Understanding and misunderstanding the Third International Mathematics and Science Study: what is at stake and why K-12 education studies matter

Alan Siegel*

Abstract. The technical portion of this paper concerns a videotape classroom study of eighth grade mathematics lessons in Japan, and how methodological design errors led to conclusions that are refuted by the actual video data. We document these errors, and trace their distillation into one- and two-sentence education policy recommendations articulated in U.S. government position papers, implemented in classrooms across the U.S. and imported by countries around the world. We also present the historical context needed to understand the misrepresentations cited in support of questionable education policy.

Mathematics Subject Classification (2000). Primary 97D40; Secondary 97-02.

Keywords. Third International Mathematics and Science Study, TIMSS.

1. Introduction

The outstanding results for the top-performing countries in the Third International Mathematics and Science Study (TIMSS) have generated widespread interest in best teaching practices around the world. In the TIMSS Videotape Classroom Study by James Stigler et al. [31], the teaching styles in Germany, Japan, and the U.S. were compared in an effort to discover what makes some programs so successful. The conclusions from this comparison are striking and have been widely cited, but often in a highly trivialized and even inaccurate manner. Moreover, this particular study, as we will show, is marred by design errors that raise serious doubt about some of its most influential conclusions. Indeed, it is these very findings that have been cited and accidentally distorted in support of the latest reform programs and education policies – both in the U.S. and elsewhere.

For example, it is widely acknowledged (cf. [31, p. 134]) that Japanese lessons often use very challenging problems as motivational focal points for the content being taught. According to the Glenn Commission[1] Report [10, p. 16],

*The author is grateful to the ICM, NSF and AMS for their support of this presentation. Disclaimer: although the assessments and statements in this paper have been made in good faith by the author, they should not necessarily be viewed as representative of or endorsed by the ICM, NSF or AMS.

[1] The commission's proper name is the National Commission on Mathematics and Science Teaching for the 21st Century. It was chaired by former U.S. Senator and astronaut John Glenn. The year-long Commission was

Proceedings of the International Congress
of Mathematicians, Madrid, Spain, 2006

> "In Japan, … closely supervised, collaborative work among students is the norm. Teachers begin by presenting students with a mathematics problem employing principles they have not yet learned. They then work alone or in small groups to devise a solution. After a few minutes, students are called on to present their answers; the whole class works through the problems and solutions, uncovering the related mathematical concepts and reasoning."

We revisit the TIMSS Videotape Classroom Study to resolve the one crucial classroom question that both the Glenn Commission and the TIMSS Classroom Study group failed to address:

> How can Japanese eighth graders, with just a few minutes of thought, solve difficult problems employing principles they have not yet learned?

We will see that the technique required to solve the challenge problem of the day will have already been taught, and that the lesson begins with a review of the fundamental method needed to solve the problem. Students begin working on these problems individually – not in groups. Sometimes group-work is allowed for second efforts on a given assignment, but only after individual seat-work. These lessons include student-presented solutions, but the presentations are closely supervised by the teacher, and the time allocated for this activity is limited so that students will be able to work on a second challenge exercise of the same type, and the teacher will have enough time to show how to apply a fundamental technique as many as ten times – all in a single lesson. Stigler's videotapes reveal master teaching of substantial content hidden within a warm and inviting teaching style. Students do indeed participate, but in moderation, and subject to the vigilant oversight of instructors who ensure that no one wanders off course.

It is also worth noting that the Videotape Classroom Study identified some of the significant differences between the current reform positions and Japanese teaching practices. For example, it pointed out that students did not use calculators in the Japanese classes, and that Japanese teaching has a far higher concentration of proofs and derivations than both reform and traditional programs in the U.S. The Videotape Study also found that Japanese teachers spend more time lecturing than even traditional U.S. teachers.

These distinctions not withstanding, the notion that Japanese teaching might be implementing U.S. reforms is given far greater emphasis in a major Government report, which flatly declares:

> "Japanese teachers widely practice what the U.S. mathematics reform recommends, while U.S. teachers do so infrequently [25, p. 9]."

---

mandated to develop a strategy to raise the quality of mathematics and science teaching in all of the nation's classrooms. Unfortunately, the cited quote was, quite possibly, the most substantive paragraph in their report to the nation. The preliminary Glenn Commission report cited Stigler and his TIMSS Videotape Classroom Study as the source of this finding, although the final version omitted the specific citation.

This report on best teaching practices worldwide makes no mention of any differences between the U.S. reforms and Japanese teaching styles. Evidently, its perspective (see also [25, pp. 40–43]) differs from that of its source of primary information, which is the more cautiously worded TIMSS Videotape Study [31]. Moreover, the differences identified in the Videotape Study – which concern direct instruction, calculators, and teacher-managed demonstrations – are all matters of contention in the U.S. debate over classroom reform.

Finally, we note that studies of individual classroom lessons – no matter how comprehensive – are necessarily incomplete. They cannot detect how coherent a curriculum might be day-by-day, much less over the course of years, and are ill-equipped to assess the completeness of a given math curriculum.

2. Background

The need for sound – and indeed first-rate – K-12 mathematics programs is well understood. In the U.S., many reform programs have been implemented over the last fifty years, but the evidence shows that on balance, we have made very modest progress toward this goal of world-class math education.

The majority of our past reform efforts can be characterized as a tug of war between traditional and student-centric education movements. Just one of these programs was sufficiently different to deserve special mention: the so-called New Math that originated in the 1950s, and which was widely implemented in the '60s. This reform was pioneered by mathematicians, and was the only program ever to attempt to teach elementary mathematics from an informal set-theoretic perspective. It failed, in part, because its implementations did not provide safeguards to ensure that mainstream American students – and teachers – could handle the material, which is an error that the current reformers have been very careful to avoid. Finally, the program has historical importance because its failure led to a fairly sharp separation between those concerned with K-12 math education and those interested in mathematics research and college teaching.

In the mid-1980s, a new version of student-centric learning and teaching began taking hold in the mathematics education community, and it is fair to say that these ideas have swept the American schools of education, and are likewise well represented by advocates in many other parts of the world.

In 1989, these ideas were codified into teaching policy when "educators ... carefully articulated a new vision of mathematics learning and curriculum in the National Council of Teachers of Mathematics' (NCTM's) *Curriculum and Evaluation Standards for School Mathematics* [6]." The 1989 Curriculum Standards [20], together with the follow-up 1991 Teaching Standards [21] and the 1995 Assessment Standards [22] called for a redirection of focus from what to teach grade by grade to new ideas about how to teach and how to assess student progress. And with the publication of these documents, the NCTM completed its transformation from an organization that began in the 1920s with ties to the Mathematical Association of America, and that had been led by content-oriented math teachers who endorsed the revolutionary New Math of the '60s, to an organization led by professors of mathematics education who endorsed a new type of revolutionary math program[2] in the '90s.

Loosely put, the theoretical core of this new vision of education is called constructivism. Like most complex social theories, constructivism is founded on a few main principles, has many interpretations and derived consequences, and a bewildering variety of implementations. A thumbnail (and necessarily incomplete) sketch of the main principles of constructivism is as follows.

The philosophical basis of constructivism is that everyone learns differently, and that we learn best by integrating new knowledge into our own core understandings and thought processes. Therefore, education is most effective when it engages the learner to become the main agent in the learning process. That is, learning should be engaging in every sense of the word. Since we learn by discovering and by doing, learning is a quintessentially social process wherein through mutual interaction, we organize, communicate, share, and thereby develop deepened understanding. Moreover, content should be based on real-world problems to reach each learner's core knowledge base, and to maximize the purposefulness of each lesson.

As stated, these objectives have merit – especially for teaching younger learners. Indeed, the author believes that the debate over abstract constructivism misses the point. However, the teaching reforms advocated by the NCTM include, in addition to abstract principles, very applied recommendations that have significant impact on curricula, pedagogy and the opportunity for students to learn mathematics.

Thus, the real questions concern the content and training provided by the reform program implementations, as well as the consequences of the derivative theories of learning and testing that are put forward as logical consequences of constructivist principles. And it is this debate about what kinds of education programs work that defines the context for the TIMSS Videotape Classroom Study and the classification of Japanese pedagogy.

The impact of reform principles on classroom structure and course content. The applied education theories advanced by contemporary reformers must be sketched out if the various assertions about Japanese teaching and the latest reform recommendations are to make sense.

The principle of discovery-based learning aims to have the students themselves discover mathematical principles and techniques. According to Cobb et al. [5, p. 28],

[2]The NCTM reform program was also endorsed by the federal department of Education and Human Resources, which provided funds to create reform-compliant textbooks, to support their use, and to support studies designed to prove that the new programs were effective. To date, more than $75 million has been allocated to produce these new mathematics textbooks, and about $1 billion has been spent on programs to foster their use. The Educational Systemic Reform programs, for example, ran for nine years with an annual budget of about $100 million, and related programs for K-12 math and science education received comparable funding. More about the history of these programs can be found in [32].

> "It is possible for students to construct for themselves the mathematical practices that, historically, took several thousand years to evolve."

In the 1999 Yearbook of the National Council of Teachers of Mathematics, the article "Teaching Fractions: Fostering Children's Own Reasoning" by Kamii and Warrington [15] advises:

> "1. Do not tell children how to compute by using numerical algorithms. . . .
>
> 2. Do not tell children that an answer is right or wrong. . . .
>
> 3. Encourage children to use their own reasoning instead of providing them with ready-made representations or 'embodiments.'
>
> 4. Ask children to estimate solutions to problems first because estimation is an effective way to build strong number sense."

To be fair to the authors, it should be pointed out that they provide alternatives to prohibitions 1, and 2. For example, they recommend that the issue of correctness be resolved by the entire class through cooperative discussion.

