Workshop Report No. 4

Research on Systemic Reform:
What Have We Learned? What Do We Need to Know?

Synthesis of the Second Annual NISE Forum

Volume 2: Proceedings

William H. Clune, Susan B. Millar, Senta A. Raizen, Norman L. Webb,
Dianne C. Bowcock, Edward D. Britton, Ramona L. Gunter,
& Ricardo Mesquita

National Institute for Science Education (NISE) Publications

The NISE issues papers to facilitate the exchange of ideas among the research and development community in science, mathematics, engineering, and technology (SMET) education and leading reformers of SMET education as found in schools, universities, and professional organizations across the country. The NISE Occasional Papers provide comment and analysis on current issues in SMET education including SMET innovations and practices. The papers in the NISE Research Monograph series report findings of original research. The NISE Conference and Workshop Reports result from conferences, forums, and workshops sponsored by the NISE. In addition to these three publication series, the NISE publishes Briefs on a variety of SMET issues.

Workshop Report No. 4

**Research on Systemic Reform:
What Have We Learned? What Do We Need to Know?**

**Synthesis of the Second Annual NISE Forum
Volume 2: Proceedings**

William H. Clune, Susan B. Millar, Senta A. Raizen, Norman L. Webb,
Dianne C. Bowcock, Edward D. Britton, Ramona L. Gunter, & Ricardo Mesquita

i

Contents

Status and Trends of Systemic Reform: Research on Systemic Reform

What Have We Learned? What Do We Need to Know?

The Role of Curriculum in Systemic Reform

The Role of Teaching and Learning in Systemic Reform

The Role of Evaluation in Systemic Reform

Forum Highlights and Looking Ahead

Appendixes

Contents of Volume 1

Preface

The goal of the Second Annual NISE Forum was to generate and share knowledge of systemic reform, including the role of curriculum, the role of teaching and learning, and the role of evaluation. To reach this goal, panelists from around the nation shared their expertise in these various areas with participants who themselves brought varied expertise to structured conversations and networking sessions.

These two volumes are the result of a collaboration among four NISE teams and, more broadly, of a collaboration among the NISE researchers, the Forum panelists, NSF staff, and the Forum participants. The Interacting with Professional Audiences (IPA) team, Policy Analysis of Systemic Reform (PASR) team, Strategies for Evaluating Systemic Reform (SESR) team, and Formative Evaluation (FE) team worked together to design, implement, and evaluate the Forum. The four teams then collaborated in the writing of Volume 1. The FE team (Susan Millar, Dianne Bowcock, Ramona Gunter, and Ricardo Mesquita) analyzed the 483 think pieces written during the Forum and produced the "Themes Articulated in Forum Participant Think Pieces." Senta Raizen and Ted Britton (IPA), William Clune (PASR), and Norman Webb (SESR) supported the FE team's analysis by framing key themes based on their readings of the think pieces. William Clune and Norman Webb drew heavily on the FE team's analysis of the think pieces as well as the panelists' papers in writing their introduction.

Given the way this document was developed, readers should view the two volumes as elements of a conversation through which a better understanding of systemic reform is emerging.

Volume 1: Analysis

- ***Introduction to the Papers and Think Piece Themes.*** In this introductory section,

William Clune and Norman Webb integrate the other two sections by developing some of the more compelling points raised by the participants' think pieces and the panelists' contributions. William Clune is Voss-Bascom Professor of Law at UW-Madison with research interests in education law and policy, systemic reform, and adequacy in school finance. Norman Webb is a senior research scientist for WCER with research interests in mathematics education, evaluation, and assessment.

- ***Themes Articulated in Forum Participant Think Pieces.*** During three breakout sessions, Forum participants were asked to write "think pieces" in reaction to the panelists' remarks. This synthesis of the themes in the think pieces, including many quotations from the writers, provides the reader a window on issues, questions, and contrasting viewpoints about systemic reform that the Forum's varied participants articulated.

Volume 2: Proceedings

- ***Papers Presented.*** The ideas, knowledge, and experience of 24 panelists involved in implementing, researching, and evaluating systemic reform appear in this section. Most of these papers were prepared ahead of time and distributed to participants. Others are transcriptions of the panelists' remarks. (Print and audio versions of several of these are available on the TEECH Web site, *http://teech. terc. edu/ modes/papers/systemic_papers.cfm*)

Both volumes include papers from the opening plenary session and from the closing plenary session.

Second Annual NISE Forum: February 24-25, 1997

AGENDA ›
Research on Systemic Reform:
What Have We Learned?
What Do We Need to Know?

Monday February 24

8:30 - 10:30   Opening Plenary
Andy Porter, chair

Status and Trends of Systemic Reform
Luther Williams, Marshall Smith

What Have We Learned? What Do We Need to Know?
Bill Clune, Susan Fuhrman, Jane Butler Kahle, Uri Treisman;
Daryl Chubin, discussant

10:30 - 10:45 *Break*

10:45 - 12:00 Breakouts

What Have We Learned? What Do We Need to Know?

12:15 - 1:30   *Lunch*

1:30 - 3:30   Afternoon Plenary
Senta Raizen, chair

The Role of Curriculum in Systemic Reform
Rodger Bybee, Midge Cozzens, Richard Greenberg,
Sylvia Johnson, Tom Romberg

The Role of Teaching and Learning in Systemic Reform
Chris Dede, LeRoy Lee, Susan Loucks-Horsley, Shirley Malcom,
Lauren Resnick

3:30 - 3:45   *Break*

3:45 - *5:oo*   Breakouts

Curriculum, Teaching and Learning in Systemic Reform

Tuesday, February 25

8:30 - 9:30   Opening Plenary
               Senta Raizen, chair

               The Role of Evaluation in Systemic Reform
               *Bernice Anderson, Tom Corcoran, Paul LeMahieu, Norman Webb,
               Iris Weiss*

9:30 - 9:45   ***Break***

9:45 - 10:45 Breakouts

               Evaluation in Systemic Reform; Discussions with Panelists

Monday's Panelists   *Rodger Bybee, Jane Butler Kahle, Bill Clune, Chris Dede,
                       Richard Greenberg, Sylvia Johnson, Lauren Resnick,
                       Susan Loucks-Horsley, LeRoy Lee, Uri Treisman*

Tuesday's Panelists   *Bernice Anderson, Iris Weiss, Tom Corcoran, Paul LeMahieu,
                        Norm Webb*

11:00 - 12:30 Closing Plenary
               Andy Porter, chair

               Forum Highlights and Looking Ahead
               *Mike Kirst, Cora Marrett;
               Larry Suter, discussant*

Status and Trends in Systemic Reform:
Research on Systemic Reform

Pascal Forgione, Jr.
Commissioner of Education Statistics,
Department of Education

I bring you greetings from Acting Deputy Secretary Marshall "Mike" Smith who really wanted to be here this morning. Mike asked me to reflect briefly on the impact of the Third International Mathematics and Science Study (TIMSS) and what it has done for the character and quality of the national conversation about systemic reform. While on the one hand, systemic reform is very complex and challenging, it is also very easy to trivialize this construct. It's easy for people to define what they are doing as systemic without actual substance or verification.

Fortunately, TIMSS is not about quick fixes. In fact, the Grade 8 data-all that's been released thus far-suggest that there are no simple solutions. TIMSS has provided data to explode three myths. I hope in June when the Grade 4 data are out that we confirm these findings. First, eighth graders in the United States have more hours of instruction in science and mathematics than German or Japanese students. Despite our shorter school year, if you take the length of the periods and the frequency per week at eighth grade compared to Japan and Germany, U.S. students in Grade 8 have more hours of instruction.

Second, on our questionnaires, American teachers reported assigning more homework than their Japanese or German counterparts. And when we looked at the after-school work, all of the work put together, American students do as much homework as their German and Japanese counterparts.

Finally, if you take three hours of TV watching per night as heavy TV watching, as many Japanese children are heavy TV watchers as children in the United States. So these three simple myths of time, homework, and TV are not what , it's about.

There are not going to be simple answers. We'll see whether the next two data sets-Grade 4 in June 1997 and Grade 12 in spring 1998— confirm these initial findings.

The TIMSS design overtly includes the key elements emphasized in a notion of systemic reform. The study had five major components that utilized a variety of methods to examine the breadth and the depth of the essential elements. Basically, one only has to look at the table of contents of our report. This report tries to give America a systemic view about achievement, curriculum, teaching, the lives of our teachers, and the lives of our students. In fact, the Grade 8 summary report called *Pursuing Excellence* was written explicitly to be read on an airplane in one hour by a businessman. Now remember, this is a statistical report. It may not appear to be a statistical report when you first look at it, because there are no standard errors; we don't talk about standard deviations. We tried to write in clear English about what these findings mean. I can assure you that every sentence went through a strict adjudication review process. Everything in the report is supported by data.

Let me highlight for two areas the kinds of questions that TIMSS is raising for this Nation. The first is the area of curriculum where Professor Bill Schmidt has developed very innovative methodology to look at textbooks and curriculum of the fifty countries. The kind of questions we asked in our report under curriculum included:

. Who sets the curriculum?
. Is the curriculum as focused as in other countries?
. Is it as advanced as in other countries?
. What do we mean by *advanced?*
▪ What is the content that really is advanced at the eighth-grade level and how much time is spent in class?

These are the kinds of powerful questions that we need to engage the Nation on and build deeper understanding through the analysis of the TIMSS database. It will take over a year to roll out the three populations of TIMSS data, from November 1996 (Grade 8), through June 1997 (Grade 4), to spring 1998 (Grade 12).

The second area that I want to emphasize is the innovative work of Professor Jim Stigler who conducted a three-country classroom video study of mathematics teaching. This is only Grade 8 mathematics and only in Germany, Japan, and the United States. It's unfortunate that we were not able to include Grade 8 science, but this study was quite an innovation five years ago. This video study answers the kinds of questions that we could never answer with questionnaires or even case studies:

- How do teachers in the three countries teach?
- How do the instructional practices differ?
- How do they organize and present their lessons?
- What are the cultural scripts of the lessons?
- What is the role of the teacher and the student during seat work and then class work?
- Are the students passive or active learners?
- What proportion of class time is spent in instructionally relevant activities? in social or housekeeping activities?
- Are teachers aware of the reform?
- Do they think they're doing it and do we see it?

This video study included a half national sample in the three countries. We have 81 lessons of a U.S. teacher. We selected the schools; we selected the teachers; and we selected the class period. It's a time warp capturing American teaching at Grade 8 in mathematics, and we've never had that before. We also have made available a CD-ROM and a videotape with six lessons, two for each country, from the study. For example, you are able to look at a Japanese algebra lesson and compare it to an American one. It

stimulates a terrific conversation about quality teaching.

