Educational Evaluation and Policy Analysis
Summer 2005, Vol. 27, No. 2, pp. 111–132

Mathematics Teaching in the United States Today (and Tomorrow): Results From the TIMSS 1999 Video Study

James Hiebert
University of Delaware
James W. Stigler
University of California, Los Angeles, and LessonLab
Jennifer K. Jacobs
University of Colorado at Boulder
Karen Bogard Givvin
LessonLab
Helen Garnier
University of California, Los Angeles, and LessonLab
Margaret Smith
Iona College
Hilary Hollingsworth
Educational Consultant, Rye, Victoria, Australia
Alfred Manaster
University of California, San Diego
Diana Wearne
University of Delaware
Ronald Gallimore
University of California, Los Angeles, and LessonLab

The Third International Mathematics and Science Study (TIMSS) 1999 Video Study examined eighth-grade mathematics teaching in the United States and six higher-achieving countries. A range of teaching systems were found across higher-achieving countries that balanced attention to challenging content, procedural skill, and conceptual understanding in different ways. The United States displayed a unique system of teaching, not because of any particular feature but because of a constellation of features that reinforced attention to lower-level mathematics skills. The authors argue that these results are relevant for policy (mathematics) debates in the United States because they provide a current account of what actually is happening inside U.S. classrooms and because they demonstrate that current debates often pose overly simple choices. The authors suggest ways to learn from examining teaching systems that are not alien to U.S. teachers but that balance a skill emphasis with attention to challenging mathematics and conceptual development.

Keywords: *international comparisons, mathematics teaching*

PROPOSED changes in school mathematics content and pedagogy (National Council of Teachers of Mathematics, 1989, 2000) have met strong resistance (Askey, 2001; Cheney, 1997; Loveless, 2001; Wu, 1997). Despite subsequent attempts to reexamine issues and achieve a working consensus (National Research Council, 2001), the debate continues. Policymakers and educators must now sort through the claims and counterclaims of this decades-old controversy in order to make decisions about the future. A contested-reform narrative is certainly not unique to mathematics education (National Research Council, 1998). As the nation moves into the next generation of proposals and implementations, what kinds of information are needed to yield thoughtful and evidence-informed policy decisions?

Student achievement data have been a common information source for many educational policy decisions, and increasingly so in recent years. The logic behind the No Child Left Behind Act, for example, is that by testing students frequently and holding teachers and administrators accountable for achievement test results, student learning will improve. The law assumes that, by examining annual achievement data, educators can divine what causes unacceptable outcomes and can correct the unproductive parts of the system. But how can *processes* be improved by inspecting only their *outcomes*? This flawed approach is not unique to the No Child Left Behind legislation. Assuming that achievement results and accountability are sufficient to inform and drive changes in the processes that improve students' learning is part of the history of education policy decisions in the United States (Adams & Kirst, 1999; Ravitch, 2002).

As observers from a range of traditions have noted, to improve (educational) products one needs to identify and understand the (educational) processes that mediate inputs and outputs (Jenkins, 1997; Sarason, 1997; Skinner, 1969; Wilson & Daviss, 1994). Only by defining more precisely what and how processes support or undermine students' learning is it possible to know what changes will increase achievement outcomes. A core process that must be understood more fully, and the one we focus on in this article, is classroom teaching: what teachers do every day in classrooms to help their students learn. Ratcheting up accountability pressures on teachers will work only if we identify changes in teaching that will help students learn more.

In this article, we describe teaching by comparing everyday classroom teaching in the United States with teaching in higher-achieving countries. We draw from results of the Third International Mathematics and Science Study (TIMSS) 1999 Video Study to explore ways in which information about teaching processes can contribute to current debates and impending decisions about mathematics education in the United States.

Following on the heels of the TIMSS 1995 Video Study (Stigler, Gonzales, Kawanaka, Knoll, & Serrano, 1999; Stigler & Hiebert, 1999), the 1999 Video Study expanded the number of countries in the sample from three to seven and revised the analytic scheme to describe more precisely the nature of teaching in each country. Videotapes of a random nationally representative sample of eighth-grade mathematics lessons in each country were analyzed to provide national-level pictures of classroom teaching.

The advantages of using TIMSS video data to inform current policy debates include the fact that nationally representative samples of classroom lessons portray ordinary teaching practices, those experienced by the majority of students in each country. Teaching rarely is studied at the national level, but education policy often is discussed nationally.

A second advantage is the comparative feature of the TIMSS data, which reveals each country's practices more clearly. Because teaching is such a common activity, one embedded within a culture, it can be difficult to notice common features, especially those that are most widely shared. Contrasts with less familiar methods used in other countries make one's own methods more visible and open for inspection. Because the six other countries in the sample all show higher student achievement in mathematics than the United States, the TIMSS video comparison is of special interest. If one wishes to improve classroom teaching in the United States, it is useful to know what teaching looks like in higher-performing educational systems.

In this article, we first outline briefly a framework for thinking about classroom teaching and its effects on learning with an eye toward informing policy discussions. Then we provide a brief overview of the TIMSS 1999 Video Study and present results from the study to portray eighth-grade mathematics teaching in several countries, focusing on the United States. The results revealed a range of systems of teaching across higher-achieving countries that balance attention to challenging content, procedural skill, and conceptual understanding in different ways. U.S. teachers employed a unique system of teaching—not because of any particular feature but because of a constellation of features that reinforced attention to lower-level mathematics skills.

We argue that these results are relevant for policy (mathematics) debates in the United States because they provide a current account of what actually is happening inside U.S. classrooms and because they demonstrate that the debates often pose overly simple choices. Choosing one feature of teaching over another will not improve the U.S.

system of teaching and does not capture the way in which teaching systems in higher-achieving countries differ from that in the United States. We suggest ways to learn from examining teaching systems that are not alien to U.S. teachers but that balance a skill emphasis with attention to challenging mathematics and conceptual development. Finally, we return to the debates regarding the future direction of school mathematics in this country and offer some recommendations informed by the evidence presented. We try to balance a respect for the complex, system nature of teaching with the desire to develop some concrete and realistic recommendations for improving classroom teaching. Although our report focuses on mathematics, we believe the principles that emerge for studying classroom teaching and formulating recommendations for change are relevant for other school subjects as well.

Classroom Teaching as a System

What kind of information about classroom teaching is most useful for understanding the nature of teaching and how it can be improved? One research tradition argues for identifying and describing individual features of teaching that correlate with gains in students' achievement (Brophy & Good, 1986). The logic underlying this tradition assumes a correspondence between a particular feature of teaching and a learning outcome. Although this tradition, often referred to as process-product research, has provided a great deal of information about teaching, we believe that isolating individual features of teaching for study and improvement provides limited, and even misleading, information for policy decisions. For example, evidence showing positive relationships between a teaching feature and achievement gains suggests engaging in the feature of teaching more often or for longer periods of time. But more is not necessarily better—timing and quality also are likely to be important. And changing just one feature of teaching does not necessarily improve the overall effectiveness of teaching; it can even produce negative, unintended consequences (Guthrie, 1990; Stigler & Hiebert, 1999).

It is more productive, in our view, to treat teaching as a system of interacting features. The core of teaching—the interactions of teachers and students around content—takes its shape from the knowledge teachers and students bring to the lesson, the tasks presented, the discourse structures and participation expectations, the assessments, the physical materials available, and so on. It is the interaction among these elements, the *system,* rather than the individual elements acting alone, that defines the learning conditions for students (Cohen, Raudenbush, & Ball, 2003; Schoenfeld, 1998; Stigler & Hiebert, 1999). It is impossible, of course, to describe all of the relevant features that constitute teaching and their interactions in the classroom. But it is possible to identify a range of teaching features and to consider how individual features work together to reinforce particular kinds of learning conditions.

An example of how features of teaching work together to define learning opportunities for students can be found in many U.S. classrooms in which teachers arrange for students to work collaboratively in small groups during part of the lesson. Research has shown that whether small groups function productively to help students achieve the learning goal depends on many surrounding features, including the knowledge and skill students acquire for working collaboratively and the kinds of tasks they are assigned (Good, Mulryan, & McCaslin, 1992; Webb, Troper, & Fall, 1995). While the first author of this article was reviewing videotaped lessons as part of the TIMSS 1995 Video Study, he observed one eighth-grade teacher asking students to break into small groups and work together on the question "What is the name for a 12-sided object?" Students quizzed each other quickly about whether anyone knew the name of the object and then visited for the remaining time about nonmathematical topics. This was not a task that lent itself to collaborative investigation. The learning opportunities afforded by collaborative small groups are shaped by the system of which they are a part.

Systems of teaching, as we have defined them, focus on the technical core of teaching: what happens inside the classroom between the teacher and students during daily lessons. Factors that shape the origins and maintenance of these systems and their individual features extend far beyond the classroom door (Dreeben, 1994; Ingersoll, 2003; Oakes, 1985). Just as with other institutions, schools and classrooms and the practices they sustain reflect the wider society. In this article, we confine ourselves to describing current systems of teaching mathematics and leave to others the question of how these social, economic, and political factors have shaped these systems.

Why are descriptions of *systems* of teaching useful for educational policy? We address this question first by presenting data from the TIMSS 1999 Video Study that describe systems of teaching in high-achieving countries and then by considering how these data can inform recommendations for improving classroom teaching in the United States. Using information on systems of teaching constitutes a relatively new model for linking research to policy (Cohen et al., 2003). Rather than isolating features of teaching and recommending their more frequent use, the model we explore in this article examines systems of teaching that support students' achievement and considers how ineffective systems could be adjusted to facilitate desired learning goals more effectively.

The TIMSS 1999 Mathematics Video Study

Overview

The TIMSS 1999 Video Study followed a simple design: Nationally representative samples of eighth-grade mathematics classrooms were selected, and one lesson from each classroom was videotaped. Videotapes, together with supplementary materials such as a teacher questionnaire and copies of relevant textbook pages, were sent back to project headquarters, where they were analyzed by a research team composed of bilingual speakers from each participating country. Coding and analysis of video was managed digitally, via software especially designed for that purpose.

Method

Sample

The countries participating in the mathematics portion of the TIMSS 1999 Video Study were Australia, the Czech Republic, Hong Kong SAR,[1] Japan, the Netherlands, Switzerland, and the United States. As can be seen in Table 1, eighth graders in all of the other countries scored significantly higher than U.S. eighth graders on the TIMSS 1995 mathematics achievement test (Beaton et al., 1996; Gonzales et al., 2000), the test used to select countries for this study.