These discovery-based policies are often implemented via the *workshop model* of teaching where students are seated in clusters of four desks facing each other with no central lecture place in the classroom. This organization is designed to foster collaborative learning and to reinforce the teacher's role as a "guide on the side" as opposed to the "sage on the stage." In some programs, the purpose of the teacher is to introduce the exercise of the day. The students then work in groups of four to discover what they can about the problem. In the next phase, the students present their findings to the class, and an active discussion typically ensues. The teacher might have a role that is confined to being a moderator to maintain order in the discussions. Likewise, some of the programs feature unsupervised group-work with the teacher serving mainly as a passive observer.

In the higher grades, the U.S. discovery-based programs feature markedly diminished content depth, and the project-based texts exhibit poor coherence in their management of topics and offering of reinforcement exercises. To date, some reform programs simply omit material that does not fit within this model. Moreover, this style of teaching, absent sufficient guidance from the teacher, is typically very time consuming, and the slow pace cannot help but limit the curriculum.

For example, on page 315 of a tenth grade reform geometry textbook [37], exercise 24 asks the student to draw an equilateral, an isosceles, and a scalene triangle, and to draw the medians and observe the outcome in each case. The assignment also asks the students to measure the lengths of the medians and the distance from the vertices of each triangle to its centroid. The problem finishes by asking, "What do you conclude?" No proofs are offered or requested, and for good reason. The study of similar triangles begins in Chapter 13 on page 737, where the final two chapters of the book present content that is less observation-based.

In 2001, I was invited to observe some of these workshop model classes at a magnet high school in lower Manhattan. In one of the ninth grade classes, the lesson problem of the day was (in mathematical terms) to determine the equation of a line through the origin that does not intersect any additional points on the integer Cartesian lattice in $\mathbb{R}^2$. The students began the exercise working unsupervised in groups of four. Then the class convened as a whole to discuss their findings with the teacher serving as moderator. The tenth (or so) student to speak observed that if the line were to intersect another lattice point, then it would have a rational slope. The teacher then called on another student, and this key observation was soon lost. The discussion devolved into an unsuccessful effort to understand the difference between rational numbers with finite decimal representations and those with repeating decimal expansions, and the math period ended with no solution to either question.

In a televised eleventh grade lesson [24] from a reform textbook series [7], students seated in groups of four were given the following problem. The teacher displayed boards of different lengths, widths, and thicknesses suspended between pairs of bricks. A karate expert, he explained, can deliver the tremendous energy necessary to break a strong board. For the first part of the lesson, the students were asked to determine a formula for the energy necessary to break a board as a function of its length and thickness. The students discussed the question with great enthusiasm. There was no evidence of any physical modeling, and it was not clear if the class knew Hooke's law or not. In the second portion of the lesson, a representative from each group presented the group's thoughts to the class. The first to speak was able to intuit that a longer thinner board would be easier to break, but nevertheless went on to opine that the formula for the energy E, as a function of the length L and thickness T, should be $E = L + T$. Another group thought that the formula should be $E = kLT$, where k is a constant that depends on the physical properties of the wood. In the next portion of the lesson, students were given strands of dried spaghetti to form a bridge between two tables, pennies to use as weights, and a paper cup plus paper clip to suspend on the strand(s) of spaghetti. They then conducted tests with different lengths and strand counts to see how many pennies were necessary to break the spaghetti – thus measuring the breaking force, which was misrepresented as energy. The use of multiple strands served to emulate different thicknesses (albeit incorrectly). Data was gathered for 1 to 5 strands, and distances of 2 to 5 inches. Then the students used their graphing calculators under the supervision of the teacher to determine the best fit for the data, which was $E = 10\frac{T}{L}$, where E is measured in pennies, T in spaghetti strands, and L in inches.

The TV program closed by noting that with the introduction of this new curriculum, grades were higher, and more students were electing to take three and four years of math classes. Of course, the stacking of spaghetti strands to model thicker pasta constitutes a fundamental conceptual error. It is no accident that plywood is manufactured with bonded layers, and as straightforward mathematical modeling shows, strength, in a simple model of deformation, is proportional to the square of a beam's thickness. Likewise, the confusion between force and energy ill serves the students, as does the lesson's implication that mathematics might be an experimental science.

The reforms seeking to maximize engagement include mandates to avoid drill and – by extension – the kind of practice necessary to instill knowledge transfer to long-term memory.

In concrete terms, the reform programs do not teach the multiplication table in elementary school. Ocken reviewed all of the printed materials produced by one of the elementary school reform programs [33] for grades K-5, and found fewer than 30 problems asking students to multiply two whole numbers, both of which contain a digit greater than five[23]. This program implements the reduced emphasis on pencil and paper calculations as recommended in the 1989 NCTM Standards, and, as recommended, supports student work with calculators even in the earliest grades.

Likewise, the standard place-based rules for multiplication of multidigit integers are no longer taught as essential material. Opponents of these reforms see the structure of place-based multiplication as precursor knowledge that helps the learner internalize the more abstract operations of polynomial arithmetic.

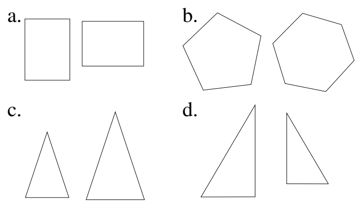
In one textbook series [16], the division of fractions was simply omitted from the curriculum. And long division is long gone from these programs.

To maximize engagement, reformers recommend that problems and content be *situated*, which means that exercises, derivations and even theorems should be presented in an applied context whenever possible. More generally, abstraction and symbolic methods are eschewed. Of course, the foregoing comments about abstraction and symbolic methods are just words. In order to understand them, we again take a few quick peeks into the reform textbooks to see how these theories and recommendations are turned into practice. For example, one ninth grade reform book [8] has, scattered among its 515 pages, only 25 pages that even contain an equal sign. Of these, only pages 435 and 436 actually concern algebra. The totality of the information about algebra is on page 436, and is as follows.

> "Some such equations are easier to solve than others. Sometimes the particular numbers involved suggest tricks or shortcuts that make them easy to solve. In each of the equations below, the letter x stands for an unknown number. Use any method you like to find the number x stands for, but write down exactly how you do it. Be sure to check your answers and write down in detail how you find them.
>
> $$\begin{array}{cccc} \frac{x}{5} = 7 & \frac{x}{6} = \frac{72}{24} & \frac{x}{8} = \frac{11}{4} & \frac{x}{7} = \frac{5}{3} \\ \frac{x+1}{3} = \frac{4}{6} & \frac{5}{13} = \frac{19}{x} & \frac{2}{x} = 6 & \frac{9}{x} = \frac{x}{16} \end{array}$$"

The preference for encouraging ad hoc "tricks" and "shortcuts" instead of teaching systematic methods is evident. Indeed, the text does not present any methods for solving these problems. The passage also illustrates how the these new programs encourage students to write expository explanations and avoid teaching students to develop and record logical solution strategies based on correct operations, problem decomposition, and the layered application of systematic methods.

On page 416 of this ninth grade text, problem 3 reads as follows.



3. Consider the following pairs of figures. In each case, state whether you consider the shape to be the same or not, and why.

The chapter goes on to explore some of the most elementary properties of similarity, but the development is probably closer to the level of sixth grade than ninth, and the overall content of the textbook is far weaker than, say, the standard sixth grade books used in Singapore [13].

The comparison with the Singapore books is worthy of elaboration. In an American Educator article [1], the mathematician Ron Aharoni writes about what he learned using the Singapore math program to teach first grade in Israel. He points out that these lessons encourage students to describe problems in words, and feature more discussion than is common in traditional programs. These characteristics are consistent with some of the constructivist principles. There are, however, fundamental differences between this teaching style and the applied recommendations and prohibitions that characterize – and indeed define – the latest reform practices. Aharoni describes how he actively teaches insights based on his mathematical knowledge – even in first grade. And he also points out that significant reinforcement is necessary to help first graders integrate this first grade content into their own thinking. Interestingly, the fifth and sixth grade Singapore texts [13] exhibit a transition from this verbal/expository approach of reasoned problem representation to an informal but precise prealgebra. The books present – with many detailed examples – a kind of pictorial algebra, where a physical segment might be used to represent an unknown length. The modeling allows graphical unknowns to be added, subtracted, and multiplied and divided by integers in physical representations of equations. Students solve many carefully constructed word problems with this modeling process and its physical representation of variables. This representation is used to strengthen intuition and understanding as preparation for variables and algebra. By the sixth grade, the students are using the method to solve sophisticated word problems that would challenge U.S. high schoolers. No U.S. reform text presents such a coherent curriculum, and none provides a systematic increase of content and problem depth chapter-by-chapter and over the course of years to build deepening layers of understanding on behalf of the learner.

In terms of pedagogy, Aharoni emphasizes the importance of deep content knowledge and a deep understanding of *what* is being taught as prerequisites for deciding *how* to teach a particular topic [1, p. 13]. He says that the understanding of fundamental mathematical principles can be taught, but this instruction requires active teaching by a very knowledgeable teacher.

The current reform programs, by way of contrast, aim to teach less, not more. In a ninth grade reform algebra text, for example, students receive enough training to solve for x in the equation $y = 3x + 2$, but there is just one equation in the book that uses variable coefficients. This one exception, which is on page 748 reads [9]:

> "Show how to derive the quadratic formula by applying completing the square to the general quadratic equation, $ax^2 + bx + c = 0$."

