EDUCATIONAL PSYCHOLOGIST, *35*(2), 87–100


# Using Video Surveys to Compare Classrooms and Teaching Across Cultures: Examples and Lessons From the TIMSS Video Studies

James W. Stigler
*Department of Psychology*
*University of California, Los Angeles*

Ronald Gallimore
*Department of Psychiatry and Biobehavioral Sciences*

James Hiebert
*School of Education*
*University of Delaware*

The video survey is a promising new approach for studying classrooms and teaching across cultures. Drawing from experience in working with two cross-cultural video surveys, the Third International Mathematics and Science Study (TIMSS) and its follow-up study (TIMSS-R), this article presents some of the challenges of studying classrooms across cultures and some of the ways that the video survey can deal with these challenges. The article begins by identifying some of the issues in cross-cultural research and classroom surveys that led to the creation of the video survey. Examples from the TIMSS and the TIMSS-R video studies are then used to illustrate some benefits and limitations of video surveys and to share some of the lessons that were learned about studying classrooms across cultures.

Studying classrooms and teaching across cultures is fraught with challenges, yet ripe with opportunity. First, this article considers the opportunities afforded by cross-cultural studies of teaching. Next, we lay out what we see as some major methodological challenges posed by such research and outline a new method that we are developing to address some of these challenges. This new method, which we call the *video survey*, evolved over the past 5 years in our work on the Third International Mathematics and Science Study (TIMSS) and its follow-up study (TIMSS-R; the *R* stands for *repeat*). Using examples from these studies (the TIMSS-R is still in progress), we identify some advantages and disadvantages of video surveys and share some lessons we have learned about studying classrooms across cultures.

Requests for reprints should be sent to James W. Stigler, Department of Psychology, 2279B Franz Hall, University of California, Los Angeles, CA 90095–1361. E-mail: stigler@psych.ucla.edu

## WHY COMPARE CLASSROOMS ACROSS CULTURES?

Perhaps the most obvious reason to study classrooms across cultures is that the effectiveness of schooling, as measured by academic achievement, differs across cultures (e.g., Peak, 1996). If cross-national achievement differences are tied to cultural variations in teaching, we may discover ways of teaching that work better than the ones our society routinely deploys. This would allow us to take advantage of the experience of others all over the world who share similar goals, at least in the domain of mathematics and science achievement, and from whom we can learn what alternatives are possible. Given the cultural diversity our teachers confront every day in their classrooms, it is surely a good idea to learn more about the diversity of teaching practices that have evolved around the world.

There is another, more subtle reason for studying teaching across cultures. Teaching is a cultural activity. Because cultural activities vary little within a society, they are often transparent and unnoticed (Geertz, 1984). Consequently, we may be blind to some of the most significant features that characterize teaching in our own culture because we take them for granted as the way things are and ought to be. Cross-cultural comparison is a powerful way to unveil unnoticed but ubiquitous practices. Consider the following example: U.S. mathematics teachers tend to teach students how to solve a particular type of problem, then ask them to solve examples on their own. Students are seldom given problems to work on by themselves before they have been taught by the teacher. This practice is so pervasive in the United States that many people would question the competence of a teacher who gave students a problem to solve that had not been covered in class (Stigler & Hiebert, 1999).

In contrast, Japanese teachers often give students problems to work on that they have not seen before, believing that it is good for students to struggle with something they have not been taught, both to develop thinking skills and to prepare them for later instruction. The comparison with Japanese teaching reveals this pervasive U.S. pattern as a commonly accepted and significant feature of U.S. mathematics teaching. Because it is the approach most U.S. mathematics teachers use, it appears natural, the way things must be, instead of a choice that was made at some point in our cultural history of teaching.

In addition, although many educators and parents perceive major variations across teachers within the United States, comparative experiences suggest that this apparent variability may be relatively insignificant compared with large differences in teaching practices across cultures (Stigler & Hiebert, 1999). Consequently, comparative research invites reexamination of the things "taken for granted" in our teaching, as well as suggesting new approaches that never evolved in our own society. It also is possible to study other high-achieving cultures to see if they follow practices similar to those of Japanese teachers. We may identify pedagogical techniques that are universally effective. Ideally, some of the knowledge generated by comparisons can be translated into U.S. classrooms.

However useful comparative research may be, it remains uncommon in part because of substantial methodological problems. Before turning to a description of the video survey as a possible solution to some of these difficulties, we briefly sketch some issues in the comparative study of classrooms and teaching in different cultural contexts.

## THE CROSS-CULTURAL STUDY OF CLASSROOMS

### Early Cross-Cultural Studies of Learning and Development

For much of the 20th century, cultural studies were the province of anthropologists who lived extended periods in (usually) distant societies described as "primitive" in earlier times and "developing" in current usage. Their method was the *ethnography*, a monograph-length description of lifeways (Erickson, 1986). Although contexts were richly described, most ethnographies devoted little attention to children, their development, socialization, enculturation, or the settings in which these processes occurred (Serpell, 1993; Weisner, 1984). For example, many ethnographies noted that siblings were often the caretakers and surrogate parents of young children, but the developmental implications of this were almost never explored (Weisner & Gallimore, 1977) despite the obvious contrast with adult-centered socialization at home and school in developed societies (Weisner, Gallimore, & Jordan, 1988). Education and schooling were rarely addressed by anthropologists (Diamond, 1971) because either formal educational institutions did not exist or access was extremely limited (Serpell, 1993). In addition, where formal schools existed they were often instruments of conquest, imposed by a colonial regime, regarded as nontraditional, and perceived as alien rather than integral parts of the culture (Diamond, 1971). However, if the definition of formal education is broad enough, then many societies can be said to employ formal schooling. For example, whereas some societies sustain continuously operating institutions such as the Western-style and Qur'anic schools (Scribner & Cole, 1981), other societies employ periodic bush schools, rites of passage, initiation ceremonies, and other means to prepare children for adulthood (Lancy, 1996; Rogoff, 1990).

Later in the 20th century, some ethnographers described the discontinuities between everyday teaching in the village and the culturally alien instruction of the school (e.g., Howard, 1970), a subject that became the focus of investigation in its own right (Gallimore, Boggs, & Jordan, 1974; Greenfield & Lave, 1982; Lancy, 1996; Rogoff, 1982; Scribner & Cole, 1973). Informal tuition was characterized as natural, incidental, participatory, and reliant on observation and imitation to achieve useful goals. Formal education was described as being set apart from the everyday routine, impersonal, broken into rational and small steps, and dependent on explicit, didactic teaching of skills for contrived rather than productive purposes. These broad-brush comparisons of informal and formal teaching contexts were well received by critics of Western-style schools favoring progressive reforms, such as experiential and project-based learning (Gladwin, 1970).