The international sampling plan followed the standards and procedures implemented for the TIMSS 1999 assessments, in which a two-stage stratified cluster design was used to produce national samples that would meet the analytical requirements necessary to allow estimates for classrooms and schools. The first stage involved selection of a probability proportionate to size (PPS) sample of schools via systematic sampling, with explicit (regions of the country) and implicit (other school characteristics) stratification, for each of the participating countries. In this PPS sample, the probability of selection assigned to each school was proportional to the number of students in the eighth grade in schools countrywide. The next stage involved random selection of one classroom within each school.[2]

TABLE 1
Average Scores on the TIMSS 1995 and TIMSS 1999 Eighth-Grade Mathematics Assessments

Country	Average score	
	1995[a]	1999[b]
Australia[c] (AU)	519	525
Czech Republic (CZ)	546	520
Hong Kong SAR (HK)	569	582
Japan (JP)	581	579
Netherlands[c] (NL)	529	540
Switzerland (SW)	534	—[d]
United States (US)	492	502

Note. Rescaled TIMSS 1995 mathematics scores are reported here. Switzerland did not participate in the TIMSS 1999 assessment.
[a]AU, CZ, HK, JP, NL, SW > US; CZ, HK, JP, SW > AU; HK > NL, SW; JP > CZ, NL, SW.
[b]AU, HK, JP, NL > US; HK, JP > AU, CZ, NL.
[c]Did not meet international sampling or other guidelines in 1995. See Beaton et al. (1996) for details.
[d]Not available.

A single 1-day lesson was videotaped in each eighth-grade mathematics classroom, without regard to the particular mathematics topic being taught or type of activity taking place. The only exception was that teachers were not videotaped on days a test was scheduled for the entire class period. Teachers were asked to do nothing special for the videotaped session and to conduct the class as they had planned. Taping in each country occurred throughout the school year so as to include lessons that contained the different kinds of topics and classroom conditions that occur over the course of a school year.

The final sample included 638 eighth-grade mathematics lessons: 87 from Australia, 100 from the Czech Republic, 100 from Hong Kong SAR, 50 from Japan, 78 from the Netherlands, 140 from Switzerland, and 83 from the United States. The Japanese lessons were the same ones collected in 1995 as part of the earlier study, but they were reanalyzed for the current study.[3] Sampling information, videotaping procedures, and other methodological notes are detailed in Appendix A of the report *Teaching Mathematics in Seven Countries: Results from the TIMSS 1999 Video Study* (Hiebert, Gallimore, et al., 2003). A more detailed discussion of the technical aspects of the study can be found in the companion technical report (Jacobs et al., 2003).

Code Development

The major challenge for the international team of researchers, representing all of the countries in the study, was to develop a reliable way of analyzing the lessons that would capture the sets of features of teaching in each country that, together, defined the mathematics learning conditions for students. The code development process began by building on the achievements of the TIMSS 1995 Video Study (Stigler et al., 1999). Although expanding the sample to seven countries and aiming to probe more deeply into the teaching systems in each country made it impossible to retain many of the exact codes from the 1995 study, it was useful to begin with the same major categories of codes: (a) structure and organization of daily lessons, (b) nature of the mathematics presented, and (c) way in which the mathematics was worked on during the lesson. Using this general framework, the research team solicited descriptions of typical lessons from experts in each country and reviewed the literature on mathematics teaching and learning. Suggestions for components or features of teaching that influence opportunities for learning in each country were transformed into codes that could be applied reliably across all countries.

The category of lesson organization was elaborated to include components such as the amount of time spent studying mathematics; the ways in which lessons were divided into segments that reviewed old material, introduced new material, and practiced new material; the grouping structures used, that is, whole-class public discussion, individual private work, and small-group work; and the ways in which lesson flow and coherence were enhanced and undermined. The nature of the mathematics presented was described by coding the mathematical topics discussed, their relative level of complexity, and the ways in which the content was connected across the lesson. The way in which mathematics was worked on during the lesson focused on the kinds of mathematical problems presented and how they were solved and discussed. In the end, the coders analyzed the lessons using more than 75 different codes organized around these three broad dimensions of teaching.

Establishing intercoder reliability for each code was a nontrivial challenge, especially when viewing lessons from seven countries with several coders from each country. Definitions for most codes included a series of general descriptors along with multiple examples and lists of specific exceptions or unusual cases. Code definitions were developed during months of identifying features to be coded, constructing initial definitions, applying codes and checking for reliability, and refining and elaborating definitions. All codes met an 85% criterion for intercoder agreement (for the full set of codes and definitions, see Jacobs et al., 2003).

In addition to the primary coding scheme, a supplementary coding scheme was developed by a group of four U.S. mathematicians and teachers of postsecondary mathematics.[4] These experts were enlisted to review the lessons for nuanced content issues not captured by the primary coding scheme, such as the kinds of mathematical reasoning engaged in by the teacher or students. The aim of this group's coding scheme was to guide a series of subjective judgments about the nature and quality of the mathematics content presented in the lessons. The group worked from detailed written records of the lessons rather than watching the videos. This procedure provided a second reason for soliciting this supplementary coding—it allowed country-identifying markers to be removed so that the group's judgments of lesson content were country blind. Judgments were the consensus opinions of the group; no intercoder reliability was assessed.

Analysis

Assessments of the quantifiable codes compared countries in regard to the presence (and duration) of teaching features using analyses of variance and two-tailed t tests at the .05 level. Bonferroni adjustments were made when more than two countries were compared simultaneously. The standard errors used in the statistical tests are reported in Appendix C of the TIMSS 1999 Video Study report (Hiebert, Gallimore, et al., 2003). All analyses involved data with survey weights that were calculated specifically for each country and each lesson and provided unbiased estimates of national means and distributions. The design effects of the overall probability of classroom selection were determined, and appropriate adjustments for nonresponse on sampling variances of the estimates were made, by using the jackknife procedure with a set of jackknife replicate weights (see Jacobs et al., 2003, for more details on weighting procedures and analyses).

The group of mathematicians and post-secondary mathematics teachers examined a random subsample of 20 lessons per country, excluding Japan.[5] Because the same group had analyzed the Japanese lessons as part of the TIMSS 1995 Video Study and because they wished to retain the country-blindness of their analysis, the Japanese tapes were not reanalyzed. As a result of the small sample size, the judgments of the group were not analyzed for statistical significance.

Results and Interpretations: A System of U.S. Mathematics Teaching as Compared With Systems in Other Countries

We begin with our central concluding observation: The results presented in this section describe a system of U.S. mathematics teaching in eighth grade characterized by frequent reviews of relatively unchallenging, procedurally oriented mathematics during lessons that are unnecessarily fragmented. This unflattering picture of U.S. mathematics classrooms is not new (National Advisory Committee on Mathematics Education, 1975; Rowan, Harrison, & Hayes, 2004; Stodolsky, 1988; Weiss, Pasley, Smith, Banilower, & Heck, 2003). What is new is the documentation of this teaching system using a nationally representative sample of videotaped lessons that provides information on how individual features work together to reinforce the system's characteristics.

The results also show that eighth-grade mathematics teaching in the United States is not different from teaching in other countries because of any one feature. Of the 117 analyses conducted on more than 75 features of teaching, the United States differed from *all* of the other countries on only 1 of the analyses (Hiebert, Gallimore, et al., 2003). On almost all features, taken individually, the U.S. mathematics classrooms were similar to those of at least one higher-achieving country.

A first indication, however, that the constellation of U.S. teaching features defines a system quite different from those in countries with high achievement is found in the number of analyses on which the United States differed from each individual country. Listing the countries in the order of achievement shown in Table 1 from high to low (based on the TIMSS 1995 assessment), along with the percentage of analyses on which the United States differed from each country, yields the following: Japan (42%), Hong Kong SAR (30%), Czech Republic (25%), Switzerland (16%), the Netherlands (27%), and Australia (8%). On a feature-by-feature basis, the United States differed most from the relatively highest-achieving countries.

Examining further the features of mathematics teaching that differentiate the United States from other countries reveals the distinctive constellation of features that defined the U.S. system of teaching. The results in the sections to follow are organized according to a set of more general characteristics that describe key aspects of the system. The United States is the only country that displayed all of these characteristics.

Characteristic 1: Low Level of Mathematical Challenge

Level of mathematical challenge surely influences students' learning opportunities, yet it is difficult to measure directly from a random sample of single lessons. Classifying lessons according to mathematical topic studied is an initially appealing strategy but quickly becomes problematic when teachers are seen to treat the same topic in very different ways. In addition, mathematical challenge for a particular classroom of students depends on the entry capabilities of the students and on experiences they had in previous lessons. Nevertheless, several coded features in the video study provided an indirect measure of challenge.

Prevalence of Routine Exercises

Figure 1 shows percentages of mathematics problems that were applications rather than exercises. Exercises were defined as straightforward problems, usually presented with little context, for which a solution procedure apparently had been demonstrated. The students' task was to execute the procedure to complete the exercise. Applications were defined as problems that appeared to require some adjustment to a known procedure, however slight, or some analysis of how to use the procedure. Applications often were presented via verbal descriptions, graphs, or diagrams. They might, or might not, be real-life situations. An average of 34% of problems per U.S. lesson were applications, a smaller percentage than in Japan and the Netherlands.

Practicing Familiar Procedures

Students spent some of the time during almost all lessons in all countries working individually or in small groups on assigned problems. Figure 2 shows average percentages of this private work time spent practicing familiar procedures versus average percentages of time spent doing something more than this, such as developing new procedures for new kinds of problems, analyzing problems to decide what procedures should be applied, or creating new problems. Many of these alternative activities are likely to pose greater mathematical challenges for students. U.S. students spent a smaller percentage of their time than students in Australia, Japan, and Switzerland doing something other than practicing familiar procedures.