TIMSS also taught us to be careful about drawing conclusions, even with statistical standards such as reliability. There's a need to continue to verify and to confirm the data. Let me give you an example. American teachers, 95% of them, said they knew the NCTM standards, and 75% of them said they were implementing them. When we examined the videotapes we found a focus on conceptual thinking in only 20% of the lessons. Again, what people called high level mathematics often ended up being the processes and not the deep mathematics. Thus, we were cautious in reporting the mathematics questionnaire data because the mathematics video data did not **confirm** the claim. In science, we had very similar data. Grade 8 science teachers said they also were implementing the standards. We held off releasing the findings because we were not able to corroborate the data. So we need to be careful in our survey reporting.

It's also ironic that in TIMSS we probably have more information about American curricula, textbooks, teaching, and instruction than in any other national database. This database is the kind that we need to build, with rich videos and curriculum components. It is a limited database in that it's cross sectional and not longitudinal, and it also doesn't **often** go to the student level. But TIMSS whets our appetite.

I'm pleased to say that Mike Smith, Neal Lane, director of the National Science Foundation, and Ernie Gibbons, the President's advisor on science, are working with the Domestic Policy Council to assure cross-agency follow through and coordination. Wouldn't it be wonderful if the federal government could be systemic about our partnerships to support the challenging reforms in science teaching and learning? I think we're beginning to have the **collegial** and cooperative conversations.

Thank you for this opportunity to share with you a brief overview of the TIMSS findings. I look forward to being your partner as we move forward in this exciting work.

Luther S. Williams
Assistant Director for Education and Human Resources,
National Science Foundation

I'm going to make a few comments bearing on the topic of status and trends of systemic reform. They are the end results of observations, experiences, and thinking about systemic reform in the context of NSF's programs. I will try to **frame** them **vis-à-vis** your conference. You asked two very important questions about research on systemic reform: What have we learned? What do we need to know? The answer to the second question is, "Substantial"; it is probably limitless. About the first question, if by *research you* mean a fundamental exercise informed by some theory or hypothesis that leads to substantial advancement of the knowledge base, the answer is, "Very little." I would argue that the size of the knowledge base is one of the challenges in an antiquality context to systemic reform. Few studies, in my judgement, configure to equal research.

There's another observation with respect to systemic reform: Today, **almost** every conceivable machination bearing on K-12 education, and now increasingly higher education, can accommodate itself under the rubric of systemic reform. That's okay at some level, and that is integral to the American culture. But it does make very challenging the maintenance of focus on systemic reform, particularly of a K-12 system. I make that observation because systemic reform is in the context of a system. In the case of NSF's programs in the K-12 domain focusing only on science and mathematics, a very powerful lesson that's been learned over the past several years is how exceedingly complex that system is. That finding of complexity, I would submit, is rarely acknowledged, whether one is talking about a state or a large urban community. Complexity is made worse by the fact that systems are highly fragmented. In effect one is attempting to promote reform in a system that is elusive and in some instances presents itself in multiple constructions in real time. The system is also exceedingly conflicted--even contradictory-owing to counter-instance agendas, roles, participants, and players. Therefore one is attempting to promote a comprehensive restructuring in a system that is highly resistant to Unitarian approaches. That recognition, in my view, from a research perspective, is one of the most important factors that should bear on any research design. There is essentially no **fixed** set that defines the problem that one is going to take up. Nonetheless, under the rubric of NSF's and others' programming, substantial progress has been made in a variety of states and in large urban communities. My sense is that progress has obtained insights that have done a very good job of controlling for a series of antisystemic or antiquality forces. I want to spend a few minutes sharing my list of those insights.

One of the most important challenges is to effect the conversion by process from a **unidimensional** to a multifaceted approach. All of the elements of the system are counter to a multifaceted approach; the system seeks to devolve to a unidimensional construction.

The second issue that has to be overcome is concerned with a very challenging nonlinear, very dynamic, ever-changing problem construction. It is made more **difficult** if emphasis is given to mathematics and science learning as opposed to educational processes. That is a very important distinction.

Third, systems that have made substantial progress have been very explicit and have **almost** elevated to a mandatory level the required infrastructure that is needed for mathematics and science learning. Pat Forgione alluded to part of this in talking about one dimension of one of the variables that equal this very, very complicated construction. Let me extend that further. All of the research he discussed has to do with eighth-grade mathematics, but his question is, "At the end of the day, what is the value of the take home?" One knows nothing about eighth-grade science; one knows nothing about the K-7 enterprise that actually informed mathematics learning. Unless one thinks eighth-grade mathematics is the end point, it is comparatively inconsequential relative to the rest of the sequence. I'm trying to empha-

size the vast domain about which one does not have comparable findings. We need to make the transition to emphasis on science learning in the context of the total system.

The fourth factor that has been very important where progress has been made is having individuals who operate from the context of a theory of reform, at least for their urban school district or their state. The reform was not haphazardly deposited in a series of machinations in a system without some reference to why, in fact, one is engaging in an explicit set of activities: What is the chronology in which one engages in those activities? What are the goals? What are the road maps? What are the strategies associated with those? What kind of formative evaluation is conducted in order to make early adjustments, and what outputs are valued? Is it the orderly progression year by year through the system, or is it the definitive outputs associated with science learning?

Progress has also been made in systems where there has been substantial attention to efforts to manage the role of others, including the federal government, which is not unitary in its contributions to standard mathematics and science education. How does one reconcile all of the conflicting inputs that bear on systems from Washington through funding and through other machinations? Similarly, state apparatus is an entity that requires some accommodation-witness the difference among the goals for learning in the variety of national or state assessment systems that are present across the states. What do they communicate differentially about the reasonable outcomes?

Another factor is higher education. Science and mathematics reform is critically dependant on the contributions from the higher education community. But there is one matter on which I would argue some thoughtful consideration, if not apprehension, needs to obtain. The suggestion that higher education as a domain is an exemplar of successful conclusion of reform or restructuring requires one to engage in a supreme fiction. There are enormous contributions to be had from the higher education community. (It is mostly scientific and technical expertise.) It is not a sector that schools should turn to for under-

standing school systems. In contrast, some school districts have done a very good job of working out arrangements with the business sector. That sector actually does have experience in this arena of systemic reform or restructuring. The business sector has taken on very challenging multifaceted problems and made progress. So there's a lot to be learned from that sector. I would argue that school systems that have made great progress beyond having very productive relationships with the business sector have forged very excellent relationships with a variety of community-based organizations that have been able to bring to bear an advocacy that is not found in other sectors. They've also catalyzed parents and employed them in the support of their efforts.

Successful sites also recognize that they're working in a very complicated, multifaceted, ever-changing, nonlinear system. Early on, these sites included "strategies for innovation replication" to develop some sense of how to scale up. Stringency is brought to documenting, communicating, and securing gains, so that they in effect become demonstrable nodes toward reform. One can continue to build on them and try to drive the system toward greater educational, financial, and intellectual economies while becoming more sophisticated in promoting reform. Many states require doing that in a systemwide fashion, converting a variety of resources whether in higher education, state departments of education, or school systems. In other words, all of these elements about which I've spoken have been configured in a system that is supporting an agenda that, if not unitary, has been reduced beyond its natural limits of replication.

There is a challenging research agenda transforming what we should learn. Some of the sites (maybe all of them in contrast, because they've had very different experiences) represent excellent source materials for mounting the research, not only in terms of documenting reform in progress, but also in trying to identify in a multifaceted domain the fmite number of overarching and control elements that are crucial. We really don't know that number. An exceedingly important and timely contribution from research would explicate within an undefined set of variables the minimum set that's really crucial to driving re-

form, leaving aside the variation that one **finds** from system to system.

Again, I acknowledge the very important work you do and very much urge you to take up these matters under your research agenda. But I would suggest you do that with deliberateness and near exclusion of the ever present and well understood, which is of little value. The **important** research to be done is that which 95% of us do not already purport to understand.

What Have We Learned? What Do We Need to Know?

William H. Clune
Policy Analysis of Systemic Reform, NISE Team Leader
University of Wisconsin-Madison

The research of the Policy Team of the NISE has concentrated in three areas: implementation of systemic reform, implementation of the national standards, and equity. I would like to report findings and key research questions in each area and then turn to what we are learning about the links between systemic reform and student outcomes and the implications for policy and research.

The Knapp Synthesis of Research on Implementation of Systemic Reform

· widespread **planning**
· incremental change in the classroom
· absence of broad and deep systemic change
. key knowledge gap: strategies leading to wholesale v. selective curriculum replacement

On the first topic of implementation, we commissioned a paper by Michael Knapp that he entitled "Between Systemic Reforms and the Mathematics and Science Classroom." This paper synthesized research not only on NSF's Systemic Initiatives but other systemic reforms such as in Michigan and California. The paper found that systemic reform had stimulated widespread planning and discussion, in other words had become a major presence on the policy landscape, but had produced only incremental change in the classroom, that is, an absence of broad and deep or systemic change. The key research question is, Which strategies lead to wholesale as opposed to

selective curriculum replacement? By *curriculum replacement* I mean the implemented curriculum, as actually taught in schools.

Findings on the National Standards and Local Implementation

- large "grain size," unguided local discretion
- tension between problem solving and formal analysis
- confusion over "deep understanding" and "habits of mind"
- political controversy over content and consequences
- key knowledge gap: alternative packages of curricula, materials, assessments, development, technology that
— fit the standards and alternative educational goals and philosophies
— draw on knowledge in NSF Systemic Initiatives, TIMSS

The findings on local implementation, or implementability, of the national standards were complementary. The standards are very general and comprehensive, which allows localities and textbook writers to claim compliance without making much real change. There is a tension between constructivists who emphasize problem solving and traditionalists who prefer formal analysis. There is real confusion about how to define deep understanding of the subject matters and so-called habits of mind, as well as political controversy over content and consequences, such

as for the traditional group of college-bound students. A key knowledge gap is the availability of packages of curricula, assessments, teacher development materials, and technology (such as interactive learning) that fit the standards, as well as alternative educational goals and philosophies, including both college prep and vocational goals. In other words, it looks like some of the disputes cannot be resolved by consensus but instead will require the design of highly developed curricular options. Relevant knowledge of this kind exists within NSF and many Systemic Initiatives but has not been pulled together adequately and connected with systemic reform.

The Case of the Virginia Standards and the Importance of Nonpoliticized Review

- quality should be recognized
- without blocking legitimate options
- without ignoring implementation in symbolic politics

It is important that we do serious quality reviews of the entire chain of developing and implementing standards, as illustrated by the politics now surrounding the Virginia standards. If these standards are indeed of high quality and are appropriate for certain educational goals and philosophies, that conclusion should be firmly established and widely publicized by independent researchers. On the other hand, if alternative models are needed, those, too, should receive appropriate support. And there is an important issue beyond disputes over competing goals and philosophies, because politicization can present dangers for any kind of standards-based reform. If a particular set of standards is not appropriate for all educational goals and philosophies, a political "victory" for one group may be costly to many students; and symbolic politics over standards may actually prevent us from even looking at any real development of the implemented curriculum and teacher capacity.

