Mathematical Features of Lessons in the TIMSS Video Study

James Hiebert, Newark, DE (USA)
James W. Stigler, Los Angeles, CA (USA)
Alfred B. Manaster, San Diego, CA (USA)

Abstract: The video component of the Third International Mathematics and Science Study marked the first time that national samples of teachers were videotaped at work in their classrooms. In this article we review some of the results of the study, with special attention to the nature of the mathematics evident in these eighth-grade lessons from Germany, Japan, and the United States. We conclude by proposing that many lessons within a country follow a cultural pattern of teaching, and that differences among countries on individual indicators of teaching must be understood in reference to these patterns.


ZDM-Classification: D13, D23, D43

1. Opportunities presented by the TIMSS Video Study

The video component of the Third International Mathematics and Science Study (TIMSS) marks the first time that video has been used to study national probability samples of teachers at work in their classrooms. In this article, we present a brief overview of this study, which involved detailed analysis and comparison of eighth-grade mathematics teaching in three countries: Germany, Japan, and the United States. We focus on the nature of the mathematics that was evident in the videotaped lessons.

Collecting national samples of teaching can serve two important purposes. First, it gives us solid information about the processes of teaching and learning inside classrooms, information that is crucial for developing sound education policies. Efforts to improve student learning succeed or fail inside the classroom, a fact that too often has been ignored by would-be reformers.

National samples of teaching also enable us to compare teaching methods across countries. This comparison allows us to see teaching in a new way. Teaching is a cultural activity (Gallimore 1996). It is an everyday event in which broad segments of society participate repeatedly and over long periods of time. Over time, we have developed norms and expectations for teaching that are widely shared within a society and passed along as this generation of students becomes the next generation of teachers (Lortie 1975; Nemser 1983). Because our models of how teaching should look are so widely shared and so familiar within each society, they become nearly invisible. We come to believe that this is the way teaching is and must be. When we observe teaching in other countries, these accepted and unquestioned practices are revealed and we see that we teach the way we do because of the options provided by our culture. This realization is crucial because it opens up new possibilities and presents new images for how we might improve teaching.

2. Conducting the TIMSS Video Study

As planning commenced for the TIMSS, there was great interest in being able to go beyond the cross-national achievement data to focus on the underlying processes that produce achievement. Instructional processes in the classroom were seen as a likely important cause of student learning. But how could something as complex as teaching be studied on a large-scale, across cultures?

One approach is to ask teachers on a questionnaire, to describe their instructional practices. Although such a questionnaire had been administered as part of the Second International Mathematics Study, there are problems with this approach. Even within a country, teachers lack shared meanings for the words they use to describe teaching. One teacher will call something “problem solving” and her colleague next door will call the same thing a “routine exercise”. This problem is compounded in a cross-cultural questionnaire study. The responses are nearly impossible to interpret.

Thus the decision was made to collect direct information on teaching by videotaping classroom instruction. Video had long been used for small-scale, in-depth qualitative research, but had never been attempted on such a large scale before. For this reason, the scope of the study was limited to one of the three grade levels studied in TIMSS (eighth grade), and to three of the 41 TIMSS countries (Germany, Japan, and the United States). Germany and Japan were chosen because they both are viewed as important economic competitors of the U.S. Japan is of special interest because it has repeatedly scored near the top in international comparisons of mathematics achievement.

Two straightforward goals guided the study:

1. To learn how eighth-grade mathematics is taught in three countries; and
2. To learn how American teachers view reform and to see whether they are implementing teaching reforms in their classrooms. To achieve these goals, a number of methodological hurdles had to be cleared.

2.1 Sampling

If we wanted to use video to paint a national portrait of eighth-grade mathematics instruction, we needed to be sure that the videotapes we analyzed were representative of instruction in each country. A number of issues were considered. First, how should classrooms be selected? Fortunately, the TIMSS sampling plan was highly sophisticated. For this reason, it was possible to construct the video sample as a random sub-sample of the full TIMSS sample. Not only were specific teachers selected, but specific class periods as well. The final video sample included 231 classrooms: 100 in Germany, 50 in Japan, and 81 in the United States.
Having chosen the classrooms, we next needed to decide how many lessons to videotape in each classroom. In the end, we videotaped one lesson in each classroom. Although this enabled us to study the largest number of teachers within our budget, it precluded us from studying some important parts of the teaching process, such as the methods teachers use for linking lessons together into units.

Finally, we wanted to be sure that our sample was representative of the entire school year. This was especially important in Japan, where a national curriculum leads to different topics being taught at different times of the year across the entire nation. Although we succeeded in videotaping evenly across the school year in the United States and Germany, we were somewhat less successful in Japan, where our sample was skewed toward geometry and away from algebra. For some analyses, therefore, we selected balanced subsamples of algebra and geometry lessons in each country.