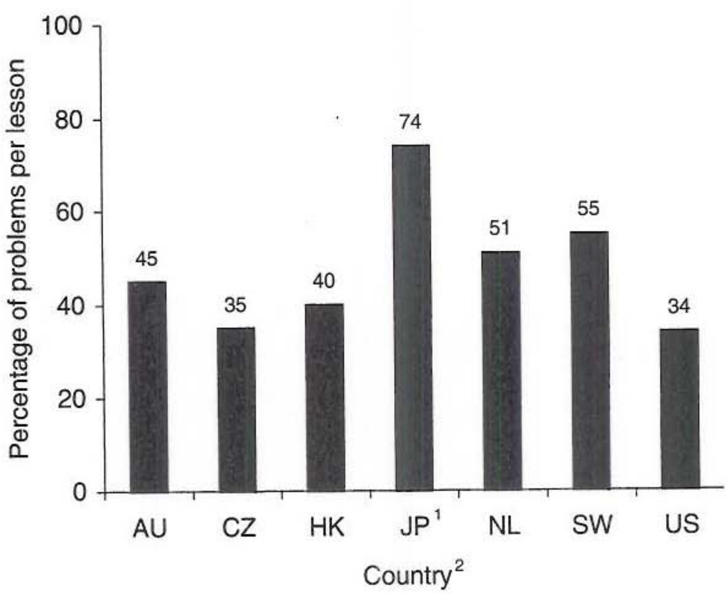

FIGURE 1. *Average percentages of problems per lesson that were applications.*
Note. All reported country differences are significant at $p < .05$. Analyses do not include answered-only problems (i.e., problems that were completed before the videotaped lesson and for which only answers were shared). JP >AU, CZ, HK, NL, US; NL > US; SW > CZ.
[1]Japanese mathematics data were collected in 1995.
[2]AU = Australia; CZ = Czech Republic; HK = Hong Kong SAR; JP = Japan; NL = Netherlands; SW = Switzerland; US = United States.

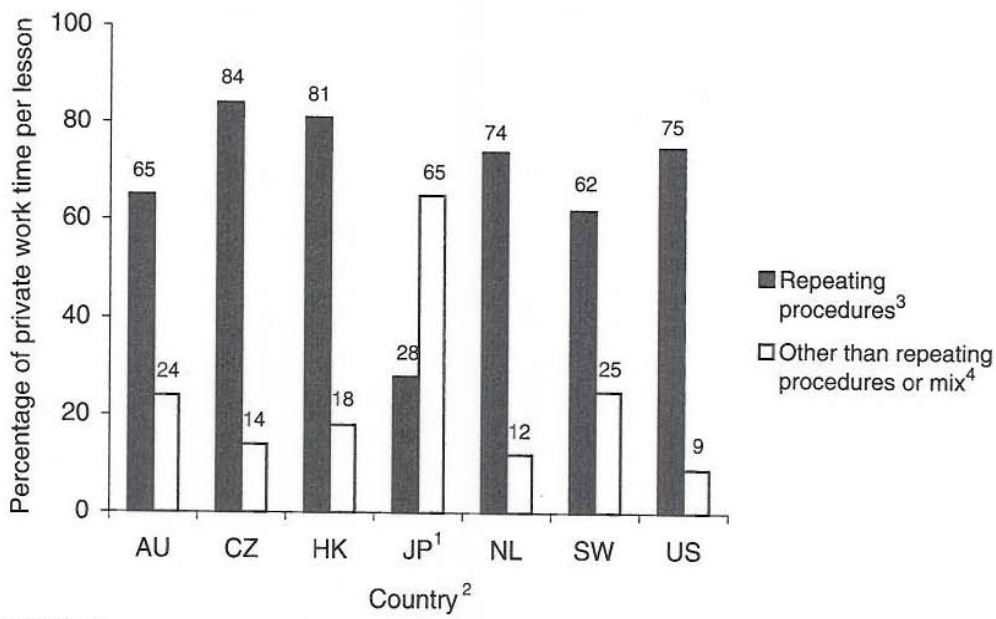

FIGURE 2. *Average percentages of private work time per lesson devoted to repeating procedures and something other than repeating procedures or mix.*

Note. All reported country differences are significant at $p < .05$. Percentages may not sum to 100 because some private work segments were marked as "not able to make judgment."

[1] Japanese mathematics data were collected in 1995.

[2] AU = Australia; CZ = Czech Republic; HK = Hong Kong SAR; JP = Japan; NL = Netherlands; SW = Switzerland; US = United States.

[3] AU, NL, SW, US > JP; CZ, HK > JP, SW.

[4] JP > AU, CZ, HK, NL, SW, US; AU, SW > US.

Relatively Elementary Content

The group of mathematicians and postsecondary mathematics teachers who reviewed a subsample of country-blind lessons rated each of the lessons, on a scale from 1 to 5, on the basis of content level. A rating of 3 was assigned to lessons in which the majority of content typically would be encountered by students just before the standard topics of a beginning algebra course common in U.S. eighth-grade classrooms. Lower ratings were assigned to more elementary lessons, and higher ratings were assigned to more advanced lessons. Averaging the ratings across the lessons within each country yielded the following: Czech Republic and Hong Kong SAR, 3.7; Switzerland, 3.0; Netherlands, 2.9; United States, 2.7; and Australia, 2.5. Recall that the lessons from Japan were not reanalyzed.

Absence of Mathematical Reasoning

The mathematics group also searched for evidence of the following special forms of mathematical reasoning, as demonstrated by the teacher or the students: deductive reasoning, developing a mathematical justification for a general rule or principle, generalizing from individual cases, and using counterexamples to show that a conjecture cannot be true. In each country, 25% or less of the lessons in the mathematics group's sample contained instances of one or more of these forms of reasoning. The United States did not appear especially different from some of the other countries in the low frequency or absence of deductive reasoning and use of counterexamples. But the United States was the only country in which no lessons contained instances of developing a mathematical justification or generalizing from individual cases.

The full sample of lessons from all countries was analyzed by the primary coding team for the presence of mathematical proofs, often considered a hallmark of more advanced mathematical reasoning as students move into secondary school. Japan was the only country in the sample that showed regular attention to proofs (26% of mathematics problems per lesson involved proofs), but some evidence of use of mathematical proofs also was found in the lessons of the Czech Republic, Hong Kong SAR, and Switzerland. All of the countries in the sample other than the Netherlands and the United States had at least one lesson with a proof. The percentages for these countries ranged from 1% of the Australian lessons to 39% of the Japanese lessons (Hiebert, Gallimore, et al., 2003).

Summary

The individual findings relating to mathematical challenge accumulate to portray U.S. lessons as presenting less of a challenge than lessons in other countries. The United States is statistically different from Australia, Japan, the Netherlands, and Switzerland on at least one feature and at opposite ends of the continuum from the Czech Republic and Hong Kong SAR on judgments by the mathematics group regarding level of content. No single feature is responsible for the effect of low mathematical challenge; it emerges through the reinforcing influence of multiple features with no countervailing feature.

Characteristic 2: Emphasis on Procedures

The debate between procedural and conceptual emphases has a long history in mathematics education (Brownell, 1935; Hiebert, 1986). Although a compelling current view is that both procedures and concepts are critical, with no trade-offs needed (National Research Council, 2001), it still is possible to ask whether classroom teachers emphasize them in different ways.

One way to determine the procedural versus conceptual emphasis of a lesson is to ask what kinds of mathematics problems are presented (Smith, 2000; Stein, Grover, & Henningsen, 1996). In the case of all of the problems in this study that were completed with some public discussion during the lesson, the statement of the problem was classified into one of three types according to its primary intent: using procedures, stating concepts, and making connections. The statement of a problem indicates the kinds of mathematical processes that apparently are intended by the problem. Using procedures and making connections are of most interest here because using procedures suggests a procedural emphasis, whereas making connections (among ideas, facts, or procedures) suggests a conceptual emphasis. A using procedures problem statement could be "Solve for x in the equation $2x + 5 = 6 - x$." A making connections problem statement could be "Graph the equations $y = 2x + 3$, $2y = x - 2$, and $y = -4x$ and examine the role played by the numbers in determining the position and the slope of the associated lines." Stating concepts problems often ask students to recall or illustrate definitions or properties. Although these problems can deal with concepts, they do so by asking students to recall or repeat them rather than construct or analyze them. A stating concepts problem could be "Show the point (3, 2) on the coordinate plane."

Figure 3 presents the average percentage of problem statements per lesson of each type by country. On this dimension, Hong Kong SAR and Japan reside at the opposite ends of the spectrum, with Japan emphasizing conceptual problems (making connections) and Hong Kong SAR emphasizing procedural problems. The United States is situated in the middle of the profiles portrayed by each country.

On the basis of these data, the United States does not appear to emphasize procedures, at least relative to other countries. But our examination of problems is not finished. The mathematics problems were coded a second time to check how they were worked on with the students. Problems can change in their nature as they are worked out during the lesson. Teachers can transform problems so that the focus shifts from one kind of mathematical process to another. For example, a problem with the apparent intent of making connections among ideas, facts, and procedures can be transformed into a problem that involves demonstrating and practicing a procedure, perhaps because students are struggling with the original problem and the teacher perceives they need additional help.

Figure 4 on page 121 shows the results of reclassifying the problems according to how they were worked on and discussed with the students. A fourth category of problem implementation—providing results only—was needed to code problems for which no mathematical work was discussed and only the answer was given.

