyashiro

Administrative Support

Reneé Beasley
Roberta Burroughs
Susan Day
Latrice Green
José Ivy
Michelle Elliott

Computer Programming

Rod Kent
Ken Mendoza

Consultants

Clea Fernandez
Alice Gill
Herbert Ginsberg
Nicolas Hobar
Christine Keitel
Magdalene Lampert
Gilah Leder
Shin-Ying Lee
Sue Ann McGraw
Kevin Miller
Eiji Morita
Michael Neubrand
Johanna Neubrand
Yukari Okamoto
Michelle Perry
Tom Rohlen
William Schmidt
Sheila Sconiers
Hidetada Shimizu
Yoshinori Shimizu
Kenneth Travers
Diana Wearne

Math Content Group

Phillip Emig
Wallace Etterbeek
Alfred Manaster
Barbara Wells

Research Assistants

Adina Ambar
Annette Aryanpour
Christine Bellordre
Dennis Childs
Sonal Chokshi
Reagan Curtis
Emi Yamaya Curtis
Stephen De Berry
Alberto de Leon
Timothy Edwards

Research Assistants (cont.)

Kiyoshi Fujimori
Cecilia Fung
Friday Genovesi
Raina Grigg
Malcom Harris
Thomas Haag
German Hercules
Emily Hsui
Sheldon Hunt
Maki Inamine
Sydjea Johnson
Mikako Kawai
Eileen Kersting
Corinna Koehnenkamp
Ken Kuwabara
Dinal Lossone
Kimberly Lee
Hiroko Lippey
Chia-Tzn Luo
William Lynch
Helen Mao
Kathryn McEwen
Kyoko Ito Meyerson
Satomi Mishina
Gregory Miller
Miri Nakamura
Tsuyoshi Oyabu
Carol Park
Christine Ringel
Simone Robinson
Eric Rodriguez
Derreck Sagara
Julie Miller Salmon
Marcelo Serrano
Natalia Serrano
Ziba Shadjaani
Richard Shannon
Lynda Stone
Lalita Suzuki
Vandana Thadani
Margaret Tsubakiyama
Makoto Yoshida

Appendix D. English Version of the Teacher Questionnaire

Videotape Classroom Study

Teacher Questionnaire

Your Name: ____________________________________ Date: ____________________

School's Name: ____________________________________

Name of Course: ____________________________________

A. In this section we will ask you a few questions about the lesson we videotaped and the students in this classroom.

1. Please describe the subject matter content of today's lesson. *(Check as many as apply)*

- ☐ 1. Whole numbers
- ☐ 2. Common and Decimal Fractions
- ☐ 3. Percentages
- ☐ 4. Number Sets and Concepts
- ☐ 5. Number Theory
- ☐ 6. Estimation and Number Sense
- ☐ 7. Measurement Units and Processes
- ☐ 8. Estimation and Error of Measurement
- ☐ 9. Perimeter, Area, and Volume
- ☐ 10. Basics of One and Two Dimensional Geometry
- ☐ 11. Geometric Congruence and Similarity
- ☐ 12. Geometric Transformations and Symmetry
- ☐ 13. Constructions and Three Dimensional Geometry
- ☐ 14. Ratio and Proportion
- ☐ 15. Proportionality: Slope, trigonometry and interpolation
- ☐ 16. Functions, Relations, and Patterns
- ☐ 17. Equations, Inequalities, and Formulas
- ☐ 18. Statistics and Data
- ☐ 19. Probability and Uncertainty
- ☐ 20. Sets and Logic
- ☐ 21. Problem Solving Strategies
- ☐ 22. Other Mathematics Content

2. For this class of students, was the content of today's lesson review, new, or somewhere in between?

- ☐ all review
- ☐ mostly review
- ☐ half review/half new
- ☐ mostly new
- ☐ all new

3. What was the main thing you wanted students to learn from today's lesson?

4. Why do you think it is important for students to learn this?

5. What was the main thing you wanted these students to learn from the previous lesson you taught them?

6. Please briefly describe that lesson.

7. What will be the main thing you want these students to learn from the next lesson you teach them?

8. Please describe what you intend to teach for that lesson.

9a. Was today's lesson planned as part of a sequence of related lessons (e.g., a unit), or was it a stand-alone lesson?