We should not have expected that the search for "universal high standards" or "hard stuff for all kids" would be easy, but neither should we leave this central question to politics. Perhaps international comparisons, such as TIMSS, can shed some light on options for implementing universally high standards.

Findings on Equity

- ethnic/gender subgroups differ in outcomes
- mixed evidence on equalization through course taking
- controversy over equity and academic courses
- controversy over how to measure the gap and expand access
- controversy over indicators as gatekeepers
- key research question: high standards, greater access

We had two papers on educational outcomes in mathematics and science by race and gender, written by Bill Tate and Alberto Rodriguez. These updated and refined similar work done in the past. They found many substantial differences across subgroups of race, ethnicity, and gender and mixed evidence for the proposition that similar course taking will produce similar results. Equivalent courses narrow the achievement gap but don't eliminate it and are not available to many students. Clearly, the effort to measure and report on gaps in equity yields valuable information.

But we also encountered serious debates about equity indicators. One is whether we should stick with college preparatory courses, or whether less academic courses are also needed for more equity. This debate appears to be occurring in the field through a conflict of contending forces, rather than through rational design and options. A related debate is whether the student assessments used to measure achievement gaps are the correct measure of equity and especially the correct criteria for selective admissions. Thus, the key research question is how to simultaneously raise standards and broaden educational access.

The Need for Theory Linking Process and Outcomes (what we need to know from research on systemic reform)

- outcomes are changed by process
- value of theory shown in example of course taking and achievement

Across all three areas, we need a better theory linking the process of systemic reform, or policies, with outcomes. Theories that link process and outcomes in education, or any area of social science, are valuable and rare. It took a long time, for example, to establish the connection between course taking and achievement to the point that we now consider course taking an intermediate outcome, as much as a process variable.

Key Process Components of Successful Systemic Reform

A change strategy controlling

- local politics, resources, and administration
- content and pedagogy
- school improvement
- public acceptance
- longitudinal data on curriculum and achievement

If we look across sites attempting systemic reform, the more successful sites have a change strategy involving control over (and through) the power structure, curriculum and teacher development, school improvement, public acceptance, and data on incremental changes in curriculum and achievement.

Critical Flaws in Real Change Strategies

- political, ideological, racial conflict
- political vision without pedagogical delivery
- school improvement without scaling up
- external accountability without internal commitment
- public rejection of professional agendas
- lack of resources, limits of volunteerism

- absence of planned sequence of classroom change

On the other hand, unsuccessful sites have one or more of a familiar list of flaws: lack of political integration and will, political vision without a pedagogical delivery structure, school improvement or teacher enhancement projects with no realistic strategy for scaling up, external accountability or technical assistance without internal buy-in, unmanageable public controversy, perhaps over a previously low-profile professional agenda, lack of resources and the limits of volunteerism, and the absence of a planned strategy for incremental change in schools and classrooms.

Implications for Research and Systemic Initiatives

- clarifying the profiles of successful process, e.g., policy plus delivery, networks plus scaling up
- setting initial conditions and incentives
- matching resources to vision: no "moonshots on a shoestring"
- cost-effectiveness of partially successful systemic reforms?

- research question: enough real success for the cost?
- skepticism plus sensitivity to small variations

If the conditions for successful systemic reform are really much more stringent than previously suspected, what are the implications for any policy seeking to expand the scope of such initiatives? First, the profile of success would need to be clarified, so that sites could decide whether the necessary elements were in place. Realistically, there probably would be several profiles, such as policy plus delivery and networks plus scaling up.

These profiles would have to be translated into initial conditions for participation, but if we want anyone to accept the conditions, the incentives for participation should match the intensity of the commitment. If we expect that a full-scale systemic reform will be built from scratch, seri-

ous attention should be given to the resources that could support this kind of effort, instead of expecting a "moonshot on a shoestring."

Perhaps the most interesting bottom-line policy question is the cost-effectiveness of partially successful systemic reforms. Suppose that nothing like full-scale systemic change will be produced, in the sense of wholesale replacement of the curriculum and substantial gains in achievement for all children. But also suppose significant progress in both process and outcomes, namely, that a system has been built that is capable of producing continuous improvement in curriculum and instruction with assurances of socially valuable change in curriculum and achievement. Then suppose that the initial investment is not very large, as is usually the case in Systemic Initiatives.

The situation represents failure of systemic reform in the large sense, but success of a systemic-type reform as judged by cost-effectiveness. The key research question would be when there are enough pieces of a complete reform to produce some significant results at a relatively modest cost, in other words, when a reform crosses some as yet undetermined threshold of cost-effectiveness.

Nothing is more common in policy than ambitious rebuilding projects being scaled down to modest home improvements (except perhaps for complete failure), and we should not lose sight of the broader objectives, because they do appear feasible in some places. On the other hand, I know of nothing in policy that should prevent us from accepting cost-effective policies, especially in this vital area of social policy. Some commentators suggest that systemic change is always long-term, and progress is recognizable from the constant vector of change over time and across institutional locations, rather than from any rapid, wholesale transformation. At base, it is the old question of the half-full or half-empty glass, with the key questions being just how half-full and what is the cost of the drink. Answering those questions will require a kind of research that is appropriately skeptical of optimistic claims yet sensitive to the importance of small, but important, variations in actual performance.

Susan Fuhrman
Director, Consortium for Policy Research in Education
University of **Pennsylvania**

My remarks focus primarily on the process of standards-based systemic reform. I trust that other speakers will address the accumulating evidence on the effects of these reforms. For example, we know that many teachers are at least aware of new policy directives and are favorably disposed toward them. In some schools and districts, important changes in teacher practice, such as increased use of real books and stories instead of basal texts and more hands-on activities in science, have been noted. CPRE researchers saw positive effects on 4,800 students in transition high school mathematics courses linked to ambitious NCTM-like standards. Students in California's Math A courses and New York's Stretch Regents courses were much more likely than general-track students to complete a minimal college-preparatory sequence by the end of high school. And students in transitional mathematics classes posted better achievement test scores than students in general mathematics (although they still lagged behind students in college-preparatory mathematics classes). We are also finding that professional development directly linked to the course content has contributed to changes in practice and improved student achievement.

My choice to focus on the process of enacting and implementing standards-based reforms reflects my background in political science, but it is also appropriate to the current state of reform development. The reforms are not yet in place in most states and districts; they are still under development. Their story is still being written. Further, many of the most important challenges facing reformers, as well as lessons about the

progress of these reforms, concern the difficulties associated with the reform process and politics.

I will focus on seven lessons about the progress of standards-based reform. I draw primarily on CPRE research in nine states (CA, CT, FL, GA, KY, MN, NJ, SC, TX) and 25 districts between 1990 and 1995 (Massell, Kirst, & Hoppe, 1997).

1. Standards-based reforms continue to make progress, despite changes in leadership and political turbulence. Although there was opposition to the reforms in each of our states, particularly in the 1994-95 period, they were not dismantled. Despite the vocal, and often virulent, objections of religious conservatives who asserted that standards interfered with the prerogatives of families and of antigovernmental forces who saw standards as infringements on the authority of local schools, standards development continued. In well-established American tradition, new reforms, such as charter schools, were developed to respond to those who advocated devolution, and they were simply added to the books while standards reforms continued to be developed and implemented. Political rhetoric focused on the newer reforms, while under the surface, and more quietly than when they were originally championed, standards documents, new assessments. and related policies were promulgated and the slow process of classroom implementation began. Standards policies were modified in response to opposition, as noted below, but the idea of standards-based reform continued to be robust.

2. Much of the continuing momentum behind standards-based reforms can be traced to the activity of nongovernmental forces. Professional associations, networks, and collaborations, some national and some state-specific, were important sources of support and expertise. For example, standards developers at both state and district levels drew on national documents and examples from other states. Sometimes nongovernmental organizations provided revenues. This is the case in eight districts undertaking standards-based reforms with the help of the Pew Charitable Trusts. Certainly, the associations and partnerships enhanced the legitimacy of standards-based reforms. Most national or regional gatherings of education policymakers became occasions to highlight the record of states and districts considered "in the lead" in these reforms and to draw implications from their experiences. These meetings, and the diversity of groups-from unions to business leaders-supporting standards-based reforms, were very reinforcing.

3. The federal government was an important source of support. Goals 2000 funds were useful in many states, particularly for subsidizing professional development. The standards framework of the Improving America's Schools Act appeared to lend greater legitimacy to state efforts, but it was just taking shape when we were last in the field. We have plans to examine its influence over time. In the context of this Forum, the critical contributions of the National Science Foundation's Statewide Systemic Initiatives must be noted. The SSIs in seven of our nine states (CA, CT, FL, GA, KY, NJ and TX) were responsible for developing the mathematics and science standards. And the SSI can sometimes be credited with giving standards reforms an essential boost. For example, Georgia's subject-matter revisions in most areas had been stalled for years, but thanks to the SSI, work in mathematics and science progressed.

4. That support for standards reforms came from many sources was essential for its political survival, but the very diversity of supporters made achieving coherent direction for education more difficult. Standards were supposed to represent agreement on what students should know and be able to do, but the many sources of standards-national associations, local and state development committees, new specifications developed by test publishers-frequently varied in instructional vision. Policymakers and educators are drawing on multiple sources in developing their own versions of standards, and an important topic for future research is the coherence of the resulting products. For example, contrary to the fears of conservative critics, local educators were not circumscribed by state standards. In most cases, they used state frameworks as only one source of their own standards and found state standards too vague to be really useful. They developed their own frameworks, using many examples and models.

5. An additional reason to worry about the coherence of emerging standards documents is the trend toward adding "'the basics" back into state standards. Part of the political and public opposition to standards centered on the new-fangled notions of learning and pedagogy incorporated in early standards documents. The new approaches seemed to threaten traditional skills and common-sense notions of what school was about-reading, writing, arithmetic, spelling, etc. To respond, many states sought more "balance" in their frameworks, for example, by incorporating phonics into language arts standards that previously had a whole language orientation. Similarly, emerging state assessments were modified to add multiple-choice items back in and to include norm-referenced as well as criterion-referenced components. Will "balance" mean more thoughtful integrations of approaches or simple aggregations, with traditional practices on one page and new ideas on another? Future research should attend to this question.

6. Early on in the standards reform movement, it became clear that the work of **standards**-based reform is extremely demanding. The idea of making change in virtually every aspect of policy-curriculum, assessment, teacher preparation, professional development-not just at once but in a coordinated fashion so they all linked to standards, represented a huge challenge to a political system that was expert in incremental, disjointed change. Over time, the overwhelming demands of standards-based reform have become even clearer, as has the need to sequence reform components to make them manageable and feasible. Teachers are furious when new assessments come on line before the standards, to which they are supposedly tied, are developed and before any curricular materials are available. They are even angrier when such new assessments are incorporated in accountability systems that carry consequences. Yet such disconnects in sequencing are not uncommon at the state and district level. In the states we studied, some that took more incremental, more step-by-step approaches to reform have had less upheaval than some of the previous reform leaders--one reason may be that the incremental states were able to

be more deliberate about each step and plan more carefully about how to bring things on line.