2.2 Videotaping procedures
At first glance, videotaping classroom instruction seems straightforward: bring in a camera and turn it on. Unfortunately, things are not so simple. Seemingly minor decisions on the part of videographers – such as where to point the camera at any given moment – can significantly affect our view of what is happening in the classroom. To get useful, comparable video in all classrooms, numerous issues had to be anticipated and resolved in advance.

Standardized procedures for camera use were developed, tested, and revised, and videographers were trained. Only one camera was used in each classroom, and it focused on what an ideal student would be focusing on – usually the teacher. After taping, teachers filled out a questionnaire describing the goal of the lesson, its place within the current sequence of lessons, how typical the lesson was, whether they had used methods recommended by current reforms, and so on.

2.3 Coding the tapes
When the tapes arrived at our research lab in Los Angeles, they were digitized (to increase durability and random access) and then translated and transcribed. The transcripts were then linked by time codes to the video in a multimedia database. These technical features helped to ease the viewing of the tapes, but they did not help in deciding what to code. Coding classroom instruction is notoriously difficult, because there is so much to choose from. We kept two goals in mind: code those aspects that might make a real difference in the mathematics the students were learning, and define codes that would yield valid and informative descriptions of instruction across the three cultures. For the first goal, we drew from the research on teaching and learning in mathematics. To meet the second goal, we convened a team of six code developers, two from each country, and spent the summer of 1994 watching and discussing 27 field test tapes. Out of these intensive discussions emerged the initial coding system, a system that included categories such as the nature of the tasks, the kind of work expected of students, and the nature of classroom discourse. The system was refined regularly as the primary coding team began applying it to the actual study tapes and as inter-coder reliability checks revealed categories that needed further definition. Only codes meeting an 80% inter-coder agreement were retained.

As the coding process unfolded, we found it essential to construct a summary table to represent each lesson. These lesson tables were skeletons of each lesson that showed, on a time-indexed chart, how the lesson was organized through alternating segments of classroom and seatwork, what pedagogical activities were used (e.g., explaining tasks, demonstrating procedures), what tasks were presented, and the solution strategies for the tasks that were offered by the teacher and the students. Some categories could be coded directly from the tables; some required re-viewing the tapes. One advantage of the lesson tables is that references to a particular country (e.g., students names, cultural objects) could be removed or altered so that the reviewer would not know the country in which the lesson was taped.

A special group of coders who we refer to as the “Math Group” used only the country-blind lesson tables for their analyses. Led by the third author, the Math Group was composed of four university and secondary school mathematics teachers. They analyzed 15 algebra and 15 geometry lessons randomly chosen from each country. The four members met frequently during the summer of 1995 to review the lesson tables, create categories describing the nature of the mathematics in the lessons, compare initial coding results, and resolve discrepancies. Full agreement among the members was reached on all coding decisions. Percentage agreement between an individual’s initial coding and the final consensus was calculated for 17 of the 90 lessons and ranged from 87% to 91%. A more complete report of the findings of the Math Group is found in Manaster (1998).

2.4 Advantages of video
The benefits of video are well worth the methodological challenges and the labor-intensive demands they impose. Video provides the researcher two kinds of data: visual images rich in descriptive power, and quantifiable indicators that summarize the contents of large numbers of images. Visual images are vivid and powerful, but they can often be misleading. They can create such a powerful impression that the viewer is seduced into thinking that a single case tells the whole story. Coding relevant indicators enables us to check the representativeness of the images. Each kind of data is significant, and each strengthens the other.

Video data also are relatively raw because they are not yet categorized or quantified. Unlike narrative observations or on-site coding, videos have not been filtered through the eyes of individual researchers, and are not as constrained by the initial hypotheses of those who design the study. For this reason they can be analyzed by researchers with different interests using different coding schemes.
3. Features of mathematics lessons in Germany, Japan, and the United States

Descriptions of teaching must be selective. There is more going on in a classroom than can be reported, even in a lengthy document. In a brief article, the problem is compounded. What follows is a selection from both the quantitative and qualitative data to illustrate the kind of information the study provides.

3.1 What kind of mathematics do students encounter?

The nature and level of students’ learning surely is influenced by the nature of their mathematical experiences in the classroom. And students’ experiences are shaped by the kind of mathematics that they encounter. One measure of the mathematics seen in the videotapes is the level of the topics compared with their average placement in the mathematics curricula of the 41 TIMSS countries.\(^5\) Matched against this scale, the average grade level for lesson topics was mid-seventh grade for the United States, mid-eighth grade for Germany, and beginning ninth grade for Japan.