Although sometimes inspired by ethnographic accounts, most psychologists interested in the effects of culture relied on methods rooted in their own discipline. This work was focused on testing the relative effects of formal and informal teaching contexts on children's development of supposedly universal cognitive structures (Rogoff, 1981; Scribner & Cole, 1973). At first, unschooled children from communities relying on informal socialization were reported to perform below the age norms on assessments, giving rise to a major controversy about the universality of cognitive developments such as Piagetian conservation or logical reasoning (Price-Williams, 1975, 1980). Although early work by Price-Williams (1961) suggested that contextual factors affected performance on conservation and other tasks, it took more than 2 decades before the implications of his pioneering experiments became widely appreciated. In a review of the literature, Rogoff (1982) concluded that formal schooling nurtures a variety of specific cognitive skills that are directly related to common classroom activities. These skills include graphic conventions for representing two-dimensional patterns, memory of disconnected bits of information, and taxonomic categorization. In addition, informal teaching embedded in everyday routines also produces specific developmental effects, so that the broadly defined distinction between formal and informal itself is an inadequate characterization of contextual variation (Nerlove & Snipper, 1981). The methodological lesson was clear: Context matters and can greatly influence conventional methods of assessment (Gallimore, 1996; Rogoff, 1982; Scribner & Cole, 1981).

## Ethnographies of the Classroom

The study of classrooms as cultural contexts was not started by psychologists but by sociolinguists and educational anthropologists (Cazden, 1986; Erickson, 1986). Originally called the ethnography of communication (Cazden, 1986), this approach was introduced into educational research by a very influential volume edited by Cazden, John, and Hymes (1972). From these diverse origins developed a line of investigation called *microethnography* by some, in contrast to the general ethnographies of the earlier era. Although some dispute whether they are truly ethnographic (see Goldenberg, 1988, for a discussion of this issue), some elements of that tradition are present in many of these richly described, detailed accounts of teacher–student communication and interaction, often focused on a single classroom, one teacher, a few students, or a single school (Duranti & Ochs, 1986; Erickson, 1986; Heath, 1983; Malcolm, 1979; Phillips, 1972; Tharp & Gallimore, 1989).

As a result of microethnographic studies, the field moved beyond generalized characterizations of contexts as formal or informal to detail-specific cultural and contextual factors that influence student behavior and performance in classrooms (Cazden, 1986; Erickson, 1986) and communities (Heath, 1983). For example, this work documented many ways cultural differences between home and school produce incompatibilities, including different assumptions about role relations between children and adults, participation structures, and appropriate interaction scripts (Cazden, 1986, 1988). For several U.S. and Canadian minority communities, microethnographies produced evidence of better student performance when classrooms incorporated cultural features of local communities into their everyday contexts and depressed performance when they did not (Erickson, 1986; Tharp, 1989).

## Early Uses of Video

In the development of microethnographic methods, audio and video recording technologies played an increasing role as investigators sought to focus on specific contextual features of classrooms that influenced student participation and performance. In most cases, however, recordings were used mainly to augment production of highly detailed, accurate transcriptions of speech. Video, when available, was used often to identify child speakers by name on transcripts so that the analysis of discourse could take account of who said what to whom. Reference to the images on the video, of course, yielded greatly enriched general descriptions, annotations of transcripts, and some limited analysis of nonverbal communication. However, a careful examination of much of microethnographic literature suggests that video was used primarily as a stenographic recording, a technique heralded as a great methodological breakthrough in teaching research at the beginning of the 20th century (Stevens, 1910).

Nevertheless, when cameras entered the classroom, the visually enhanced recordings generated rich, qualitative descriptions. A new problem surfaced: With so much rich information, it became difficult to study large numbers of classrooms. By the 1990s, the annual program of the American Educational Research Association was filled with qualitative investigations in a single classroom or a small number of classrooms. The small sample sizes made it difficult to generalize the findings to other classrooms, and it was not easy to get education policymakers to pay serious attention to studies that did not allow generalization to a wider population of schools and classrooms.

## The Alternative Tradition of Survey Research

Apart from the tradition of cultural studies previously outlined, a separate line of cross-cultural classroom research has had its own history. In the 1960s, an international organization called the International Association for the Evaluation of Educational Achievement (IEA) started the work of comparing academic achievement across cultures (Husen, 1967). By the 1970s, IEA was moving well beyond the study of achievement to collect data on teachers and teaching as well (McKnight et al., 1987). These researchers came primarily from a survey research tradition: They aimed for large, nationally representative samples of students, teachers, and schools. However, because of the difficulties involved in large-scale cross-national research, their studies of classroom processes relied mostly on teacher reports on questionnaires or on simple observational coding schemes.

At this point, the cross-cultural study of classrooms was at a problematic juncture. The purely qualitative approach that was associated with video analysis could not be used as the foundation for more general models of teaching and learning in classrooms or to inform policy on a wide scale. On the other hand, questionnaires and checklist-type observational schemes, especially across cultures, lacked validity. By the end of the 1980s, researchers in both camps were calling for an integration of qualitative and quantitative approaches. Thus, the video survey was born.

## A NEW INTEGRATION: THE VIDEO SURVEY

The first video survey was developed at the University of California, Los Angeles (UCLA), for the TIMSS video study. The concept was to marry videotaping, heretofore used on a small scale for mostly qualitative analysis, with national sampling, commonly used in survey research. Video surveys are especially promising ways to integrate the qualitative and quantitative study of classroom teaching across cultures, increasing the investigator's chances of capturing not only powerful universal or etic quantitative indicators but culturally particular or emic qualitative categories (Pike, 1966). Pike distinguished between analyses derived from within a cultural system (i.e., emic) and those that are constructed from outside the system (i.e., etic). Most emic approaches tend to be small sample, qualitative, and relatively particular in their descriptions, whereas etic approaches are larger sample, quantitative, and more universal in their descriptions. Each approach is constrained by obvious limitations and infrequently used in concert.

Video surveys combine videotaping (typically used for small-scale qualitative studies) with national probability sampling (most often used in large-scale survey research). Qualitative analyses of video can be validated against a national sample of videos. Quantitative analyses are rendered more interpretable by being efficiently linked to specific video examples of the categories coded. The key enabler for this kind of research has been the revolution in multimedia computer technology over the past 5 years. Now it is possible to archive large quantities of video information in a way that can easily be indexed and retrieved.

### Advantages of Video Surveys

Video surveys, like any video-based method, greatly expands the researcher's ability to analyze complex human interactions such as those found in classrooms. Live observations are limited to whatever an observer can record. Checklists can be useful, but it is possible for a live observer to make only a limited number of reliable judgments at the speed required for classroom research. There is simply too much going on. Video, on the other hand, can be paused, rewound, and watched again. Two observers can watch the same video, independently, and go back to replay and discuss those parts that they saw differently. Videos can be coded multiple times, in passes that require only limited judgments by an observer on any single pass. This makes it easier to train observers and enables reliable coding of complex events.