7. The most important factor influencing the progress of reform development and ultimately whether standards reforms can support meaningful changes in teaching and learning is the capacity of the system. Capacity is essential at every level. States must manage complex processes of standards development that balance public and professional input, choose or create assessments that balance needs for adequate information for parents and teachers and the desire to model good instruction through challenging items; revise licensing systems; support schools and districts in curriculum development and professional development-and this is just some of what is required. Districts must do much of the same, focusing on daily support to schools going about the difficult process of improving instruction. Teachers and administrators must develop new approaches to teaching and learning, while frequently managing new site-based governance policies and developing ways of relating to one another and to parents. Parents are expected to support new, more challenging expectations for their children, understand and respond to new, very complex accountability systems, and see that schools are supported with adequate resources. Everyone needs help.

There are some signs that the need for greater capacity is recognized. For example, Connecticut, Florida, and Minnesota were among the states actively revising their teacher certification processes to support more challenging visions of instruction. Many states have orchestrated or supported teacher networks that group teachers by grade level or subject in order to provide continuing support; some states and districts are facilitating school access to reform designs and technical assistance.

But capacity-building efforts to date are dwarfed by the need and undercut by omissions and contradictions. For example, in our states, state and local central agencies continue to be downsized. At the state level, a number of agency budgets were cut by about 25% during this period, coming on top of earlier cuts during the 1980s. Little has been done to see to the quality

of professional development available to teachers. The prevailing mode is to push money down to the school or district (that is, where the state is making a new investment in professional development at all) assuming that demand for professional development will be generated by new standards and assessments and that the demand will in turn generate supply. This turns out to be true; suppliers do show up. But much of what is available is of very low quality, and few places have tried to develop criteria to guide selection of professional development providers or to educate consumers about good choices. And, finally, a crying need is for curriculum and materials, reflecting the standards, that can be used for daily instruction. Increasingly, teachers themselves are questioning the romantic notion that day-to-day curricula should be developed by collaborative school or department-level groups. They have neither the energy nor the desire to create from whole cloth when others might have invented

very usable and good materials. We need better methods of sharing what is available and more attention paid to development.

Systemic reform has proven to be both hardy and difficult. Challenges multiply, but the basic logic has enormous appeal and staying power. It is important to act on the lessons we are learning about the challenges, for example, by supporting more extensive efforts to develop and disseminate excellent curricula. It is also important to keep tracking the results. Evidence of positive effects will be critical to the reform's staying power over time. We must show that the hard work is worth it.

Reference

Massell, D., Kirst, M. W., & Hoppe, M. (1997). *Persistence and change: School reform in nine states.* Unpublished manuscript, University of Pennsylvania, Philadelphia, Consortium for Policy Research in Education.

*Jane Butler **Kahle****
Condit Professor of Science Education
Miami University

Although much has been written about the nature and policies of systemic reform by policymakers, those papers provide only a partial vision of systemic reform-one from the outside in. Less has been written from the field, i.e., from those who are actively trying to promulgate reform either in the classroom or at the state level. Even less has been written about changes in teacher practices and in student learning-yet, without those changes the reforms eventually will wither and fade away. The encompassing nature of systemic reform provides critical roles for national and state leaders, for professional

groups, and for individual teachers. Experience with several systemic initiatives in one state, Ohio, forms the basis for the following discussion of the challenges to reform and the changes needed for success.' My discussion focuses on the challenges faced by one state's efforts, on evidence of changes in teaching practice and in student learning, and on the meanings that may be drawn from that evidence. Specifically, Ohio's reform was characterized by the following parameters. It

- developed a regional infrastructure to support and sustain reform.
- focused on teacher professional development that emphasized content, provided in-depth experiences in inquiry and problem solving, and extended over one year.
- targeted middle schools for equity and economic reasons.

- included substantive involvement of university scientists and mathematicians.
- avoided K-12 curriculum development.

Four years into the reform, changes in parents' perceptions of science, in principals' support for inquiry instruction, in teachers' practice of inquiry, and in students' learning of science were assessed. These findings articulated both the challenges and changes of the systemic initiative's efforts.

What Are the Challenges?

The challenges describe five aspects of systemic reform that must be addressed and aligned. Because each one poses risks to a part of the education community, varying levels of success have been reached. However, the lessons learned in attempting to meet the challenges provide insights and directions for the future.

Challenge One: Sustained professional development of a validated model can produce a culture shift in participants, but it is costly and time intensive. Given a state cohort of over 7,500 teachers per grade level, there are neither the human nor the financial resources to reach more than a small fraction of the target audience within a five-year period.[2]

Lesson Learned Not only are human and financial resources limiting, but the pool of teachers who can, or will, undertake sustained professional development is limited. Teachers who may be characterized as needing professional development the most---ones in poor schools, ones with general licenses, ones with few courses in science or mathematics, ones teaching out of certification areas, ones who are disenchanted or disenfranchised-do not readily volunteer for a rigorous summer of mathematics or science. Rather, they must be reached in their communities and schools, the academic program must be at the level at which they teach science and/or mathematics, and the materials used must be directly applicable in their classrooms and with their students.

To meet this challenge, research validated curricula were identified. Next, teachers, who had had at least one year of professional development, offered local, 40-hour workshops for their peers. Districts supported the teacher-instructors and often required all science or mathematics teachers at the targeted grade to attend. Ostensibly, the workshops were to help teachers learn to use standards-based curricula; in reality, much mathematics and science was learned.

Challenge Two: Any reform has a limited and unique function. Although it must offer resources that are not available during its lifetime, those resources eventually must be assimilated into the ongoing educational system.

Lesson Learned At the beginning of the reform, both the Systemic Initiative and the Ohio Department of Education divided the state into eight professional development regions. Two sets of centers were established that tested two different paradigms for professional development. The Systemic Initiative insisted upon regional collaboration before identifying and supporting its regional centers, which, then, delivered the six-week institutes. They were taught by regional academic leaders (both outstanding teachers and Ph.D. scientists and mathematicians) who were available during the school year to assist teachers in their classes, to work on curriculum teams, and to provide local workshops. The Systemic Initiative's model was successful beyond anyone's expectations. The Department's less costly model involved short-term, usually three-day, workshops with limited classroom follow-up. The teaching staff rotated, and many were imported for a few days work. Furthermore, the selection process for the center was competitive, resulting in antagonism between the units funded (districts, colleges, county offices) and those passed over.

It was obvious that the two regional units needed to merge, and merger has occurred. The mergers have been slow and fraught with difficulties; for example, how to retain the sustained professional development model within the merged regional center. However, systemic reform is about taking chances, about building consensus, and about moving ahead with the results. In a large complex state, regionalization is needed, and it is better to have one unit than

none, and two units are not sustainable. Furthermore, the merged center is fully assimilated into the state's educational system.

Challenge Three: Systemic Initiatives have underestimated the difficulty of getting mathematicians and scientists-at all levels-to work together, the difficulty of shifting university faculty from teaching by lecture to inquiry, as well as the difficulty in communicating between campuses and across disciplines. In Ohio, two separate groups of mathematicians developed mathematics by inquiry courses, because that process was more efficient in terms of time than the collaboration necessary to develop one course. Later, when middle school mathematics and science teachers were both in institutes on the same campuses (often in adjacent classrooms), both groups actively resisted working on integrated units or comparing strategies across disciplines.

Lesson Learned It takes time and effort to encourage collaboration. Yet, without it, a reform is only pockets of change, not systemic. Therefore, the Systemic Initiative instigated collaborations with Ohio's three Urban Systemic Initiatives and the Appalachian Rural Systemic Initiative. Further, it identified collaborative relationships with Ohio's Mathematics and Science Coalition, the Parent Teachers Association, and many regional and local businesses and foundations.

Challenge Four: Although Ohio's Systemic Initiative focused on individual (or groups of) teachers, a school is a more viable unit of change. Teachers need a support system for the reforms they are initiating in their classrooms. A school focus also produces the visibility to attract external market-driven resources that may continue the reform after the funding period. Changes in a school's science program or science department-with documented improvement in results-is a phenomenon that may be quickly communicated to parents and policymakers.

Lesson Learned As part of the assessment, described below, brief site visits were made to 12 schools, primarily in poor urban or rural areas. Both the observational and questionnaire data gathered suggested that there were greater changes in learning environments, in teaching

practices, and in student outcomes in schools that had a critical mass of reform teachers (up to one-third of the mathematics/science faculty), compared to schools with only a few teachers, or an isolated teacher with the sustained professional development. A supportive group of teachers is especially important because of the high mobility of principals.[3]

Systemic or standards-based reform requires a critical and self-sustaining mass of teachers in a school. The solution was twofold: first, districts were requested to require or provide incentives so that all appropriate teachers would be involved in local professional development; and, second, the intensity and depth of the professional experiences were moderated with caution and some trepidation. (See Challenge One.)

Challenge Five: All initiatives that are systemic in nature have important research/development and dissemination/support roles. Well-researched and validated professional development packages, such as Physics by Inquiry (McDermott, Shafer, & Rosenquist, 1996) have sustainability independent of the instructors. Such packages can be assimilated very quickly into existing delivery systems, such as a state's regional centers. Where such packages do not exist (or lack research validation), the Systemic Initiative must take the professional development packages through carefully controlled field tests and refinement activities to document their value and sustainability through research studies.

Lesson Learned There is neither time nor money to do it all. Ohio has learned two lessons. First, find and use the expertise of others, and, second, assess progress and outcomes in order to refine and improve your efforts. As mentioned earlier, in the fourth year of the reform, the Systemic Initiative undertook a major study to describe the landscape of science and mathematics education in the state. The intent of the assessment was to tell the reform story in terms of changes in learning environments, in teaching practices, and in student learning. Because of the focus on equity, the schools selected to gather student achievement data were in poor urban or poor rural districts.

What Are the Outcomes?

Four years into the reform, a comprehensive assessment of learning environments; teaching practice; principal, teacher, and student attitudes as well as student learning was undertaken. The study consisted of two levels and involved the collection of both quantitative and qualitative data. Level one consisted of a random sample of 126 schools, drawn from all the schools in the state that had at least one teacher who had completed the initiative's professional training. At level one, principals and all teachers who taught either science or mathematics (grades 5 through 9) completed questionnaires concerning classroom instruction, administrative support, and parent influence as well as issues of school change. Level two consisted of brief site visits to 12 (from the original random sample of 126) selected schools. In those schools, students and parents also completed questionnaires, students completed achievement tests, and principals, teachers, and randomly selected students and parents were interviewed. At each site visit school, a teacher who had had the sustained professional development (reform teacher) was matched with a teacher who had not had that experience (nonreform teacher). In addition, a randomly selected class of the reform teacher was matched with a comparable class of the matched nonreform teacher. Changes were identified by comparing the responses of these matched groups of teachers and students.