- ☐ stand-alone lesson (**skip to 10**)
- ☐ part of a sequence (**go to 9b**)

9b. What is the main thing you want students to learn from the whole sequence of lessons?

9c. How many lessons are in the entire sequence? ________________

9d. Where did today's lesson fall in the sequence (e.g., number 3 out of 5)? __________

10. How is the topic of this lesson related to other topics in the mathematics curriculum?

11a. Did you previously assign mathematics homework that was due for today?

- ☐ no (**skip to 12**)
- ☐ yes (**go to 11b**)

11b. Please describe the content of this homework.

11c. How long would it have taken the typical student to complete this homework?
________________ minutes.

12. Was this class formed on the basis of students' mathematics ability? (Choose one):

- ☐ Yes, this is a low ability class
- ☐ Yes, this is an average ability class
- ☐ Yes, this is a high ability class
- ☐ No, this is a mixed ability class

13a. During today's lesson, were all students given the same work to do, or was different work given to different students?

- ☐ same work for all students (**skip to 14a**)
- ☐ different work for different students (**go to 13b**)

13b. How was the work tailored for different students?

14a. Did you divide your class into groups for any part of today's lesson?

- ☐ no (**skip to 15**)
- ☐ yes (**go to 14b**)

14b. Were the groups formed to be relatively homogeneous or heterogeneous with respect to ability and/or mathematical proficiency?

- ☐ homogeneous
- ☐ heterogeneous

B. In this section we want to compare what happened in today's lesson with what normally happens in your classroom.

15. The teaching methods I used for today's lesson were:

- ☐ very similar to the way I always teach
- ☐ similar to the way I always teach
- ☐ somewhat different from the way I always teach
- ☐ very different from the way I always teach

16. What, if anything, was different from how you normally teach?

17. How would you describe your students' behavior during today's lesson?

- ☐ very similar to their usual behavior
- ☐ similar to their usual behavior
- ☐ somewhat different from usual
- ☐ very different from usual

18. What, if anything, was different about the nature and amount of your students' participation during today's lesson?

19. How would you describe the tools and materials (e.g., worksheets, manipulatives, models, pictures, calculators) used during today's lesson compared to those you normally use?

- ☐ very typical
- ☐ mostly typical
- ☐ not typical
- ☐ completely atypical

20. What, if anything, was not typical about the tools and materials used during today's lesson?

21. How would you describe today's lesson as a whole? Was it typical/representative of the lessons you normally teach?

- ☐ very typical
- ☐ mostly typical
- ☐ not typical
- ☐ completely atypical

22. How nervous or tense did you feel about being videotaped?

- ☐ very nervous
- ☐ somewhat nervous
- ☐ not very nervous
- ☐ not at all nervous

23. Do you think that having the camera present caused you to teach a lesson that was better than usual, worse than usual, or about the same as usual?

- ☐ better than usual
- ☐ same as usual
- ☐ worse than usual

24a. Was there anything about today's lesson that did not go according to plan or that you would have wanted to be different?

☐ no (**skip to 25**) ☐ yes (**go to 24b**)

24b. Please describe what did not go according to plan.

C. In this section we want to find out about the ideas that influence your teaching.

25. How aware do you feel you are of current ideas about the teaching and learning of mathematics?

- ☐ very aware
- ☐ somewhat aware
- ☐ not very aware
- ☐ not at all aware

26. How do you usually hear about current ideas about the teaching and learning of mathematics?

27. What written materials (e.g., reform documents, publications, books, articles) are you aware of that describe current ideas about the teaching and learning of mathematics? Please list up to three, and indicate whether or not you personally have read each one.