A confirmation of this result is the Math Group’s finding that 41% of the lessons in the U.S. were actually Pre-Algebra, whereas 13% of the German lessons and 3% of the Japanese lessons fit this lower level description.

One feature of lessons on which the Math Group focused was mathematical reasoning, both inductive and deductive. Only one-fourth of the 90 lessons contained instances of reasoning. 19 lessons showed deductive reasoning and 3 showed inductive reasoning. Reasoning was much more prevalent in geometry than algebra or pre-algebra. Only 2 of the 45 algebra and pre-algebra lessons showed mathematical reasoning, whereas 20 of the 45 geometry lessons showed such reasoning. The almost complete lack of explicit mathematical reasoning in algebra classes suggests that many beginning algebra students in all three countries are encouraged to approach algebra in an exclusively procedural fashion. When analyzed by country, mathematical reasoning was evident in 20% of the German lessons, 53% of the Japanese lessons, and 0% of the U.S. lessons.

Together, these indicators suggest that the level and nature of mathematics encountered was different for students in different countries. But this is not the whole story. The way in which teachers construct lessons and engage students can influence how and what students learn.

3.2 How are lessons organized?

In order to understand how lessons are constructed, it is useful to know what goals teachers set. Teachers were asked in the questionnaire what “main thing” they wanted students to learn from the lesson. Most teachers focused either on mathematical skills (solving specific kinds of problems or using specific formulas) or mathematical thinking (exploring, developing, and understanding mathematical ideas, or inventing new ways to solve problems). As shown in Fig. 1, there were large differences between countries. Japanese teachers emphasized thinking; German and U.S. teachers emphasized skills. To understand how these goals are translated into classroom lessons, it is helpful to consider a few additional indicators on lesson organization.

A primary feature of lesson organization is coherence – the connectivity of the mathematics across the lesson. Imagine the lesson as a story. Well-formed stories consist of a sequence of events that fit together to reach the final conclusion. Ill-formed stories are scattered sets of events that don’t seem to connect. As readers know, well-formed stories are easier to comprehend than ill-formed stories, and well-formed stories are like coherent lessons. They offer students greater opportunities to make sense of what is going on.

Coherent lessons are achieved by weaving together ideas and activities. One way to help students notice how ideas are related is to point them out explicitly. For example, several minutes into a German lesson, the teacher said, “Next is a step you really need to pay close attention to because we’re dealing here with different numbers from those we dealt with yesterday.” Explicit connections also can be made among ideas within a lesson. The majority of teachers in all countries made explicit connections from one lesson to another, but only Japanese teachers routinely linked together the parts of an individual lesson. In fact, 96% of Japanese lessons contained statements linking one part of the lesson with another whereas about 40% of both German and U.S. lessons contained such statements.

The Math Group developed a method of representing the links among the mathematical ideas within a lesson using a directed graph. One result of this analysis showed that in 92% of Japanese lessons, 76% of German lessons, and 45% of U.S. lessons, all parts of the lesson were connected by at least one appropriate mathematical relationship (e.g., one segment was dependent on another, or extended another). A summary score also was calculated for each lesson to indicate the degree to which the parts of a lesson were interrelated. By this measure, German lessons scored four times as high as U.S. lessons; Japanese lessons scored six times as high as U.S. lessons.

3.3 Are mathematical concepts and procedures developed?

Mathematical concepts and procedures either can be simply stated by the teacher or they can be developed through examples, demonstrations, and discussions. Suppose the topic is the area of right triangles. Teachers can state that

![Fig. 1: Teachers’ responses, on the questionnaire, to the question, “What was the main thing you wanted students to learn from today’s lesson?” Source: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994-95]
the area is found by measuring the base, measuring the height, and dividing by 2. Alternatively, teachers can develop this procedure showing, for example, how the formula 1/2 base x height can be derived by halving rectangles (and parallelograms) to form two triangles. Of course, the teacher might ask students to develop some of this themselves. We coded “developed” if teachers made any attempt to motivate a procedure or explain why it worked.

As shown in Fig. 2, concepts and procedures usually were developed in German and Japanese lessons but usually were just stated in U.S. lessons.

Fig. 2: Average percentage of topics in each lesson that contained concepts that were developed versus only stated.