This question requires a tremendous leap in skill given the text's limited use of equations with variable coefficients. Moreover, the presentation on completing the square is so weak that it is inconceivable how any but the most exceptional student could learn enough to solve this problem. The totality of the exposition reads:

> "Here's an example of how to use completing the square to solve the quadratic equation $x^2 + 6x - 2 = 5$.
>
> Since -2 doesn't make $x^2 + 6x$ a perfect square, it is in the way. Move it to the other side: $x^2 + 6x = 7$.
>
> Add 9 to both sides to make the left side a perfect square: $x^2 + 6x + 9 = 16$.
>
> Write the left side as a perfect square: $(x + 3)^2 = 16$."

There is no attempt to teach a systematic approach for completing the square, or to explain how the magical 9 was selected for use in this particular case.

The avoidance of abstraction and symbolic coefficients, and the recommendations against teaching systematic methods have undermined the quality of the textbook. This instance of teaching by one explicit example cannot instill wide-spread understanding. And the inclusion of the exercise to derive the quadratic formula (which is just about the last problem in a very long book) would appear to be based less on it being an appropriate exercise than on the need to include the topic in the curriculum.[3]

Ralston recommends the outright abandonment of pencil and paper calculations in favor of mental arithmetic supplemented by calculators [26]. Non-reformers disagree, and suggest that proficiency in arithmetic is not taught for its own sake but rather to strengthen the learner's core knowledge and intuition as a prerequisite for understanding fractions. Arithmetic fluency is even more important for a mastery of and fluency in algebra, where the rules of arithmetic are revisited at an abstract level with the introduction of variables and exponents. Many teachers report that those who lack a grounding in the concrete operations of arithmetic experience great difficulty with algebra and its manipulation of symbols. Other non-reformers argue that the written record of pencil and paper problem solving documents a student's approach to a problem, which can be reviewed by the student and the teacher for conceptual errors as well as computational mistakes. Non-reformers also argue that it is the use of the written record that allows learners to combine fundamental steps into more

[3] It is also fair to say that some of the most project-based reform texts are designed around sequences of typically unrelated projects, which result in a disorganized and incomplete curriculum with very few review and reinforcement exercises (cf. [33], [8], [7], [9]).

complex solutions that are too detailed to retain as mental calculations. In addition, it is argued that the written representations of algebra bring a precision of expression, of computation and of modeling that surpasses the written word in accuracy, clarity, and simplicity.

The purpose of this inside review of American mathematics education was to identify the controversies arising from the latest reforms in concrete (i.e. situated) – as opposed to abstract – terms. It is time to explain why Japanese pedagogy has become a topic of worldwide interest, and to investigate how well it aligns with the latest reform principles.

The Third International Mathematics and Science Study. TIMSS is an enormous umbrella project that seeks to measure academic achievement around the world, and which includes many subsidiary studies that analyze a host of related issues in an effort to determine how best to improve math and science education. TIMSS began in 1994–95 with the testing of 400,000 students worldwide at grades four, eight, and twelve. It has grown into a quadrennial program that conducted additional testings and data acquisitions in 1999 and 2003, and has already begun to lay the groundwork for the next round in 2007. The program now includes nearly fifty countries, and the studies cover a large number of independent projects with publications in the many thousands of pages.

Although there have been some fluctuations in the TIMSS rankings over the last decade, and the participating countries have varied to some degree over time, the overall results have been much more consistent than not. This fact is probably a testament to the meticulous effort to maintain balanced student samples from the participating countries, and the care that is exercised in the testing protocols and data analyses. The project also deserves very high marks for adhering to a wonderfully high standard of scholarship. The research projects produce not only reports of findings but also detailed documentation of the data acquisition and analysis procedures and indeed every aspect of project methodology. When feasible, these studies even publish enough raw data for independent researchers to review every step of the research effort for independent assessment.

Despite the wealth of information provided by the TIMSS publications, it is fair to say that two specific TIMSS findings have captured the majority of the headlines, and have had the greatest influence on classroom practice and education policy.

The most eye-opening results come from the achievement scores of students around the world. For example, in the little multicultural, multilingual, top-performing country of Singapore, some 46% of the eighth graders scored in the top 10% of the world. And 75% of their students placed among the top 25% of all eighth graders worldwide. Just 1% of their students placed among the bottom 25% of all eighth graders around the world. This is a stunning achievement. Singapore has indeed shown what it really means to have an education system where no child is left behind.

Moreover, these performance results have held up with remarkable consistency in each of the TIMSS testing rounds. Just a notch down from Singapore, the next group of top performers have been Korea, Hong Kong, Chinese Taipei (formerly known as Taiwan) and Japan (mostly in this order) with Flemish Belgium trailing somewhat behind, but consistently next in line.

The U.S. scores are also worth mentioning. Roughly put, American fourth graders and eighth graders scored somewhat above the international average. But at the twelfth grade, the U.S. scored at the bottom of the industrialized world, and only significantly out-performed two countries: South Africa and Cyprus. No other country fell so far so fast. There was also a more sophisticated twelfth grade test that was reserved for twelfth graders in advanced math programs in the participating countries. On that test, the U.S. was next-to-last; even Cyprus performed significantly better.

For completeness, it should be noted that the twelfth grade testing has not been repeated since 1995 and the U.S. plummet just described. This is unfortunate because the lack of follow-up testing forces us to infer whether the American mathematics programs have recovered from the results documented in 1995. Moreover, the real purpose of a K-8 program is to prepare students for subsequent study as opposed to an eighth grade TIMSS test. So our understanding of mathematics education around the world would be greatly enhanced by a schedule of testing that includes grade twelve as well as grades four and eight.[4]

In view of the absence of follow-up twelfth grade testing, one could speculate that the American TIMSS scores might show that the newest programs are beginning to make a difference. After all, the latest math reforms are often introduced at the earlier grades first, and then extended by one grade level per year. Could it be that U.S. high school students are performing better now because more of them are participating in reform math programs? The answer seems to be a clear no. A variety of studies[5] have documented very little progress in high school math achievement over the last decade. To date, the NAEP scores, for example, have been most notable for their lack of improvement.

In short, TIMSS testing shows that the US, and indeed most of the world have K-12 mathematics programs that are nowhere near the quality of the best programs worldwide. These results constitute a compelling argument for continued testing on an international scale. Simply stated, TIMSS is one of our best mechanisms for identifying unforeseen weaknesses in national programs, and for discovering exemplary programs that can be investigated in an effort to improve domestic teaching.

The other finding that has generated enormous impact can be traced to "TIMSS Videotape Classroom Study: Methods and Findings from an Exploratory Research Project on Eighth-Grade Mathematics Instruction in Germany, Japan, and the United States" [31]. For convenience, we condense the TIMSS Videotape Classroom Study's name to TVCS.

---

[4] For countries such as Singapore, which do not have a twelfth grade, the testing might well be given at the completion of the secondary education system.

[5] See, for example, Too Little Too Late: American High Schools in an International Context by William H. Schmidt. In *Brookings Education Policy papers 2003* (ed. by Diane Ravitch), pp. 253–277.

The Videotape Classroom Study documentation. During 1994–95, the TVCS team recorded 231 eighth-grade mathematics lessons in Germany, Japan and the U.S. The TVCS project report by Stigler et al. [31] contains an extensive analysis of these tapes and a description of the data acquisition and analysis methodologies. Stigler and James Hiebert subsequently conducted a joint study of Japanese training in pedagogy, which has strong cultural traditions that are surprisingly different from the programs of teacher development in the U.S. [30]. In 1999, Hiebert and Stigler began a second TIMSS videotape classroom study [11] that covered a broader selection of higher performing countries.

These videotape study projects produced a variety of supporting documentation [34], [35], [36], [14], [12], but the follow-up study did not record a new series of Japanese lessons and instead relied on the earlier tapings. We cover the main findings from the second study and the differences in its methodology and conclusions (which may well have resulted from criticisms of the earlier project), but will focus primarily on the 1995 TVCS, which remains the far more influential of the two publications.

The 1995 project produced a publicly available videotape [34] that begins with Stigler presenting an overview of the Japanese lessons that is very similar to the description already quoted from the Glenn Commission Report. It then shows carefully selected representative excerpts of the geometry and algebra lessons recorded in Germany, Japan, and the U.S. The German and American lesson samples were produced in addition to the original 231 recordings, which are not in the public domain due to confidentiality agreements. The Japanese excerpts were selected from the original 50 tapings recorded in Japan, and disclosure permissions were obtained after the fact.

The TIMSS videotape kit includes a guide to the excerpts [36] and a CD ROM [35] is available with the same excerpts, but without Stigler's introduction.

3. What the Japanese video excerpts show

Geometry. The tape shows the Japanese geometry lesson beginning with the teacher asking what was studied the previous day. After working to extract a somewhat meaningful answer from the class, he himself gives a summary: Any two triangles with a common base (such as AB in Figure 1) and with opposing vertices on a line parallel to the base (such as the line through D, C and P) have the same area because the lengths of their bases are equal, and their altitudes are equal. The teacher states this principle and uses his computer graphics system to demonstrate its potential application by moving vertex P along the line CD. The demonstration

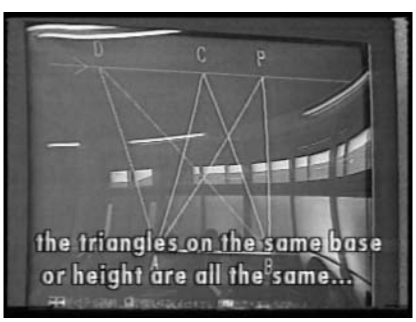


Figure 1

shows how to deform triangle ABP in a way that preserves its area. Next, he explains that this principle or method is to be the "*foundation* [36, p. 136]" for the forthcoming problem, which he then presents. It is the following.