The most important advantages of video derive from its concrete, vivid, and raw, unanalyzed nature (i.e., the categories can be derived from these data rather than vice versa, leaving these data open to a vast array of analyses). There are at least five major opportunities that arise:

1. Video records of classroom lessons provide opportunities to discover unanticipated ideas and alternative analytic categories. Checklists and other live coding schemes imply that the events of interest are fully known and that coding categories can be defined in advance. Video permits researchers to capture events more fully and to take advantage of serendipity whenever possible.
2. The concrete nature of video means that it is not as theory bound as other methods of data collection. The same video data is useful to a far wider range of researchers than would be the case for questionnaires or live coda observation systems. Video data, therefore, are amenable to analysis from multiple perspectives and provide a natural focal point for interdisciplinary collaboration. Psychologists, anthropologists, sociologists, and others interested in understanding classroom processes can all make good use of a single video data set.
3. Video is not only interesting to researchers from different perspectives; it also has a longer shelf life than other kinds of data. Researchers of today would have little interest in reanalyzing most of the process–product data generated by classroom researchers in the 1970s and 1980s, mainly because the theoretical context that motivated the collection of those data was so different then. However, imagine if videos of teaching had been collected during these earlier periods. The videos would be of great interest and easily appropriated by the theories of today.
4. Video provides concrete referents for the words and concepts used to describe instructional processes. Teachers and educators often lack a shared language for describing teaching. Certain terms, such as *problem solving*, are frequently used but rarely defined. Video images make it possible for multiple observers from multiple backgrounds to agree on the meanings of such commonly used words. Not only does this advance the scientific understanding of classroom processes, but it also facilitates the communication of research results to various constituencies.
5. Perhaps the greatest advantage of video is that it allows us to integrate qualitative and quantitative methods of analysis. Some ways in which this can be done are illustrated in the video studies described later in the article.

Some of these five advantages apply to almost any use of video for the study of teaching without the cost of a national sample. Why expend the resources to do a video survey of a national sample of classrooms? In addition to the benefits of a larger and representative sample, there are at least three other specific reasons to believe a national sample is critical, whether the audience includes researchers, policymakers, or teachers.

First, only with a national sample can evidence be provided about the average experiences of students in classrooms. This kind of information is needed to begin constructing grounded explanations for the achievement levels of national samples of students.

Second, for policymakers seeking to improve education, a national sample provides the only credible source of information about the national-level effects that education policies are or are not having inside classrooms. For example, an organization like the National Council of Teachers of Mathematics engages in a variety of efforts to improve the quality of mathematics teaching nationwide. Without a national sample, it is not possible to estimate the nature and magnitude of influence these efforts have (Stigler & Hiebert, 1999).

A third and more specific reason for a national sample involves the ultimate application of the results. We believe that typical lessons, which only can be identified in the context of a national sample, play an important role in teacher professional development. Teachers often are shown videos of only exemplary practice, a strategy with some potentially serious consequences: More than a few teachers may believe that they could never teach like the exemplary teacher in the video, or they may copy some features and not others, yielding an incoherent system of teaching. Using typical lessons produces a very different effect: It reminds teachers of their own classrooms and provides a context for reflecting on and critiquing their own teaching methods and for generating and thinking through alternative approaches to teaching.

### Limitations of Video Surveys

Despite the advantages of video and the potential of new technologies to simplify the task of organizing and analyzing large video data sets, video surveys have their limitations.

Given finite resources, the decision to study a national sample of teachers usually places severe limits on the length or the number of times that each classroom can be videotaped. In the TIMSS and TIMSS-R video studies to be described shortly, only one lesson was videotaped in each classroom. Thus, the information collected about any particular teacher's classroom is scant and decontextualized. Although sampling one lesson per teacher can yield a reliable picture of teaching at the national level, it is not a reliable picture of an individual teacher's teaching. Also, it is important to realize that video is not a complete picture of reality. Much of what is going on in the situation being videotaped is not visible on the screen, so even video provides limited data.

The concrete nature of video images can be problematic, even if the camera is pointed in the ideal direction. Concrete images can be quite persuasive to the human information processing system, even if they turn out to be completely unrepresentative of what typically occurs. This fact is well known by cognitive psychologists: Humans are easily misled by anecdotes, even when they are told to ignore them (Nisbett & Ross, 1980). Researchers must be aware of this extraordinary power of video when engaged in qualitative analysis and when interpreting and reporting the results.

Another potential problem with video is the camera effect. Will students and teachers behave as usual with the camera present? May a teacher, knowing that he or she is to be videotaped, even prepare a special lesson just for the occasion that is unrepresentative of normal practices? This problem may not be as serious as it appears initially. First, this problem is not unique to video studies. Teachers' questionnaire responses, as well as their behavior, may be biased toward cultural norms. It is 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 of that. Students may look puzzled, comment about the differences, or may not be able to follow routines that are new for them.

It also should be noted that changing the way a teacher teaches is not accomplished easily, 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 obviously will try to do an especially good job and may do some extra preparation for a lesson that is to be videotaped. Therefore, the video is likely to show a somewhat idealized version of what the teacher normally does in the classroom.

In spite of the limitations, we believe video surveys provide a promising new design that opens new windows on classroom practice. The following two examples illustrate the benefits of this approach by presenting some sample findings and identifying some lessons we have learned. In Examples A and B, we provide detailed descriptions of how some aspects of etic and emic analyses were combined in the TIMSS and TIMSS-R video studies.

## EXAMPLE A: THE TIMSS VIDEO STUDY

The TIMSS is the largest cross-national study of educational achievement ever conducted. It is the third in a series of studies organized by the IEA (Husen, 1967; McKnight et al., 1987). Forty-one countries participated in the TIMSS, which included mathematics and science achievement testing of 4th-, 8th-, and 12th-grade students (complete information about TIMSS can be obtained at http://nces.ed.gov/timss).

Organizers of the TIMSS wanted to go beyond simple comparison of mathematics and science achievement to consider the contextual factors that may help to explain international achievement differences. Three supplementary studies were conducted with this aim in mind: Case studies of the educational systems in each country (Stevenson & Nerison-Low, 1999); an analysis of curriculum materials from the TIMSS countries (Schmidt, McKnight, & Raizen, 1997); and the TIMSS video study. The TIMSS video study was carried out in three countries: Germany, Japan, and the United States. Germany and Japan were chosen because they are viewed as economic competitors to the United States. The goals of the study were to create a portrait of eighth-grade mathematics teaching in the three countries and to assess the effects of reform policies on U.S. mathematics teaching practices.

The design of the study was relatively simple: National samples of eighth-grade mathematics classes were selected and videotaped once during the school year for a single class period. As far as we know, this is the first time that videotapes have been collected from a national sample of anything, and certainly of classroom teaching. Numerous methodological and analytic challenges had to be met, some by designing new technological aids. The study is described briefly in the next section; for more complete accounts see Stigler, Gonazales, Kawanaka, Knoll, and Serrano (1999) and Stigler and Hiebert (1999).

## Method

The TIMSS video sample was a subsample of the full TIMSS sample of eighth-grade mathematics classes. Not only were specific teachers selected randomly, but specific class periods were as well. The final sample included 231 mathematics lessons: 100 in Germany, 50 in Japan, and 81 in the United States.

Videotaping procedures were standardized so that all videographers followed the same decision rules. Only one camera was used in each classroom, and in general it focused on what an ideal student would be focusing on—the teacher. After taping, the teacher filled out a questionnaire describing the goal of the lesson, its place within the current sequence of lessons, and so on.

When the tapes arrived at the TIMSS video research laboratory at UCLA, they were digitized (to increase durability and random access) and then translated and transcribed. The transcripts then were linked by time codes to the video in a multimedia database. The technical features facilitated the viewing of the tapes.

The next task, to develop a coding scheme that would turn the visual and auditory images into quantitative data, required the intensive and prolonged attention of multiple researchers with different backgrounds and expertise. We convened a team of six code developers, two from each participating country, and spent the summer of 1994 watching and discussing 27 field test tapes. From these discussions emerged the initial coding system. This system, built on theories of mathematics learning and teaching as well as on an interest in defining codes that would yield valid descriptions across the three cultures, included categories such as the organization structure of the classroom and the kind of work expected of students.