Note that the relative emphases within countries appear different now. In the United States, 69% of the mathematics problems were *presented* as using procedures (Figure 3), but 91%

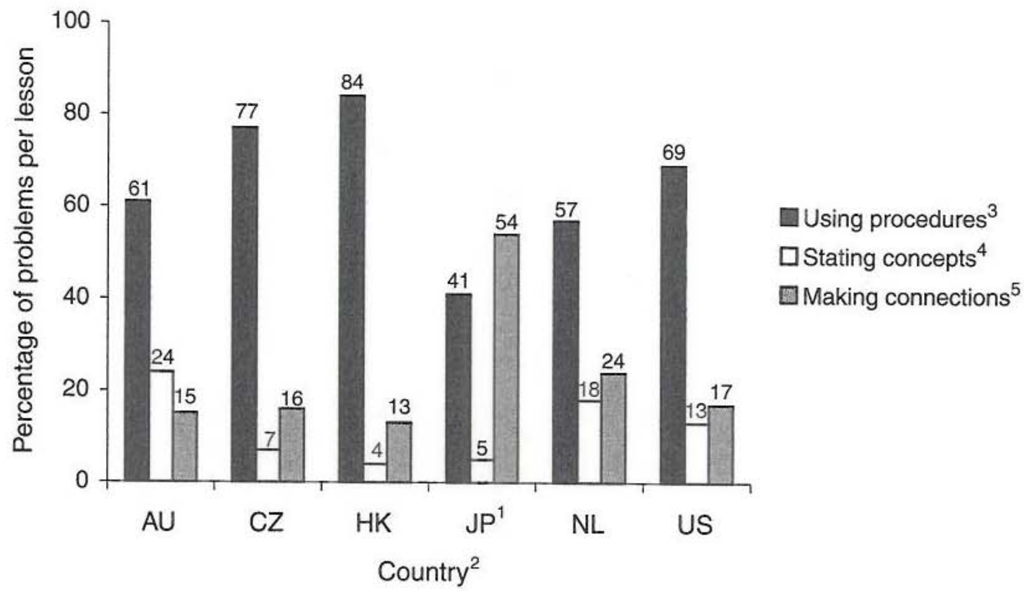

FIGURE 3. *Average percentage of problems per lesson presented as each type.*
Note. All reported country differences are significant at $p < .05$. Analyses include only problems with a publicly presented solution. They do not include answered-only problems (i.e., problems that were completed before the videotaped lesson and for which only answers were shared). English transcriptions of Swiss lessons were not available for these analyses. Percentages may not sum to 100 owing to rounding.
[1] Japanese mathematics data were collected in 1995.
[2] AU = Australia; CZ = Czech Republic; HK = Hong Kong SAR; JP = Japan; NL = Netherlands; US = United States.
[3] CZ > JP, NL; HK > AU, JP, NL, US; US > JP.
[4] AU > CZ, HK, JP; NL, US > HK, JP.
[5] JP > AU, CZ, HK, US.

were *worked on* by using procedures or by providing results only (Figure 4). By contrast, in Hong Kong SAR 84% of the problems were presented as using procedures (Figure 3), but 63% were worked on by using procedures or providing results only. Clearly, teachers in all countries transformed some of the problems so that students' actual experiences were somewhat different than what might have been predicted from looking only at the statements of the problems.

The picture of how problems were worked on comes into clearer focus by following the implementation of a particular kind of problem. Making connections problems are of special interest. Recall that Hong Kong SAR and Japan were at opposite ends of the distribution with regard to problems stated as using procedures and making connections. Recall also that 17% of the problem statements in the United States suggested a focus on mathematical connections, a percentage within the range of many higher-achieving countries (Figure 3). What happened to these problems when they were worked on in class?

Figure 5 shows how the making connections problems were worked on with students. Hong Kong SAR and Japan now appear quite similar, along with the Czech Republic and the Netherlands. At least 37% of the making connection problems in these countries retained their original intent. In contrast, virtually none of the making connections problems in the United States were discussed in a way that made the mathematical connections or relationships visible for students. Mostly, they turned into opportunities to practice procedures, or they were implemented as problems in which even less mathematical content was visible—only the answer was given.

A plausible conclusion from these results is that teachers in the higher-achieving countries attended more to the conceptual development of mathematics than teachers in the United States. Even teachers in Hong Kong SAR, who appeared

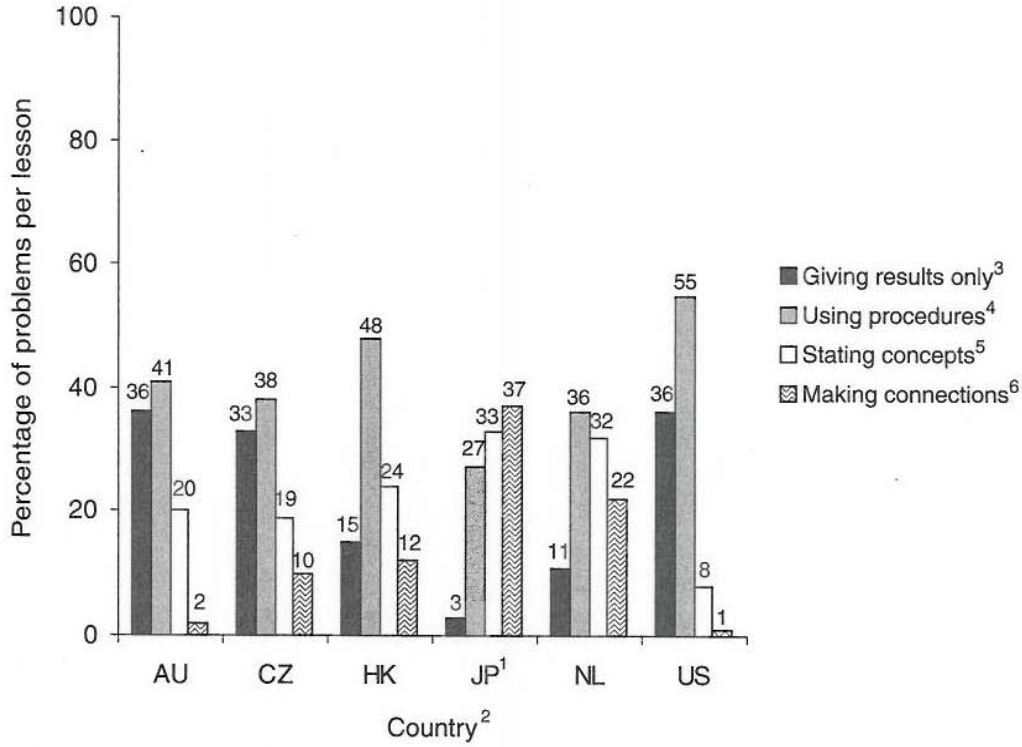

FIGURE 4. *Average percentages of problems per lesson solved by explicitly using processes of each type.*
Note. All reported country differences are significant at $p < .05$. Analyses include only problems with a publicly presented solution. They do not include answered-only problems (i.e., problems that were completed before the videotaped lesson and for which only answers were shared). English transcriptions of Swiss lessons were not available for these analyses. Percentages may not sum to 100 owing to rounding.

[1]Japanese mathematics data were collected in 1995.

[2]AU = Australia; CZ = Czech Republic; HK = Hong Kong SAR; JP = Japan; NL = Netherlands; US = United States.

[3]AU, CZ, US > HK, JP, NL; HK, NL > JP.

[4]HK > JP; US > CZ, JP, NL.

[5]AU, CZ, HK, JP > US; NL > CZ, US.

[6]CZ, HK, NL > AU, US; JP > AU, CZ, HK, US.

to focus on procedures when presenting problems (Figure 3), were found to examine conceptual underpinnings in an explicit way (Figures 4 and 5). The significance of this finding comes, in part, from the fact that the kinds of mathematical processes that are highlighted and made visible for students while working on the problems, rather than those implied by the initial statements of the problems, affect the nature and level of students' learning (Stein & Lane, 1996).

Additional information about the conceptual development of the content was provided by the mathematics group's qualitative analysis of the subsample of lessons. One of the codes the group created for this study was the degree to which mathematical concepts *or* procedures were developed during the lesson. Development required that mathematical reasons or justifications be given for the mathematical results presented and used. A rating of 1 indicated that a lesson contained little mathematical justification, by the teacher or students, for why things work like they do. A rating of 5 was assigned to a lesson in which the concepts and procedures were mathematically motivated, supported, and justified by the teacher or students. Figure 6 shows the percentage of lessons placed into each category. Note that 40% of the U.S. lessons received a rating of 1 (i.e., undeveloped); no other country received a rating of 1 on more than 15% of its lessons.

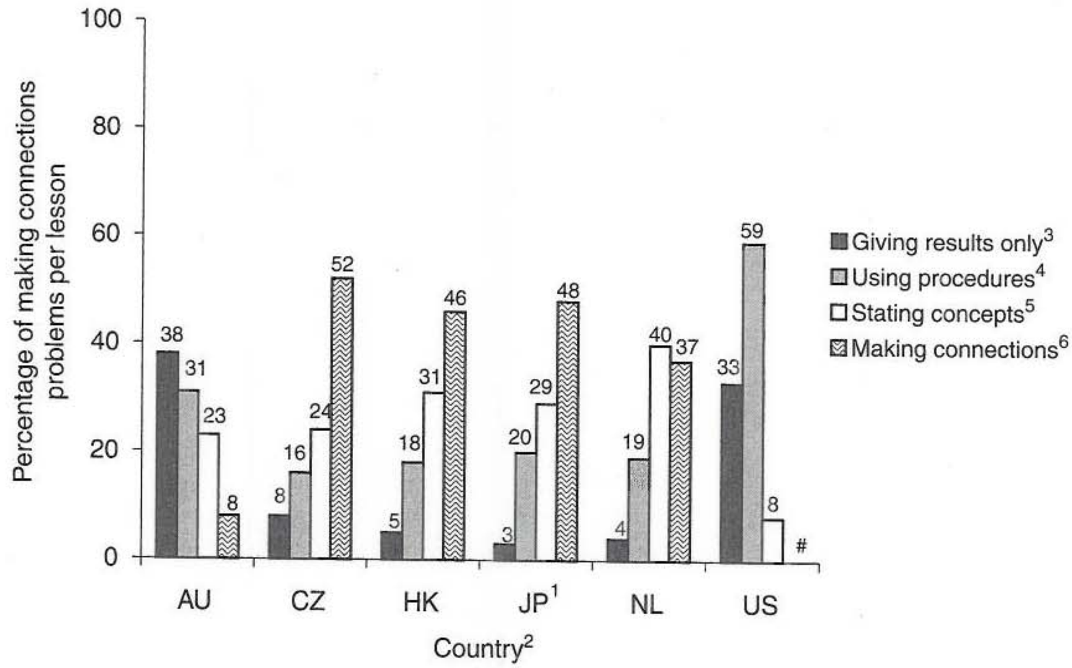

FIGURE 5. *Average percentages of making connections problems per lesson solved by explicitly using processes of each type.*

Note. All reported country differences are significant at $p < .05$. Analyses only include problems with a publicly presented solution. Analyses do not include answered-only problems (i.e., problems that were completed prior to the videotaped lesson and only their answers were shared). English transcriptions of Swiss lessons were not available for these analyses. Lessons with no making connections problem statements were excluded from these analyses. Percentages may not sum to 100 because of rounding.

#Rounds to zero.