Eda and Azusa each own a piece of land that lies between the same pair of lines. Their common boundary is formed by a bent line segment as shown.

The problem is to change the bent line into a straight line segment that still divides the region into two pieces, each with the same area as before.

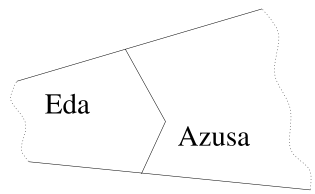


Figure 2

Despite the previous review, the problem is still going to be a challenge for eighth graders, and it is fair to infer that the teacher understands this very well. In geometry, one of the most difficult challenges in a construction or proof is determining where to put the auxiliary lines. These lines are needed to construct the angles, parallel lines, triangle(s), etc. that must be present before a geometry theorem or principle can be applied to solve the problem. For the exercise in Figure 2, the key step is to draw two crucial auxiliary lines. One defines the base of a triangle that must be transformed in a way that preserves its area. The other is parallel to this base, and runs through its opposing vertex.

So what should a master instructor do? The answer is on the tape.

After explaining the problem, the teacher asks the students to estimate where the solution line should go, and playfully places his pointer in various positions that begin in obviously incorrect locations and progress toward more plausible replacements for the bent line. Now here is the point. With the exception of two positions held for about one second (which come shortly after the frame shown in Figure 4), none of his trial placements approximate either of the two answers that are the only solutions any student will find.

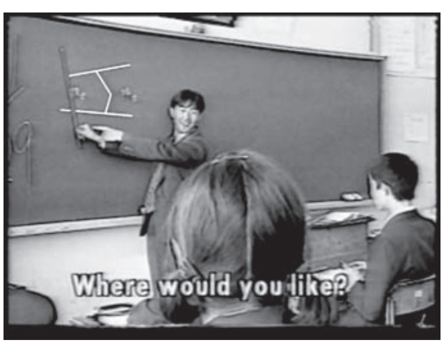


Figure 3

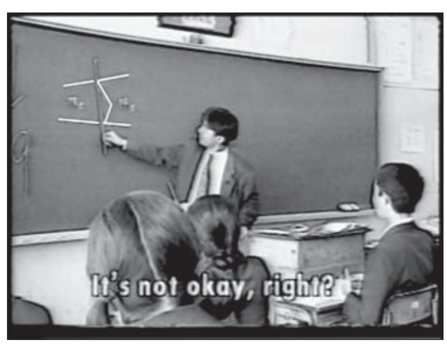


Figure 4

Rather, they are all suggestive of the orientation for the auxiliary lines that must be drawn before the basic method can be applied. He is giving subtle hints, and calling the students' attention to the very geometric features that must be noticed if the problem is to be solved. It is surely no accident that the teacher pauses with his pointer placed in two particular locations far longer than anywhere else. One of the locations is shown in Figure 4. The other is parallel to this placement, but located at the opposing vertex, which forms the bend in the boundary between Eda and Azusa.

Only after this telling warm-up – the heads-up review of the solution technique necessary to get the answer, and the casual discussion loaded with visual cues about what must be done – are the children allowed to tackle the problem.

But this is not the end of the lesson, and the students only get an announced and enforced three minutes to work individually in search of a solution.

As the children work, the teacher circulates among the students and gives hints, typically in the form of leading questions such as: "Would you make this the base? [The question is] that somewhere there are parallel lines, okay [36, p. 140]?"

He then allocates an additional 3 minutes where those who have figured out the solution discuss it with the other teacher. Weaker students are allowed to work in groups or to use previously prepared hint cards. The excerpt does not show what happens next. The TIMSS documentation [36] reports that students prepare explanations on the board (9 minutes).

Then a student presents his solution. The construction is clearly correct, and he starts out with a correct explanation. But when the time comes to demonstrate the solution, he gets lost and cannot see how to apply the area preserving transformation that solves the problem. The teacher then tells him to use "the red triangle" as the target destination.

Figure 5

The advice turns out to be insufficient, and the teacher *steps in* to redraw the triangle that solves the problem, and calls the student's attention to it with the words, "over here, over here." The student seems to understand and begins the explanation afresh. But he soon winds up saying, "Well I don't know what I am saying, but . . . " He then regains his confidence, and the presentation comes to an end without additional explanation.

A number of students say that they do not understand.

Then another student explains her answer, but the presentation is omitted from the tape. According to the Moderator's Guide [36, pp. 139–41], these two student presentations take altogether less than three minutes. Next, the teacher explains how to solve the problem. There are two equivalent answers that correspond to moving the middle vertex in Figure 1 to the left or right. Both directions solve the problem, and he shows this.

For completeness, we also show the two ways that the triangle transformation technique can be used to solve the problem. In order to make the connection between the review material and the challenge problem absolutely clear, the problem and its two answers have been rotated to present the same perspective as the triangle transformations in Figure 1, which began the day's lesson.

Evidently, no one devised an alternative solution method.

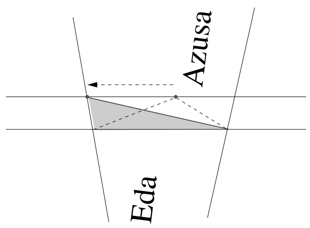


Figure 6

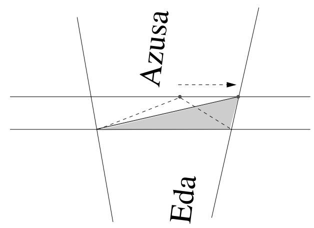


Figure 7

In his discussion of the solution, the teacher points out [36, p. 141] that this line straightening technique eliminates one of the two corners at the base of the triangle in Figures 6 and 7. This observation exposes a subtlety in that the corner that is eliminated is not the apex of the triangle, which is the point being moved to straighten out the line.

The lesson then continues with the teacher posing a new problem that can be solved with the same technique. This time the figure is a quadrilateral, and the exercise is to transform it into a triangle with the same area. At this point, the basic solution method should be within a student's reach, although the problem still requires a sound understanding of the basic method. There is also added difficulty due to the need to recognize that two consecutive sides of the quadrilateral should be viewed as representing the bent line of Figure 2, and that the other two sides should be extended as auxiliary lines to recast this new problem into a version of the Eda–Azusa exercise. The basic line straightening method can be applied so that any one of the four vertices can serve as the point where the line bends, and this designated vertex can be shifted in either of two directions to merge one of its two connecting sides with one of the auxiliary lines. The students again work individually for three minutes, and then are allowed to work in groups, use hint cards or ask the teacher.

Figure 8

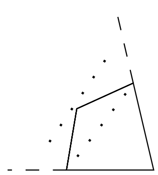


Figure 9

The TIMSS documentation indicates that this joint phase lasts for 20 minutes, and includes student presentations of their answers. There are apparently eight such presentations, which were selected to illustrate all eight ways the basic method can be applied: there are four vertices that can each be moved two ways. Then the teacher analyzes these eight ways in greater depth, and explains how they all use the same idea. All students remain seated during this portion of the lesson, and he controls the discussion very carefully and does almost all of the speaking. For homework, the teacher asks the students to transform a five-sided polygon[6] into a triangle with the same area.

An analysis of the teaching and its content. This lesson is nothing less than a masterpiece of teaching, and the management of classroom time is remarkable. Although many students did not solve the first problem of the day, the assignment certainly succeeded in engaging the attention of everyone. The second problem was no give-away, but it gave students the chance to walk in the teacher's footsteps by applying the same ideas to turn a quadrilateral into a triangle. The teacher-led study of all possible solutions masked direct instruction and reinforcement practice in an interesting and enlightening problem space.

Evidently, no student ever developed a new mathematical method or principle that differed from the technique introduced at the beginning of the lesson. Altogether, the teacher showed how to apply the method 10 times. Yet the lesson is an excellent example of how to teach problem solving, because each successive problem required a complete understanding of the basic proof technique.

The homework assignment is yet another application of the same method, and gives everyone a chance to revisit the lesson of the day once more. It also hints at the use of induction.

It is also worth pointing out that this geometry lesson, which is a specific application of measure preserving transformations, has additional uses. It appears, for example, in Euclid's proof of the Pythagorean Theorem (cf. Book I Prop 47 of Euclid's Elements).[7] More advanced exercises of this type appear on national middle school mathematics competitions in China and regional high school entrance examinations in Japan. And it is not much of a stretch to suggest that measure preserving transformations lie at the heart of those mysterious changes of variables in the study of integration.

All in all, the lesson is a wonderful example of the importance of a deep understanding of fundamental mathematics.

Algebra. The Japanese algebra lesson begins with student-presented answers for each of the previous day's six homework problems [36, p. 114]. These activities, along with the accompanying classroom discussion are omitted from the excerpts.

Then the teacher presents a more challenging problem that uses the same basic calculation method that the students have been studying, but needs one common-sense extension. The problem is this.

[6] The problem probably should be restricted to convex figures; otherwise it includes irregular cases that are difficult to formalize. On the other hand, this concern is just a minor technicality that has no effect on the pedagogical value of the problem.

[7] In fact, the technique is central to Euclid's development of area in general, which is based on transforming any polygon into a square with the same area. And the natural extension of this problem became a question for the ages: how to square the circle.

There are two kinds of cakes for sale. They must be bought in integer multiples; you cannot buy a fraction of a cake. The most delicious cake costs 230 yen, and a less tasty one is available for 200 yen. You wish to purchase 10 cakes but only have 2,100 yen. The problem is to buy 10 cakes and have as many of the expensive cakes as possible while spending no more than 2,100 yen.