A separate group of coders, referred to as the math group, were asked to code a subset of the lessons. There was general agreement that some important but subtle differences in the mathematics content of the lessons could not be defined fully and coded reliably by the primary coders. For these analyses we needed a group more sophisticated in mathematics. The math group, composed of four postsecondary mathematics teachers, developed their own coding scheme and reached consensus about more global aspects of subject matter quality.

Two key aspects of the code development process foreshadow two important lessons that have been learned. Meaningful coding of a classroom lesson is based on segmenting the lesson into meaningful chunks. This requires identifying a unit of classroom practice that can be identified reliably so that its beginning and end points can be marked. Choices about which units to use are crucial because different units afford different analyses. In some ways, this is the most challenging aspect of coding video lessons, both conceptually and practically. The lesson we have learned here is developed later as we describe the TIMSS-R video study.

The second key aspect of code development was the need for multiple kinds of expertise. Representatives from each country were essential for creating a coding scheme that fairly described teaching in each country, and the members of the math group were essential for making judgments about the quality of the content. This work requires a true interdisciplinary (and intercultural) research team. Again, this lesson is developed more fully in later sections of this article.

## Findings

A few results are presented to illustrate the nature of the findings that can be generated by the video survey approach.

*Quantitative results.* Many interesting quantitative results address the nature of the mathematics engaged during the lesson and the way that the teacher and the students treated the mathematics.

Mathematical concepts and procedures either can be simply stated by the teacher or developed through examples, demonstrations, and discussions. Suppose the topic is the area of right triangles. Teachers can state that 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 "½ base × height" can be derived by halving rectangles (and parallelograms) to form two triangles. Of course, the teacher may ask students to develop some of this themselves. We coded this *developed* if teachers made any attempt to motivate a procedure or explain why it worked. As shown in Figure 1, concepts and procedures were usually developed in German and Japanese lessons but usually just stated in U.S. lessons.

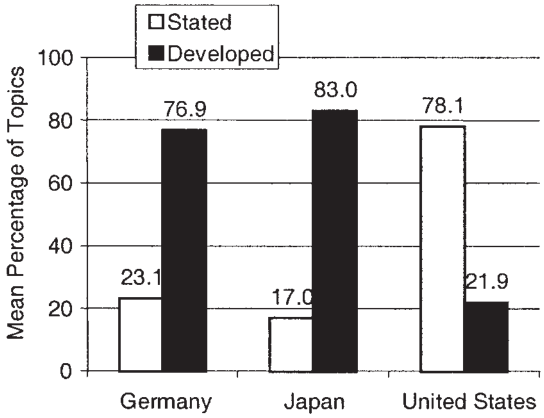


FIGURE 1 Average percentage of topics in eighth-grade mathematics lessons that contained concepts that were developed versus those that were stated. *Note.* From the U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994–1995.

When we examined what students 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 freedom students had in solving the problems. In some lessons, a procedure was demonstrated or developed by the teacher, and then students were asked to apply this procedure to solve the assigned problems. In other lessons, students were asked to develop procedures themselves, based on what they had learned in previous lessons. The math group described this distinction with the terms *task controlled* and *solver controlled*. Results showed that in 63% of Japanese lessons the tasks were solver controlled. Percentages of comparable lessons in Germany and the United States 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 because 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 problem (e.g., 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 (e.g., finding the areas of nonright triangles). The third category requires even more of students because they are asked to invent solution methods, analyze mathematical situations, or generate mathematical proofs. For example, students may be asked to work out 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 Figure 2. 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 their time practicing routine procedures.

A summary 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 previously reported. Although judgment was subjective, the members of the math group reached agreement for all the lessons they coded. Figure 3 shows the results. Again, Japanese and U.S. lessons are the most different, with German lessons falling in between.

*Qualitative descriptions.* What do these individual findings add up to? We noted earlier that their meaning could not be divined by studying just the findings themselves. By looking back at the tapes, however, it was possible to detect patterns across the lessons within each country that provided the larger framework in which the individual results could be interpreted.

The patterns can be briefly described as follows. In the United States, most lessons moved through the following four activities, in sequence: (a) Review previous material, either through a warm-up activity or by checking homework; (b) demonstrate how to solve problems for the day, with the teacher leading the students through a relatively quick demonstration of the relevant procedures; (c) practice the same procedures on a set of similar problems; and (d) correct the assigned problems and assign additional, similar problems for

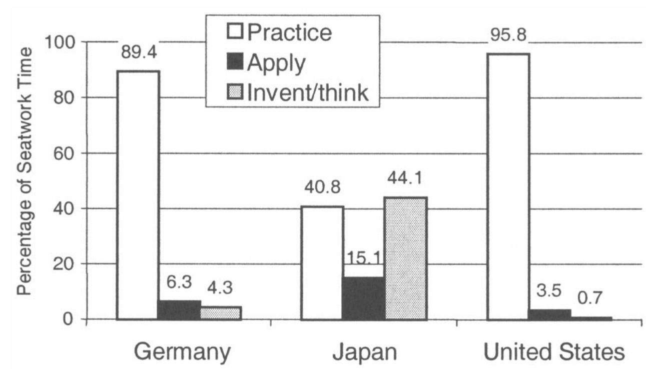


FIGURE 2 Average percentage of seatwork time spent in three kinds of tasks. *Note.* From the U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994–1995.

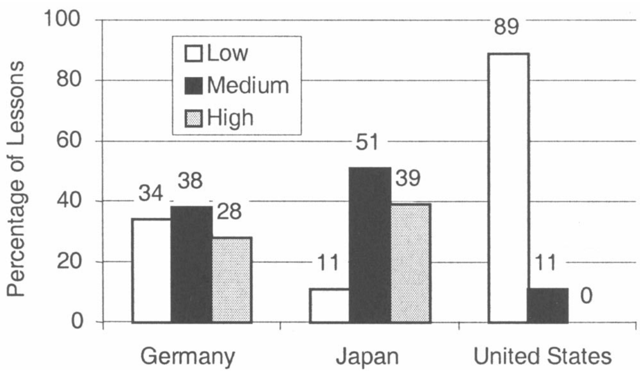


FIGURE 3 Percentage of lessons rated as having low, medium, and high quality of mathematical content. *Note.* From the U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Videotape Classroom Study, 1994–1995.

homework. Usually some class time is allowed for beginning the homework.

German lessons often were consistent with the following sequence: (a) Review previous material, either by reviewing homework or reminding students what has been accomplished up to this point; (b) present the topic and problems for the day; (c) develop the procedures to solve the problems, often with the teacher carefully guiding the students through the details; and (d) practice the procedures together or individually on a set of similar problems.

A typical Japanese lesson was characterized by the following sequence of activities: (a) Review previous lesson, usually through a brief summary by the teacher; (b) present the problem for the day, often a problem that builds on the previous day's work; (c) students try to solve the problem, individually or in small groups; (d) students share solution methods they have tried with comments and suggestions from the teacher and other students; (e) summarize the major point of the lesson, frequently through a brief lecture by the teacher. Often, students attempt to solve a second problem before the lesson ends with a summary.