Using 1990 and 1992 public release items from the National Assessment of Educational Progress (NAEP), science and mathematics achievement measures were developed by teams of faculty, teachers, and regional leaders. Test items focused on process, not product, because the reform's goals were to increase conceptual understanding as well as skills needed to interpret and use scientific and mathematical information. In Miller's (1996) discussion of barriers to systemic reform, he notes that there is often a disconnect between the practical paradigm of reform (focused on process) and the technical paradigm of education (focused on product). This disconnect was avoided by developing new achievement measures. When possible the questionnaires for principals, teachers, students, and parents contained the same questions, phrased appropriately. That strategy allowed comparison of responses across groups. For example, did both teachers and students respond that manipulatives were used at least once a week? The results indicate that students responded similarly to their teachers concerning instruction in reform and nonreform classes. That is, students in reform classes significantly more often talked with each other about the subject, had to support their claims, and were encouraged to ask questions. Interestingly, significantly more students and teachers in reform classes reported that their principal had learned to accept classroom noise.

When the ways in which students learn were examined, interesting and significant differences were found between reform and nonreform classes. Students in classes taught by reform teachers significantly more often wrote about how they solve problems, solved problems in small groups, and used hands-on manipulatives. Those strategies are recommended both by the *National Science Education Standards* (*NSES*; National Research Council, 1996) and by the research literature concerning strategies to improve the participation, attitudes, and achievement levels of girls and minority students. One of the six-week content courses incorporated the computer as a learning tool; another heavily used graphing calculators. Further, both the NCTM standards and *NSES* argue for the incorporation of technology into science and mathematics lessons (National Council of Teachers of Mathematics, 1989; NRC, 1996). However, neither practice with, nor information about, the efficacy of technology in promoting learning affected the use of calculators or computers. The lack of appropriate equipment and software remains a major challenge to implementing the reform agenda.

Briefly, there were significant differences in science achievement in favor of reform classes, as measured on the Discovery Inquiry Tests. It is important to note that it was a low-stakes test (grades were not affected) and that it focused on a student's ability to interpret information and on conceptual, not factual, understanding. Because of the reform's focus on equity, the results also were examined for any gender, race, and/or

group (reform versus nonreform) differences. Those analyses revealed interesting patterns of achievement; for example, minority students (in this case, African American) in reform classes scored significantly higher than their peers in nonreform classes. In fact, African American seventh and eighth graders in reform classes scored as well as white students in nonreform classes. Because the data were collected from reform and nonreform students taking the same type of class (e.g., general science, introductory life science) with a "matched" teacher in the same school, economic differences were not a major factor.

Further, in science classes taught by teachers involved in the systemic reform, both African American and white females scored higher than the males in their racial group. In reform classes, white females, shown by other research to be the group most socialized away from science (Campbell, 1991; Campbell & Connoly, 1987; Kahle & Damnjanovic, 1996), scored higher than white males or African American females or males. When science test scores were subdivided into physical science and life science items, females in classes taught by the reform teachers scored higher than males on the physical science items. These fmdings contrast with those of other studies. Using a large national data set (NELS: 88), gender differences have been found in achievement in physical science, but not in biology (Burkham, Lee, & Smerdon, 1995). These results may be the first time that a gender achievement gap in physical science, favoring girls, has been reported. Clearly, the type of teaching observed and recorded in responses to student and teacher questionnaires-more use of manipulatives, more time to talk about science, more opportunities to write about science, increased use of cooperative learning groups-has affected achievement, particularly for students who have been underrepresented in science.

A Further Challenge

Once the data were collected and analyzed, the challenge was to distribute the findings widely in an accessible way. Over 10,000 copies of a small, easy-to-read publication, the Pocket Panorama (Kahle & Rogg, 1996), have been dis-

tributed across the state and nation. Because most state departments of education do not have either the time or the expertise to perform large-scale research and dissemination activities, documentation, validation, and dissemination of change and of best practice provide unique and important roles for Systemic Initiatives. Such activities are critical for the public's understanding and acceptance of systemic reform. Indeed, research and dissemination may be the key roles for externally funded reform initiatives within a state. The final lesson learned is that research or assessment without dissemination benefits only those who are already involved in the reform. Dissemination of findings in practical and easy-to-use ways informs others of the initiative's success and invites them to become active participants in it. Further, assessment, coupled with wide-spread dissemination, provides the basis for successful reform strategies to become sustained through the existing educational system.

All parts of a reform must be addressed and work together if the results are to be systemic. The challenges in one state led to alterations in its reform strategies; those alterations, in turn, led to wider participation and acceptance of the reform. Although the changes described in the assessment cannot be directly attributed to the professional development and support strategies that were part of the reform, the findings suggest that improved learning is associated with improved practice that is initiated through sustained professional development.

Notes

1. Ohio was one of the first ten states to receive National Science Foundation funding for a Statewide Systemic Initiative. Further, Ohio's three cities that were eligible for Urban Systemic Initiative funds, Cincinnati, Cleveland, and Columbus, have been awarded grants. Ohio also has five counties in the Appalachian Rural Systemic Initiative.
2. NSF Systemic Initiative awards are for up to five years, although contracts are renewed annually. Three of the first cohort of states were terminated during the five-year cycle.
3. In Ohio, over 50% of principals are in that position in a particular school for four years or less.

References

Burkham, D. T., Lee, V. E., & Smerdon, B. A. (1995). *Gender and science learning early in high school: Subject matter and laboratory experiences.* Unpublished paper, University of Michigan at Ann Arbor.

Campbell, J. R. (199 1). The roots of gender inequity in technical areas. *Journal of Research in Science Teaching, 28,* 25 1-264.

Campbell, J. R., & Connoly, C. (1987). Deciphering the effects of socialization. *Journal of Educational Equity and Leadership, 7*(3), 208-222.

Kahle, J. B., & Rogg, S. R. (1996). *A pocket panorama of the landscape study,* 1995. Oxford, OH: Miami University.

Kahle, J. B., & Damnjanovic, A. (1996, April). *Effects of inquiry teaching on the achievement levels of* urban middle school science students by race and sex. Paper presented at the annual meeting of the American Association for the Advancement of Science, Baltimore, MD.

McDermott, L. C., Shafer, P. S., & Rosenquist, M. L. (1996). *Physics by inquiry.* New York: Wiley.

Miller, K. W. (1996). Paradigmatic school philosophies as barriers to school reform. *Science Educator 5*(1), 1-6.

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

National Research Council. (1996). *National science education standards.* Washington, DC: National Academy Press.

Uri Treisman
Professor of Mathematics
Director; Charles A. Dana Center for Mathematics and Science Education and the Texas *SSI* University of Texas at Austin

My purpose in these remarks is to comment not only on what we, the Dana Center/SSI, have learned, but also on what we are trying to learn about systemic reform as it is practiced in Texas and other large states. I have organized my remarks around the first four drivers of systemic reform as enshrined in the NSF's assessment and monitoring system for its systemic initiatives. In my view, these drivers do in fact capture the essential dimensions of the Dana Center/SSI's work and provide quite a useful framework for describing the complex of interlocking projects and initiatives that we have found necessary for moving the system forward in ways that respect our fundamental commitment to equity.

I. Policy

The extent to which state or local educational policies affect practice is determined by the nature and sharpness of the teeth associated with them in the relevant accountability system. In this sense, managing the policy side of systemic reform can be thought of as a kind of educational orthodontics-requiring lots of steady pressure at just the right place constantly applied. In the work in question, generating and maintaining the required steady pressure requires the development and maintenance of relationships with individuals and groups whose interest in mathematics and science is quite minor. But more on this later, in my discussion of public engagement.

It is my observation that in many states the policy focus of the mathematics and science education communities has been almost exclusively on shaping particular policy documents, i.e., on setting the content of state curriculum frameworks. Too little attention has been paid to the mechanisms through which these documents have their influence-an essential issue in systemic reform.

In Texas, for example, the processes for successfully developing state curriculum frameworks and for getting them adopted by a conservative State Board of Education were remarkably similar for mathematics and science. Unusual and sometimes unnatural coalitions had to be built; a broad perspective had to be inculcated in individuals who are both skilled in and fond of fighting over arcane and often politically meaningless issues. Incentives had to be created for various leaders to support positions that differed

in minor ways from those of the professional organizations that give them authority and special influence. Most challenging, of course, was productively engaging the opponents of **standards**-based reform, who often argued vociferously about the fine details of the proposed curriculum frameworks but did so in ways that were actually intended to shake confidence in public education.

Yet, these now-adopted mathematics and science frameworks have very different effects on practice at the school and district levels. By law, each constitutes the legal base of any state examinations in its subject area. But, herein lies the crux. While mathematics is tested at many grade levels, science is tested only at grade eight. Moreover, state ratings of school performance, which are widely published and attended to **by**—among others-the real estate industry, are influenced heavily by mathematics scores but not at all by science scores. The effect is profound. School administrators spend heavily to help their students learn mathematics; they spend no more than is absolutely necessary on science unless it is a special interest of theirs, their spouse's, or of an all-too-rare group of their teachers who are both science people and effective negotiators.

This unpleasant reality has consequences for the kinds of networks one builds, the state-level professional development one offers, and so on. In mathematics, one can count on felt need to drive action. In science, one dependson arguments about the general good, the future, and so on-arguments that, in the absence of a clear threat to the nation's security, are very, very hard to make.

Whereas, in mathematics, the Texas SSI can support (and **find** support for) the implementation on a large scale of high quality curricula such as those developed with NSF support, in science, the primary challenge is to organize political support for the inclusion of science in the accountability system. In practice, one confronts a Catch-22 situation. The legislature will only move to include science in the accountability system if it believes that the school system can meet the challenges it will generate. Few legislators want their constituents' schools to face yet another challenge they cannot overcome. These same legislators will only support increases in

funding for science if they feel an overwhelming demand from their constituents. This demand is not there because it is not generated by the accountability system. And so on.

In such an environment, the natural but very long-term strategy is capacity building-creating and supporting the kinds of statewide professional development networks, **collaboratives,** etc., that will generate reasonable confidence that higher standards can indeed be met. Until then, science standards will serve as a banner and a goal. Mathematics standards will be drivers of local action.

Another and even more important question centers on equity and the allocation of scarce resources in the educational system. It is widely assumed in our community-indeed it is the mantra of systemic reform-that standards-based educational policy is a friend of, if not a prerequisite for, equity. Roughly, the most common argument for this position is that explicit standards **demystify** the system by making public exactly what children must know and be able to do in each curricular domain. In being public, standards enable the transformation of a system putatively based on ability to one that is manifestly based on effort. The needed catalyst (which, of course, is rarely present) is adequate resources so that the playing field is approximately level for all students, no matter their family's financial or educational resources.