The reproduction of the six homework exercises as shown in the TIMSS Moderator's Guide [36, p. 114] confirms that the class was already experienced with the technical mechanics necessary to solve problems with inequalities. Evidently, prior lessons had also covered word problems and the translation of word problems into equations and inequalities. Indeed, the teacher introduces the problem with the remarks, "Today will be the final part of the sentence problems [36, p. 159]." Thus, it is fair to infer that the only difference between the cake problem and the material they had just reviewed is the requirement that the solution use whole numbers of cakes.

After making sure that the students understand the problem, he asks them to devise a way to solve it. They get an announced and enforced three minutes.

Next, the teacher solicits solution approaches from the students. A student volunteers that she tried all possibilities. Her approach was to try 10 cheap cakes, then 9 cheap ones and 1 expensive one, etc., until she had the best answer. However, she was unable to finish in the three minutes that the teacher allocated for the problem. The teacher emphasizes the point, and it will soon become clear that part of the lesson is to show that this unstructured approach is unsound.

He then briefly discusses another way to solve the problem. The approach, which is quite inventive, uses a notion of marginal cost. If we buy 10 of the most expensive cakes, we exceed our budget by 200 yen. Trading in an expensive cake for a cheaper cake gives a net savings of 30 yen. Evidently, seven cakes have to be traded in, which shows that the answer is three expensive cakes and seven cheaper ones. As the teacher expected [36, p. 164], no student solved the problem this way.

Then he calls on another student, who explains how she set up the problem as an inequality, solved it as an equality, and then rounded the number of expensive cakes down to the nearest lesser integer. As she explains the equation, he writes it on the board. Only a few students understand the explanation, and he asks for another explanation of the same process. In subsequent activities that are only summarized on the tape and in the Moderator's Guide, the teacher then passes out a worksheet and works through a detailed analysis of the solution for the class.

After the detailed presentation, another problem of the same type was assigned, but with larger numbers. The teacher's words are telling:

> "If you count one by one, you will be in an incredibly terrible situation. *In the same way that we just did the cake situation, set up an inequality equation by yourself and find out . . . [the answer].* Because finding the answers one by one is hard, I wonder if you see the numerous good points of setting up inequality equations . . . "

The students worked on the problem individually. After 11 minutes, the teacher went over the problem with the class. The class ended with the teacher summarizing the solution technique that constituted the lesson of the day.

The video excerpts contain no group-based problem solving in this algebra lesson, and the Moderator's Guide confirms that none of the class time included problem solving in groups.

An analysis of the teaching and its content. Students never developed new solution methods. In the algebra class, the students were given the opportunity to learn first-hand why ad hoc trial-and-error approaches (which are encouraged by some of the latest reform recommendations) do not work. Although the tape does not explicitly show how many students were able to solve the original cake problem in the allotted time, the student responses suggest that no more than five could have possibly succeeded. But the three minutes of struggle might well have served to make the lesson more purposeful.

From a mathematical perspective, the cake problem was designed to require a deep understanding of inequality problems and their solutions. Mathematicians would say that when we solve a problem, we find all of the answers. If the cake problem had allowed fractional purchases, and had simply required that altogether any mix of ten cakes be purchased for at most 2100 yen, then the algebraic formulation would read,

$$230x + 200(10 - x) \leq 2100,$$

where x is the number of expensive cakes purchased, and $10 - x$ is the number of the inexpensive ones. The problem would also require that x be non-negative, since you cannot buy negative quantities of cake. A little manipulation gives:

$$0 \leq x \leq \frac{10}{3}.$$

Now, the point is that every x in this interval is a solution to the simplified problem, and every solution to the problem is in this interval. So if we want a special answer, the interval $\left[0, \frac{10}{3}\right]$ is the place to look. If we want the largest x, it is $\frac{10}{3}$. If we want the largest integer x, it is 3. And if we wanted the largest even integer, for example, we would look nowhere else but into $\left[0, \frac{10}{3}\right]$ to conclude that this answer is $x = 2$. Incidentally, a complete answer must also observe that the number of inexpensive items must be non-negative.

This problem variant is more than a matter of common sense; it exposes students to a deep understanding of solutions to inequalities and the implications of real world constraints. Moreover, the problem illustrates the idea of decomposing a complex exercise into a more basic problem whose solution can then be adapted to achieve the original objective.

Evidently, the video excerpts feature challenge problems that cover fundamental principles, techniques, and methods of systematic thought that lie at the heart of mathematics and problem solving. As such, they ought to provide experiences that build a powerful foundation of intuition and understanding for more advanced material yet to come. As a derivative benefit, these problems are so rich they can be readily transformed into follow-up exercises for use as reinforcement problems in class and as homework.

Both lesson excerpts exemplify a multi-round teaching and reinforcement pedagogy that begins with review of the fundamental (and systematic) principle that is the key to solving the challenge problem. The review is followed by two or three rounds (when homework is counted) that feature equivalent problems, often with additional educational content. Between each round, the teacher guides the students through the solution process to open the eyes of each learner to the basic idea, and to give the students yet another chance to apply the technique by themselves and to integrate the material into their own understanding – all in an engaging style without rote or tedium.

4. What can be deduced about Japanese teaching?

Many publications claim that the Japanese lessons teach students to invent solutions, develop methods and discover new principles. For example, this view is expressed in the Glenn Commission report [10, p. 4], and is clearly stated in TVCS as well: "[In Japan, the] problem ... comes first [and] ... the student has ... to invent his or her own solutions [31, p. vi]." In fact, TVCS reports that the 50 Japanese lessons averaged 1.7 student-presented alternative solution methods per class [31, Figure 22, p. 55]. Yet the excerpts exhibit no signs of such activity. They contain just one student-devised solution alternative, and it failed to produce an answer.

These differences are fundamental, and they should be reconciled. Part of the problem is that students are unlikely to devise their own solutions when the time is limited, the problems are so difficult that hints are needed, and the exercises are (clearly) designed to teach the value and use of specific techniques. Students would presumably have a better chance of finding alternative solution methods for less challenging exercises. And they would have an even better chance with problems that can be solved by a variety of methods that have already been taught. Examples might include geometry problems where different basic theorems can be used, and studies of auxiliary lines where the exercises are designed so that different auxiliary lines build different structures that have already been studied. TVCS illustrates alternative solution methods with the U.S. assignment to solve $x^2 + 43x - 43 = 0$ by completing the square and by applying the quadratic formula [31, p. 97]. Of course, this problem directed students to use different methods they already knew. The example contains no hint of any discovery.

So the question remains: where are the alternative solution methods, and when do they demonstrate signs of student-discovery?

The answers are in TVCS. It presents the actual examples that were used to train the data analysts who counted the "Student Generated Alternative Solution Methods"

(SGSM1, SGSM2, ...) in each lesson. The training lessons, it turns out, were the Japanese excerpts that we have just analyzed. The two student presentations for the Eda–Azusa problem are coded as SGSM1 and SGSM2 [31, p. 26–27]. Similarly, the second problem, where each of four vertices could be moved in two directions, has the codings SGSM1–SGSM8. *Altogether, this lesson is counted as having 10 student-generated alternative solution methods, even though it contains no student-discovered methods whatsoever*. And the failed try-all-possibilities approach in the Japanese algebra excerpt is counted as yet another student-discovered solution method. (See also "Teacher and Students Presenting Alternative Solution Methods [36, pp. 161–163].")

TVCS also contains a partial explanation for the source of these judgments. It reports that the data coding and interpretation procedures were developed by four doctoral students – none of whom were in mathematics programs [31, p. 24]. Moreover, TVCS states that the project's supporting mathematicians only saw coder-generated lesson tables, and were denied access to the actual tapes [31, p. 31]. It is reasonable to infer, therefore, that they did not participate in the design of these coding practices. As for the question of invention, TVCS explains: "When seat-work is followed by students sharing alternative solution methods, this generally indicates that students were to invent their own solutions to the problem [31, p. 100]." Altogether, there appears to have been a sequence of misinterpretations that counted student presentations as alternative solution methods, which became student-generated, and then invented and which ultimately evolved into invented discoveries that might even depend on new principles the students had not yet learned ([31], [25], [10]).

On the other hand, the contributions by the Japanese teachers received much less generous recognition. Yet in the defining examples of student discovery, the teachers – not the students – manage the ideas and lead the education process.

Additional statistics from the TIMSS projects. It is worth reiterating that in the sample Japanese lessons, students began working individually – and not in groups – on each of the four representative exercises. Similarly, the Stigler–Hiebert analysis [30, p. 79] states that "Students rarely work in small groups to solve problems until they have worked first by themselves." TVCS contains no comparable statement, and even implies otherwise: "[After the problem is posed, the Japanese] students are then asked to work on the problem ... sometimes individually and sometimes in groups [31, p. 134]." However, not one of the 86 figures and bar charts documents instances where problems began with students working in groups. Chart 41 [31, p. 78] indicates that of the seat-work time spent on problem solving, 67.2% of the time comprised individual effort and 32.8% of the time was spent in group-work.

Another TIMSS study addressed this issue in the statistics it gathered for a carefully balanced sampling of 3750 or so eighth graders from each participating country. One of its questionnaires asked teachers about their classroom organization and whether most of their lessons included students working in small groups, individually, as a class, etc. The results, which were weighted by the number of students in each responding teacher's class, are reproduced below for the U.S. and Japan [3, pp. 154–155].

Country	Percent of Students Whose Teachers Report Using Each Organizational Approach "Most or Every Lesson"					
	Work Together as a Class with Students Responding to One Another	Work Together as a Class with Teacher Teaching the Whole Class	Work Individually with Assistance from Teacher	Work Individually without Assistance from Teacher	Work in Pairs or Small Groups with Assistance from Teacher	Work in Pairs or Small Groups without Assistance from Teacher
Japan	22	78	27	15	7	1
United States	r 22	r 49	r 50	r 19	r 26	r 12

An "r" indicates teacher response data available for 70–84% of students.