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. 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 United States 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.

Indeed, one of our primary observations from integrating the quantitative and qualitative results is that teaching is a system, and this fact must be understood and appreciated to interpret research results and to design effective programs for improvement (Stigler & 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. It makes little sense, for example, to compare the amount of time students spend solving problems unless one knows how that activity is functioning with the system. Moreover, cross-cultural differences in individual features of teaching do not, by themselves, suggest how to improve teaching. It would be foolish, for example, to recommend simply replacing one feature by another or adding some features to an existing system. The systems are likely to change the features rather than vice versa.

## Lesson 1: Integrate Qualitative and Quantitative Analysis

As we claimed earlier, one of the primary benefits of video survey research is that it affords an integration of qualitative and quantitative analysis. We learned this lesson many times while conducting the TIMSS video study, perhaps no more forcefully than when we detected the way in which the patterns of teaching within each country wove together the individual quantitative results.

Integration often can be achieved by what we call the *circle of analysis* that links qualitative images and descriptions with quantitative coding and analysis. The summary linking, described previously, emphasized qualitative analyses later in the process, but the circle can, and often does, begin with qualitative analysis, sometimes of one or a select few lessons. In-depth discussion of single lessons leads to hypotheses either about the nature of teaching in a country in general or about the relations among different parts of lessons in that country. A hypothesis is then tested empirically by coding the full sample of lessons, producing a quantitative result that either supports or undermines the hypothesis. Once coding is complete, the hypothesis is reevaluated in light of the coding. Often the hypothesis is refined once the coding is complete, and a more intelligent question is asked. In this way, the circle revolves from observing and generating to coding and evaluating.

An example of this process comes from a memorable discussion at an early stage of the TIMSS video study that involved the prevalence of outside interruptions during the math lesson. The Japanese representatives were shocked to see such interruptions in U.S. classrooms and claimed never to have witnessed something similar in their own country. This observation generated a hypothesis, one that could be tested by coding the full sample of lessons. Are interruptions really as prevalent in the United States and as rare in Japan as suggested? It turned out that our Japanese colleague was right: There were no outside interruptions in the Japanese sample, whereas 30% of U.S. lessons were interrupted.

The important point is that neither qualitative or quantitative analysis alone is sufficient for comparing classroom practices across cultures. Quantitative coding is necessary to validate insights gained from close qualitative analysis. It is too easy to be misled by anecdotes remembered from individual cases. On the other hand, qualitative descriptions are essential because they lend substance and coherence to the results of quantitative coding.

## EXAMPLE B: THE TIMSS-R VIDEO STUDY

The TIMSS-R, a follow-up study to the TIMSS, started in September 1998. The main purpose of the TIMSS-R is to conduct mathematics and science achievement testing of eighth-grade students 4 years after the TIMSS. Students in eighth grade for the TIMSS-R were in fourth grade for the TIMSS. We hope that testing a new cohort 4 years later will shed some light on trends over time in cross-national achievement differences.

The TIMSS-R includes a video component as well, the TIMSS-R video study, which we are conducting in collaboration with a number of colleagues around the world. As we write this article, videotapes and related data are being collected in nationally representative samples of classrooms in seven countries: Australia, the Czech Republic, Hong Kong, Japan, the Netherlands, Switzerland, and the United States. Like the TIMSS video study, this follow-up study is funded by the U.S. Department of Education, National Center for Education Statistics. However, some participating countries are paying their own expenses or are being subsidized by the National Center for Education Statistics.

The TIMSS-R video study is important for several reasons. First, the sample has been expanded to include more high-achieving countries than in the TIMSS video study. In the previous study, only Japan was high achieving. Because Japanese teaching practices differed so markedly from those in both Germany and the United States, many inferred that Japanese classrooms are prototypical of those used to produce high-achieving students. The TIMSS-R video study includes countries with high achievement but with different cultural traditions. This will expand the vision of what teaching looks like in classrooms where students reach high levels of achievement.

A second important attribute of the TIMSS-R video study is that science classrooms, as well as mathematics classrooms, are videotaped. This allows researchers to compare the teaching of two subjects within and between cultures.

Finally, the TIMSS-R video study allows us to expand and extend the methodology of video surveys. The lessons that we began learning in conducting the TIMSS video study are being elaborated as we conduct the TIMSS-R video study, and new lessons are being learned. We highlight these as we sketch the methods of this new video survey.

The goals of the TIMSS-R video study can be summarized as follows:

- Investigate and compare mathematics and science teaching practices among the seven participating countries.
- Compare teaching practices in the lower achieving countries (i.e., United States and Australia) with those found in higher achieving countries.
- Discover new ideas about teaching mathematics and science.
- Develop new teaching research methods and tools for teacher professional development.
- Create a digital library of images of teaching to inform United States educational policy.
- Stimulate and focus discussion of teaching practices among educators, policymakers, and the public.

### Sampling

Our goal is to videotape a representative sample of eighth-grade mathematics or science lessons in each participating country. Four countries—Australia, the Czech Republic, the Netherlands, and the United States—will participate in both the mathematics and science components of the study. Hong Kong and Switzerland will participate in only the mathematics component and Japan in only the science component.

Sampling is done in two steps: One hundred schools in each country are selected, and then one mathematics and one science class within each school is selected. Sampling of schools in each country uses the same procedures used in the TIMSS-R achievement study. Most countries, however, are not videotaping in the same schools in which the TIMSS-R testing is being conducted. Both public and private schools are eligible for inclusion in the sample. A replacement school is chosen for each of the 100 schools in each country and is used if the original school refuses to participate.

Within each school, one mathematics class, one science class, or both is randomly selected for videotaping. Classes are selected, if possible, so that they can be videotaped on the same day. No substitutions of teachers or class periods are allowed once the sampling has been completed. Thus, principals or administrators are not able to select their best teachers or students for videotaping.

Each class is videotaped only once. Every effort is being made to ensure that at least 8 months of the academic year are represented in the full set of videotaped lessons within each country. This is important, especially in countries with national curricula in which teaching techniques may vary systematically with changes in topic over the course of 1 school year. Individual lessons are videotaped, however, without regard to topic being taught or type of activity. In some countries, students take two separate science classes in eighth grade—one in biology and the other in physical sciences. One of these is randomly chosen in each school.

### Data Collection Procedures

One complete lesson—mathematics or science—is videotaped in each classroom. What counts as a lesson is determined according to what is standard in each participating country. As in the TIMSS video study, videotaping procedures are standardized, and videographers are trained to follow the standard procedures. New procedures have been developed to account for the unique characteristics of science lessons compared with mathematics lessons.

In the TIMSS-R video study, two cameras are used to videotape each lesson. One camera focuses primarily on the teacher and is operated manually. This camera also captures closeups of students' work during periods when students are working independently. The second camera is stationary, positioned at the front of the room facing the students to capture students' interactions with the teacher or with each other during the lesson.

In addition to videotaping, the following data are being collected to help us understand each lesson more fully:

- A teacher questionnaire in which the teacher provides additional information about the videotaped lesson, background and experience, attitudes, and professional development.
- A student questionnaire designed to elicit basic demographic characteristics of the students and their expectations for education.
- Samples of student work from the lesson.
- Samples of text pages, worksheets, and other materials used in the lesson.
- Samples of tests that the teacher may use to evaluate students' learning of the curricular content of the lesson.

