A Practical and Theoretical Agenda for Progress in Mathematics Education

Alan H. Schoenfeld
University of California
Berkeley, CA, USA
Alans@Berkeley.edu

July 28, 2011

A Practical and Theoretical Agenda for Progress in Mathematics Education[1]

Executive Summary

This paper offers an action agenda for the improvement of mathematics education. After briefly describing the current and not very favorable situation, I focus on the following main goal:

> The goal for mathematics instruction should be *meaningful engagement* with *powerful mathematics* for *all children* – resulting in children's development of the ability to engage in sense-making in and with mathematics, a deeper understanding of mathematical ideas, the ability to use mathematical ideas productively in solving problems, and a more positive view both of mathematics and of themselves as sense-makers in mathematics.

The benefits of attaining this goal are obvious. The result would be a more quantitatively literate society better suited than the current population to live in an increasingly technological world and to contribute to its advancement. A series of conditions are necessary to attain this goal:

1. Mathematically rich content and process standards, which have at their core the notion of mathematics as a sense-making activity.
2. Curriculum development and refinement consistent with these standards.
3. Assessments that are consistent with these standards.
4. Professional development for teachers consistent with these standards – and the opportunity for teachers to develop the understandings necessary for "teaching for understanding."
5. Consistency and stability, to allow for steady improvement in the system.
6. A solid body of research to understand and facilitate items 1 through 5, and to provide a solid basis for continuing progress.

These six items constitute an agenda for action. Although simply stated, they are far from easy to realize. In the final sections of this paper I describe some of the complexities one faces in trying to make them a reality.

---

[1] This paper is derived from my opening plenary presentation "A practical and theoretical agenda for the future of mathematics education in Europe, which was given at a conference, *The Future of Mathematics Education in Europe*, promoted by the Acadaemia Europaea in the framework of the Portuguese Presidency of the European Union, December 16-18, 2007.

A Practical and Theoretical Agenda for Progress in Mathematics Education

Preliminary comments

My assigned task for this paper and the conference talk from which it was is derived, was to suggest the outline of a possible practical and theoretical agenda for the future of mathematics education in Europe. I was honored by this assignment, but humbled by it as well. I cannot pretend to understand the diverse European context, much less the context of mathematics education worldwide. I know enough to know that there are significantly different educational policies and practices across the globe, and that it would be foolish to believe that I have either a comprehensive or deep understanding of those policies and practices. But perhaps there is an advantage to seeing some things from a distance. The situations world-wide and in the United States may not be as different as would seem on the surface. In particular,

- The so-called "United" States are in many ways 50 semi-independent entities with widely varied policies, standards, and assessments, so there may be nearly as much diversity in the American Union as in the European Union;

- The U.S. has lived through:

  - the "democratizing" of higher education, with increasing collegiate enrollments;
  - falling mathematics enrollments at the same time;
  - "math wars;" and
  - less-than-desirable levels of teacher preparation and professionalism.

Thus, we have faced (and are facing) many of the challenges that currently face nations around the world. It may be possible to learn from our experience in the United States, and perhaps to profit from our mistakes.

The major goal of this paper is to describe a set of conditions that, if met, will result in a significant improvement of the quality of mathematics learning and instruction – in any jurisdiction that takes these conditions seriously. I begin with a discussion of the "image problem" faced by mathematics. I then proceed to the main substance of the paper. I outline a set of goals for mathematics instruction, and a set of necessary and perhaps sufficient conditions to attain those goals. Having outlined those conditions, I provide some evidence to justify my claims of their importance. Setting out to achieve these conditions, and thus improving mathematics education, is the agenda that I propose.

The Challenge

I begin with a simple observation. Mathematics has an image problem. Figure 1 is reproduced from the Wednesday, January 3, 2001 *Times of London*. Simultaneously light-hearted and serious, it captures a school child's view of mathematicians.

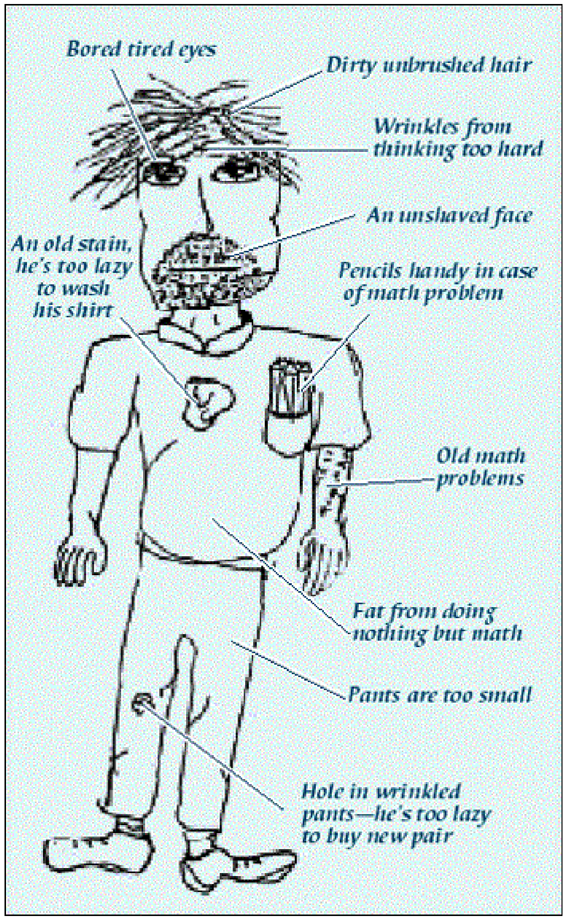


Figure 1.
"Unflattering portrait: how one child in the survey sees mathematicians"
Times of London, Wednesday, January 3, 2001

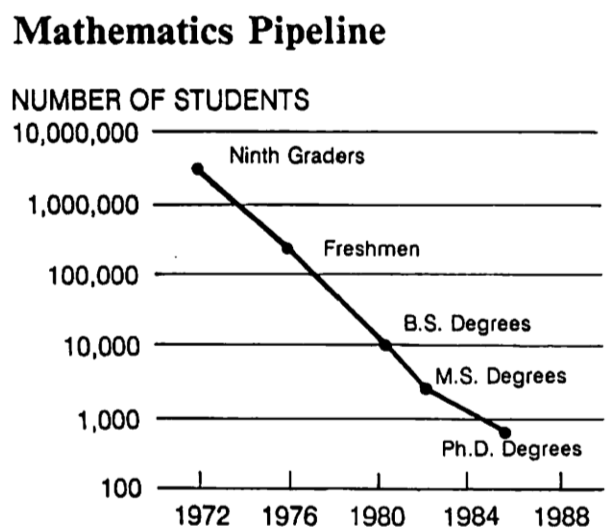
This perspective is widespread. For example, the "Yahoo hot jobs" section of the San Francisco Chronicle on December 9, 2007 has a lead, full-page article on "hot jobs" for mathematicians (Cadwell & Berman, 2007). The article itself is upbeat, in an attempt to sell mathematics to an unreceptive public. Yet, even in promoting mathematics, the article acknowledges the negative stereotype:

> Einstein was one of them. So was Euclid. But mathematicians in the Bay Area today don't have job titles like "E=MC^2 guy" or "father of geometry." On the other hand, math whizzes have plenty of opportunities to make differences outside of traditional academic settings. *So while the word*

> "*mathematician" might conjure up images of ratty tweed jackets and chalkboards festooned with arcane symbols*, the reality is that they are helping to design products, make hedge funds successful, cure cancer and even make cell phone calls more reliable. (emphasis added).