Figure 10

The table shows that Japanese lessons do not have significant numbers of small-group activities. In fact, American classes evidently contain about 4 times as many such lessons. Of course, it should be noted that the data is based on questionnaires and depends, therefore, on the judgment of each respondent. The meaning of "most or every lesson" might have cultural biases, as might the definitions of "small groups" and "teacher assistance." Still, these TIMSS statistics support the notion that the Japanese style of teaching is substantially different from many of the U.S. reform practices.

Placing Japanese teaching in the context of U.S. reform. The video excerpts show Japanese lessons with a far richer content than the corresponding offerings from the U.S. and Germany. TVCS reports that the eighth-grade lessons recorded in Japan, Germany, and the U.S. covered material at the respective grade levels 9.1, 8.7, and 7.4 by international standards [31, p. 44]. We suspect that the interactive nature of the teaching style, the coherent, concept-based exercises with disguised reinforcement problems, the motivated direct instruction, and the deep understanding of the teachers all contribute to the quality of the Japanese curriculum.

Additional analysis shows that 53% of the Japanese lessons used proof-based reasoning, whereas the comparable statistic for the US lessons – which included both traditional and reform programs – stood at zero [31, p. vii]. And comparisons evaluating the development of concepts – including their depth and applicability – and the overall coherence of the material likewise judged the Japanese programs to be vastly superior [30, p. 59]. By all evidence, the use of proof-based reasoning as reported in Japan is not at all representative of the reform programs in the U.S., and the use of such remarkably challenging problems is beyond the scope of any American program past or present.

When comparing U.S. reform practices and Japanese teaching methods, TVCS offers somewhat guarded conclusions that are sometimes difficult to interpret:

> "Japanese teachers, in certain respects, come closer to implementing the spirit of current ideas advanced by U.S. reformers than do U.S. teachers. For example, Japanese lessons include high-level mathematics, a clear

> focus on thinking and problem solving, and an emphasis on students deriving alternative solution methods and explaining their thinking. In other respects, though, Japanese lessons do not follow such reform guidelines. They include more lecturing and demonstration than even the more traditional U.S. lessons [a practice frowned upon by reformers], and [contrary to specific recommendations made in the NCTM Professional Standards for Teaching Mathematics] [8]we never observed calculators being used in a Japanese classroom [31, p. vii]."

Subsequent elaboration on the similarities between U.S. reform and Japanese pedagogy recapitulates these ideas in the context of various reform goals, but again offers no statistical evidence to compare with the data accumulated from the analysis of Japanese teaching practices [31, pp. 122–124]. Consequently, it is difficult – absent additional context – to compare these reform notions in terms of mathematical coherence, depth, international grade level, or the preparation of students for more advanced studies and challenging problems. And no matter what "the spirit of current reform ideas" may mean, it is clear that Japanese and U.S. reform pedagogies differ in their management of classroom time, their use of proof-based reasoning, their tradeoffs between student-discovery and the use of direct instruction, as well as their use of individual and small group activities.

For completeness, we note that TVCS makes a distinction between the idealized goals as prescribed in the NCTM Professional Standards for Teaching Mathematics, and as embodied in actual classroom practices of some reform programs. In particular, TVCS discusses two reform-style lessons. One involved students playing a game that was purported by the teacher as being NCTM compliant, but happens to have very little mathematics content: "It is clear to us that the features this teacher uses to define high quality instruction can occur in the absence of deep mathematical engagement on the part of the students [31, p. 129]." The other lesson was deemed compliant with the spirit of NCTM reforms. It began with the teacher whirling an airplane around on a string. The eighth graders then spent the period working in supervised groups to determine the speed of the plane, and came to realize that the key issues were the number of revolutions per second, and the circumference of the plane's circular trajectory. The problem also required a realization that units conversions would be needed to state the speed in miles per hour. The problem engaged the class, and a variant to compute the speed of a bird sitting on the midpoint of the string was evidently a challenge. The homework for this math class was a writing assignment: *the students were asked to describe the problem, to summarize their group's approach, and to write about the role they played in the group's work* [31, p. 127]. TVCS did not evaluate this lesson or the homework in terms of international grade level or its coherence within a curriculum.

[8]The bracketed additions are elaborations from page 123 of TVCS, where the discussion of calculator usage is reworded and thereby avoids the slight grammatical misconstruction we have caused with the unedited in-place insertion.

Other characterizations of Japanese classroom practices. Studies that use human interaction as a primary source of data must rely on large numbers of interpretations to transform raw, complex, occasionally ambiguous, and even seemingly inconsistent behavior into meaningful evidence. Given the complexity of the lessons, it is not surprising that different interpretations should arise. TVCS – to its credit – documents an overview of these decision-making procedures, although the actual applications were far too numerous to publish. Moreover, TVCS actually contains widely diverse observations, ideas, and conclusions that sometimes get just occasional mention, and that are necessarily excluded from the Executive Summary. Understandably, this commentary is also missing – along with any supporting context – from the one-sentence to one-paragraph condensations in derivative policy papers (cf. [25], [10]). Perhaps the seventh and eighth words in the opening line of the TVCS Executive Summary explain this issue as succinctly as possible: "preliminary findings [31, p. v]." It is now appropriate to explore these larger-picture observations and to place them within the context of actual lessons.

TVCS even offers some support for our own observations:

> "[Japanese] students are given support and direction through the class discussion of the problem when it is posed (figure 50), through the summary explanations by the teacher (figure 47) after methods have been presented, through comments by the teacher that connect the current task with what students have studied in previous lessons or earlier in the same lesson (figure 80), and through the availability of a variety of mathematical materials and tools (figure 53) [31, p. 134]."

Unfortunately, these insights are located far from the referenced figures and the explanations that accompany them. The words are effectively lost among the suggestions to the contrary that dominate the report. It is also fair to suggest that the wording is too vague to offer any inkling of how powerful the "support and direction through class discussion" really was. Similarly, the value of the connections to previous lessons is left unexplored. This discussion does not even reveal whether these connections were made before students began working on the challenge problems, or after. For these questions, the video excerpts provide resounding answers: the students received masterful instruction.

The Math Content Group analyzed a representative collection of 30 classroom lesson tables. Their assessments, as sampled in TVCS, agree with our overall observations, apart from the use of hints, which were mostly omitted from the tables. These analyses are highly stylized with abstract representations for use in statistical processing and were, presumably, not intended to be a reference for the actual teaching.[9]

[9]For example, the analysis of the excerpted geometry lesson consists of a directed graph with three nodes, two links and nine attributes. The first node represents the basic principle (attribute PPD) illustrated in Figure 1. The node's link has the attributes NR (Necessary Result) and C+ (Increased Complexity). It points to a node representing the first challenge exercise. The representations were used to get a statistical sense of various

Another sentence in TVCS begins with teachers helping students, but ends with students inventing methods.

> "The teacher takes an active role in posing problems and helping students examine the advantages of different solution methods [*however, rather than elaborating on how this takes place, the sentence changes direction with the words*], but the students are expected to struggle with the mathematical problems and invent their own methods. [31, p. 136]."

This interpretation of student work as inventive discovery appears throughout TVCS. In its analysis of the excerpted Japanese geometry lesson, TVCS categorizes the teacher's review of the basic solution method (shown in Figure 1) as "APPLYING CONCEPTS IN NEW SITUATION [31, Figure 63, p. 101]," but inexplicably switches tracks to count the student applications as invented student-generated alternative solution methods. Another such instance reads, "students will struggle because they have not already acquired a procedure to solve the problem [31, p. 135]." Similarly, TVCS never explains how teachers participate in the problem solving by teaching the use of methods and by supplying hints. Its only discussion about hinting is to acknowledge the offer of previously prepared hint cards [31, pp. 26–30]. And by the time the Glenn Commission finished its brief encapsulation of student progress, even the struggle had disappeared along with proper mention of extensive teacher-based assistance.

5. The matter of pedagogy

Having sequenced through the Japanese lesson excerpts to determine exactly what took place in the classrooms, we now compare these applied teaching practices with current reform principles.[10] One of the most important differences between these two approaches to teaching concerns discovery-based learning. As with any idealized theory, the real issue is how well it works in practice. Discovery-based lessons can make sense – in moderation – provided suitable safeguards are in place. In particular:

- Judgments must resolve how much time is needed for students to discover the mathematics, and the necessary tradeoffs among time for guided discovery, time for additional (or deeper) lessons, and time for practice.
- There must be detection/correction mechanisms for incomplete "discoveries".
- There must be allowances for the fact that in even the best of circumstances, only a few students will succeed in discovering non-trivial mathematical principles.

The lesson excerpts reveal a teaching style that is surprising and very different from the U.S reforms – in theory and practice. In the Japanese classes, the time allotted for the first round of grappling with problems is remarkably modest. Consequently, the

broad-brush characteristics of the lessons [31, pp. 58–69].

[10] See [2] for an enlightening albeit jargon heavy exposition on the differences between the theories of learning advanced by educators and by cognitive psychologists.

remaining time is sufficient for teacher-assisted student presentations to help identify conceptual weaknesses, and for direct instruction to present new insights, as well as for follow-up problems designed to solidify understanding. Due to the time limitations and the difficulty of the more challenging problems, many students will be learning via a model of "grappling and telling." That is, most students will struggle with a tough problem in class, but not find a solution. They will then learn by being told how to solve it, and will benefit by contrasting their unsuccessful approaches against methods that work [27]. There is no question that preliminarily grappling with a problem is both motivational and educational (cf. [4, p. 11] and [27]). And discussions to understand why some approaches fail, to understand why a solution might be incomplete, and to explore alternative problem solving techniques are all sound investments of class time. However, the use of grappling and telling raises the implementation question:

Who should do the telling?