The collected data are sent to our research facility in Los Angeles. Tapes are digitized and stored on a video server in MPEG-1 format. They are transcribed and translated into English, and transcriptions are linked by time code to the video using the vPrism multimedia database software. Questionnaires and other supporting data (such as textbook pages) are scanned and stored digitally in PDF format. They are linked to the video, so that coders have rapid and convenient access to all materials relevant to the analysis of each lesson.

### Coding and Analysis

In the first TIMSS video study we constructed coding categories by watching and discussing lesson tapes with our colleagues from Germany and Japan. Agreement eventually was reached on how the lessons could be segmented and analyzed using the same analytic categories for all three nations. Theories of learning and teaching entered the discussions, mostly as tools to make sense of individual videotapes rather than as overarching principles for shaping the coding scheme. The initial segmentation of the tapes, for example, was in terms of shifts in classroom organization and activities (e.g., from classwork to seatwork), primarily because these were the distinctions on which agreement was reached most easily.

The detection of culture-specific patterns of teaching was made relatively late in the first TIMSS video study. As described earlier, after the bottom-up quantitative coding was underway, we began a more top-down qualitative analysis from which we derived an understanding of cultural systems of teaching in each country (Stigler & Hiebert, 1999). To better facilitate the design of bottom-up codes, we began code development in the TIMSS-R video study using a more global top-down approach. The construction of lesson patterns was led by the representatives from each country, who viewed more than 50 field-test videotapes; they applied their own knowledge of classroom teaching and solicited feedback from educators in their own countries. Given the conclusions of the first video study, we expected to find cultural patterns of teaching in each country and wished to use these as an analytic framework.

All aspects of the coding and analysis process can benefit from exploiting the framework provided by the top-down constructed patterns. The country patterns derived from the top-down approach serve as initial working hypotheses to guide and interpret bottom-up coding and analysis. For example, by comparing the early descriptions of lesson patterns within each country, we agreed on six dimensions of classroom practice that could be used for all descriptions: purpose of activity, classroom routines, content, actions of participants, classroom talk, and climate. Using the same dimensions enabled the patterns to serve two purposes. First, they provide a basis on which to identify key universal variables for quantitative coding. Second, they provide the context in which the results of quantitative coding can be interpreted. Using the circle of analysis described earlier in this article, we expect to revise and refine the country patterns as the project progresses as well as use the patterns to guide and refine bottom-up category development.

### Lesson 2: Engage in Simultaneous Top-Down and Bottom-Up Analyses

In developing a coding and analysis scheme, there is always a question of whether it is better to start with global, theoretically driven questions about the nature of teaching or with the coding of individual units, letting the bigger questions suggest themselves as analysis proceeds. On one hand, it is not possible to code small units and then hope that the results have clear implications for the larger questions about teaching (Reese, Kroesen, & Gallimore, 1998). The whole is greater than the sum of the parts, and the whole will not necessarily emerge once the parts are coded. On the other hand, larger questions must be continually refined and modified as analysis proceeds. Sometimes what appears to be a well-defined and significant question at the beginning turns out to be ambiguous or unimportant once results have been coded. Close inspection of the results of coding can suggest alternative ways of thinking about or formulating a research question. In many cases, the method of video survey allows both kinds of analysis to proceed simultaneously.

## Lesson 3: Continually Reevaluate the Units of Coding and Analysis

Multiple units of coding and analysis are needed to capture the images of teaching caught on videotape. The classroom lesson is an important unit of analysis; it is the unit of sampling in both the TIMSS and the TIMSS-R, but lessons must be parsed into smaller units for many analyses, and selecting these units is challenging. A major shift in our coding and analytic approach in the TIMSS-R video study is our decision to use content (e.g., mathematical content) rather than classroom organization as the basis for our initial segmentation of the videos. Although segmenting lessons based on content is more difficult than segmenting based on classroom activities, we believe it better captures key features of teaching because it is generated by current theories of how students learn school mathematics and because it affords a more direct analysis of what we believe are important differences in practice.

There are two critical issues to consider when selecting units for parsing classroom lessons: One concerns the size of the units, and the other concerns the basis on which the units are defined. Lessons can be parsed into pieces that are too small or too large for the analytic task at hand. An example comes from the coding of discourse in the TIMSS video study. We parsed discourse into utterances, then coded each utterance into an elaborate system of categories: teachers' questions, students' responses, and so on. Although we were able to calculate the frequencies of each type of utterance within each lesson and produce graphs of the average frequencies of each type across countries, we were not able to aggregate these data into a picture of how teachers in different cultures use language to teach mathematical content. The problem is that our unit was too small, and there was no clear path to larger, more meaningful units. The key, of course, is not how frequently utterances are made but when they are used and how their use relates to the instructional context. These relations were not captured by our coding.

The basis on which lessons are parsed is just as important as the size of the units. Theories of learning and teaching provide likely sources for identifying and justifying units that capture important features of the classroom. In mathematics education, many current theories claim that students' opportunities to learn mathematics are related to the kinds of mathematical problems they solve. In the TIMSS video study we tried, unsuccessfully, to parse lessons into mathematical problems. It simply proved too difficult to gain agreement among researchers from the three countries on what constituted a problem. What one person called a problem another called an exercise, and the enterprise collapsed in a sea of ill-defined words. Instead, we parsed the lessons based on shifts in interaction, from classwork to seatwork and back again. As it turned out, many indirect measures of mathematical problems still were developed, and these revealed significant differences among the three countries.

In the TIMSS-R video study, we have returned to the goal of parsing lessons on the basis of content and, in particular, using mathematical problems as a primary unit. Although it has been difficult to gain agreement among representatives of six countries on how to define a problem, the theoretical importance of the unit and the prevalence of mathematical problems in almost all lessons compelled us to work through the difficulties. Once problems are marked in each lesson, we can ask a range of questions about the nature of the problems and how they were treated during the lesson. We expect the payoff to be substantial when we reach the analysis phase of the project.

## Lesson 4: Build a Research Team With Diverse and Specialized Skills

We alluded earlier to the fact that coding the videotapes in the TIMSS video surveys was enhanced by the work of researchers with different kinds of knowledge—cultural knowledge of classrooms and specialized knowledge of mathematics. In fact, no single researcher, or small group of researchers, is likely to have the specialized knowledge and skills required to conduct cross-cultural video surveys. A team of researchers must be assembled, each member bringing special skills to the project. The areas of expertise that must be represented include traditional research skills, knowledge of teaching and learning, inside knowledge of classrooms within each culture, and subject matter knowledge.

We have found that researchers often find themselves working in nontraditional roles. In more traditional projects on classroom practice, the principal researchers define the codes, then train assistants to implement them. The results are brought back to the researchers, who then plan the next analysis. In cross-cultural classroom research using video survey data, these traditional roles do not work. On-site researchers usually do not have the cultural knowledge required to develop the codes. They must rely on native informants who may have extensive knowledge of classrooms in their country but not the kind of research training required for such a project. Similarly, ways must be found to consider the content being studied as the coding scheme is being developed and implemented.