______________________ I have read

- ☐ all of it
- ☐ most of it
- ☐ some of it
- ☐ none of it

______________________ I have read:

- ☐ all of it
- ☐ most of it
- ☐ some of it
- ☐ none of it

______________________ I have read:

- ☐ all of it
- ☐ most of it
- ☐ some of it
- ☐ none of it

28a. To what extent do you feel that the lesson that you taught today is in accord with current ideas about the teaching and learning of mathematics?

- ☐ not at all (**skip to end**)
- ☐ a little (**go to 28b**)
- ☐ a fair amount (**go to 28b**)
- ☐ a lot (**go to 28b**)

28b. Please describe one part of today's lesson that you feel exemplifies current ideas about the teaching and learning of mathematics and explain why you think it exemplifies these ideas.

THANK YOU!!!
for your cooperation

Appendix E. Standard Errors

Table E1—Standard errors for TIMSS Videotape Classroom Study estimates shown in figures, by country

Figure	Category	Germany	Japan	United States
Figure 6	Percentage very nervous	3.93	3.75	4.13
	Percentage somewhat nervous	4.40	5.86	4.18
	Percentage not very nervous	4.25	2.73	5.86
	Percentage not at all nervous	3.14	3.63	6.17
Figure 7	Better than usual	1.58	2.52	2.55
	Same as usual	5.03	6.60	4.82
	Worse than usual	4.82	7.07	4.80
Figure 8	Teaching methods	0.07	0.09	0.09
	Students' behavior	0.09	0.05	0.10
	Tools/materials	0.06	0.08	0.09
	Lesson as whole	0.05	0.06	0.08
Figure 10	Grade level of content	0.24	0.19	0.29
Figure 11	All review	2.62	0	4.79
	Mostly review	6.70	1.90	4.43
	Half/half	6.04	8.09	4.99
	Mostly new	6.09	7.59	4.61
	All new	4.10	5.74	5.54
Figure 12	Skills	7.85	6.19	8.07
	Thinking	6.32	6.49	5.77
	Social	1.80	0	3.85
	Test prep	0	0	2.58
	Can't tell	3.15	1.75	3.38
Figure 13	Number of topics	0.07	0.08	0.21
	Number of topic segments	0.09	0.08	0.33
Figure 15	Concepts only	3.75	5.41	1.15
	Applications only	4.03	4.74	6.84
	Both concepts and applications	4.80	7.26	6.92
Figure 18	Developed	5.26	7.11	8.56
	Stated	5.26	7.11	8.56
Figure 21	Increase	3.69	9.54	5.54
	Same/decrease	3.69	9.54	5.54

Table E1.—Standard errors for TIMSS Videotape Classroom Study estimates shown in figures, by country—Continued

Figure	Category	Germany	Japan	United States
Figure 22	Percentage containing teacher-presented alternative solution methods	2.84	3.08	5.72
	Percentage containing student-presented alternative solution methods	4.29	8.69	2.90
	Number of teacher-presented alternative solution methods	0.04	0.04	0.15
	Number of student-presented alternative solution methods	0.14	0.36	0.07
Figure 24	Number of principles/properties	0.17	0.21	0.24
	Number of definitions	0.11	0.15	0.38
Figure 27	Number of nodes	0.61	0.35	0.72
	Number of links	0.74	0.39	1.02
Figure 28	1 Component	6.14	5.73	7.48
	2 Components	5.41	3.09	10.65
	More than 2 components	5.95	4.83	7.88
	1 Leaf	8.99	9.81	8.44
	2 Leaves	8.73	8.73	9.37
	More than 2 leaves	8.71	4.83	7.46
Figure 29	Percentage containing illustrations	7.95	9.72	13.45
	Percentage containing motivations	9.96	10.75	4.79
	Percentage containing increase in complexity on nodes	8.02	3.89	2.80
	Percentage containing deductive reasoning on nodes	8.47	6.54	0
Figure 30	Percentage containing links coded as increase in complexity	5.38	12.41	2.80
	Percentage containing links coded as necessary for result/process	7.54	10.11	8.37
Figure 31	Number of codes per node	0.04	0.09	0.04
	Number of codes per link	0.06	0.13	0.04
Figure 32	More single-step tasks	9.01	4.83	8.96
	Equal numbers of single-and multi-step tasks	8.67	7.04	6.79
	More multi-step tasks	10.41	7.90	11.13