In Texas there is a particular and, in my view, all-important feature of the state's educational accountability system that may turn out to be a necessary and sufficient condition for greater equity to be an outcome of standards-based reform. Specifically, in Texas, schools and districts are given one of four ratings ranging from "low performing" to "exemplary," based mostly on student performance on state examinations (Texas Assessment of Academic Skills) whose content is directly defined by the state curriculum **frame**works. School data are reported for all students and separately for four subgroups: African American, Hispanic, White, and Economically Disadvantaged students. To achieve a given rating, the scores of every subgroup must exceed a particular cut-off score, which rises over time on a predetermined schedule. Sanctions for low **per**-

formance include public hearings and possible takeover by the state. In short, school and district ratings depend on the success of everybody's children.

The effects of this rating system in Texas have been extraordinary and are increasingly well known. Ten years ago, Texas student performance mirrored that of the South-it was abysmal. Today, the average scale scores of each of the above mentioned populations on NAEP mathematics tests are near or at the top of national rankings. Indeed, on NAEP Grade 4 Mathematics scores, Texas is tied for sixth place-only slightly behind Maine, Minnesota, Connecticut, Wisconsin, and North Dakota. Texas **fourth-** grade scores now exceed those of such traditionally high performing states as New Jersey, Montana, Michigan, Colorado, Vermont, and Utah. Amazing. The mechanism of action is clear and easily visible from the ground floor where I spend most of my days. Today, building administrators in Texas are far more likely than in the past to direct their resources to supporting high achievement among groups of students who they believe might endanger their school accountability ratings. The focus on mathematics (still the dominant reason for low school ratings) startles visitors from out of state. Indeed, this feature of the Texas accountability system, in concert with the recent changes in the Improving America's Schools Act, has set a new benchmark for the speed at which a large state's performance data can change.

In systems in which school ratings depend on average (nondisaggregated) scores, the natural strategy for a building administrator would likely be to direct resources to those children whom he or she believes can be most easily educated. This strategy would keep the performance gap among ethnic groups large and might increase it. Indeed, Texas is one of the few states where differences in performance in mathematics among ethnic groups is decreasing-at least at the K-8 levels, where the accountability system is strongest.

In short, there is no evidence that **standards-** based accountability by itself is a friend of equity. There must be significant incentives for addressing equity, but unless there are explicit and substantive consequences for not educating to

high levels everybody's children, unless the parts of the accountability system that deal with equity have teeth, it will continue to be minorities who are underserved by the educational system.

II. Curriculum and Instruction

Perhaps the greatest surprise to emerge from our work in the Texas SSI has come from our research on effective school responses to our **high-** stakes standards-based accountability system. We have now studied quite carefully 26 high poverty/high minority elementary schools whose students score well above the state average at every grade level on state examinations in both mathematics and reading. These are schools whose performance profiles are similar to those of typical schools in the wealthiest suburbs of Texas. We have also studied the 13 Texas high schools in Title I feeder patterns, i.e., in high poverty communities, with the highest average scores on the state's End-of-Course Algebra examination.

What did we **find?** First, we found enormous variations both in how schools organized themselves to succeed and in the particular curriculum and instructional approaches they adopted. Yes, some of the high-performing high schools use Saxon's Algebra. Others use books that would make any NCTM leader happy. At the elementary level, we found as many examples of schools that espoused direct instruction/explicit **phonics-** based approaches as we did schools that identified themselves with "constructivist" approaches. The eclectic pragmatism of these schools impressed us. The teachers were clearly more committed to nurturing students than to climbing on bandwagons.

It is, of course, one thing to espouse or to identify with a philosophy and quite another to actually practice it. In so many cases we observed mathematics teachers supplementing their textbooks in ways that compensated for perceived limitations. I took some pleasure in observing "Saxon" teachers assigning worksheets of interesting and challenging problems to prepare their students to do well on state and AP examinations. I also observed in several schools UCSMP (University of Chicago School Mathematics Project) teachers supplementing their

texts with structured drill and practice. Good teachers in both settings had a finely **tuned** appreciation for the value of many instructional strategies.

Second, we were impressed by the quality of school leadership and its ability to focus instructional staff energy on achieving performance. School leadership created the conditions that allowed teachers to organize themselves to help students master challenging material. It is true that the school staff had "high expectations," but these expectations were as much for their own practice as for their students' achievement. Teachers' high expectations were manifested in carefully thought out procedures for proactively addressing problems and in their readiness to take whatever steps were necessary to ensure each student's success.

Good administrators facilitated effective teacher organization by conveying to their teaching staffs the belief that collaboration would be respected and built upon. Their teachers knew that a certain amount of time working together on curricular alignment and other curricular, school, or student problems would help them achieve their individual teaching goals. They knew that collaboration would yield identifiable benefits for themselves and for their students. In schools with weak administration, teachers responded rationally. They knew that little could come of cooperative work and, thus, they retreated to their individual classrooms where, at least, they had control.

If a particular curriculum choice mattered in these effective schools, they mattered in that the choice made supported a coherent school culture that reflected certain ambient community values. The choices were not controversial ones in the context of the community setting of the school. Indeed, the school leadership understood that effective instruction depended on minimizing the distance between school and home.

The recognition that there are many ways for schools to help their students master the challenging content defined in the state's curriculum frameworks has allowed us to build extremely strong and diverse coalitions in support of the state's standards-based accountability system. It has allowed us to avoid unnecessary fights with commercial interests. We trust that ratcheting up content standards and leaving them unfettered by pedagogical theories will produce incremental improvements both in what is taught and in what is learned. We suggest that it is this approach whose effects are reflected in Texas's current NAEP scores.

My experience in managing systemic reform has led me to believe that those who devote their time to determining whether or not district X's curriculum is more constructivist than district Y's are not helping us. Even more dangerous are those who advocate for extreme forms of performance assessment as part of the state's accountability system, despite the fact that there is a zero probability that such assessments would survive legal or political challenges. I am suggesting that some in the "reform" community seem more committed to constructivism than they are to equity. These individuals need to reflect on their goals. We can mobilize a broad coalition in support of educating everyone's children to high standards. We cannot build a strong coalition in support of constructivism or any educational philosophy. The rhetoric of "more constructivist than thou" must end.

This is not to say that the adoption of **high**-quality curriculum a la NSF cannot lead to great improvements in student learning. In our efforts to implement the Connected Mathematics Project in Texas districts, we have seen impressive gains in student test scores on many kinds of tests in dozens of schools. What must be said, however, is that these schools chose CMP and that they receive very high levels of sophisticated outside support. Time will tell whether the changes these schools are making can be sustained when concentrated external support vanishes and when teacher mobility and retirements bring growing numbers of new teachers into these schools.

III. The Convergence of Resources

Systemic reform depends on the adoption of state policies that (1) direct the attention of a system's decision makers to what is deemed important and (2) define the incentives and disincentives that shape the priorities of these decision makers. But good policy is not in and of it-

self enough to bring about change in a system as complex as education. One must also create mechanisms for directing and focusing the system's resources, financial and human, in ways that reflect that system's new priorities. This is the heart of the NSF driver concerned with the convergence of resources.

One might think that this focusing task would be easier if the agent of systemic reform were located within the bureaucracy that is the target of reform. These bureaucracies do have, at least in theory, some discretion in how money is spent and how professionals are deployed. And, indeed, many school districts and state education agencies are the recipients of NSF support to lead systemic reform in their regions.

I suggest, rather, the counterintuitive notion that certain entities outside the formal education system may in fact be better positioned than those within the formal system to shape the distribution and allocation of resources. Moreover, I assert that this will be particularly true in periods of significant political change such as our nation experienced in the mid-1990s. Changes in who controls resources may allow policy and reform entrepreneurs to gain significant, if not effective, control of these resources. Changes in the scope and responsibilities of governing bodies may allow reformers to form new alliances that would be difficult to establish in times of status quo.

In Texas, for example, the downsizing of the state education agency and the decentralization of many functions of state government created ideal conditions for the Dana Center/SSI. The SSI was able to take over the management of critical state programs and to successfully win grants and contracts to provide technical assistance to state and federal programs that have a collective budget of over $2 billion. In particular, the Dana **Center/SSI** now manages, with various partners, the state's Internet system for schools, the Federal Region Eight (Texas) Comprehensive Assistance Center for Elementary and Secondary Education Act Programs, the Head Start Collaborative, the **McKinney** Homeless Education Program, the Discretionary Eisenhower Program for Mathematics, and other programs that shape the way schools use categorical funds and support access and curricular improvement goals.

The co-location of these technical assistance support structures in one organization outside the formal K-12 education system has allowed a blending and focusing of resources unlike any in the state's history. The effects of this arrangement on mathematics performance are, of course, hard to nail down exactly, but a rough estimate can be made by examining student performance on state and NAEP reading and mathematics examinations. The improvement of reading scores in Texas has been a major priority of the governor and of many civic and business leaders and leadership groups. And a thousand flowers have bloomed.

Of special interest is the fact that the Texas accountability system puts on schools the same pressure to strengthen reading instruction as it does to strengthen mathematics instruction. There is now a plethora of well-funded **state-level and local literacy initiatives with very little** coordination among them. Yet, while reading scores in Texas have risen significantly, the gains do not approach those in mathematics. The difference in gains in reading and mathematics scores should be a rough approximation of the focusing strength and power of the SSI lens.

The above suggests that it would be useful to study the differential effectiveness of Statewide Systemic Initiatives located outside the formal **K-**12 system, such as those in Texas, Connecticut, Louisiana, Puerto Rico-as well as Urban Systemic Initiatives, such as **the one in El Paso-and** those within the system, i.e., where the district or state agency is the fiscal agent.

One more issue of special importance to both researchers and **funders** is that of scale in systemic reform. In some states, no teacher is more than three hours away by car from the SSI headquarters. But, alas, in Texas, there are schools that are more than twelve travel hours away from my office; London, England, is only eleven hours away. In small states, local leadership development may suffice for shaping resource allocation. In large states, one needs more complex plans that focus strategically on the intermediate agencies within the state system. Those interested in this issue should read our various progress reports to the NSF.

IV. Public Engagement

A systemic approach to thinking about the mobilization of public support for educational improvement requires a careful analysis of exactly which publics one needs as allies to accomplish one's goals. Then, of course, the challenge is to figure out just how to mobilize and support those publics that will ultimately constitute your authority in the public sphere. Again, I refer those interested in these matters to our reports to the NSF. I do, however, want to make to two special points.

First, the most critical public constituency for reform is the community of professional educators. Teachers and administrators in their everyday lives as citizens and as individuals speak with authority to their neighbors about the conditions and realities of public education. How often important legislation has failed because educational leaders have failed to involve and organize classroom teachers, aides, building administrators, and other educators to support change. Properly situated systemic initiatives can play a role that the formal system cannot play for itself without appearing self-serving. The use of systemic initiatives to support educator leadership is a new art, well practiced in Connecticut and some other SSI states. It is an art worthy of study.