3.4 What are students expected to do?
When we examined what students actually did during the lessons, we found additional differences among countries that are likely to affect students’ opportunities to learn mathematics. In all three countries, in almost all lessons, students were asked to solve problems. Lessons differed, however, in how much creative mathematical work was possible in solving the problems presented. In some lessons, problems were presented that afforded only one method of solution, often a method demonstrated by the teacher. In other lessons, the problems afforded a variety of methods and students were allowed to use (or develop) a method of their choice. The Math Group described this distinction in terms of locus of control – task-controlled versus solver-controlled. Results showed that in 63% of Japanese lessons the tasks were predominantly solver-controlled. Percentages of comparable lessons in Germany and the U.S. were 30 and 14, respectively.

In a related analysis, the primary coding team classified the nature of the work expected of students during seatwork into three categories:
- Practice routine procedures,
- Apply procedures in new situations, or
- Invent new procedures and analyze new situations.

The first category is familiar: the teacher demonstrates or develops a procedure, such as solving a linear equation for $x$, and then assigns a number of similar problems on which students are to practice the same procedure. The second category includes cases in which a procedure is demonstrated or developed for solving one kind of prob-

lem, say finding the area of a right triangle by adjoining an identical triangle to form a quadrilateral and calculating half its area, and then students are asked to apply the same procedure to another kind of problem, say finding the areas of non-right triangles. The third category required even more of students: students were asked to invent solution methods, analyze mathematical situations, or generate mathematical proofs. For example, students might be asked to develop a general method for finding the sum of the interior angles of an $n$-sided polygon after measuring the sums for 3-, 4-, and 5-sided polygons.

Coding seatwork into these three categories resulted in the differences shown in Fig. 3. Japanese students spent less time practicing routine procedures and more time inventing, analyzing and proving than their peers in the other countries. German and U.S. students spent almost all of their time practicing routine procedures.

Fig. 3: Average percentage of seatwork time spent in three kinds of tasks.

3.5 Overall lesson quality
A final analysis conducted by the Math Group assessed the overall quality of the mathematics in each lesson with regard to its potential for helping students understand important mathematics. This judgment took into account many of the specific indicators reported above. Although subjective, the members of the Math Group reached agreement on this judgment for all 90 lessons. Fig. 4 shows the results. Again, Japanese and U.S. lessons are the most different with German lessons falling between.

3.6 The teacher’s role
Based on the information presented thus far, the reader may have developed the impression that teachers play a far more active role in Germany and the U.S. than in Japan. For example, whereas German and U.S. students usually practice methods developed or presented by the teacher, Japanese students often are asked to develop methods themselves. But to assume that Japanese teachers are less active or directive than German or U.S. teachers would be a mistake. For example, although it is true that Japanese teachers give students time to struggle with challenging problems, they often follow this up with direct explanations and summaries of what the students have learned.
This is why Japanese teachers were coded as engaging in more direct lecturing than either German or U.S. teachers. Although the time devoted to lecturing was minimal in all three countries, 71% of Japanese lessons contained at least some lecturing, compared with only about 15% of German and U.S. lessons. Japanese teachers also control the direction of the lesson in subtle ways, such as creating conditions in the classroom that will govern the kinds of solution methods students are likely to invent. For example, they often select problems to begin a lesson that can be solved by modifying methods that have been developed during the previous lesson.

Fig. 4: Percentage of lessons rated as having low, medium, and high quality of mathematical content.

4. Cultural patterns of teaching
We noted earlier that teaching is a cultural activity. Cultural activities often have a "routineness" about them that ensures a degree of consistency and predictability. Lessons are the daily routine of teaching and often are organized according to a cultural pattern, a commonly accepted and predictable way of structuring a lesson and sequencing the instructional activities. Features of cultural patterns can be difficult to detect when examining teaching within a single culture because the features are so common. But, these features become apparent when comparing teaching across cultures.

4.1 The German pattern
German lessons often moved through the following four activities, in sequence:
(1) Review previous material, either by reviewing homework or reminding students what has been accomplished to this point;
(2) Present the topic and problems for the day;
(3) Develop the procedures to solve the problem(s), often with the teacher carefully guiding the students through the details; and (4) Practice the procedure(s) together or individually on a set of similar problems.

4.2 The Japanese pattern
A typical Japanese lesson was characterized by the following sequence of activities:
(1) Review previous lesson, usually through a brief summary by the teacher;
(2) Present the problem for the day, often a problem that builds on the previous day’s work;
(3) Students try to solve the problem, individually or in small groups;
(4) Students share solution methods they have tried with comments and suggestions from the teacher and other students;
(5) Summarize the major point of the lesson, frequently through a brief lecture by the teacher.
Often, activities 2-4 are repeated for a second problem before the lesson ends with a summary.