1 Japanese mathematics data were collected in 1995.

2 AU = Australia; CZ = Czech Republic; HK = Hong Kong SAR; JP = Japan; NL = Netherlands; US = United States.

3 AU, US > CZ, HK, JP, NL.

4 US > CZ, HK, JP, NL.

5 JP, NL > US.

6 CZ, HK, JP, NL > AU, US.

Averaging the ratings for each country yielded the following, in order of mathematical development: Hong Kong SAR (3.9), Switzerland (3.4), the Czech Republic (3.3), Australia (3.0), the Netherlands (2.8), and the United States (2.4). The fact that Hong Kong SAR ranked first in this analysis, together with the findings presented in Figures 4 and 5 regarding problem implementation, suggests a balance within the Hong Kong SAR system between attention to procedural skill and development of conceptual underpinnings. Although the U.S. lessons momentarily appeared to show a balance among procedural and conceptual emphases, on the basis of the types of problems presented, follow-up indicators pointed to a uniquely heavy emphasis on procedures.

Characteristic 3: Emphasis on Review

Researchers analyzing classroom practices in the United States, especially in mathematics, have frequently noted the large amount of review and the minimal development of new material. The emphasis on practicing old material rather than developing new material is a feature of both the commonly used curricula (Flanders, 1987; Schmidt, McKnight, Valverde, Houang, & Wiley, 1997) and classroom pedagogy (Good, Grouws, & Ebmeier, 1983). Educators have argued that students would learn concepts and procedures more

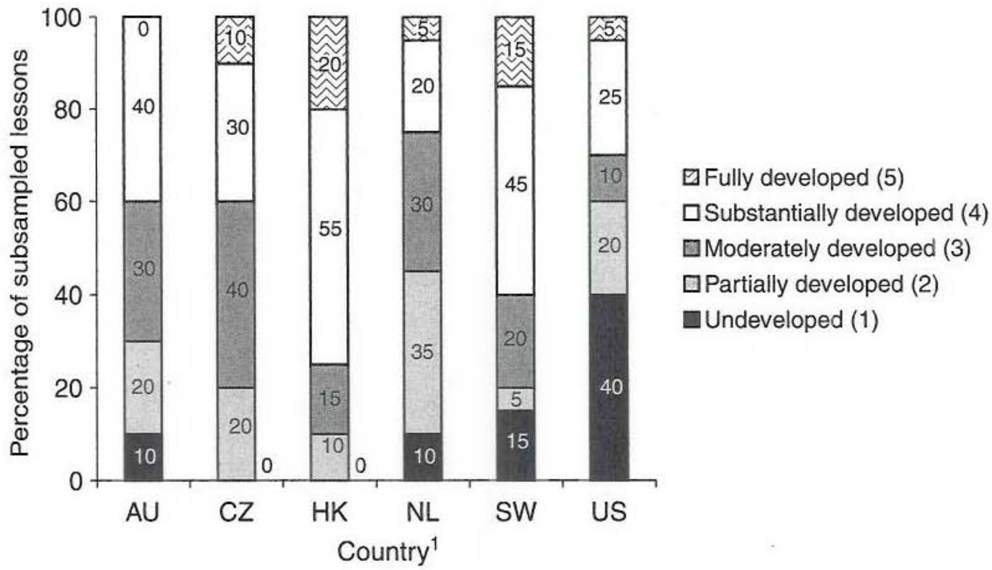

FIGURE 6. *Percentages of lessons in subsample rated at each level of mathematical development.*
Note. Lessons included here represent a random subsample of lessons in each country. The number in parentheses is the ranking number for that category.
[1]AU = Australia; CZ = Czech Republic; HK = Hong Kong SAR; NL = Netherlands; SW = Switzerland; US = United States.

deeply, thereby requiring less review, if time was reallocated so that more time was spent developing new material (National Research Council, 1989). Data supporting this claim have been reported (e.g., Good et al., 1983; Hiebert & Wearne, 1993).

Despite the arguments in favor of allocating more time to the development of new material, a relatively strong emphasis on reviewing old material was found in U.S. lessons. As shown in Figure 7, 53% of lesson time, on average, was devoted to review. Except for the Czech Republic, all of the higher-achieving countries spent more time working on new material than reviewing old material. The time devoted to new content included introducing the new content and practicing the new content (for instance, solving problems using procedures newly introduced in the lesson). The high percentage of review time in the United States comes both from the fact that 28% of the U.S. lessons were entirely review (a larger percentage than in Hong Kong SAR and Japan) and the fact that 94% of U.S. lessons contained at least one review segment (a larger percentage than in the Netherlands and Switzerland) (Hiebert, Gallimore, et al., 2003).

Of interest is the fact that the Czech Republic shared with the United States an emphasis on review. However, as we argue later, review in the Czech Republic takes on a different character when it is placed together with the earlier findings of a higher level of mathematical challenge and more frequent development of concepts.

Characteristic 4: Fragmented Lessons, Mathematically and Pedagogically

It is reasonable to assume that the coherence of lessons influences the learning opportunities for students, with coherent lessons enabling students to abstract more easily the key points of the lesson (Fernandez, Yoshida, & Stigler, 1992). Coherence, like mathematical challenge, is difficult to measure. However, some indirect indicators developed for this study addressed the mathematical coherence and the pedagogical coherence of the lessons.

Mathematical Coherence

Lessons that focus on a single mathematical topic might appear more coherent to students than lessons that address several topics. Figure 8 shows the percentages of lessons in which all of the mathematics problems focused on a single topic. The topics defined for this study were three subcategories each for number (whole numbers/

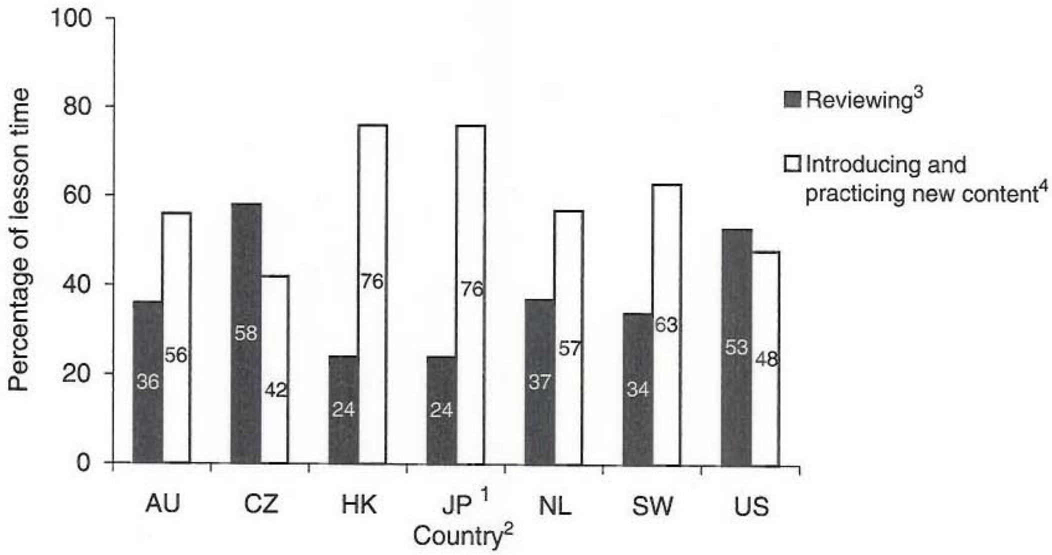

FIGURE 7. *Average percentages of time per lesson devoted to each purpose.*
Note. All reported country differences are significant at $p < .05$. Percentages may not sum to 100 owing to rounding and the possibility of portions of lessons being coded as "not able to make a judgment about the purpose."
[1] Japanese mathematics data were collected in 1995.
[2] AU = Australia; CZ = Czech Republic; HK = Hong Kong SAR; JP = Japan; NL = Netherlands; SW = Switzerland; US = United States.
[3] CZ > AU, HK, JP, NL, SW; US > HK, JP.
[4] HK > AU, CZ, US; JP > CZ, US; SW > CZ.

fractions/decimals, ratio/proportion/percent, integers), geometry (measurement, two-dimensional, three-dimensional), and algebra (linear expressions, linear equations and inequalities, higher-order functions) along with statistics and trigonometry. In the United States, 34% of lessons focused on a single topic, a smaller percentage than in Hong Kong SAR, Japan, and Switzerland.

The group of mathematicians and post-secondary mathematics teachers examined the coherence of lessons, defined by the group as the implicit and explicit interrelation of all mathematical components of the lesson, and assigned each (country-blind) lesson in their subsample to one of five levels of coherence ranging from thematic (5) to fragmented (1). The average ratings of the countries were as follows: Hong Kong SAR, 4.9; Switzerland, 4.3; Australia, 4.2; the Netherlands, 4.0; the Czech Republic, 3.6; and the United States, 3.5.

Pedagogical Coherence

Lesson coherence is likely to be affected not only by how well the mathematics content is connected across the lesson but by actions of the teacher that can either support or undermine coherence. For example, a teacher might provide an explicit goal statement for the lesson, alerting students to the important ideas that will be developed, or a teacher might interrupt students by making an off-task announcement, thereby breaking the flow of the lesson. The results showed that U.S. teachers were similar to many of their international colleagues in the frequency with which they supported coherence by providing goal or summary statements for the lessons (Hiebert, Gallimore, et al., 2003). However, U.S. lessons were more likely than lessons in some of the other countries to be interrupted in various ways.