Second, we must use our knowledge of informal education to enlist the support of the tens of thousands of individuals who volunteer every day in our public schools. Volunteers can also, as disinterested witnesses, speak with conviction and special power to their neighbors about the reality of schools; they are an antidote to the demagogues who present extreme cases of bad practice as typical of the whole of education. In Texas we have built an extraordinarily powerful structure for supporting volunteerism and, more broadly, service. The opportunities created by the President's Summit for America's Future, by the follow-on program America's Promise, and more broadly by the decision of the new volunteerism movement to focus on children's well being and education are golden opportunities for systemic initiatives. We just must be careful never to confuse charity with justice nor the responsibilities of good citizenship with those of good government.

*Daryl Chubin**
Division Director, Research, Evaluation and Communication
Education and Human Resources Directorate, National Science Foundation

This is a most learned panel and I am not going to do justice to the observations that they have made. What I have tried to do is organize some of what I've heard (and some of what I read in advance) under three headings.

What Have We Learned about Systems and Their Reform?

First, all systems tend toward equilibrium, stasis, or inertia. We know this and there are more disincentives to change than incentives.

Two components that have been singled out this morning repeatedly are teacher unions and higher education, suggesting that we have a couple of strategic sites we might focus on. Despite all of this, as Susan Fuhrman observed, in the face of inertia and criticism, standards-based reform as a movement continues and maybe even continues to pick up steam. And while that's happening, as Uri Treisman just reminded us, systemic reform must leave no one behind. If it does, then it's not systemic, and it's something less than what we aspire to accomplish.

Second, the value of strategic thinking among educational systems leaders can not be overestimated. We at NSF have observed this repeatedly; I think some of you have experienced

that first hand. Bill Chme's list of strategies and flaws ·in them or in their execution is a nice summary of the value of strategic thinking. Let me give you one very brief example: a simple but bold policy change in some localities has been the elimination of remedial courses. For me it is the best expression of the phrase "all children can learn." It is an operationalization of what we mean when we say that we have high expectations for all students.

Third, it's pretty clear by now that systemic reform must be approached both **from** outside classrooms, about which teachers have virtually no control, and from inside, where teachers may lack either preparation, and therefore confidence to teach the kind of content that we talk about, or opportunities to learn about that content and the quality materials and other resources they need to do the job, including technology and technical assistance. It's what Jane Kahle has referred to as "pressure points" and what Susan Fuhrman wrapped into the nice word of "momentum."

What Do We Need to Do in Research on Systemic Reform?

First, we must be vigilant to measure and reward what children learn in the classroom and not punish them for what **they** lack in preparation that they bring to the classroom. This is the fear that is invoked as one criticism of any voluntary national exam. Testing is a totem in this society. The real issue is not the test, but how we intelligently use assessment to raise the achievement of all students. While there has been a call for a national voluntary eighth-grade mathematics exam and fourth-grade reading exam, the idea here is to change the conditions, change the environment, bring to bear all the resources that all of you in this audience have on that effort. That's one of the jobs of the Executive **Office** of the President, and it's one of the things that I'm going to be spending time on. No agency can do this alone. The federal government can not do it alone. We need your help.

Second, we may have learned how to recognize improvements in the performance of a sys-

tem, but we are far from developing robust measures of that performance. I was encouraged by John Witte's paper, which was produced as a draft for the Institute, because I think he has it right. He said, "Simple reporting of student achievement does not provide necessary information about the quality of the system." We need to be able to measure the value added by the system, i.e., the change in achievement the system has produced over time. We can argue about which measures, which tests, but not about which interpretations.

Third, research on systemic reform must occur at the sites of reform, with researchers as partners of other system participants, not as drop-in observers. This partnership will blur the distinction between research and practice, between internal and external, between top down and grass roots, between government and **non**-government. I think that's healthy. There's a great contradiction between systemic reform, which acknowledges everyone is supposed to be working together, sharing credit, and sharing blame, and things like the Government Performance and Results Act, which instructs agencies to trace their impact through the system and then be rewarded or punished accordingly. We may not like that, but it's the law, so we must be responsive.

Fourth, theories of systemic reform are overdue. NSF's systemic initiatives through the six drivers represent a test of one theory of process and outcome. We need to look outside of education to other models and conceptualizations of systemic change. There are other very smart people who have looked at systemic change. From a potpourri of research, hypothesis testing and scenario building, we should develop a **finer** sense of complexity, of relationships between levels of a system and time **frame** for the emergence of different kinds of system changes, barriers to scaling up and a number of other things that have already been mentioned. How many profiles or models of reform are there? Can we identify the distinguishing characteristics of each? At one point we had twenty-five Statewide Systemic Initiatives. We need some number of models of reform less than twenty-five. How do we get that reduced set?

Some Questions for Future Consideration

How do we learn from transitory failure, without stigmatizing the reform site? Many of you may have seen a little note in *Education Week* reporting on the first report card in Philadelphia, even with a picture of Superintendent David Hornbeck, in which the headline was, "All but One School Fails." Think of how rare that event is, reporting failure. We have to be able to deal with that without penalty. How a strategic thinking superintendent responds to that report card is a challenge indeed. We know that reform may be taking root. It will take many years, however, for reform to be recognized and hailed by the public as a success.

How do we adjust or align the expectations of different system participants? This is Jane Kahle's point about the media and Uri Treisman's about explaining to policymakers why something makes a difference. We've got to get better at doing this. Researchers aren't good at doing that because they think this is a research issue. It is not a research issue. Call it communication, call it dissemination, call it translation. We need to do it better.

How typical is the experience of Ohio or Texas that we've heard about here? In this weekend's *Washington Post* there was a front page story by Rene Sanchez about Milwaukee, which was really quite encouraging. There was also a back page letter about Baltimore, which was an interesting strategic piece of thinking. We have to learn from states' and from cities' experiences.

To take one of Jane's pressure points a step further, can researchers skilled at classroom level research, assessment, and evaluation on teaching and learning become systemic researchers? Is this a capability, encouraging the creation of reformed researchers, that funding agencies should expressly cultivate? In the cooperative agreement for the National Institute for Science Education, the expectation is that fellows who come through the Institute will never do business the same way. The Institute is intended to be an incubator for change, change in the way people practice. What is a reasonable expectation, then, for what we should be learning about systemic reform from NSF's Systemic Initiatives and Goals 2000 efforts? Can we walk the walk as well as talk the talk of systemic reform?

Finally, let me end this parade of interrogatives with this thought attributed, much to my chagrin, to the president of the International Association of Professional Bureaucrats. He said, "Sharing ignorance may not lead to wisdom, but it spreads responsibility." I thank the panel for specifying our ignorance and expanding our responsibilities.

The Role of Curriculum in Systemic Reform

Rodger W. Bybee
Executive Director
Center for Science, Mathematics, and Engineering Education
National Research Council

Science and mathematics educators have worked for over a decade on contemporary reform. In that decade, the National Science Foundation (NSF) funded Systemic Initiatives and, perhaps more importantly, introduced *systemic* into our perspective of reform. The National Council of Teachers of Mathematics and the National Research Council achieved a consensus on standards for mathematics and science, respectively. The science and mathematics education community is poised for a major transition from work on establishing a systemic perspective of education and identifying standards for development and implementation of curriculum to achieve science and mathematics literacy for all students (Bybee, 1997).

A Perspective for Curriculum

Most would agree that curriculum has a significant role in systemic reform. Upon further review, however, one often discovers significant variation in the use of the term curriculum. For some, curriculum is a framework or syllabus; for others, it includes instructional materials; and for still others, it is the course of study. Curriculum may include what some intend, what teachers teach, or what students learn. Here, I will try to clarify what I mean by curriculum.

The science and mathematics curriculum includes a series of constructed relationships among conceptual schemes, procedural strategies, and contextual factors; that is, the concepts, methods, and topics that define the respective disciplines of science and mathematics. Compared to commonly used definitions (Jackson, 1992), my definition presents a broader and more systemic view that includes the fundamental concepts of disciplines, the actions and behaviors of teachers and learners, the various technologies of education, and the contexts within which the context and processes of science and mathematics may be learned.

This definition of curriculum includes dimensions of structure, function, and feedback. The structural aspects include the extant relationship among concepts, procedures, and contexts in materials such as textbooks, scope and sequence charts, curriculum frameworks, videodisks, software, and standardized tests. The curriculum structure is what is usually thought of as instructional materials, and some refer to it as the intended curriculum. It is rationally thought out, has an organization, and by itself is static (Murnane & Raizen, 1988; Cuban, 1992; TIMSS, 1996).

The functional dimension involves the actions and behaviors of teaching. This dimension combines the structure with various ways science and mathematics teachers adapt instructional materials to accommodate myriad classroom situations involving individuals and groups of students. This dimension includes what the classroom teacher contributes to the curriculum including his or her understanding of science and mathematics, the various pedagogical abilities and strategies, and understanding of contexts, such as history and society.

Any viable curriculum in science and mathematics must include feedback. That is, the assessment of student attainment and the opportunities for learners have to develop the understandings and abilities identified in the structural and functional aspects of the curriculum. In one sense this is the achieved curriculum (Murnane & Raizen, 1988) and in another I use the term in a more systemic sense as the results should serve to modify the structural and functional aspects of the curriculum.

A Perspective on Systemic Reform

Despite common use of the terms educational system and systemic reform, the meaning of systemic is often vague. Systemic perspective requires an understanding of the whole in terms of interacting components (e.g., subsystems), boundaries (e.g., critical factors and leverage points), flow of resources (financial and intellectual resources in education), feedback (e.g., assessment of achievement and opportunities to learn).

Intuitively, most educators recognize the systemic perspective. For example, if one addresses one component such as curriculum they will often point out the necessity of another component such as administrative support. The fact that one can list numerous such instances when one component is juxtaposed to another provides ample evidence of the degree to which there is coordination among components and coherence in the system. I would argue, along with others (Fuhrman, 1993; NCES, 1996), that achieving greater coherence is one of the major challenges facing science and mathematics educators. The good news is that for over a decade we have been preparing for this task through the NSF Systemic Initiatives and development of the *National Science Education Standards (NRC,* 1996).