In some teaching practices, the theory of discovery-based learning is extended to include the notion of *cooperative learning*, which holds that the students should teach one another because they "understand" each other. In contrast, the TIMSS videotape and the data in Figure 10 show that Japanese teaching is by no means purely or principally based on cooperative learning. Although students get to explain their solutions, the video excerpts show that Japanese teachers are by no means passive participants. Student explanations frequently need – and get – supervision, and students can be remarkably incoherent (cf. Figure 5) even when their solutions are absolutely perfect. When all is said and done, the teachers do the teaching – and the most important telling – but in an interactive style that is highly engaging and remarkably skillful.

Stigler and Hiebert report that the lessons do not adhere to a fixed organization. Some lessons feature more direct instruction or extended demonstrations, whereas others demand that the students memorize basic facts [30, pp.48–51]. Students might even be asked to memorize a mandate to think logically [30, p. 49].

Aharoni's article on experimental math programs in Israel deserves mention in this context. In the late 1970s, Israel developed a unique and nearly unrecognizable adaptation of the 1960s New Math, which is still in use to this day. The curriculum has been controversial; Israel had placed first on the original 1964 precursor to the TIMSS exams, and had fallen to 28th place on TIMSS 1999. Of course, this small country has experienced demographic shifts and many other sources of instability, so this drop in rank is by no means proof that the curriculum has failed, but there were other concerns about the program, and the TIMSS results gave little reason to believe that all was well.

Israel was just months away from adopting the latest U.S. reform standards when circumstances led to a reconsideration and the decision to test a program based on translations of the Singapore textbooks (from English). Aharoni is participating in this experiment, and writes about his experiences with these textbooks.

He argues that teachers must have a deep knowledge of fundamental mathematics if they are to instill a sound understanding of elementary arithmetic. His first-grade teaching uses deep insights to provide a purposeful understanding of the most basic arithmetic operations. For example, he guides first-graders through story problems designed to open their eyes to the many different ways that a single operation – such as subtraction – can be used in the modeling of problems so that all students will enter the higher grades with the intuition and core knowledge necessary to master the translation of word problems into the native language of algebra. Only time will tell if the program is successful, but if so, his observations would have implications about best practices and teacher training.

This perspective places high demands on teachers and – by extension – on schools of education. Currently, most education programs allocate modest resources for courses on mathematics content, and very few programs are prepared to offer the kind of deep applied understanding that Aharoni describes. Instead, schools of education typically emphasize courses on developmental psychology, learning theory, and related topics such as *authentic assessment*, which is a grading practice based on portfolios of student work such as a study of how ancient Greek geometry was used 2000 years ago, or on real-life applications of periodicity – as opposed to exams. Similarly, very few mathematics departments feature course offerings on deep knowledge for K-12 instruction. This problem is further compounded by the certainty that most education majors would not have attended K-12 programs where such deep understanding would have been taught.

A small, but highly respected and widely cited comparative study by Liping Ma gives additional insight into this problem. In her study, American and Chinese elementary school teachers were asked to compute $1\frac{3}{4} \div \frac{1}{2}$, and to give a physically meaningful problem where the answer is determined by this computation. In the U.S., only 43% of those questioned performed the calculation correctly, and just one of the 23 teachers provided a conceptually correct story problem. In China, all 76 teachers performed the calculation correctly, and 80% came up with correct story problems [17].

In contrast, Hiebert and Stigler came to very different conclusions about how best to foster world-class teaching. They began with the TVCS tapes and findings, and conducted new investigations into Japanese teaching traditions. Their findings are published in *The Teaching Gap: Best Ideas from the World's Teachers for Improving Education in the Classroom* [30]. According to the authors, "differences" such as "teaching techniques, . . . and [teaching] basic skills [versus teaching for] conceptual understanding . . . paled" in comparison to the differences they observed in the culture of teaching. In their view, the Japanese tradition of life-long reflection on how to teach, and the culture of teachers sharing these ideas among each other in a continuing process of professional development was more significant than any of these other issues, which comprise the entirety of the debate over education reform in the U.S. and elsewhere. That is, they opined that the Japanese practices of ongoing collaborative- and self-improvement were even more important than the current state of the Japanese art of teaching as well as the curriculum differences reported in their book.

However, in a follow-up videotape classroom study of teaching in Australia, the

Czech Republic, Hong Kong, Japan, the Netherlands, Switzerland, and the United States, Stigler and Hiebert came to different conclusions [11]. For this study, new data coding schemes were developed to replace those used in the 1995 TVCS. Two of the findings are particularly noteworthy. First, the new study does not mention student-invented or student-discovered solution methods, and instead of reporting an average of 1.7 student-presented solution alternatives per Japanese lesson, the new study reports that 17% of the Japanese problems featured presentations of alternative methods [11, p. 94], and that students had a choice of methods in 31% of the lessons. Second, the study found no unifying theme to explain why the stronger countries perform so well. According to the authors:

> "A broad conclusion that can be drawn from these results is that no single method of teaching eighth-grade mathematics was observed in all the relatively higher achieving countries participating in this study [12, p. 11]."

> "It was tempting for some people who were familiar with the 1995 study to draw the conclusion that the method of teaching mathematics seen in the Japanese videotapes was necessary for high achievement [11, p. 119]."

Evidently, this positional retreat (see also [11, p. 1]) must include Stigler, Hiebert, and the Glenn Commission, among others. And the fact that the follow-up videotape study did not report student-discovered mathematics suggests that the earlier finding of student discoveries was inaccurate.

These changes in understanding notwithstanding, the earlier TVCS and the follow-up *The Teaching Gap: Best Ideas from the World's Teachers for Improving Education in the Classroom* will almost certainly outlive the more recent Hiebert–Stigler classroom study. These earlier publications continue to make must-read lists on education, and continue to inspire calls for reforms based on their findings. For example, on November 21, 2005, a New York Times editorial titled "Why the United States Should Look to Japan for Better Schools" cited the Teaching Gap book, and issued a call to reconsider

> "*how* teachers are trained and *how* they teach what they teach" (emphasis added).

Not one word was spent on the importance of *what content* is taught, and *what* a teacher should know in depth [29].

6. Conclusions

Mathematicians often ask what they can do to help preserve the integrity of K-12 math programs. In 1999, a letter protesting the new textbooks was signed by more than 200 leading American mathematicians and scientists and was published in the Washington Post. It had some positive results, but failed to stop the latest reforms. A similar protest in Israel was successful – but just barely. In California, protests supported by grassroots parents organizations, mathematicians, scientists, concerned journalists, and politicians were able to secure a sound revision of the State K-12 math standards in 1997 – after more than five years of struggle.

In many countries, mathematics societies will probably be most effective by lobbying as a group and by seeking a role in the textbook adoptions and in overseeing the assessment programs. In the U.S., reform curricula have often been introduced in conjunction with new testing programs designed and even managed by the publishers of the newly adopted textbooks. This practice eliminates the opportunity to compare pre- and post-reform student achievement. And publishers seldom provide in-depth testing on the weakest aspects of their own programs.

It is also worth pointing out that program validation tests should cover an entire curriculum. Whereas achievement tests should concentrate on the most important material that can be covered in the allotted time, the testing of education programs should use sampling to achieve comprehensive coverage at a nominal marginal cost in the overall testing process. Needless to say, the oversight required for these assessment programs should be of the highest caliber.

Some tests use closely guarded questions. The secrecy allows the same questions to be used year after year to maintain consistency in the scoring. For example, one of the more widely cited validation studies relied mainly on a test that to the best of my knowledge has had only three of its questions appear in the literature. This achievement test was devised to align with the new math reforms, but is also reported to assess basic computational skills. It is given over a period of three days with the teachers retaining custody of the materials after school. So its questions are not really secret, and the administrative procedures lack safeguards to protect the integrity of the assessment program. Sometimes students were even allowed to rework questions from the previous day. Moreover, the test manufacturer does not require the test to be given with time limits, which are optional even for the testing of basic skills. The validation project reported year-by-year improvement of fourth-grade scores with the new reform program, but this progress was not matched by the scores for the more securely administered state testing of fifth graders.

In the U.S., the government-mandated No Child Left Behind (NCLB) testing (with state-determined tests) shows good progress for the majority of our states year by year, whereas the National Assessment of Educational Progress (NAEP) math testing shows that the net achievement of our twelfth graders has been unchanged nationwide for more than a decade. Something does not quite add up. The NAEP uses a mix of new and secret questions but is designed to be free of the biases that result from test-specific instruction and cramming. It is given to randomly selected schools, and the performance results are reported at the state level with additional results for subcategories based on gender and socio-economic status. Each student is given a randomly selected subset of test questions, and no performance results are released for students, schools, school districts, or education programs. Consequently, there is little incentive to teach to the test. The majority of the California achievement test questions are released and retired each year, and state law forbids the use of these materials in classroom preparation for forthcoming tests. There are programs in place to detect cheating, but it is not possible to know how effective they are, and students can always use these questions for practice independently of their school assignments. In New York, there are no such prohibitions, and many New York City schools use old tests routinely in required after-school preparation sessions held during the six weeks prior to the State and City testing.

But although the NAEP may be our most uncompromised testing program, it is far from perfect. The test is consensus-based, with an oversight committee that has limited authority and where only about 10% of its members are mathematicians. The web-released sample questions suggest that the twelfth grade test is probably at a sixth grade level, on average. A representative question on fractions might be to compute two-thirds of 12 marbles. Evidently, the NAEP Governing Board (NAGB) has not reached a consensus about the benefits of knowing if an American high school education enables seniors to evaluate, say, $1/6 - 1/9$, much less $2\frac{1}{9} - 4\frac{1}{6}$.