Interruptions to lessons can come from the outside (e.g., an announcement over the speaker system) or can be introduced by the teacher (e.g., asking students to raise their hand if they are going on the field trip). The percentage of U.S. lessons interrupted from the outside (29%) was substantial but not significantly different from that of other countries. However, as shown in Figure 9 on page 126, U.S. teachers were more likely than teachers in the Czech Republic and Japan to shift the focus away from mathematics during

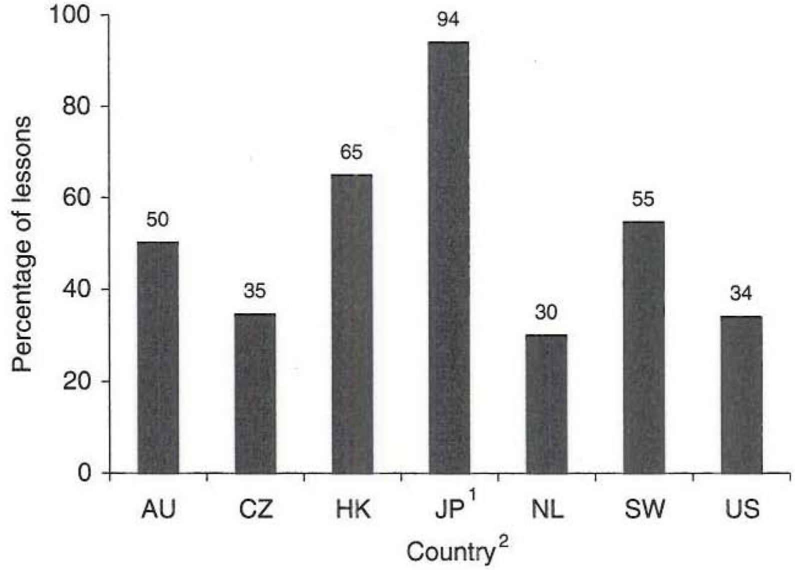

FIGURE 8. *Percentages of lessons that contained problems related to a single topic.*
Note. HK > CZ, NL, US; JP > AU, CZ, HK, NL, SW, US; SW > US. All reported country differences are significant at $p < .05$.
[1]Japanese mathematics data were collected in 1995.
[2]AU = Australia; CZ = Czech Republic; HK = Hong Kong SAR; JP = Japan; NL = Netherlands; SW = Switzerland; US = United States.

the lesson, and they were more likely than teachers in the Czech Republic to make an unrelated public announcement while students were working privately on assigned problems. The strikingly large percentage of lessons in the Netherlands in which teachers made unrelated public announcements can be explained, in part, by the fact that the Netherlands was the only country in which a majority of lesson time was devoted to private work, so there simply was more time for Dutch teachers to make these announcements.

Contrasting Systems of Teaching: What Can Be Learned?

We began the results section with the claim that U.S. eighth-grade mathematics teaching is characterized by frequent review of relatively unchallenging, procedurally oriented mathematics during lessons that are unnecessarily fragmented. We also proposed earlier in the article that studying alternative systems of teaching can inform policy discussions by suggesting changes to relatively ineffective systems. What can be learned by comparing the U.S. system with those of other countries?

Because a number of countries in the sample are consistently high achievers, a first lesson from the results is that different systems of teaching can support similar learning goals. No one system is necessary for high achievement on international tests. From a policy point of view, this opens the possibility of studying the variation across systems in order to understand how different combinations or clusters of teaching features might work together within different systems to promote high achievement. By comparing the U.S. system with a variety of more effective systems, educators can envision adjustments that would align the system more directly with desired learning goals and that realistically could be enacted by practicing teachers.

Japan provides the most striking contrast with the United States, just as it did in the 1995 TIMSS Video Study (Stigler et al., 1999). The Japanese system of teaching, as seen in the sample of lessons, is different from the U.S. system on each of the four characteristics, often in dramatic ways. This makes the Japanese system so different from that of the United States that it is difficult to imagine how the U.S. system could be adapted, if teachers wished to do so, to align it closely with

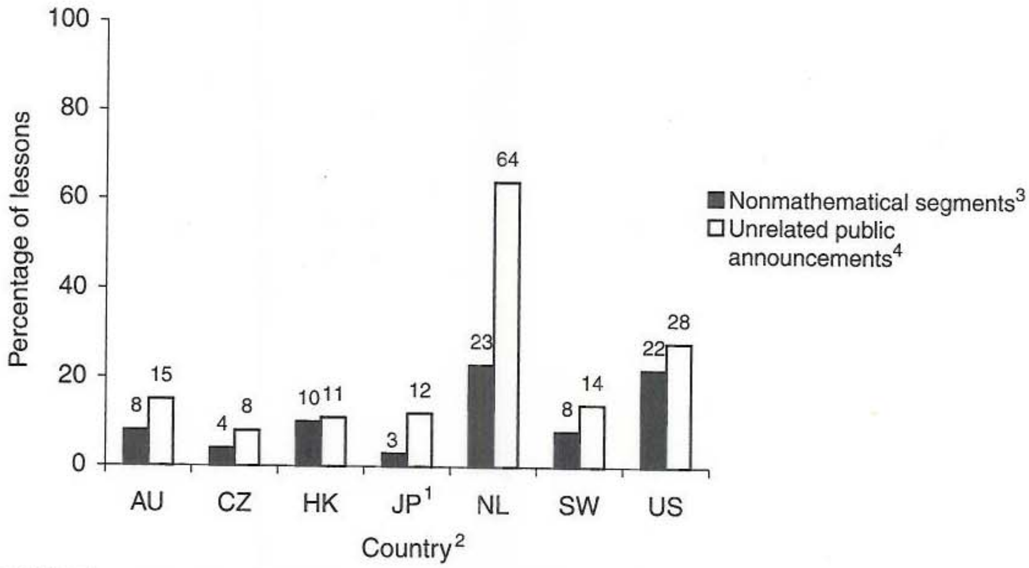

FIGURE 9. *Percentages of lessons with at least one nonmathematical segment at least 30 seconds in length within the mathematics portion of the lesson and percentages of lessons with at least one public announcement by the teacher during private work time unrelated to the lesson.*

Note. All reported country differences are significant at $p < .05$.

[1] Japanese mathematics data were collected in 1995.

[2] AU = Australia; CZ = Czech Republic; HK = Hong Kong SAR; JP = Japan; NL = Netherlands; SW = Switzerland; US = United States.

[3] NL, US > CZ, JP.

[4] NL > AU, CZ, HK, JP, SW, US; US > CZ.

the Japanese system. A nearly complete replacement of the system would be necessary. Although very attractive to some mathematics educators, there are questions about whether this kind of transformation on a large scale is realistic (see Fullan, 2001, and Elmore, 1996, for descriptions of barriers to this kind of change). The Japanese system probably is best interpreted as a completely different system, as a distinct and instructive alternative to the U.S. system.

In many ways, systems of teaching in countries other than Japan provide equally instructive alternatives because they share some characteristics with the United States and therefore are somewhat familiar to U.S. teachers. The Czech Republic and Hong Kong SAR provide two interesting contrasts. As noted earlier, the system in the Czech Republic shared with the United States an emphasis on review. In the Czech Republic, however, review was part of a system that included more challenging content, greater attention to the conceptual aspects of mathematics, and more coherence within lessons. A number of Czech videotaped lessons showed students at the chalkboard, working through complex mathematics problems and being asked to justify the steps they were using to solve the problems. A key feature of these segments is the teachers' persistent questions about why mathematical procedures work. Impressions of well-structured and demanding review segments in Czech lessons are supported by the mathematicians' relatively high ratings of these lessons, presented earlier, in regard to level of content, degree of mathematical justification, and extent of mathematical development.

Our impression is that Czech students experienced different mathematics than U.S. students during their respective review segments. This comparison suggests that students' learning opportunities are not defined simply by whether review is provided, or even how much time is spent on review, but rather how the reviews are conducted and for what purpose. In other words, what matters is the role the review plays in the system. Review is a familiar part of the U.S. system; the comparison with the Czech system suggests that U.S. educators might want to rethink the goals that are set for reviews and the way in which they are conducted.

Hong Kong SAR provides a second instructive contrast to the United States because Hong Kong SAR emphasized procedures, at least as measured by the types of problems presented (see Figure 3), to an even greater extent than did the United States. But other indicators showed that the procedures used in Hong Kong SAR lessons were at a higher level of mathematical challenge and that, at least some of the time, teachers developed the conceptual underpinnings of the procedures. The Hong Kong SAR videotapes show teachers, in these cases, working through a procedure deliberately, discussing at each step why the procedure works.[6] Attention to the conceptual underpinnings of procedures was nearly absent from the U.S. lessons. Again, the relative focus on procedures suggests that, in some ways, the Hong Kong SAR system would be familiar to U.S. teachers. The differences derive in how procedures are worked out with students. Although the differences are nontrivial, they should be comprehensible to U.S. teachers.

The systems of teaching in Hong Kong SAR and the Czech Republic show that systems of teaching that share some similarities with the U.S. system can support high achievement. The current system of teaching in the United States and the dramatically different system of Japan are not the only choices. Given the characteristics of the U.S. system that relentlessly reinforce attention to lower-level skills, there are good reasons to want to improve the system. If U.S. teachers wished to change their teaching by studying the systems in higher-achieving countries, the findings just presented suggest that they have more options than replacing their system with an entirely different one. They could study, in detail, the ways in which systems in other countries balance attention to lower-level skills with attention to more challenging, conceptual work.

It is important to note that we focused our comparison with the Czech Republic and Hong Kong SAR on the features of review and executing procedures, respectively. This could be viewed as violating our cautions against isolating features. In fact, our argument for considering systems of teaching rather than individual features does not mean that individual features should not be identified or examined. Rather, it means that the effects of individual features must be understood within the context of the system. The systems of teaching in the Czech Republic and Hong Kong SAR provide instructive examples because the features of review and executing procedures, respectively, function quite differently in these systems than in the United States and consequently appear to provide quite different learning opportunities.

Informing Policy Debates on Mathematics Teaching

Having reviewed results from the TIMSS 1999 Video Study that portray different systems of teaching operating in different countries, we now ask what the results mean for the continuing national debates on how mathematics should be taught in U.S. schools. What teaching processes should be examined in addition to student assessment outcomes and other accountability indicators?

Aligning Teaching With Learning Goals

A first premise is that different systems of teaching will be more or less effective in terms of achieving different learning goals. Systems of teaching are not simply effective or ineffective, they are more or less effective *for* something. It is foolish to discuss whether one system of teaching is better than another until learning goals are clearly specified. The teaching systems of the higher-achieving countries in the TIMSS 1999 Video Study clearly support the learning goals assessed in the TIMSS 1999 achievement tests. The interpretations we have offered assume that these goals are at least a part of those U.S. mathematics educators endorse. However, it is likely that other goals are valued as well (National Research Council, 2001). To make cumulative progress in improving a system of teaching, it is essential to be explicit about and prioritize the learning goals.

Improving Systems of Teaching

Once learning goals are clearly specified, how can teachers change their practices to help students accomplish these goals? Policy debates often have focused on specific features of teaching. One example is the current debate in the mathematics education community between emphasizing conceptual understanding versus basic skills as a route to higher achievement (Loveless, 2003; National Council of Teachers of Mathematics, 2000). The debate often leads to recommendations either to present more problems that emphasize procedures or to present more problems that emphasize concepts. This simplistic distinction isolates features of teaching, and, as the data show, it does not capture what separates teaching in higher-achieving countries from that in the United States. In fact, the two highest-achieving countries in the sample (Japan and Hong Kong SAR) were at the opposite ends of this dimension. Despite the enormous energy that goes into debating issues such as this in the United States, these simplifications ignore the theoretical arguments for understanding the effects of features within systems and ignore the evidence that systems of teaching in higher-achieving countries are not characterized by such bifurcated choices.