The challenge faced by the mathematical community is that the kind of attitude toward mathematics and mathematicians reflected in Figure 1 has serious consequences. In the U.S., and more generally in developed countries around the world, mathematics enrollments have dropped as alternative educational and career options have become available. We face a shortage of quantitatively trained personnel for the workplace; and equally important, we face populations that, in general, do not possess the kinds of quantitative literacy that are necessary for full participation in an increasingly technological society.

Figure 2 shows the year-to-year drop in mathematics enrollments for the cohort of ninth graders enrolled in mathematics courses in the U.S., starting in 1972. Roughly speaking, half of the students enrolled at any given year opted out of the mathematical pipeline.



Data from enrollments in mathematics courses reveal that roughly half the students leave the pipeline each year.

Figure 2.
Mathematics Pipeline. *Everybody Counts* (National Research Council, 1989, p. 6)

Since then, the population of the United States has become more diverse, with Latinos and African Americans constituting an increasing percentage of the total population. Unfortunately (see Figure 3), attrition rates from mathematics for these groups are even larger than those shown in Figure 2. African Americans and Latinos, who comprised 12% and 7% respectively of the 8th grade students enrolled in mathematics in 1990, comprise only 2% each of those who earned doctorates in mathematics[2].

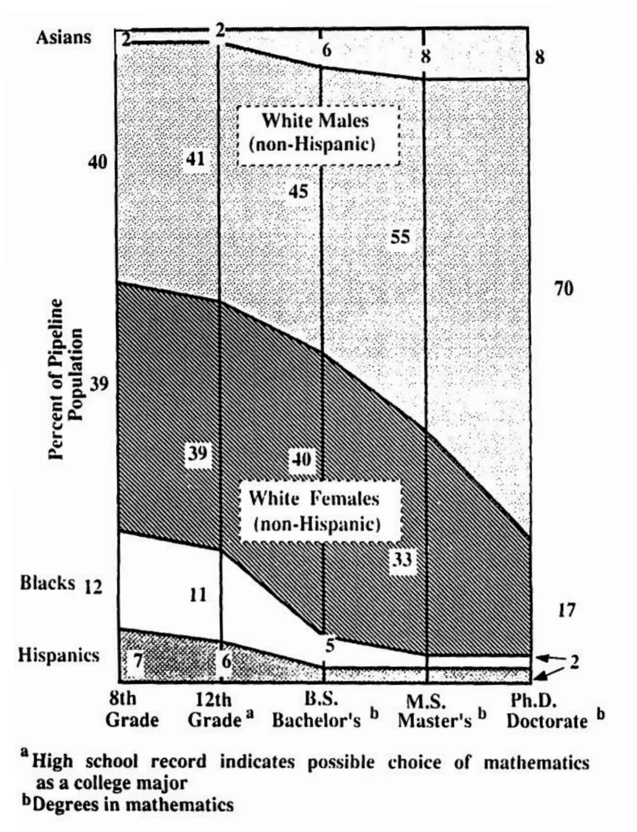


Figure 3.
Mathematics trajectories of various ethnic and gender groups. *A Challenge of Numbers* (National Research Council, 1990, p. 36)

We are, then, faced with somewhat of a contradiction. On the one hand, mathematics is a subject of extraordinary beauty and power. It was that beauty that attracted me to mathematics (my Ph.D. is in topology and measure theory) and, I suspect, that appealed to the majority of participants in the conference *The Future of Mathematics Education in Europe*. But that appreciation is not broadly shared. The vast majority of people world-wide dislike mathematics; indeed, the very existence of the term

[2] Although the data in Figures 2 and 3 are somewhat old, I use them for two reasons. First, they represent trends in American curricula immediately prior to the "reform" and "standards" movements (see below), which were catalyzed by the publication of the *Curriculum and Evaluation Standards for School Mathematics* in 1989 by the National Council of Teachers of Mathematics (NCTM). Thus they present an accurate picture of the impact of the "traditional" curriculum on enrollments. Second, the enrollment picture from the mid-1990s onward is difficult to assess. "Standards-based" materials, which slowly began to enter the marketplace in the mid-to-late 1990s, currently have perhaps 20-25% of the textbook market in the U.S. (Precise figures are difficult to obtain because publishers consider the data to be proprietary, and they are reluctant to share them.) In addition, the first versions of standards-based materials began to appear in the mid-1990s, so virtually no students have studied only those curricula in K-12; there are no reliable data on persistence in mathematics for students who studied from those curricula.

"mathematics anxiety" is evidence that many people not only dislike mathematics, but fear it.

The sad truth is that most people's dislike of mathematics – and their belief in the stereotype represented in Figure 1 – comes from their experience with mathematics in school. One of the beauties of mathematics is how it coheres: mathematical ideas fit together in ways that make sense, requiring very little by way of memorization. Few students experience it this way, however. For most students of my generation and the generation that followed it, the mathematics in school *was* dry and boring. School mathematics was largely focused on the implementation of procedures, with a great deal of rote memorization. Although students were told that mathematics is beautiful and powerful, they did not get to experience that power themselves. Mathematics instruction has always been the embodiment of delayed gratification – "you need to learn this year's mathematics in preparation for next year's mathematics, and when you've learned that, you'll be able to use it." Those few who made sense of the mathematics on their own, or who were fortunate enough to have teachers who helped them to learn the discipline in meaningful ways, found mathematics to be beautiful and exciting. The rest provided the data for figure 2. We have, in large measure, brought our current problems upon ourselves. And we must tackle those problems head on. The balance of this paper addresses how we might do so.

Goals for Mathematics Instruction

From 1997 through 2000, I was part of the team that worked to create *Principles and Standards for School Mathematics,* NCTM's (2000) successor to the 1989 NCTM *Standards*. The 1989 *Standards* had been created in response to a number of pressures: a general perception of the decline of American scientific and technical prowess (see, e.g., National Commission on Excellence in Education, 1983) and the need to reverse it; the problems with American curricula reflected in Figures 2 and 3; the wish to serve a broader set of mathematical goals, including the creation of a quantitatively literate citizenry; and, the fact that research on mathematical thinking and problem solving had expanded the field's conception of goals for mathematics learners. The *Standards* broke with tradition, specifying not only content goals but process goals as well: at all grade levels, instruction was to attend to problem solving, mathematical reasoning, mathematics as communication, and making mathematical connections. The U.S. National Science Foundation issued requests for proposals to develop curricula in line with the ideas in the *Standards*. The *Standards* themselves contained broad desiderata rather than detailed curriculum specifications. the result was a collection of curricula that varied substantially in style from each other, but were consistently different in style and content from the more traditional skills-oriented curricula. Ultimately, the new curricula would become controversial, and stimulate a backlash known as the "math wars" (see Schoenfeld, 2004). By 1997, it was clear that the *Standards* needed to be revised. *Principles and Standards* took into account the "lessons learned" from the original curriculum development and advances in both research and technology, in order to offer a newer and more elaborated statement of curricular goals.