To date, just one of the released algebra problems is categorized as solving a system of equations. This twelfth grade multiple choice question reads:

What number, if placed in each box below, makes both equations true?

$4 \times \square = \square$ and $3 \times \square = \square$: A) 0 B) 1 C) 2 D) 3 E) 4

A "hard" problem reads:

For what value of x is $8^{12} = 16^x$? A) 3 B) 4 C) 8 D) 9 E) 12

Only 34% of our high school seniors found the correct answer even though calculators were available for use on this problem. The NAEP testing also asked students if they used a calculator for this question, but this data, unfortunately, does not appear to have been released on the web.

Needless to say, the TIMSS test questions and testing procedures, unlike many U.S. practices, stand out as a beacon of hope. But we must take care to ensure that all of the TIMSS analyses are well documented, are open to external review, and are as accurate as possible. And with so many challenges in the search for sound education reform, we may all be able to contribute somewhere in this complex of vital activities.

We close with the following summary assessments.

1. The undisciplined appeal to constructivist ideas has produced American programs that are more a betrayal of true constructivism than an advance of its principles. The result is an unprecedented reduction in the transmission of mathematical content.

2. The reform books and classroom curricula focus on examples, tricks, and experiments rather than fundamental mathematical principles, systematic methods, and deep understanding.

3. The justification for these "reforms" is based on mostly inaccurate interpretations of the best teaching practices in other countries. In particular, paradigmatic classroom examples from Japan have been misconstrued by researchers to suggest that students discover mathematical principles. In fact, the teacher conveys these

principles quite explicitly, albeit engagingly and through examples.

4. As a consequence of these misinterpretations, "exemplary" math lessons in the U.S. convey little content, take too much time, and can even lead to false "discoveries" of mathematical principles.

5. A proper understanding of best practices suggests that

i. teachers must be trained to understand, at a deep level, the mathematics they are teaching

ii. teachers should encourage individual work, but must ensure that important principles are conveyed in an orderly and cumulative manner.

6. Mathematicians, guided by proven programs such as those in Singapore, should be involved in determining the principles that are taught, the examples that help convey them, and the exercises that reinforce the net learning.

7. Mathematicians must play an active role in overseeing the quality of achievement tests in an effort to determine where our education programs are succeeding and where they are not.

Acknowledgments. The author is greatly indebted to Professor Michiko Kosaka for her tutoring on Japanese culture, and her verification of translated and mistranslated passages (cf. Figure 1) in the TIMSS Videotape Study documentation. It is also a pleasure to thank Clilly Castiglia and Kevin Feeley of the NYU Center for Advanced Technology and the Media Research Lab, who graciously provided the VHS frame processing.

Portions of this paper are adapted from *Testing Student Learning, Evaluating Teaching Effectiveness*, edited by W. M. Evers and H. J. Walberg [28].

References

[1] Aharoni, R., What I Learned in Elementary School. *Amer. Edctr.* **29** (3) (2005), 8–13.

[2] Anderson, J., Reder, L., Simon, H., Applications and Misapplications of Cognitive Psychology to Mathematics Education. *Texas Education Review* (Summer, 2000). http://act-r.psy.cmu.edu/people/ja/misapplied.html

[3] Beaton, A. E., et al., *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study*. International Association for the Evaluation of Educational Achievement (IEA), Chestnut Hill, MA, 1996.

[4] Bransford, J. D., et al., *How People Learn: Brain, Mind, Experience and School*. National Research Council, National Academy Press, Wash., DC, 2000.

[5] Cobb, P., Yackel, E., Wood, T., A constructivist alternative to the representational view of mind in mathematics education. *JMRE* **23** (1992), 2–33.

[6] Cook, C., The Significance of the NCTM Standards to the Pathways Critical Issues in Mathematics. North Central Regional Educational Laboratory, 1995. www.ncrel.org/sdrs/areas/issues/content/ctnareas/math/ma0/htm

[7] Coxford, A. F., et al., *Contemporary mathematics in context: a unified approach*. Core-Plus, Janson Publications, Dedham, MA, 1997.

[8] Fendel, D., et al., *Interactive Mathematics Program: Integrated High School Mathematics 1,2 3, 4*. Key Curriculum Press, Emeryville, CA, 1997.

[9] Garfunkel, S., et al., *Mathematics Modeling Our World, 1,2,3,4*. COMAP, Inc., W.H. Freeman and Co., New York, NY, 1998.

[10] Glenn, J. et al., *Before It's Too Late, A report to the Nation from the National Commission on Mathematics and Science Teaching for the 21st Century*. 2000. www.ed.gov/inits/Math/glenn/report.pdf

[11] Hiebert, J., et al., *Teaching Mathematics in Seven Countries: Results From the TIMSS 1999 Video Study*. National Center for Education Statistics (NCES), Wash., DC, 2003.

[12] Hiebert, J., et al., *Highlights From the TIMSS 1999 Video Study of Eighth-Grade Mathematics Teaching*. NCES, NCES 2003-011, Wash., DC, 2003.

[13] Hong, K. T., et al., *Primary Mathematics 5A, 5B, 6A, 6B*. Curriculum Planning & Development Division, Ministry of Education, Federal Publications, Singapore, 2000.

[14] Jacobs, J., et al., *Third International Mathematics and Science Study 1999 Video Study Technical Report Volume 1: Mathematics*. NCES, Wash., DC, 2003.

[15] Kamii, C., Warrington, M. A., Teaching Fractions: Fostering Children's Own Reasoning, Developing Mathematical Reasoning in Grades K-12. In *1999 Yearbook of the NCTM* (ed. by L. Stiff and F. Curcio). NCTM, Reston, VA 1999, 82–92.

[16] Lappan, G., et al., *Connected Mathematics 6, 7, 8*. Connected Mathematics Project, Prentice Hall, Upper Saddle River, NJ, 1999.

[17] Ma, L., *Knowing and Teaching Elementary Mathematics: Teachers' Understanding of Fundamental Mathematics in China and the United States*. Lawrence Erlbaum Associates, Mahwah, NJ, 1999.

[18] Martin, M. O., et al., *School Contexts for Learning and Instruction IEA's Third International Mathematics and Science Study*. IEA, Chestnut Hill, MA, 1999.

[19] Mullis, I. V. S., et al., *TIMSS 1999 International Mathematics Report Findings from the IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. IEA, Boston College, Chestnut Hill, MA, 2000.

[20] NCTM, *Curriculum and evaluation standards for school mathematics*. NCTM, Reston, VA, 1989.

[21] NCTM, *Professional Standards for Teaching Mathematics*. NCTM, Reston, VA, 1991.

[22] NCTM, *Assessment Standards for School Mathematics*. NCTM, Reston, VA, 1995.

[23] Ocken, S., personal communication.

[24] Public Broadcasting Service, *Stressed to the Breaking Point*. High School Math Project – Focus on Algebra, TV show, MathLine series, J. Peters, Producer, 1997. See also the lesson plan at www.pbs.org/teachersource/mathline/lessonplans/pdf/hsmp/stressedtobreaking.pdf.

[25] Peak, L., et al., *Pursuing Excellence: A Study of U.S. Eighth-Grade Mathematics and Science Teaching, Learning, Curriculum, and Achievement in International Context*. NCES, NCES 97-198, Wash., DC, 1996.

[26] Ralston, A., Let's Abolish Pencil-and-Paper Arithmetic. *Journal of Computers in Mathematics and Science Teaching* **18** (2) (1999), 173–194.

[27] Schwartz, D. L., Bransford, J. D., A time for telling. *Cognition and Instruction* **16** (4) (1998), 475–522.

[28] Siegel, A., Telling Lessons from the TIMSS Videotape. In *Testing Student Learning, Evaluating Teaching Effectiveness* (ed. by W. M. Evers and H. J. Walberg), Hoover Press, Stanford, CA, 2004, 161–194.

[29] Staples, B., Why the United States Should Look to Japan for Better Schools, Editorial Observer, *New York Times*, Section A, Page 22, Column 1, Nov. 21, 2005.

[30] Stigler, J. W., Hiebert, J., *The Teaching Gap: Best Ideas from the World's Teachers for Improving Education in the Classroom*. Free Press, New York, NY, 1999.

[31] Stigler, J. W., et al., *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, NCES 99-074, Wash., DC, 1999.

[32] Stotsky, S., ed., *What's at Stake in the K-12 Standards Wars A Primer for Educational Policy Makers*. Peter Lang, New York, NY, 2000.

[33] TERC, *Investigations in Number, Data, and Space*. K-5 book series, Dale Seymour Publications, Parsippany, NJ, 1998.

[34] U.S. Dept. of Education, *Eighth-Grade Mathematics Lessons: United States, Japan, and Germany*. Video Tape, U.S. Dept. of Education, NCES, Wash., DC, 1997.

[35] U.S. Dept. of Education, *Eighth-Grade Mathematics Lessons: United States, Japan, and Germany*. CD ROM, U.S. Dept. of Education, NCES, Wash., DC, 1998.

[36] U.S. Dept. of Education, *Moderator's Guide to Eighth-Grade Mathematics Lessons: United States, Japan, and Germany*. U.S. Dept. of Education, Wash., DC, 1997.

[37] Usiskin, Z., et al., *Geometry*. University of Chicago School mathematics Project, Prentice Hall, Upper Saddle River, NJ, 1999.

Courant Institute, NYU, 251 Mercer St., New York, NY 10012, U.S.A.
E-mail: siegel@cims.nyu.edu