It is reasonable to conclude, on the basis of accumulating data, that the feature-by-feature approach to improving teaching simply does not work. When trying to implement particular features, teachers and administrators often focus on *what* to implement rather than *how* to implement it, focusing on the presence of a feature rather than its purpose and how it interacts with other features to achieve or block learning goals. District leaders (Spillane, 2000) and classroom teachers (Guthrie, 1990; Spillane & Jennings, 1997) often make sure that the form of the feature is in place and worry less about the function. Form focuses attention, for example, on whether students are using specific materials, whether conceptually challenging problems are presented, and whether students are presenting solution methods. Function focuses on what role these features play in the teaching system and how the features work together to facilitate students' achievement of particular learning goals. Form over function leads to superficial implementation that does not achieve the intended improvement (Spillane, 2000).

With the understanding that teaching is a system and that individual features work together to support specific learning goals, the function of a feature takes center stage. It is not enough to simply include a new feature, such as presenting more challenging problems or spending more time on new material. What matters is how these features together are enacted with students. Although these conclusions are not new (e.g., Brown & Campione, 1996; Guthrie, 1990; Stein & Lane, 1996), the ways in which they were supported in the TIMSS video data provide compelling evidence for them.

Much time has been wasted in the United States studying achievement scores and guessing what individual features of teaching should be changed to improve these scores. In mathematics, a growing set of data indicates that classroom practice currently is tailored to support students' execution of low-level skills. Popular calls for change (National Council of Teachers of Mathematics, 1989, 2000) have not (yet) affected the ordinary eighth-grade mathematics classroom. As policy discussions consider future directions for U.S. school mathematics, and as the next generation of recommendations are formed, educators should take into account what is currently happening inside classrooms and should consider how the current system of teaching, rather than individual features, could be improved. This requires a realistic assessment of how the current system, with which U.S. teachers are familiar, could be adjusted to increase its alignment with more ambitious learning goals. The systems summarized in this article provide some promising options that warrant further detailed examination.

The difficulty of making educationally significant changes in a system of teaching should not be underestimated. Earlier we alluded to forces outside the classroom that are likely to shape and sustain systems of teaching. A comprehensive solution would address these forces, including development of (a) a consensus on standards of practice that signal a true profession (Yinger, 1999), (b) a teacher professional learning system that is coherent and consistent from the early days of preparation (Hiebert, Morris, & Glass, 2003; Nemser, 1983) through the teacher's entire career (Darling-Hammond & Sykes, 1999; Loucks-Horsley, Hewson, Love, & Stiles, 1998) and that would reduce professional isolation (Lortie, 1975), and (c) a system of testing, accumulating, and sharing of knowledge bases for teaching (Cochran-Smith & Lytle, 1999; Hargreaves, 1998; Hiebert, Gallimore, & Stigler, 2002). In our view, the goal toward which these changes should be directed is a teaching system well aligned with clear and widely accepted student learning goals. Although work remains on developing a consensus on learning goals (Loveless, 2001; National Council of Teachers of Mathematics, 2000; Schmidt et al., 1997), the contrasts among systems presented in this article provide information that can be used to work toward a teaching system that is more effective in helping students achieve the more ambitious goals around which consensus is building (National Research Council, 2001).

Notes

The TIMSS 1999 Video Study was funded by the National Center for Education Statistics and the Office of Educational Research and Improvement of the U.S. Department of Education, as well as the National Science Foundation. It was conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA). Support was also provided by each participating country through the services of a research coordinator who guided the sampling and recruiting of participating teachers. In addition, Australia and Switzerland contributed direct financial support for data collection and processing of their respective samples of lessons.

We would like to thank the editors and three anonymous reviewers for their comments on previous versions of the article. The views expressed are those of the authors and do not necessarily reflect the views of the IEA, the funding agencies, or the reviewers and editors.

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

[2] It should be noted that, in most of the countries, the video sample did not include the same classrooms as those used to collect achievement data. Because it is impossible to link directly teaching methods and levels of achievement in a sample that includes only one lesson in each classroom and during only the year of achievement testing, most countries did not draw the video sample from the classrooms tested for achievement. That direct links between teaching and achievement cannot be made does not, however, eliminate interest in studying the teaching methods used in other countries because, in part, these comparisons help one see one's own practices more clearly and reveal alternative practices that one might not have imagined otherwise.

[3] The Japanese mathematics lessons collected for the TIMSS 1995 Video Study were reexamined according to the revised and expanded coding scheme developed for the present study. As noted in reports of the 1995 study, the Japanese sample was filmed over a part of the school year rather than the whole year (e.g., Stigler et al., 1999).

[4] The group was led by Alfred Manaster and included Phillip Emig, Wallace Etterbeek, and Barbara Wells. This is the same group that analyzed a subsample of the TIMSS 1995 Video Study lessons. The group developed a different set of constructs for this study than those used in the 1995 study.

[5] The group of mathematicians and postsecondary mathematics teachers examined only 20 lessons per country because of limitations in regard to time and resources. Considerable time was required for all four members to analyze, discuss, and reach consensus on the multiple judgments for each lesson.

[6] Related features of Hong Kong SAR teaching, such as the careful sequencing of problems that highlight key conceptual aspects of the topic, can be seen in the videotapes (Hiebert & Handa, 2004) and have been reported by others using independent data sets (Gu, Huang, & Marton, 2004; Leung, 1995).

References

Adams, J. E., Jr., & Kirst, M. W. (1999). New demands and concepts for educational accountability: Striving for results in an era of excellence. In J. Murphy & K. S. Louis (Eds.), *Handbook of research on educational administration: A project of the American Educational Research Association* (pp. 463–489). San Francisco: Jossey-Bass.

Askey, R. (2001). Good intentions are not enough. In. T. Loveless (Ed.), *The great curriculum debate* (pp. 163–183). Washington, DC: Brookings Institution.

Beaton, A., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study.* Chestnut Hill, MA: Boston College.

Brophy, J. E., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328–375). New York: Macmillan.

Brown, A. L., & Campione, J. C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education* (pp. 289–325). Mahwah, NJ: Erlbaum.

Brownell, W. A. (1935). Psychological considerations in the learning and teaching of arithmetic. In W. D. Reeve (Ed.), *The teaching of arithmetic: Tenth yearbook of the National Council of Teachers of Mathematics* (pp. 1–31). New York: Teachers College, Columbia University.

Cheney, L. V. (1997, June 11). President Clinton's mandate for fuzzy math. *Wall Street Journal,* p. A22.

Cochran-Smith, M., & Lytle, S. L. (1999). The teacher researcher movement: A decade later. *Educational Researcher, 28*(7), 15–25.

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25,* 119–142.

among systems presented in this article provide information that can be used to work toward a teaching system that is more effective in helping students achieve the more ambitious goals around which consensus is building (National Research Council, 2001).

Notes

The TIMSS 1999 Video Study was funded by the National Center for Education Statistics and the Office of Educational Research and Improvement of the U.S. Department of Education, as well as the National Science Foundation. It was conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA). Support was also provided by each participating country through the services of a research coordinator who guided the sampling and recruiting of participating teachers. In addition, Australia and Switzerland contributed direct financial support for data collection and processing of their respective samples of lessons.

We would like to thank the editors and three anonymous reviewers for their comments on previous versions of the article. The views expressed are those of the authors and do not necessarily reflect the views of the IEA, the funding agencies, or the reviewers and editors.

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

[2] It should be noted that, in most of the countries, the video sample did not include the same classrooms as those used to collect achievement data. Because it is impossible to link directly teaching methods and levels of achievement in a sample that includes only one lesson in each classroom and during only the year of achievement testing, most countries did not draw the video sample from the classrooms tested for achievement. That direct links between teaching and achievement cannot be made does not, however, eliminate interest in studying the teaching methods used in other countries because, in part, these comparisons help one see one's own practices more clearly and reveal alternative practices that one might not have imagined otherwise.

[3] The Japanese mathematics lessons collected for the TIMSS 1995 Video Study were reexamined according to the revised and expanded coding scheme developed for the present study. As noted in reports of the 1995 study, the Japanese sample was filmed over a part of the school year rather than the whole year (e.g., Stigler et al., 1999).

[4] The group was led by Alfred Manaster and included Phillip Emig, Wallace Etterbeek, and Barbara Wells. This is the same group that analyzed a subsample of the TIMSS 1995 Video Study lessons. The group developed a different set of constructs for this study than those used in the 1995 study.

[5] The group of mathematicians and postsecondary mathematics teachers examined only 20 lessons per country because of limitations in regard to time and resources. Considerable time was required for all four members to analyze, discuss, and reach consensus on the multiple judgments for each lesson.

[6] Related features of Hong Kong SAR teaching, such as the careful sequencing of problems that highlight key conceptual aspects of the topic, can be seen in the videotapes (Hiebert & Handa, 2004) and have been reported by others using independent data sets (Gu, Huang, & Marton, 2004; Leung, 1995).

References

Adams, J. E., Jr., & Kirst, M. W. (1999). New demands and concepts for educational accountability: Striving for results in an era of excellence. In J. Murphy & K. S. Louis (Eds.), *Handbook of research on educational administration: A project of the American Educational Research Association* (pp. 463–489). San Francisco: Jossey-Bass.

Askey, R. (2001). Good intentions are not enough. In. T. Loveless (Ed.), *The great curriculum debate* (pp. 163–183). Washington, DC: Brookings Institution.

Beaton, A., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study.* Chestnut Hill, MA: Boston College.

Brophy, J. E., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328–375). New York: Macmillan.

Brown, A. L., & Campione, J. C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education* (pp. 289–325). Mahwah, NJ: Erlbaum.

Brownell, W. A. (1935). Psychological considerations in the learning and teaching of arithmetic. In W. D. Reeve (Ed.), *The teaching of arithmetic: Tenth yearbook of the National Council of Teachers of Mathematics* (pp. 1–31). New York: Teachers College, Columbia University.

Cheney, L. V. (1997, June 11). President Clinton's mandate for fuzzy math. *Wall Street Journal,* p. A22.

Cochran-Smith, M., & Lytle, S. L. (1999). The teacher researcher movement: A decade later. *Educational Researcher, 28*(7), 15–25.

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25,* 119–142.