4.3 The U.S. pattern
Most U.S. lessons were consistent with the following sequence:
(1) Review previous material, either through a “warm-up” activity or by checking homework;
(2) Demonstrate how to solve problems for the day, with the teacher leading the students through a relatively quick demonstration of the relevant procedures;
(3) Practice the procedures individually on a set of similar problems;
(4) Correct the assigned problems and assign additional, similar problems for homework. Usually some class time is allowed for beginning the homework.

4.4 Comparing the lesson patterns
The three patterns share some basic features: the class reviews previous material, the teacher presents problems, and students solve problems at their desks. On closer inspection, however, it becomes apparent that these activities play different roles within the different cultural patterns. Presenting the problem in Germany sets the stage for a relatively detailed development of the preferred solution procedure, a whole-class activity guided by the teacher. Presenting the problem in Japan sets the stage for students working on their own to develop, share, and analyze solution procedures. Presenting the problem in the U.S. sets the stage for a relatively quick demonstration by the teacher of the preferred procedure and then practice by students in executing the procedure. The fact that similar lesson activities can function in different ways is not surprising because the individual activities are embedded in different systems of teaching.

5. Conclusion
Indeed, one of our primary conclusions is that the systemic nature of teaching must be understood and appreciated, both to interpret research results and to design effective programs for improvement (Stigler and Hiebert, 1999). From a research perspective, cross-cultural differences in individual features of mathematics teaching must be understood within the cultural system of teaching of which they are a part. Any given feature of classroom teaching might be positive or negative, depending on the context in which it appears. It is possible, for example, that lecturing by Japanese teachers works well because students are prepared, through individual problem solving, to hear the information teachers have to offer. Lecturing may have a very different effect in the U.S. or in Germany, where it is more often used as a means of introducing new topics or procedures that students have not yet experienced using.
In the TIMSS Video Study, we did not collect data on student learning, and so we could not begin the job of understanding how systems of teaching work to produce learning in students. We cannot say, for example, if lecturing does increase student learning in Japan, or if it does not in Germany and the U.S. Even without such data, however, understanding and documenting culturally distinct systems of teaching gives teachers and researchers new ways of thinking about the classroom processes that are at the core of education. Even if we could identify features common to high-achieving classrooms, it is unlikely that teaching will improve simply by communicating to teachers what those features are. Because teaching is a complex system, and because it is cultural, simply replacing one feature by another, or adding some features to an existing system, is unlikely to produce the desired results.

The systems are likely to change the features rather than vice versa. That is a characteristic of complex systems, and it also is confirmed by our experience in the U.S. (Education Evaluation and Policy Analysis 1990; McLeod, Stake, Schappelle, Mellissinos, & Gierl 1996).

Long-term improvement depends on implementing mechanisms that change the system of teaching, not just selected features. Teachers must play a major role in this process, as they are the gatekeepers of classroom life. We hope that studies such as the TIMSS Video Study will provide one input into the process of improving education in the classroom.

Acknowledgments
This article was adapted, in part, from Stigler, J.W. & Hiebert, J. (1997): Understanding and improving classroom mathematics instruction: An overview of the TIMSS Video Study. Phi Delta Kappan. The research reported in this article was funded by the U.S. Department of Education’s National Center for Education Statistics. The study was conducted in collaboration with Jürgen Baumert and Rainer Lehmann in Germany, and Toshiro Sawada in Japan. The authors would like to gratefully acknowledge the assistance of Patrick Gonzales, Takako Kawanaka, Steffen Knoll, and Ana Serrano, who functioned as primary researchers on the study, and of Eric Dergzharian, Gundula Huber, Fumiko Ichikawa, and Nicole Kersting, who served as primary coders.

6. Annotations
1. A complete description of sampling procedures may be found in J. W. Stigler, P. Gonzales, T. Kawanaka, S. Knoll & A. Serrano (1999).
3. The Math Group included Alfred Manaster, University of California, San Diego; Phillip Emig, California State University, Northridge; Wallace Etterbeek, California State University, Sacramento; and Barbara Wells, University of California, Los Angeles.
4. The TIMSS video data are currently available for secondary analysis. Interested researchers should contact the TIMSS Video Data Center, Takako Kawanaka, Director, e-mail: kawanaka@psych.ucla.edu.

7. References
Educational Evaluation and Policy Analysis (1990): 12(No. 3), Special Issue

Authors
Hiebert, James, Dr., School of Education, University of Delaware, Newark, DE 19716, USA. E-mail: hiebert@udel.edu
Manaster, Alfred B., Dr., Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112, USA. E-mail: amanaster@ucsd.edu
Stigler, James W., Dr., Department of Psychology, University of California, Los Angeles, Los Angeles, CA 90025, USA. E-mail: stigler@psych.ucla.edu

201