Table E1.—Standard errors for TIMSS Videotape Classroom Study estimates shown in figures, by country—Continued

Figure	Category	Germany	Japan	United States
Figure 33	All task controlled tasks	10.96	9.13	7.78
	Task and solver controlled tasks	9.08	14.13	6.11
	All solver controlled tasks	5.99	15.25	4.69
Figure 34	Low quality math content	8.33	7.49	5.63
	Medium quality math content	7.58	11.04	5.63
	High quality math content	7.73	8.51	0
Figure 35	Desks in rows	3.66	3.50	4.17
	Desks in groups	3.42	3.50	4.25
	Desks in U-shape	2.38	0	1.91
	Desks in U with rows	3.62	0	1.31
	Desks in other configuration	8.29	0	2.03
Figure 36	Lessons with outside interruptions	2.61	0	6.75
Figure 37	Number of organizational segments	0.23	0.55	0.52
Figure 38	Number of classwork segments	0.12	0.25	0.25
	Number of seatwork segments	0.11	0.31	0.26
Figure 39	Percentage of time spent in classwork	1.69	2.23	2.66
	Percentage of time spent in seatwork	1.87	2.25	2.61
Figure 40	Duration of classwork segments	0.93	0.72	1.36
	Duration of seatwork segments	0.49	0.56	0.72
Figure 41	Percentage of seatwork time spent working individually	3.16	4.68	4.66
	Percentage of seatwork time spent working in groups	2.33	4.68	4.16
	Percentage of seatwork time spent working in mixed individual/group segments	2.15	0	2.28
Figure 42	Percentage of lessons in which only individual seatwork occurred	4.65	5.40	6.24
	Percentage of lessons in which only seatwork in groups occurred	0.07	5.65	3.37
	Percentage of lessons in which both individual and group seatwork occurred	4.05	7.09	5.61
	Percentage of lessons in which no seatwork occurred	2.25	0	1.58
Figure 44	Number of activity segments	0.35	0.77	0.93

Table E1.—Standard errors for TIMSS Videotape Classroom Study estimates shown in figures, by country—Continued

Figure	Category	Germany	Japan	United States
Figure 45	Percentage of time spent on unrelated activities	0.08	0.13	0.40
	Percentage of lessons containing segments of unrelated activities	2.11	4.62	6.26
Figure 46	Percentage of lessons containing segments of working on homework	1.56	0	5.12
	Percentage of lessons containing segments sharing homework	4.35	3.19	7.00
Figure 47	Percentage of time spent on teacher talk/demonstration	0.25	1.13	1.25
	Percentage of lessons containing time spent on teacher talk/demonstration	3.75	6.55	5.60
Figure 48	Percentage of time spent working on task and/or situations	2.48	1.35	2.76
	Duration of time spent on working on task/situation	0.78	0.43	0.94
Figure 49	Percentage of time spent working on task/situation in classwork	2.74	2.62	3.56
	Percentage of time spent working on task/situation in seatwork	1.80	2.34	2.49
Figure 50	Percentage of time spent on setting up task and/or situations	0.44	1.02	0.63
	Percentage of time spent on sharing task and/or situations	1.29	1.57	1.43
Figure 51	Percentage of topics including development with seatwork	4.91	6.15	20.05
	Percentage of topics including development that include actual development of concepts during a seatwork	4.80	6.44	8.36
Figure 52	Percentage of lessons with chalkboard used	2.58	0	5.84
	Percentage of lessons with overhead projector used	3.04	1.95	5.53

Table E1.—Standard errors for TIMSS Videotape Classroom Study estimates shown in figures, by country—Continued