Darling-Hammond, L., & Sykes, G. (Eds.). (1999). *Teaching as the learning profession: Handbook of policy and practice.* San Francisco: Jossey-Bass.

Dreeben, R. (1994). The sociology of education: Its development in the United States. In A. Pallas (Ed.), *Research in sociology of education and socialization* (Vol. 10, pp. 7–52). Greenwich, CT: JAI Press.

Elmore, R. F. (1996). Getting to scale with good educational practice. *Harvard Educational Review, 66,* 1–26.

Fernandez, C., Yoshida, M., & Stigler, J. W. (1992). Learning mathematics from classroom instruction: On relating lessons to pupils' interpretations. *Journal of the Learning Sciences, 2,* 333–365.

Flanders, J. R. (1987). How much of the content in mathematics textbooks is new? *Arithmetic Teacher, 35*(1), 18–23.

Fullan, M. (2001). *The new meaning of educational change* (3rd ed.). New York: Teachers College Press.

Gonzales, P., Calsyn, C., Jocelyn, L., Mak, K., Kastberg, D., Arafeh, S., Williams, T., & Tsen, W. (2000). *Pursuing excellence: Comparisons of international eighth-grade mathematics and science achievement from a U.S. perspective, 1995 and 1999* (NCES Publication No. 2001-028). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Good, T. L., Grouws, D. A., & Ebmeier, H. (1983). *Active mathematics teaching.* New York: Longman.

Good, T. L., Mulryan, C., & McCaslin, M. (1992). Grouping for instruction in mathematics: A call for programmatic research on small-group processes. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 165–196). New York: Macmillan.

Gu, L., Huang, R., & Marton, F. (2004). Teaching with variation: A Chinese way of promoting effective mathematics learning. In L. Fan, N. Y. Wong, J. Cai, & S. Li (Eds.), *How Chinese learn mathematics: Perspectives from insiders* (pp. 309–347). Singapore: World Scientific.

Guthrie, J. W. (Ed.). (1990). [Entire issue]. *Educational Evaluation and Policy Analysis, 12*(3).

Hargreaves, D. H. (1998, August). *The knowledge-creating school.* Paper presented at the annual meeting of the British Educational Research Association, Belfast, Northern Ireland.

Hiebert, J. (Ed.). (1986). *Conceptual and procedural knowledge: The case of mathematics.* Mahwah, NJ: Erlbaum.

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., Chui, A. M., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeek, W., Manaster, C., Gonzales, P., & Stigler, J. W. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 Video Study* (NCES Publication No. 2003-013). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Hiebert, J., Gallimore, R., & Stigler, J. W. (2002). A knowledge base for the teaching profession: What would it look like and how can we get one? *Educational Researcher, 31*(5), 3–15.

Hiebert, J., & Handa, Y. (2004, April). *A modest proposal for reconceptualizing the activity of learning mathematics procedures.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Hiebert, J., Morris, A. K., & Glass, B. (2003). Learning to learn to teach: An "experiment" model for teaching and teacher preparation in mathematics. *Journal of Mathematics Teacher Education, 6,* 201–222.

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

Ingersoll, R. (2003). *Who controls teachers' work? Power and accountability in America's schools.* Cambridge, MA: Harvard University Press.

Jacobs, J., Garnier, H., Gallimore, R., Hollingsworth, H., Givvin, K. B., Rust, K., Kawanaka, T., Smith, M., Wearne, D., Manaster, A., Etterbeek, W., Hiebert, J., & Stigler, J. W. (2003). *Third International Mathematics and Science Study 1999 Video Study technical report: Vol. 1. Mathematics* (NCES Publication No. 2003-012). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Jenkins, L. (1997). *Improving student learning: Applying Deming's quality principles in classrooms.* Milwaukee, WI: ASQC Quality Press.

Leung, F. K. S. (1995). The mathematics classrooms of Beijing, Hong Kong, and London. *Educational Studies in Mathematics, 29,* 297–325.

Lortie, D. C. (1975). *Schoolteacher: A sociological study.* Chicago: University of Chicago Press.

Loucks-Horsley, S., Hewson, P. W., Love, N., & Stiles, K. E. (1998). *Designing professional development for teachers of science and mathematics.* Thousands Oaks, CA: Corwin Press.

Loveless, T. (2001). A tale of two math reforms: The politics of the new math and the NCTM standards. In. T. Loveless (Ed.), *The great curriculum debate* (pp. 184–209). Washington, DC: Brookings Institution.

Loveless, T. (2003, February). *Trends in math achievement: The importance of basic skills.* Paper presented at the Secretary's Summit on Mathematics, Washington, DC.

National Advisory Committee on Mathematics Education. (1975). *Overview and analysis of school mathematics, Grades K–12.* Washington, DC: Conference Board of the Mathematical Sciences.

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

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Research Council. (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

National Research Council. (2001). *Adding it up: Helping children learn mathematics.* Washington, DC: National Academy Press.

Nemser, S. (1983). Learning to teach. In L. S. Shulman & G. Sykes (Eds.), *Handbook of teaching and policy* (pp. 150–170). New York: Longman.

Oakes, J. (1985). *Keeping track: How schools structure inequality.* New Haven, CT: Yale University Press.

Ravitch, D. (2002). A brief history of testing and accountability. *Hoover Digest,* No. 4. Retrieved September 23, 2004, from http://www.hooverdigest.org/024/ravitch.html

Rowan, B., Harrison, D., & Hayes, A. (2004). Using instructional logs to study mathematics curriculum and teaching in the early grades. *Elementary School Journal, 105,* 103–127.

Sarason, S. B. (1997). *How schools might be governed and why.* New York: Teachers College Press.

Schmidt, W. H., McKnight, C. C., Valverde, G. A., Houang, R. T., & Wiley, D. E. (1997). *Many visions, many aims: A cross-national investigation of curricular intentions in school mathematics.* Dordrecht, the Netherlands: Kluwer.

Schoenfeld, A. (1998). Toward a theory of teaching-in-context. *Issues in Education: Contributions from Educational Psychology, 4*(1), 1–94.

Skinner, B. F. (1969). *The technology of teaching.* Englewood Cliffs, NJ: Prentice Hall.

Smith, M. (2000). *A comparison of the types of mathematics tasks and how they were completed during eighth-grade mathematics instruction in Germany, Japan, and the United States.* Unpublished doctoral dissertation, University of Delaware.

Spillane, J. P. (2000). Cognition and policy implementation: District policymakers and the reform of mathematics education. *Cognition and Instruction, 18,* 141–179.

Spillane, J. P., & Jennings, N. E. (1997). Aligned instructional policy and ambitious pedagogy: Exploring instructional reform from the classroom perspective. *Teachers College Record, 98,* 449–481.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal, 33,* 455–488.

Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation, 2*(1), 50–80.

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* (NCES Publication No. 1999-074). 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.

Stodolsky, S. S. (1988). *The subject matters: Classroom activity in math and social studies.* Chicago: University of Chicago Press.

Webb, N. M., Troper, J. D., & Fall, R. (1995). Constructive activity and learning in collaborative small groups. *Journal of Educational Psychology, 87,* 406–423.

Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. J. (2003). *Looking inside the classroom: A study of K–12 mathematics and science education in the United States.* Retrieved July 19, 2004, from http://horizon-research.com

Wilson, K. G., & Daviss, B. (1994). *Redesigning education.* New York: Holt.

Wu, H. (1997). The mathematics education reform: Why you should be concerned and what you can do. *American Mathematical Monthly, 103,* 946–954.

Yinger, R. (1999). The role of standards in teaching and teacher education. In G. Griffin (Ed.), *The education of teachers: Ninety-eighth yearbook of the National Society for the Study of Education* (pp. 85–113). Chicago: University of Chicago Press.

Authors

JAMES HIEBERT is the Robert J. Barkley Professor of Education, School of Education, University of Delaware, 16 West Main Street, Newark, DE 19716; hiebert@udel.edu. His areas of specialization are mathematics education and classroom teaching and learning.

JAMES W. STIGLER is a Professor, Department of Psychology, University of California, Los Angeles, and CEO of LessonLab, 3330 Ocean Park Boulevard, Suite

115, Santa Monica, CA 90405; jims@lessonlab.com. He specializes in studies of classroom teaching and teacher learning.

JENNIFER K. JACOBS is a Research Associate, Department of Education, University of Colorado, 1099 Fairway Court, Boulder, CO 80303; jenniferj@davidslane.com. Her areas of specialization are teaching practices and beliefs and mathematics professional development.

KAREN BOGARD GIVVIN is a Research Scientist, LessonLab Research Institute, 3330 Ocean Park Boulevard, Suite 115, Santa Monica, CA 90405; kareng@lessonlab.com. Her interests are mathematics education, the international examination of teaching practice, and motivation-related beliefs.

HELEN GARNIER is Chief Analyst, LessonLab, and a Research Statistician, University of California, Los Angeles, 1094 La Mirada, Laguna Beach, CA 92651; hgarnier@ucla.edu. Her area of specialization is educational research methods and evaluation.

MARGARET SMITH is an Assistant Professor, Education Department, Iona College, 715 North Avenue, New Rochelle, NY 10801; msmith@iona.edu. Her area of specialization is mathematics education.

HILARY HOLLINGSWORTH is an Educational Consultant, 1 Geraldine Street, Rye, Victoria 3941, Australia; hilary@hilaryhollingsworth.com. Her interests focus on mathematics education, teacher professional learning, and the use of video cases for professional learning.

ALFRED MANASTER is a Professor Emeritus of Mathematics, Mathematics Department, University of California, San Diego, 9500 Gilman Drive, Department 0112, La Jolla, CA 92093-0112; amanaster@ucsd.edu. His areas of specialization are mathematics education and mathematical logic.

DIANA WEARNE is a Professor Emerita of Mathematics Education, School of Education, University of Delaware, 16 West Main Street, Newark, DE 19716; wearne@udel.edu. Her areas of interest are the teaching and learning of mathematics.

RONALD GALLIMORE is a Distinguished Professor, Departments of Psychiatry and Biobehavioral Sciences and Education, University of California, Los Angeles, and Chief Scientist, LessonLab Research Institute, 3330 Ocean Park Boulevard, Suite 115, Santa Monica, CA 90405; ronaldg@ucla.edu. He investigates teaching and its improvement.

Manuscript received December 5, 2003
Revision received February 1, 2005
Accepted March 3, 2005