For the past half-dozen years I have been part of a project called Diversity in Mathematics Education (DiME), whose goal has been, in part, to address the challenges reflected in both Figures 2 and 3 – to find ways to redress the inequities reflected in Figure 3, and, while maintaining high mathematical standards, to ameliorate the flight from mathematics reflected in Figure 2. The DiME team crafted the following framing, which I find helpful.

> **Goal**
> **The goal for mathematics instruction should be *meaningful engagement* with *powerful mathematics* for *all children* – resulting in children's development of the ability to engage in sense-making in and with mathematics, a deeper understanding of mathematical ideas, the ability to use mathematical ideas productively in solving problems, and a more positive view both of mathematics and of themselves as sense-makers in mathematics.**

I stress that all three of the italicized items in the goal statement are essential. Those who love mathematics know that mathematics is something you *do*. Whether one is exploring, conjecturing, proving, solving problems or applying ideas, one is actively engaged with mathematical ideas. Mathematics instruction should foster such engagement. That engagement should be with "big ideas," so that students can and understand see the patterns of powerful thinking that pervade mathematics. The broad range of thinking and problem solving skills (see, e.g., Schoenfeld, 1985) should be learned. And, I stress, such experiences should not be reserved for "the happy few" (cf. Stendhal), but for all students. This is not to say that one expects all students to become mathematicians – as in any field, some will excel while others do not – but that the opportunity for such engagement will attract a larger number of talented people to mathematics, and will serve both mathematics and society well by producing a more quantitatively literate and "mathematics-friendly" population.

Claims

I claim that substantial progress toward the goal of meaningful engagement with powerful mathematics for all children is possible when the five conditions highlighted in Table 1 are all in place:

1. Mathematically rich content and process standards, which have at their core the notion of mathematics as a sense-making activity;
2. Curricula aligned with those standards;
3. Assessments aligned with those standards;
4. Professional development consistent with these standards – and the opportunity for teachers to develop the understandings necessary for "teaching for understanding."
5. Consistency and stability, to allow for steady improvement in the system.

Table 1. Conditions for moving toward classrooms that provide meaningful engagement with powerful mathematics for all children.

Each of the conditions described in Table 1 is a challenge in its own right. I will address each of these challenges in what follows – but first, I want to make a plausibility case that if all of these conditions do obtain, there is a significant likelihood of progress (and conversely, that if any of them are not in place, long-term progress is unlikely).

Evidence of Plausibility, Part 1: The Pittsburgh Schools.

Pittsburgh, Pennsylvania, has a population of slightly more than 300,000 people. At the tail end of the 20th century, it had 97 schools (59 elementary, 19 middle, 11 secondary, and 8 other) with a school population of about 40,000 students. Of those, 56.4% were African American and 43.6% White or other ethnicity. 62.2% of the students qualified for free or reduced price lunches, an indication of low socioeconomic status.

Pittsburgh is unusual in that, for a short time, all five of the conditions discussed above were in place. The district's mathematics specialist, Diane Briars, had been in charge of professional development for some year. Briars, a past member of the NCTM Board of Directors, was firmly committed to the mathematical values embodied in the NCTM *Standards*. Her professional development sessions were consistent with those values, as were the assessments employed by the district, the New Standards assessments. The one missing element, until 1998, was curriculum: "standards-based" curricula were not widely accessible until that point, so teachers had been using a "traditional" curriculum that they tried to supplement with their own "standards-oriented" materials. But, as we all know, developing curricula on one's own (especially on top of a full-time job!) is hard!

In 1998 Pittsburgh adopted a new, standards-based elementary curriculum. In a sense, this was like the last piece of a jigsaw puzzle fitting into place: rather than developing their own materials, teachers could now rely on a coherently organized standards-based product. (And, because they had been "primed" for such texts, the impact could be felt almost immediately. Had they begun from scratch, there would have been a much longer period of accommodation, during which the other conditions fell into place.) The story is told in Figures 4, 5, and 6.

Figure 4 demonstrates one aspect of the impact of the new text adoption, across the whole district. With the new text in place, there was a substantial increase (from roughly 30% of the student population to more than 50%) in the percentage of students scoring proficient or better on "skills," and an approximate doubling of the % of students who scored proficient or better on concepts and problem solving. All these scores leave a great deal to be desired, but they do indicate the impact of the new curricula. Especially noteworthy, given the "math wars" in the U.S., is the fact that scores on skills rose dramatically with the new, "standards-based" curriculum. Traditionalists' fears have been that a decreased emphasis on skills in the newer curricula would result in a decrease in students' proficiency on basic mathematical skills, and these data indicate that, at least in this case, that fear was unwarranted. Also noteworthy, re condition 5 (stability) is that as teachers became more familiar with the curricula, the scores continued to rise.

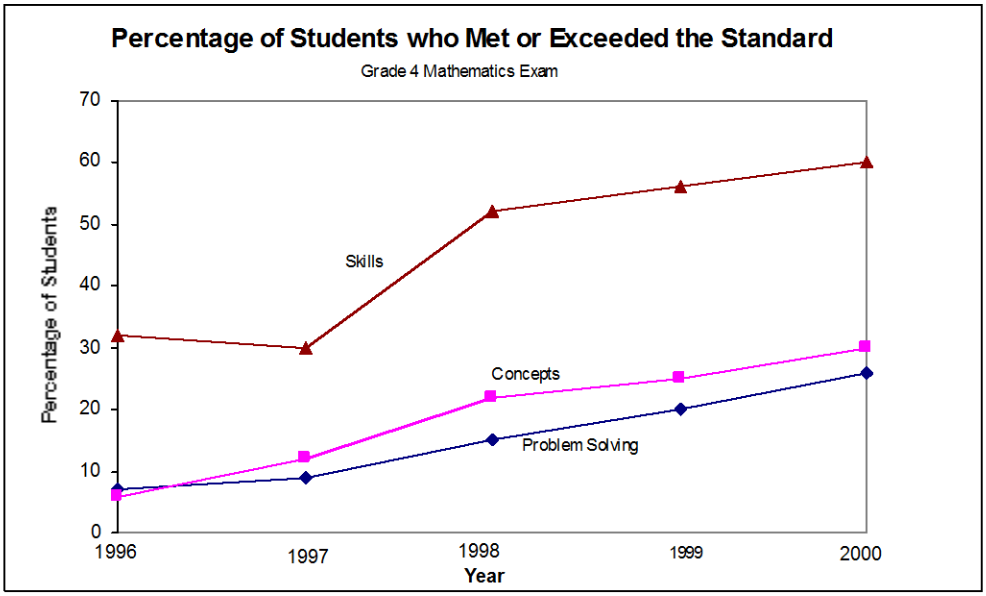


Figure 4.
Increases in proficiency following standards-based text adoption.
(Adapted from Briars, 2001)

One way of reading Figure 4 is to say that "more students did well." But, might it be the case that the curriculum only helped those students who were "close to" proficiency? What about the others? Figure 5 shows outcomes for students at the low end of the distribution. There was a significant decrease in the number of very low-performing students, an indication that the curriculum was effective with those students as well.