Figure	Category	Germany	Japan	United States
Figure 53	Percentage of lessons with worksheets	4.18	7.62	4.94
	Percentage of lessons with textbook	4.68	1.83	5.18
	Percentage of lessons with computers	0	4.13	2.45
	Percentage of lessons with calculator	2.04	0	6.94
	Percentage of lessons with manipulatives	3.30	7.14	3.74
	Percentage of lessons with math tools	5.55	5.74	4.62
	Percentage of lessons with posters	1.01	4.07	0.76
Figure 54	Percentage of lessons with chalkboards used by students	4.50	8.02	5.91
	Percentage of lessons with overheads used by students	9.49	0	8.76
Figure 56	Percentage of T/S/PPD remaining on chalkboard	5.45	2.76	8.07
	Percentage of T/S/PPD erased from chalkboard	5.45	2.76	8.07
Figure 57	Percentage of manipulatives in each lesson used by teacher only	9.85	15.07	11.06
	Percentage of manipulatives used by students only	8.75	4.18	14.98
	Percentage of manipulatives used by both teachers and students	9.71	14.97	14.05
Figure 61	Percentage of time in seatwork/working on task/situation segments spent working on one task/one situation	4.98	5.30	6.64
	Percentage of time in seatwork/working on task/situation segments spent working on multiple task/one situation	3.55	3.26	2.49
	Percentage of time in seatwork/working on task/situation segments spent working on one task/multiple situations	5.16	5.08	6.40
	Percentage of time in seatwork/working on task/situation segments spent working on multiple tasks/multiple situations	4.18	3.94	5.99

Table E1.—Standard errors for TIMSS Videotape Classroom Study estimates shown in figures, by country—Continued

Figure	Category	Germany	Japan	United States
Figure 65	Percentage of seatwork spent in practice procedure	2.53	3.59	2.11
	Percentage of seatwork spent in application of concepts	1.67	6.43	2.05
	Percentage of seatwork spent in inventing new solutions and thinking	1.76	6.47	0.70
Figure 69	Percent of utterances made by the teacher	1.20	1.91	1.23
	Percent of words spoken by the teacher	1.54	2.18	1.20
Figure 70	Utterances coded as information	0.27	0.99	0.54
	Utterances coded as elicitation	0.24	0.58	0.31
	Utterances coded as direction	0.31	0.42	0.24
	Utterances coded as uptake	0.18	0.27	0.23
	Utterances coded as teacher response	0.11	0.06	0.14
	Utterances coded as provide answer	0.03	0.01	0.03
Figure 71	Utterances coded as student response	0.29	0.49	0.31
	Utterances coded as student elicitation	0.13	0.08	0.14
	Utterances coded as student information	0.22	0.19	0.21
	Utterances coded as student direction	0.03	0	0.05
	Utterances coded as student uptake	0.03	0.06	0.03
Figure 72	Length of student response measured by number of words	0.38	0.85	0.19
Figure 73	Number of utterances coded as content teacher elicitation	0.23	0.41	0.28
	Number of utterances coded as interactional teacher elicitation	0.10	0.21	0.15
	Number of utterances coded as metacognitive teacher elicitation	0.08	0.11	0.12
	Number of utterances coded as evaluative teacher elicitation	0.05	0.23	0.09
	Number of utterances coded as other teacher elicitation	0.06	0.08	0.11
Figure 74	Number of utterances coded as Name/State content elicitation	0.28	0.37	0.30
	Number of utterances coded as Yes/No content elicitation	0.10	0.05	0.17
	Number of utterances coded as describe/explain content elicitation	0.13	0.28	0.03

Table E1.—Standard errors for TIMSS Videotape Classroom Study estimates shown in figures, by country—Continued