Video survey permits a much different approach with certain advantages afforded by having so many vivid images of the teaching to be coded and analyzed. For example, the TIMSS-R video is assembling a large network of researchers. The on-site code development is being led by a mathematics educator (for mathematics) and a science educator (for science) with a team of country associates (one from each participating country). The country associates bring with them knowledge of classroom practices in their own culture and, to varying degrees, research expertise, teaching expertise, and subject matter expertise. The country associates are full-time participants for the duration of the project and work on code development for both mathematics and science.

Within each participating country, a country collaborator is coordinating the implementation of the study and consulting with the on-site team with respect to coding and analysis. In addition, the country collaborators have assisted in assembling a small group of country experts in each country that serve as informants and reviewers for code development. Reviews are conducted using a special section of the TIMSS-R video Web site. As codes are developed they are posted, together with video examples, on a special section of the Web site. This section of the Web site is secure and accessible only to those with a user ID and password. Country experts log into the Web site, study the codes and examples, then provide feedback through a linked Web discussion forum. One advantage of this approach is that it can be expanded easily to include a wider range of experts when needed.

Additional participants on the team include a U.S.-based steering committee, a reconvened math group for supplementary coding, and perhaps several teacher focus groups to provide additional observations of the videotapes.

One example illustrates how the work of these diverse individuals can be configured to boost the capabilities of the team. As mentioned previously, mathematical problems provide a primary unit for parsing the lessons in the TIMSS-R video. Many questions can be asked about the nature of the problems, including how conceptually challenging they are and how they are related to each other. These judgments require a level of mathematical expertise that not all our coders possess. Our strategy for dealing with this issue is to solicit the assistance of a researcher with expert knowledge of mathematics and a familiarity with our coding scheme to construct a comprehensive, detailed, and structured list of all mathematical topics covered in the eighth grade in all participating countries. Each problem marked in a lesson is then connected to a topic in the list. Analyses of conceptual challenge and relations then can be made by referring to the lists that are structured with these constructs in mind. In this way, the coders are able to generate data that are much richer than they could have generated on their own.

### Lesson 5: As the Number of Comparisons Increase, Challenges Rise Exponentially

The move from the TIMSS video to the TIMSS-R video has brought numerous challenges, some expected and some unexpected, due to the increase in number of nations included in the survey. Doubling the nations in the survey did not simply double the challenges with which we were confronted; it increased them many times over, and the nature of the challenges shifted as well. The logistics of implementing the study in more countries required new methods of communication and record keeping. One new challenge that we did not anticipate emerged during the code development process.

In the TIMSS video study, codes were developed by a team of three country associates. Although work toward agreement of code definitions and implementation rules was filled with periods of debate and temporary regress, the team was able to reach consensus on most codes within several months. Three additional coders, one from each country, were added to the team, and the six members perfected their coding skills. They worked from a basic coding manual but held many face-to-face discussions and constructed a great deal of implicit knowledge that supported their ability to code the lessons in similar ways.

In the TIMSS-R video study, codes are developed by a team of six country associates (for mathematics and for science) led by a seventh member. Not only has it proven to be more difficult to reach consensus on code definition and implementation rules, but it has been almost impossible to develop a substantive body of shared implicit knowledge that can be used to support intercoder agreement. Consequently, we have had to rely, to a much greater extent in this study than the previous one, on explicit and clearly articulated written records of code definitions and rules. This has become even more essential as coders (two from each country) were trained. Discussions to clarify the meaning of codes are not sufficient; for the code to exhibit sufficient reliability it has to be described fully and in great detail with many supporting examples. This, in turn, leads to a coding manual that is nearly overwhelming in its scope and detail but is essential for handling the scale of the TIMSS-R video study.

## CONCLUSIONS

We have learned two kinds of things from using comparative video surveys. One pertains to the nature of classroom teaching in different cultures. This was the intended set of outcomes, and the interested reader can consult other publications for the full set of findings (Stigler et al., 1999; Stigler & Hiebert, 1999). We only illustrated these findings in this article.

The second set of findings was methodological. We learned lessons about conducting comparative video surveys that served as the focus for this article.

The advantages of video surveys of teaching can be summarized as follows:

- Enables the study of complex classroom processes.
- Allows a relatively large sample chosen through standard sampling techniques.

- Increases interrater reliability and decreases observer training problems.
- Amenable to coding from multiple perspectives.
- Facilitates integration of qualitative and quantitative information.
- Provides referents for teachers' descriptions.
- Facilitates communication of the results of research.

These advantages must be considered within the limitations of this approach. From one perspective, the limitations can be seen as a compromise within the emic–etic tension. On one hand, there is the anthropological injunction to seek insider or fully emic perspectives when investigating cultural matters. Although the TIMSS-R is employing experts from each participating culture, what will be accomplished falls short of a full emic analysis. Such a full analysis would include, in addition to the videos, interviews of participating teachers and students, classroom observations, and extensive input from culture experts to secure within-culture understanding and perspectives (see Kawanaka, 1999, for such an effort).

Perhaps the two TIMSS video studies come closer to an acceptable etic analysis with their use of national samples. However, they fall short here as well on several accounts. A sample of 100 classrooms in a given subject matter area is small in national terms (however huge in video logistics it may seem), and it is further restricted by sampling only one lesson from each teacher. The latter is a severe limitation given that mathematics curricula are not topically segmented into daily units but rather into units of longer intervals. Thus, it is likely that some systemic features of teaching in some cultures are missed because there is no way to capture the connections represented in the longer intervals (e.g., teaching units that last 5 or more days).

The comparative video survey is a hybrid of sorts, capturing some of the spirit of both emic and etic designs. One metaphor that has been used to describe this work is that the individual features captured in specific quantitative indicators are like the tops of a mountain range poking above the surface of the water (Stigler & Hiebert, 1999). Specific features are like islands in the sea because they indicate that something bigger and important lies beneath the surface and they can reveal something of its nature.

Are mountaintops in the sea an apt metaphor for the tension between etic and emic analyses? In one respect the metaphor is appropriate, for it conveys the idea that one must grasp the whole system (i.e., emic analysis) to correctly interpret individual parts (i.e., etic analysis), just as one cannot fully describe an island in the sea without an analysis of the topography of the hidden mountain. In another way, the metaphor falls short. Islands in the sea are instantly recognizable as tops of things hidden below the water. We cannot know the specific topography of the drowned mountain from the visible top, but we know the island means that something significant may lie below the surface. The visible island may be the top of a still-growing volcano resting on the bottom of the Pacific or a misleading indicator of what cannot be seen, such as a coral atoll resting on a sinking crater. The same is true for many parts of a teaching system. Japanese teachers used chalkboards and U.S. teachers did not, but the significance of this difference was impossible to judge until the underlying structure was mapped. Differential use of chalkboards could mean nothing much, or it could mean a lot. Until we know as much about teaching and learning in different cultures as we do about the geology of the earth, there is a severe limit on what we can infer from statistically significant differences in specific indicators. New vistas may be just over the horizon as new technologies make comparative expeditions more feasible and appealing.