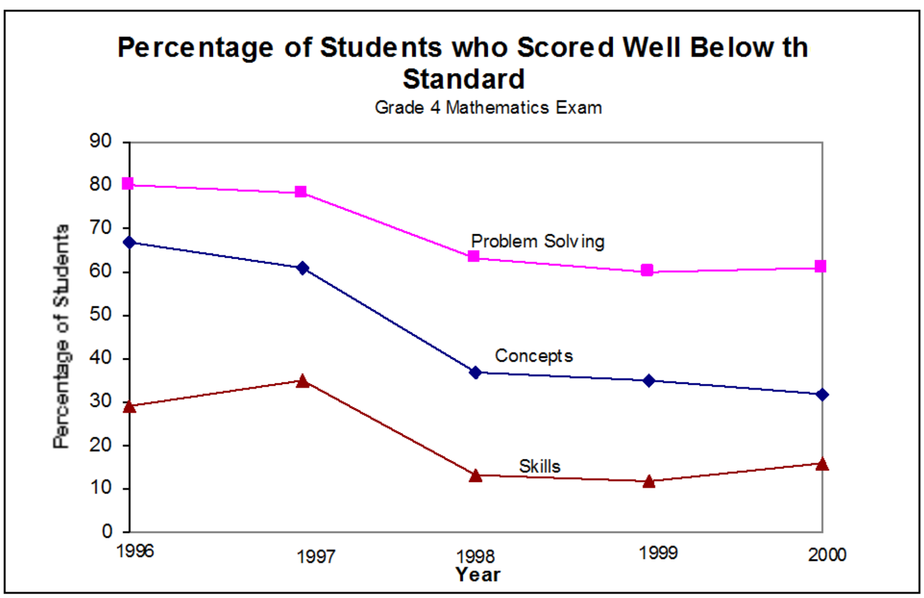


Figure 5.
Decrease in lowest-scoring students following standards-based text adoption.
(Adapted from Briars, 2001)

Finally, one can ask whether the new curriculum (in combination with the other factors already in place) is really what made the difference. A more fine-grained analysis of district data addresses this question. In any district, there will be some schools that embrace new materials and some schools that take a "this too shall pass" attitude. School visits identified a sample of "strong implementation" schools, in which the five conditions described in Table 1 were in place, and a series of demographically matched "weak implementation" schools, where the new curricula and support materials were still on the shelves in their plastic wrappings, and traditional texts were on student desks.

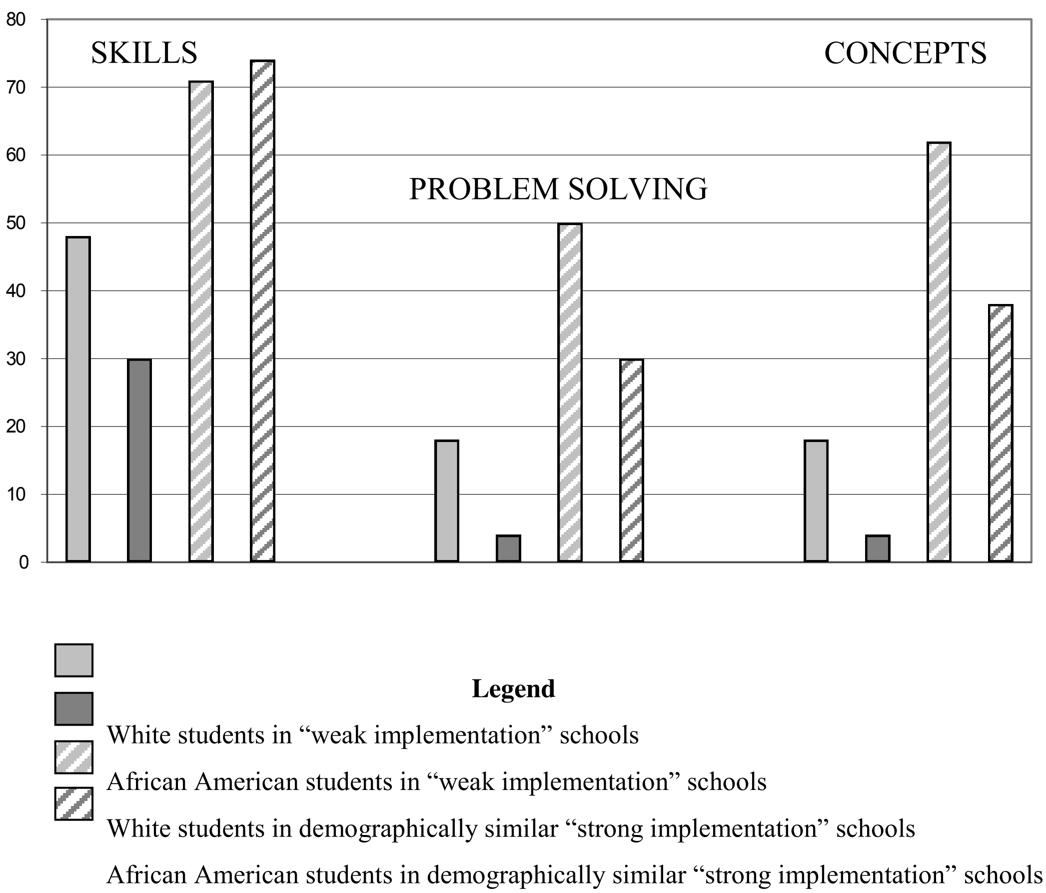


Figure 6.
Percentage of 4th grade students in demographically matched
"weak implementation" and "strong implementation" schools
who achieved the skills, problem solving, and concepts standard in 1998
(Derived from Briars and Resnick, 2000)

As Figure 6 indicates, skills-related "racial performance gaps" were eradicated in the strong implementation schools. There were still such gaps in performance on concepts and problem solving, but a much higher proportion of African American students in the strong implementation schools did well in those categories; and in all categories, the African American students in the strong implementation schools outperformed the White students in the weak implementation schools. In sum, the constellation of five conditions in Figure 1 did indeed move the schools toward the main goal of meaningful engagement with powerful mathematics for all children.

Evidence of Plausibility, Part 2: The 90/90/90 Schools

The 90/90/90 Schools are so named because they have the following characteristics:

- More than 90 percent of the students are eligible for free and reduced lunch;
- More than 90 percent of the students are from ethnic minorities
- More than 90 percent of the students met or achieved high academic standards, according to independently conducted tests of academic achievement.