Figure	Category	Germany	Japan	United States
Figure 77	Number of codes per minute of classwork	0.52	1.05	1.51
	Number of Elicitation-Response sequences per minute of classwork	0.15	0.12	0.24
Figure 78	Percentage of initial elicitations, Name/State Fact	3.58	2.64	2.00
	Percentage of initial elicitations, Yes/No Fact	0.83	2.10	2.27
	Percentage of initial elicitations, Describe/Explain Fact	3.64	0.79	0.88
Figure 79	Percentage of initial elicitations, Name/State Individual Ideas	0.52	0.77	0.07
	Percentage of initial elicitations, Yes/No Individual Ideas	0	0.64	0.07
	Percentage of initial elicitations, Describe/Explain Individual Ideas	0.67	1.00	1.61
Figure 80	Percentage of external links	7.97	6.17	10.47
	Percentage of internal links	11.73	3.88	8.88
Figure 81	Not at all aware	2.18	4.66	0
	Not very aware	3.78	8.37	2.29
	Somewhat aware	3.84	12.32	6.60
	Very aware	3.41	3.84	6.57
Figure 82	Percentage of teachers who mentioned School-Based Programs	6.44	5.94	6.96
	Percentage of teachers who mentioned External Seminars	5.62	0	7.19
	Percentage of teachers who mentioned Information from Colleagues	5.66	4.69	6.18
	Percentage of teachers who mentioned Publications	3.20	5.10	4.13
	Percentage of teachers who mentioned other sources	3.04	4.57	1.12
Figure 83	Not at all in accord	0	4.37	3.22
	A little in accord	7.03	4.10	5.88
	A fair amount in accord	7.03	3.10	7.63
	A lot in accord	0	0	5.14

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Table E2.—Standard errors for figure 84, TIMSS Videotape Classroom Study, by reform position

Category	Reformers	Non-Reformers
Individual seatwork only	11.70	10.93
Group seatwork only	6.52	10.01
Both individual and group seatwork	13.78	6.23
No seatwork	0	0

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95.

Appendix F. Transcription Conventions

The transcription system developed for this study was intended to capture, as accurately as possible, the discourse of mathematics classrooms in Germany, Japan, and the United States. The system was designed to represent speech only, and not to capture the actions and activities that take place around the speech. German and Japanese data were translated into English. Our goal was to capture the meaning of the original language without sacrificing readability.

1. Speaker Codes

T	Teacher
S	Single student
Ss	Multiple students, but not the entire class
E	Entire class (or sounds like the entire class); used to indicate choral responses
O	Other; used to indicate when a non-member of the classroom speaks, such as school personnel, office monitors, or talk from public address systems
B	Blackboard; used to indicate the translation of foreign words/phrases written on the blackboard

2. Punctuation, Diacritical Marks, and Other Conventions

Period	.	A period marks the end of a turn at talk that is not to be understood as being a question.
Question mark	?	A question mark indicates that the utterance is to be understood as a question (usually determined through intonation).
Hyphen	-	A hyphen indicates that a speaker has "cut-off" (or self-interrupted) his/her speech.
Three dots	. . .	A series of three dots, separated by a blank space before and after, is used to indicate a pause.
Parentheses	( )	Empty parentheses or single parentheses surrounding a word(s) indicate that some interactant has spoken, but the exact utterance is unclear (hard to hear).
Numerals		Numbers are always written out as words, in the way in which they are said.
CAPS		Capital letters are only used with proper nouns (names, cities, countries, languages, etc.), at the beginning of a new turn at talk, or after a period or question mark. When speakers refer to points, lines, angles, etc. by their alphabetical label (e.g., angle ABC, line DE), the labels are written in capital letters, even if in the notation system used in the classroom it would otherwise appear as a lowercase letter.
C A P S		When speakers spell out words, each letter is placed in capital letters, with a single space between each letter.
Overlap	//	When one participant speaks over the talk of another participant, it is noted in the transcript using a double backslash (//) to indicate where the overlap begins. Overlap brackets are used in sets to indicate the lines of talk which overlap.

3. Turns at Talk

Turns are separated when: 1) there is a change in speakership, and 2) there is a "gap" in the talk such that it would be possible for someone else to speak in that silence.

United States
Department of Education
Washington, DC 20208-5650

Official business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G-17

Standard Mail (B)