More is at stake than improving researchers' tools. Comparative study of teaching and learning using the results of video survey can be the basis for a powerful addition to the armamentarium of teacher professional development. Throughout we have noted some possibilities, including building digital libraries of teaching, multiple examples of teaching the same concepts, and others. One of the most important lessons is what we can learn about ourselves from comparative study of teaching and learning. Like the TIMSS research team, many U.S. educators are likely to be surprised to learn how similar teachers teach in our culture. What they previously saw as great variation in our classrooms will pale in comparison to how U.S. practices differ from others around the world. Because new methods seem easier to implement and less like utopian dreams when they are seen as commonplace in another culture, U.S. educators will be encouraged to try new ways of teaching.

## REFERENCES

Cazden, C. B. (1986). Classroom discourse. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 432–464). New York: Macmillan.

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

Cazden, C. B., John, V., & Hymes, D. (Eds.). (1972). *Functions of language in the classroom.* New York: Teachers College Press.

Diamond, S. (1971). Epilogue. In M. L. Wax, S. Diamond, & F. O. Gearing (Eds.), *Anthropological perspectives on education* (pp. 298–303). New York: Basic Books.

Duranti, A., & Ochs, E. (1986). Literacy instruction in a Samoan village. In B. B. Schieffelin & P. Gilmore (Eds.), *Acquisition of literacy: Ethnographic perspectives* (pp. 213–232). Norwood, NJ: Ablex.

Erickson, F. (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119–161). New York: Macmillan.

Gallimore, R. (1996). Classrooms are just another cultural activity. In D. Speece & B. Keogh (Eds.), *Research on classroom ecologies: Implications for children with learning disability* (pp. 229–250). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Gallimore, R., Boggs, J. W., & Jordan, C. (1974). *Culture, behavior, and education: A study of Hawaiian-Americans.* Beverly Hills, CA: Sage.

Geertz, C. (1984). "From the native's point of view": On the nature of anthropological understanding. In R. A. Shweder & R. LeVine (Eds.), *Culture* *theory: Essays on mind, self, and emotion* (pp. 123–136). Cambridge, England: Cambridge University Press.

Gladwin, T. (1970). *East is a big bird.* Cambridge, MA: Belknap.

Goldenberg, C. N. (1988). Methods, early literacy, and home–school compatibilities: A response to Sledge et al. *Anthropology and Education Quarterly, 19,* 425–432.

Greenfield, P. M., & Lave, J. (1982). Cognitive aspects of informal education. In D. A. Wagner & H. W. Stevenson (Eds.), *Cultural perspectives on child development* (pp. 180–207). San Francisco: Freeman.

Heath, S. B. (1983). *Ways with words: Language, life, and work in communities and classrooms.* New York: Cambridge University Press.

Howard, A. (1970). *Learning to be Rotuman: Enculturation in the South Pacific.* New York: Teachers College Press.

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

Kawanaka, T. (1999). *Field test report: A study of mathematics and science teaching in seven cultures.* Unpublished manuscript, University of California, Los Angeles, Department of Psychology.

Lancy, D. F. (1996). *Playing on the mother-ground: Cultural routines for children's development.* New York: Guilford.

Malcolm, I. (1979). The West Australian Aboriginal child and classroom interaction: A sociolinguistic approach. *Journal of Pragmatics, 3,* 305–320.

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

Nerlove, S. B., & Snipper, A. S. (1981). Cognitive consequences of cultural opportunity. In R. H. Munroe, R. L. Munroe, & B. B. Whiting (Eds.), *Handbook of crosscultural human development* (pp. 423–474). New York: Garland.

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

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

Phillips, S. U. (1972). Participant structures and communicative competence: Warm Springs children in community and classroom. In C. B. Cazden, V. John, & D. Hymes (Eds.), *Functions of language in the classroom* (pp. 370–394). New York: Teachers College Press.

Pike, K. L. (1966). Etic and emic standpoints for the description of behavior. In A. G. Smith (Ed.), *Communication and culture: Readings in the codes of human interaction* (pp. 152–163). New York: Holt, Rinehart & Winston.

Price-Williams, D. R. (1961). A study concerning concepts of conservation of quantities among primitive children. *Acta Psychologica, 18,* 297–305.

Price-Williams, D. R. (1975). *Explorations in cross-cultural psychology.* San Francisco: Chandler & Sharp.

Price-Williams, D. R. (1980). Anthropological approaches to cognition and their relevance to psychology. In H. Triandis (Ed.), *Handbook of cross-cultural psychology* (Vol. 3, pp. 155–184.). Boston: Allyn & Bacon.

Reese, L. J. , Kroesen, K., & Gallimore, R. (1998). Cualitativos y cuantitativos, no cualitativos vs. cuantitativos [Qualitative and quantitative, not qualitative vs. quantitative]. In R. Mejía Arauz & S. A. Sandoval (Eds.), *Tras las vetas de la investigación cualitativa: Perspectivas y accercamientos desde la prática* (pp. 41–75). Tlaquepaque, Mexico: Instituto Tecnológico y de Estudios Superiores de Occidente.

Rogoff, B. (1981). Schooling and the development of cognitive skills. In H. Triandis & A. Heron (Eds.), *Handbook of cross-cultural psychology* (Vol. 4, pp. 233–294). Boston: Allyn & Bacon.

Rogoff, B. (1982). Integrating context and cognitive development. In M. E. Lamb & A. L. Brown (Eds.), *Advances in developmental psychology* (Vol. 2, pp. 125–170). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context.* New York: Oxford University Press.

Schmidt, W. H., McKnight, C. C., & Raizen, S. (1997). *A splintered vision: An investigation of U.S. science and mathematics education.* New York: Kluwer.

Scribner, S., & Cole, M. (1973). Cognitive consequences of formal and informal education. *Science, 182,* 553–559.

Scribner, S., & Cole, M. (1981). *The psychology of literacy.* Cambridge, MA: Harvard University Press.

Serpell, R. (1993). *The significance of schooling: Life-journeys in an African society.* Cambridge, England: Cambridge University Press.

Stevens, R. (1910). Stenographic reports of high school lessons. *Teachers College Record, 11,* 166.

Stevenson, H. W. & Nerison-Low, R. (1999). *To sum it up: Case studies of education in Germany, Japan, and the United States.* Washington, DC: National Center for Education Statistics.

Stigler, J. W., Gonzales, P., Kawanaka, T., Knoll, S., & Serrano, A. (1999). *The TIMSS videotape classroom study: Methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan, and the United States.* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom.* New York: Free Press.

Tharp, R. G. (1989). Psychocultural variables and constants: Effects on teaching and learning in schools. *American Psychologist, 44,* 349–359.

Tharp, R. G., & Gallimore, R. (1989). *Rousing minds to life: Teaching, learning, & schooling in social context.* Cambridge, England: Cambridge University Press.

Weisner, T. S. (1984). Ecocultural niches of middle childhood: A cross-cultural perspective. In W. A. Collins (Ed.), *Development during middle childhood: The years from six to twelve* (pp. 335–369). Washington, DC: National Academy of Sciences Press.

Weisner, T. S., & Gallimore, R. (1977). My brother's keeper: Child and sibling caretaking. *Current Anthropology, 18*(2), 169–190.

Weisner, T. S., Gallimore, R., & Jordan, C. (1988). Unpackaging cultural effects on classroom learning: Native Hawaiian peer assistance and child-generated activity. *Anthropology and Education Quarterly, 19*(4), 327–353.