As noted above, eligibility for free or reduced lunch means that the students are of low socioeconomic status (that is, poor). That fact, combined with the more than 90% minority enrollment rate, would mark the schools as stereotypically bound for failure. Thus, the third bullet – a remarkably high achievement rate – is a sign that something very special is taking place at those schools. According to Douglas Reeves,

> We found five characteristics that were common to all 90/90/90 Schools. These characteristics were:
>
> - A focus on academic achievement
> - Clear curriculum choices
> - Frequent assessment of student progress and multiple opportunities for improvement
> - An emphasis on nonfiction writing
> - Collaborative scoring of student work
>
> (Reeves, 2000, p. 187.)

This is not a perfect match with the criteria given in Table 1, but it is close. The first three bullets speak to standards, curriculum, and assessment. The fourth, it is worth noting, is a vehicle for sense-making – both in terms of what one has to say and how one says it. And the fifth speaks to the teaching community: collaborative scoring is a way of sharing and enforcing values that are important to the community (i.e., professional development consistent with standards and expectations, curricula, and assessment) and providing the kind of stability that allows for teachers' ongoing professional growth.

Evidence of Plausibility, Part 3: The ARC Center Tri-State (Illinois, Massachusetts, & Washington) Student Achievement Study

Parts 1 and 2, immediately above, offered evidence of the impact of coherent systemic approaches to improvement, exploring contexts embodying (to some degree) the five desiderata in Table 1. Here and in Part 4, I turn to more limited data – the sad fact is that the are few examples that satisfy all five of the desiderata. However, having even a subset of the conditions in Table 1 in place can make a difference.

Some of what follows may be particular to the United States, which does not have a culture of teacher professionalism. In the U.S. teacher preparation programs typically take place either during the year after a candidate teacher obtains a bachelor's (4-year undergraduate college) degree, or during the candidate's undergraduate career. After this, teachers enter the profession – where they are typically given a large amount of autonomy in their classrooms, and provided little by way of organized professional development. It should not be surprising, then, that many teachers stick close to their textbooks, which structure their work for them Classical "traditional" texts offered students and teachers each day's work in a "two page spread," with a given day's work described in two facing pages (and the teacher's edition containing suggested assignments and worked-out examples, all around the margins of those two pages).

In this kind of climate, improved curricular materials can make a significant difference. In comparison to many of the traditional curricular materials, the Standards-based materials had at least the first two of the five desiderata in Figure 1 working to their advantage. First, they were based on a rich set of standards aimed at sense-making. Second, they were, in general, more coherent and better designed. Large-scale comparisons of standards-based and traditional curricula are hard to come by, because there are in general no mechanisms for such comparisons. One exception is the ARC Center Tri-State Student Achievement Study, which was funded by the National Science Foundation. Here is the abstract from the study:

> In 2000–2001, the ARC Center, located at the Consortium for Mathematics and Its Applications (COMAP) [http://www.comap.com/elementary/projects/arc/], carried out a study of reform mathematics programs in elementary schools in Illinois, Massachusetts, and Washington. The study examined the performance of students using three elementary mathematics curricula—Everyday Mathematics; Math Trailblazers; and Investigations in Number, Data, and Space—on state-mandated standardized tests administered in spring 2000. The study included over 100,000 students, 51,340 students who had studied one of the three reform curricula for at least two years and 49,535 students from non-using comparison schools matched by reading level, socioeconomic status, and other variables. Small differences on the SES variables remaining between the reform schools and the matched comparison schools were further controlled by adjustments based on regression analyses. Usage of the reform curricula was verified by a telephone survey of schools and districts.
>
> Results show that the average mathematics scores of students in the reform schools are significantly higher than the average scores of students in their matched comparison schools. The results hold across five different state-mandated tests, and across topics ranging from computation, measurement, geometry, and algebra to

> problem solving and making connections. The study compared the scores on all the topics tested at all the grade levels tested (grades 3–5) in each of the three states. Of 34 comparisons across five state-grade combinations, 28 favor the reform students, six show no statistically significant difference, and none favor the comparison students. The results also hold across all income and racial/ethnic subgroups, except for Hispanic students, where there are no significant differences between the scores. (Arc Center, undated).

Evidence of Plausibility, Part 4: Senk & Thompson (2003) on the NSF Curricula.

Smaller in size and scope than the ARC Center study, but more broadly based, is the set of studies found in Senk & Thompson's 2003 volume *Standards-based school mathematics curricula: What are they? What do students learn?* This volume reports on the performance of all of the NSF-supported standards-based curricula. Speaking as a researcher, there are things I wish were different in the volume. Many of the curriculum evaluations were done by colleagues of the curriculum developers, often using locally developed rather than independent assessments; and many of the teachers, unlike those in the ARC Center study, had received extensive professional development. Nonetheless, the overall findings are remarkably consistent, both across the varied NSF-supported curricula and in comparison with the ARC Center data. A typical summary, this one of the elementary curricula, is as follows:

> Students in these new curricula generally perform as well as other students on traditional measures of mathematical achievement, including computational skill, and generally do better on formal and informal assessments of conceptual understanding and ability to use mathematics to solve problems. (Putnam, 2003, p. 161).

In Summary:

When instruction offers a balance of skills, concepts, and problem solving (that is, they include a fair dose of sense-making, at the cost of some time on practicing skills), students will do as well on tests of skills as students whose instruction focused on skills only – and they will do much better on tests of conceptual understanding and problem solving. As parts 3 and 4 of this section indicate, curricula focused on sense-making can make a positive difference. As parts 1 and 2 of this section indicate, that difference is likely to be larger, and more long-lasting, if there is systemic coherence. If standards are focused on sense-making; if curricula, assessments, and professional development are aligned with those standards; and if there is enough stability for teachers to develop the deep understandings that enable them to make the best of those curricula, one can expect change for the better.

An Agenda

The claims made above lead to an agenda for practical action. Practice, however, should inform and be informed by research; hence the agenda should include as a sixth item the conduct of research that helps both researchers and practitioners to understand attempts at improvement. Table 2 provides the outline of the agenda. Concisely stated as in Table 2, those items look simple and straightforward – but they are not. In the balance of this article I describe some of the issues that need to be dealt with, item by item, if progress is to be made.

Agenda

1. The development of mathematically rich content and process standards, which have at their core the notion of mathematics as a sense-making activity.
2. Curriculum development and refinement consistent with these standards.
3. The development of assessments that are consistent with these standards.
4. Professional development for teachers consistent with all of the above – and the opportunity for teachers to develop the understandings necessary for "teaching for understanding."
5. Consistency and stability, to allow for steady improvement in the system.
6. A solid body of research to understand and facilitate items 1 through 5, and to provide a solid basis for continuing progress.

Table 2.
A practical and theoretical agenda for the improvement of mathematics education

Challenge 1: The development of mathematically rich content and process standards, which have at their core the notion of mathematics as a sense-making activity.

The key point to understand is that the establishment of standards is a political process, requiring consensus and collaboration. If things go wrong, there is little rationality in the process.

Consider, for example, the "math wars." They roiled the U.S. in the late 1990s, to the degree that the U.S. Secretary of Education felt compelled to address the Joint Mathematics Meetings and ask for civility and reasoned discourse surrounding issues of mathematics curricula. And, alas, math wars seem to be a new American export, to other nations around the world.

I will not go into detail here – see Schoenfeld, 2004, 2008 – but I will note two critically important facts:

The math wars in the United States were fought in the complete absence of meaningful evidence.

The NCTM *Standards* were issued in 1989. The National Science foundation issued an RFP for the development of standard-based curricula shortly afterward, and it took a while to award the curriculum development grants. It takes approximately three to five years to develop a three-to-five-year curriculum, so the first, "alpha" versions of the new curricula became available in 1995 or later. Beta versions, which were ready for robust testing, came out later. In consequence, virtually no cohorts of students had studied from a complete elementary, middle school, or high school “reform” curriculum until roughly the year 2000. Note that the math wars had reached a fever pitch by 1998, before such data were available. The main salvos in the math wars were fired on the basis of anecdote and opinion, with no hard data to back them up.

Subsequent evidence consistently favors “teaching for understanding” (“reform”).

The evidence in the previous section of this paper is clear. A few isolated studies may favor a particular “traditional” text in a comparison with a standards-based approach. (But, see challenge 3, below: the choice of assessment in such comparison studies makes a difference.) However, the mast majority of studies tell a very clear story: students who learn from standards-based curricula tend to perform about the same on tests of skills as students who studied from skills-oriented curricula, but do much better on assessments of concepts and problem solving. That is: globally speaking, students are better off with standards-based teaching for understanding.[3]

The lesson to be learned from these two points is clear: *in order to arrive at a asset of standards that actually can make a difference, political fighting must be minimized and evidence should be insisted upon*. The challenge is to engage all of the relevant constituencies (mathematicians, mathematics educators, psychologists, teachers, policy representatives) in productive ways, which build on their expertise and which result in the gathering and use of meaningful evidence.

Challenge 2: The development and refinement of curricula consistent with high mathematical standards.

I suspect that this too is a political process, differing from country to country; attunement to the subtleties of this political process, either within nations or across the world, will be essential.

Curriculum development is difficult, and it does not come about by chance. Thus, unless there are effective commercial or governmental mechanisms for the creation and refinement of mathematically powerful curricula, the result is likely to be the perpetuation of the status quo. Changes in the U.S. came about when the National Science Foundation, recognizing that the commercial marketplace would not produce standards-based textbook series on its own – representatives of the publishing industry claimed that it cost $25 Million to create a brand new series, and they were unwilling to make that kind of investment without knowing that it would pay off – issued a request for proposals that catalyzed the curriculum development process. Obviously, policies differ widely from nation to nation. However, there will have to be some form of incentive (one

[3] One must issue a series of caveats here. First, the evidence is weaker than one would like: there has not been the funding for the kinds of studies that would address questions of curricular impact in more definitive ways. Second, a much more refined notion of “curriculum” is needed than is typically used; context matters, and there are sure to be interactions between curricula and context. Thus my statement is a broad generalization, and inaccurate as all generalizations are, but it stands nonetheless. See the discussion of Challenge 6, and Schoenfeld (2007) for additional detail.

hope, tied to actual data-gathering!) to induce people to devote the kind of time and attention required to produce high quality curricula.

I note that there does not have to be "one correct answer" to the curriculum development problem – there are many ways to provide mathematical riches. Indeed, the National Science Foundation made the wise decision, in the U.S. political context, to fund the initial development of a number of different curricula. The non-negotiable desiderata are that all designs should offer *meaningful engagement* with *powerful mathematics* for *all children*. How they do so may vary.

Challenge 3: The development of assessments consistent with high mathematical standards.

The naïve view is that tests are neutral mechanisms for indicating what students know. The reality is much more complex.

Tests are not value-free: they reflect what the test designers think is important. As a result, different tests often measure different things. The fact that a particular nation can score low on TIMSS and high on PISA (or vice-versa) is clear evidence that two carefully designed tests of "mathematics" do not measure the same knowledge of mathematics.

Moreover, assessments are not neutral instruments: measuring the system has an impact on the system being measured. The acronym WYTIWYG – "What You Test Is What You Get" – is a useful reminder that testing, especially "high stakes testing" (tests where the outcomes have serious consequences for students, teachers, or schools), can have an fundamental impact on what is taught (and therefore learned). These effects can be dramatic and unexpected. In the 1990s, for example, the state of California began mandatory high-stakes testing in English literacy and in mathematics. One result of this testing program was completely unexpected. Prior to the high stakes testing program, California students had scored about the median on the (low stakes) science tests on the National Assessment of Educational Progress (NAEP). Shortly after the high stakes testing program in English literacy and in mathematics was implemented, California student scores on the NAEP science plummeted, with California landing near the bottom of the 50 states. Why? There was so much pressure to teach English literacy and mathematics that much less science was taught.

It is thus critical to develop and use assessments that are aligned with curriculum and standards. Whatever understandings are considered important in standards and curricula – including things such as problem solving processes – had better be on the assessments. If they are not, they may well not be emphasized in instruction.

As described in the examples below, testing has the potential to reveal or obscure important information. It can be a positive or negative systemic influence, and some of the effects can be subtle.

Example 1: Drilling on skills can give the illusion of competency.

Consider a simple subtraction problem such as

$$\begin{array}{r} 87 \\ \underline{-\ 24}, \end{array}$$

and a slight variant such as "subtract 24 from 87."

Suppose both of these tasks are used as test items for young students. One would expect student performance on them to be close. Perhaps students would do slightly worse on the second problem, because it calls for reading text and then lining the numbers up before performing the operation.

Items such as these were used in both "low stakes" and "high stakes" testing contexts. In the low stakes district, teachers taught the curriculum (for better or worse) as designed. In the high stakes district, teachers focused on the algorithm, and on practice problems that looked just like the test problems.

In the low stakes district, test scores were as one would predict: 77% of the students obtained the correct answer to the standard task, and 73% of the students obtained the correct answer to the variant. The 4% drop in performance is plausible, given the slight difference in the task statements.

In the high stakes testing district, 83% of the students answered the first question correctly. On its own, this datum suggests that having students practice extensively on test-related items has a clear payoff with regard to skills-related proficiency. However, only 66% of the students obtained a correct answer to the problem variant! This 17% drop in proficiency, resulting in a performance rate much lower than that of the low stakes district, calls the students' understanding into question. How much did they really understand, when so large a proportion of the students could not do an essentially identical problem stated in words? In sum, their performance on the standard problem gives the *illusion* of competency. (Flexer, 1991; Shepard, 2001).

Example 2: Tests can misrepresent what students know and can do.

As noted above, TIMSS and PISA focus on different aspects of mathematical proficiency. Thus, a nation whose curriculum was aimed at the kind of mathematical understandings rewarded on the PISA exam might feel that TIMSS mis-represented student competencies, and vice-versa.

An example of a the kinds of dramatically different information that tests can reveal is given in Ridgway, Crust, Burkhardt, Wilcox, Fisher, and Foster (2000). In that study, two sets of mathematics tests were administered to more than 16,000 students at grades 3, 5, and 7. The first test was a standardized skills-oriented test known as the SAT-9, which is used by the State of California as a high-stakes examination. The second was the Balanced Assessment test produced by the Mathematics Assessment Resource Service (MARS). MARS tests examine a broad range of skills, concepts, and problem solving. Table 3 shows scores aggregated across the three grade levels. Based on their test scores, each student was assigned a score of either "proficient" or "not proficient" according to each test.

	SAT-9	
MARS	Not Proficient	Proficient
Not proficient	29%	22%
Proficient	4%	45%

Table 3.
Comparison of student proficiency ratings on two examinations.

Table 3 shows that 74% of the students were rated the same on both examinations. Of course, this should happen with two mathematics tests at the same grade level. What is of interest, however, are the two cells of the table where students were not rated the same. Twenty-two percent of the total population were rated as proficient on the skills-oriented test (the SAT-9), but not proficient on the test that assessed skills, concepts, and problem solving (MARS). This was approximately one-third of all the students labeled proficient on the SAT-9. (In contrast, 4% of the total population, less than one in twelve who were labeled proficient on the MARS exam, were labeled not proficient on the SAT-9.) What this means, in essence, that that early one third of those students labeled as proficient on the SAT-9 were "false positives." The narrow-band skills-oriented test gave an incorrect impression of these students' competency.

As these data suggest, problems can also be caused by a mis-match between assessments and curricula. Imagine a comparison study that examined two curricula – one curriculum being skill-oriented, the other focusing on skills, concepts, and problem solving. A narrow skills-oriented assessment might show no difference between the two groups, because the "value added" of the more broad curriculum was not assessed. In contrast, a more broad-based assessment would capture performance difference.

In sum, the challenge is to develop assessments that reflect one's mathematical values (i.e., one's standards, or goals for student learning) *and* that provide useful information to teachers and students, so that the system is aligned and so that the information from assessment helps improve the system.

Coda: My colleagues on the Mathematics Assessment Project (2011) have been working to develop mathematically powerful summative assessments, and compatible formative assessment lessons. These assessments and lessons can be downloaded for non-commercial use at no cost from the project web site, http://map.mathshell.org/materials/.

Challenge 4: The creation of professional development for teachers consistent with high mathematical standards – and the opportunity for teachers to develop the understandings necessary for "teaching for understanding."

Teaching for understanding in the ways envisioned here – teaching in ways that provide meaningful engagement with powerful mathematics for all children – is difficult. It demands a broad and deep knowledge of mathematics, a curricular knowledge of where one's students have been and where they will be going mathematically, a substantial base of pedagogical content knowledge, knowledge of student thinking and learning, and the ability to create productive learning communities in classrooms. No teacher has all of this knowledge, or even a very large fraction of it, when he or she begins teaching. The challenge, then, is to find ways to develop and refine these kinds of understandings as teachers evolve in their professional careers. Some models for such professional development exist, e.g., the Japanese practice of lesson study (Fernandez & Yoshida, 2004; Stigler & Hiebert, 1999). Lesson study has the significant advantage that studying one's practice, and reflecting on it, are defined as part of teachers' regular work responsibilities – thus, time is set aside for such practices and the are officially sanctioned. Every nation has its own traditions of professional development, and of teacher professionalism; hence the idea of "importing" any practice such as lesson study is untenable. Any approach toward enhancing professionalism of teachers and providing them with enhanced opportunities to learn and to reflect on their work must be grounded in national cultural traditions.

It should be obvious, but to repeat a recurrent theme: the more that work on teachers' professional development is aligned with robust mathematical standards, curricula, and assessment; the more it is grounded in teachers' experience of and reflection on those standards, curricula, and assessment; and the more such work is embedded in the ongoing work of teachers, the more effective it will be.

Challenge 5: Consistency and stability, to allow for steady improvement in the system.

Let the United States serve as a warning to everyone: when fads and fashions in education result in major directional changes every few years, the result is chaos and little progress. Educational systems are complex, and complex systems move slowly. Significant progress is made when there is thoughtful planning and enough time for ideas and methods to take hold. This does not imply stasis – nobody gets things right the first time, and systems need to evolve – but it does imply thoughtful evolution in response to feedback, rather than attempts to implement the next "great new idea" every few years.

In short, stability (and the alignment of standards, curricula, assessment, and professional development) provides a context that allows for professional growth and change, and for learning from experience. As with all of the other challenges, this is a political issue and an essential one. (In the U.S., politicians and administrators all want documentation of dramatic results, thanks to their new policies of course, before the next election takes place – coherence and reality be damned!). Short-term changes and/or instability are guaranteed to undermine progress.

Challenge 6: The development of a solid body of research to understand and facilitate items 1 through 5 in the agenda, and to provide a solid basis for continuing progress.

I have framed Challenges 1 though 5 as practical issues, because the community knows enough to begin to work on them now – and that work will make a difference. However, it is essential to note that each of those challenges can and should be approached as a fundamental research issue.

Challenge 1, the creation of rich mathematical standards, is in large measure a political process. We understand relatively little about how to facilitate this kind of process, engaging all of the relevant constituencies (mathematicians, mathematics education researchers, teachers, administrators, and policy-makers) productively. The existence until this past year of a wide range of standards among the 50 states in the United States indicates how contentious and uneven this process can be – partly because much of the process has not been undergirded by research on what we do know[4].

We also know very little about Challenge 2, curriculum design. Some years ago one of my students, who was interested in design, wrote all of the design teams creating the NSF-supported curricula, asking if they had written about their design processes. Uniformly, the response she got was, "we're so busy writing curricula that we don't have time to write about how we do it." For the most part, instructional materials design is an art form, learned by apprenticeship. We have neither standards for instructional materials nor, in truth, effective mechanisms for evaluating them. (See Burkhardt & Schoenfeld, 2003, for a discussion of impediments to progress in this area; see also Schoenfeld, 2007, for a specific research proposal.) There are some signs of progress, for example the existence of the International Society for Design and Development in Education (ISDDE; http://www.isdde.org/) and a research focus on "design experiments" (Cobb, Confrey, diSessa, Lehrer & Schauble, 2003; Schoenfeld, 2006); but there is much to be done.

Challenge 3, assessment, also needs a significant amount of research in order to provide a solid base for the development of assessments that produce reliable and valid information about the processes that matter in mathematical thinking and problem solving. For some time, we have had frameworks that provide robust descriptions of the processes involved in mathematical problem solving (e.g., Schoenfeld, 1985). However, assessment lags far behind because there do not exist, in general, reliable psychometric techniques to capture the processes that we care about. That is: we tend to test what we can measure reliably, rather than developing measures to test what we care about. There is a long way to go theoretically, but there are also some practical advances than can be made now.

[4] The situation is changing dynamically. A "Common Core State Standards Initiative," supported by the U.S. National Governors Association, has produced a set of standards for mathematics that has been adopted by 44 states (Common Core State Standards Initiative, 2011). It is possible that mathematics education in the U.S. will be more coherent in the years to come. A lot depends on the assessments that will be used to judge student progress toward the standards. The development of those assessments is taking place as I write.

Challenge 4, professional development, is both under-researched and under-theorized. On the one hand, most studies of professional development have been "local" in the sense that a particular (presumably successful) intervention has been documented and analyzed. But, absent a larger theoretical framework, the sum of the parts does not add up to a whole – so much varies from context to context that there is little that applies broadly. Although there has been some progress toward modeling of teacher cognition and decision-making, we do not, for example, have robust understandings of teachers' developmental trajectories, or of the kinds of knowledge discussed above (a broad and deep knowledge of mathematics, a curricular knowledge of where one's students have been and where they will be going mathematically, a substantial base of pedagogical content knowledge, knowledge of student thinking and learning, and the ability to create productive learning communities in classrooms) and how they grow. Although there are some apparently successful models of professional development (e.g., some point to lesson study as a model practice), it is clear that professional development is deeply embedded in national culture and tradition; a significant research question is how to uncover the underlying principles supporting successful professional development, so those principles can be adapted to other cultural contexts.

Challenge 5 could, potentially, be a fascinating and productive arena for political scientists and policy analysis. Once again, the United States serves as a negative example. Across the 50 states, one finds sets of standards that are almost incommensurate; some focus on skills, some focus on problem solving and understanding. Some states moved from one emphasis to the other, with radical shifts in policy and instructional materials. As noted in the brief discussion of the math wars, many of these changes were uninformed by research – politics and opinion, rather than data, drove the process. This is exacerbated, in the U.S., by the fact that almost every elected or appointed official wants to make his or her mark in education, with new policies whose impact is demonstrated before the next election. The question of how one might understand this process, and how one might make it more sane (and reliant on data) is critically important.

Finally, I turn to Challenge 6, the issue of research itself. I have not, in this paper, discussed the excellent research in mathematics education presented at this conference; and that, of course, is just the tip of the proverbial iceberg. Enough is known, at present, for the European Community (and the United States for that matter) to do far better than at present, and to make significant strides toward practical agenda items 1 through 5. But, much more needs to be known. As noted above, each of items 1 through 5 should be the locus of major research efforts. In addition, the research enterprise needs to be nurtured. Industrialists and industry researchers will tell you that in order for a corporation to survive in the long term, it must invest a minimum of 2 to 5% of its proceeds into research and development; and some industries, such as pharmaceuticals, typically invest 15 to 20% of their income into R&D. In the United States, the federal government invests far less than .1% of annual educational expenditures for educational research and development. Similar calculations regarding the European Union might be informative.

We have a saying, “you get what you pay for.” It should be clear that an increased investment in the educational research infrastructure, along with a clear focus, both in practical and research terms, on the agenda highlighted above, will pay for itself many times over.

References

ARC Center. (Undated). Tri-State Student Achievement Study. Downloaded on January 10, 2005, from http://www.comap.com/elementary/projects/arc/tri-state%20achievement%20full%20report.htm.

Briars, D. (March, 2001). Mathematics performance in the Pittsburgh public schools. Presentation at a Mathematics Assessment Resource Service conference on tools for systemic improvement, San Diego, CA.

Briars, D., & Resnick,. (2000). Standards, assessments – and what else? The essential elements of standards-based school improvement. Unpublished manuscript.

Burkhardt, H., & Schoenfeld, A. H. (2003). Improving educational research: toward a more useful, more influential, and better funded enterprise. *Educational Researcher 32(9), 3-14*.

Cadwell, C., & Berman, R. (December 9, 2007). All in the numbers: Jobs for mathematicians. San Francisco Chronicle. Downloaded February 24, 2007 from http://www.sfgate.com/cgi-bin/article.cgi?f=/g/a/2007/12/09/JOBScadwellberman.DTL.

Cobb, P., Confrey, J. diSessa, A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher*, 32(1), 9-13.

Common Core State Standards Initiative. *Common Core State Standards for Mathematics*. (2011). Downloadable from the Common Core State Standards Initiative website, http://www.corestandards.org/.

Fernandez, C., & Yoshida, M. (2004). *Lesson study: A Japanese approach to improving mathematics teaching and learning*. Mahwah, NJ: Erlbaum.

Flexer, R. J. (1991, April). Comparisons of student mathematics performance on standardized and alternative measures in high-stakes contexts. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Mathematics Assessment Project (2011). Summative Assessments and Formative Assessment Lessons. Downloadable from the Mathematics Assessment Project web site, http://map.mathshell.org/materials/.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government printing office.

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

National Council of Teachers of Mathematics. (1989) *Principles and Standards for School Mathematics*. Reston, VA: NCTM.

National Research Council. (1989). *Everybody counts*. Washington, DC: National Academy Press.

National Research Council. (1990). *A challenge of numbers*. Washington, DC: National Academy Press.

Putnam, R. (2003). Commentary on four elementary mathematics curricula. In S. L. Senk & D. R. Thompson (Eds.). *Standards-based school mathematics curricula: What are they? What do students learn?* (pp. 161-191). Mahwah, NJ: Erlbaum.

Reeves, D. (2000). *Accountability in Action: A Blueprint for Learning Organizations*. Denver, CO: Advanced Learning Press.

Ridgway, J., Crust, R., Burkhardt, H., Wilcox, S., Fisher, L., & Foster, D. (2000). *MARS report on the 2000 tests*. San Jose, CA: Mathematics Assessment Collaborative.

Schoenfeld, A. H. (1985). Mathematical Problem Solving> Orlando, FL: Academic Press.

Schoenfeld, A. H. (2004). The math wars. *Educational Policy, 18(1)*, 253-286.

Schoenfeld, A. H. (2006). Design experiments. In P. B. Elmore, G. Camilli, & J. Green (Eds.), *Handbook of Complementary Methods in Education Research* (pp. 193-206). Washington, DC & Mahwah, NJ: American Educational Research Association and Lawrence Erlbaum Associates.

Schoenfeld, A. H. (2007). Method. In F. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning,* pp. 69-107. Charlotte, NC: Information Age Publishing.

Schoenfeld, A. H. (2008). Problem Solving in The United States, 1970-2008: Research and Theory, Practice and Politics. In G. Törner, A. H. Schoenfeld, & K. Reiss (Eds.), *Problem solving around the world – Summing up the state of the art*. Special issue of the *Zentralblatt für Didaktik der Mathematik*: Issue 1, 2008.

Senk, S. L., & Thompson, D. R. (Eds.) (2003). *Standards-based school mathematics curricula: What are they? What do students learn?* Mahwah, NJ: Erlbaum.

Shepard, L. (2001, April). Protecting Learning from the Harmful Effects of High-Stakes Testing. Presented in a symposium, “Three Blueprints for a Revolution: How to Halt the Harm Caused by High-Stakes Tests,” at the annual meeting of the American Educational Research Association, Seattle, WA.

Stigler, J, & Hiebert, J. (1999). *The teaching gap*. New York: Free Press.