A Strategy for the Curriculum in Systemic Reform

Table 1 identifies a strategy that centers on curriculum and uses a systemic perspective in

Table 1

The Role of Curriculum in Systemic Reform: Strategies for State and Local Educators

Activity	Goal	Central Questions
Continuing the Discussion	Avoid the misunderstanding that systemic reform consists of what we have accomplished	Since we are not finished, what do we have to do now?
Defining Our Goals	Clarify the major goal for curricular reform of science and mathematics	What do we wish to achieve through curricular reform?
Committing to Standards	Define the understandings and abilities that the curriculum should achieve	What has the nation already agreed that all students should know and be able to do?
Deepening Understanding	Challenge and change fundamental miscomceptions of science, mathematics, and education	Have we thought about the role of curriculum in systemic reform?
Increasing Coherence	Achieve greater alignment among components of the curriculum and between the curriculum and the educational system	Is there greater alignment among components of the educational system?
Monitoring Progress	Provide feedback about the fole of curriculum in systemic reform	How are we doing so far? What do we have to do now?

order to move the community of science and mathematics educators through the period of transition. The following paragraphs address the six points and provide some recommendations. Curriculum reform provides the context for the strategy.

Continuing the Discussion

American education has a long history of large and small innovations that have influenced policies, programs, and practices. Recently we have witnessed numerous such innovations, for example, cooperative learning and performance-based assessment, which hold promise of improving science and mathematics education. Unfortunately, we have also developed a perspective that all such innovations carry the same importance, and once we have implemented the new ideas our work is finished. First, the innovations of national standards and sys-

temic reform cannot be equated with other innovations because they are dominant organizers that influence curriculum development and implementation, but they are not the curriculum. The important point here is that we cannot assume that because we have standards and systemic initiatives we are finished. We are not. The steady work of reform is nearer the beginning than the conclusion.

The first step in the proposed strategy is to recognize and accept the fact that reform is a continuing process and to engage each other is a dialogue of central issues of science and mathematics education. I suggest that we begin with discussions that center on our goals.

Defining Our Goals

When asked about the purposes of science and mathematics education, we easily answer with slogans such as scientific literacy or politi-

cally oriented goals such as being first in the world by the year 2000. We have to ask the second and third questions, What do these terms mean? How are they translated into curriculum, instruction, assessments and teacher preparation and professional development?

I recommend that the community of science and mathematics educators begin with focused discussion of the National Education Goals, specifically Goal 3.

> By the year 2000, American students will leave grades four, eight, and twelve having demonstrated competency in challenging subject matter including . . . mathematics, science, . . . and every school in America will ensure that all students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment in our modern economy.

This goal poses several questions worthy of discussion within our community.

I.   How do we **define** competency?
II. What is challenging subject matter?
III. Why grades four, eight, and twelve?
IV. What experiences in mathematics and science will help students use their minds well?
V.   What experiences will help students become responsible citizens?

Committing to Standards

It is time to recognize the place of standards and their significance in guiding decisions about the science and mathematics curriculum. In the *National Science Education Standards* (NRC, *1996)* and *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989) we have consensus documents that should inform curriculum decisions. I realize that placing trust in these documents is a major deviation from usual practices. What the community should understand is that the commitment is to students and learning-what all students should know and be able to do-and this defines the core content of the curriculum. Other decisions, including other content, are left to state and local educators and policymakers.

Deepening Understanding

Decisions to improve science and mathematics programs through the design and implementation of curriculum can facilitate discussions that result in deeper understanding of content and pedagogy. These discussions can center on fundamental misunderstandings about standards, curriculum, and systemic reform. I have identified several such misconceptions in Table 2. School personnel can address these, and probably other, misunderstandings through professional development that accompanies curricular reform. I should point out that in this case curriculum has assumed another role in systemic reform; namely, an opportunity for educators to deepen their understanding of science, mathematics, and education.

Table 2

Misconceptions about the Role of Curriculum in Standards-Based Systemic Reform

---

Curriculum and instructional materials are the same.

Standards and curriculum are the same.

Standards and other educational innovations have equal value.

Science as inquiry and mathematics as problem solving are only instructional strategies.

Standards can be met by selecting the right instructional materials.

Whether materials align with standards or are standards-based is an either/or issue.

Standards are designed to be used directly by teachers as they design lessons.

Inquiry as a mode of instruction ensures that inquiry as content is learned.

Standards provide a menu from which to select the portions to be implemented.

Systemic reform requires a policy that establishes only standards and assessment.

Increasing Coherence

The November 1996 release of the Third International Mathematics and Science Study (TIMSS), in particular the extensive analysis of curriculum that complemented achievement results, clarified a problem in the educational system. I am referring to the lack of coherence among essential components of the system. For example, the content of contemporary instructional materials is not aligned with widely used assessments; teacher preparation and professional development are not aligned with state and local frameworks and practices. Further, some initiatives, such as vouchers, focus attention on issues that vary from the central components that support teachers and teaching and students and learning.

What can be done to make science and mathematics more coherent? My answer centers on the role of curriculum opportunities it provides students to learn the content in the *National Science Education Standards* (*NRC,* 1996) and the *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989). Curriculum provides the concrete and practical entry into systemic reform. However, there must be logical connections and orderly relationships among the instructional materials, teaching practices, and assessment strategies. We need, to use a biological metaphor, a nervous system that coordinates-brings coherence-to basic educational components. Such a view proposes standards as a central organizing guide for school science and mathematics programs.

Using the national standards in this manner leaves considerable latitude for state and local decision making. I will restate an earlier point. These documents thoroughly elaborate what all students should know and be able to do. In a systemic perspective, they define the systems or student outcomes and the content of the curriculum. Educators and communities can make decisions about the way content is organized, emphasized, presented, and assessed.

Monitoring Progress

My final step in the strategy seems obvious. We need to monitor our progress and provide feedback among various components of the system. Our usual approach emphasizes assessment of student learning. Many states have implemented assessments; at the national level we have the National Assessment of Educational Progress (NAEP); and at the international level we have TIMSS. A very important complement to assessments of student achievement is the evaluation of opportunities students have had to learn the valued content.

Conclusion

Curriculum has a very important role in systemic reform. Many educators have the natural inclination to identify curriculum as a critical leverage point for improving student learning. Although I support this view, I have argued that reform requires a more systematic approach, one that centers on standards and sets in process a strategy that attends to the varied components and their interactions in the educational system.

References

Bybee, R. W. (Ed.). (1996). *National standards and the science curriculum: Challenges, opportunities, and recommendations.* Dubuque, IA: Kendall Hunt.

Bybee, R. W. *(1997). Achieving scientific literacy: From purposes to practice.* Portsmouth, NH: Heinemann.

Cuban, L. (1992). Curriculum stability and change. In Philip W. Jackson (Ed.), *Handbook of research on curriculum.* New York: Macmillan.

Fuhrman, S. H. (Ed.). *(1993). Designing coherent education policy: Improving the system. San* Francisco: Jossey-Bass.

Gabel, D. T. (Ed.). (1994). *Handbook of research on science teaching and learning. New* York: Macmillan.

Glatthorn, A. A. (1987). *Curriculum renewal.* Alexandria, VA: Association for Supervision and Curriculum Development.

Goodlad, J. I., & Zhixin, S. (1992). Organization of the curriculum. In Philip W. Jackson (Ed.), *Handbook of research on curriculum. New* York: Macmillan.

Jackson, P. W. (Ed.), (1992). *Handbook of research on curriculum.* New York: Macmillan.

Murnane, R., & Raizen, S. (1988). *Improving indicators of the quality of science and mathematics*

education in grades K-12. Washington, DC: National Academy Press.

National Center for Education Statistics (NCES). (1996, November). *Pursuing excellence* (NCES 97-198). Washington, DC: U.S. Government Printing Office.

National Council of Teachers of Mathematics (NCTM). (1989). *Curriculum and evaluation*

standards for school mathematics. Reston, VA: Author.

National Research Council (NRC). (1996). *National science education standards.* Washington, DC: National Academy Press.

Taba, H. (1962). *Curriculum development: Theory andpractice. New* York: Harcourt, Brace, & World.

Margaret B. Cozzens
Division Director; Elementary, Secondary, and Informal Education
Education and Human Resources Directorate, National Science Foundation

This paper provides the supporting text for the Instructional Materials Development Review that took place at NSF in March 1996.

Background

Instructional materials influence what students are taught and how teachers teach. An innovative, comprehensive, and diverse portfolio of instructional materials that implement standards-based reform in mathematics, the natural and social sciences, and technology education is required for pre-K-12+ education. These materials must be of high quality and consistent with state and national standards to be widely adopted and used in schools nationally. They must provide students with the skills, knowledge, and attitudes necessary to function in a high performance workplace and to continue their education. High quality materials that are accurate in content, age-appropriate, and accessible to all students are essential to raising levels of student achievement and are one of the major building blocks of systemic reform. The development of a wide range of quality instructional materials is fundamental, but so too is the dissemination, adoption, and implementation of these materials.

The curriculum and materials development of the Sputnik Era in the late 1950s and 1960s was a response to both competition with the Russians and what we had learned from researchers like Howard Gardner and Jerome Bruner about how young minds grow and the knowledge about how that growth can be stimulated through the use of new methods and materials. This era saw

the birth of "hands-on" science. The politics and the culture of the time, however, dampened this effort in the 1950s and 1960s to introduce real change into classrooms across the country. Also, these materials were designed on the premise that children can know and do much more than we thought they could, but these materials were predominantly used in a tracking environment and only reached the top 20% of students. However, in both mathematics and science they paved the way for the new materials of today by addressing age-specific needs and individual learning styles. We also learned from these earlier prototypes that their success is highly dependent on the professional development of all teachers, not just a self-selected group of teachers with excellent mathematics and science backgrounds who are interested in using innovative materials. It also became abundantly clear that not everyone wants to teach children how to think independently. This issue still remains today!

The NSF/Federal Role in Materials Development

The Instructional Materials Development (IMD) Program at NSF supports the development of strategies and materials that promote the improvement of science, mathematics, and technology instruction for students at all ability levels. As an agency that has a broad mandate to support the vitality of basic science, engineering, and mathematics in the United States, NSF develops materials broadly and comprehetisively across the mathematics, science, and technology education disciplines and the grade level continuum.

According to the Federal Coordinating Council for Science, Engineering and Technology (FCCSET) FY 1994 Budget Summary, NSF spent $43.9 million of the total federal expenditure of $54.3 million for curriculum improvement.

This broad mandate makes NSF unique among federal agencies engaging in science, mathematics, and technology education activities. In FY 1994, only four other Federal agencies were engaged in the development of instructional materials in these areas. The National Aeronautics and Space Administration (NASA) was second in expenditure with $4.4 million, followed by the Environmental Protection Agency (EPA) with an expenditure of $3 million, the Department of Energy (DOE) with $2.2 million, and the Department of Health and Human Services (NIH), with $.8 million (Appendix A). Each of the mission agencies (NASA, Energy, EPA, and NIH) develops short modules in disciplines that advance the mission of the agency.

IMD Goals

The IMD goal is to develop instructional materials, aligned with standards for content, teaching and assessment, that:
- enhance the knowledge, thinking skills, and problem-solving abilities of all students;
- apply the latest research on teaching and learning;
- are content accurate and age-appropriate;
- incorporate the recent advances in disciplinary content and educational technologies;
- assist teachers in changing practices; and
- ensure implementation in broadly diverse settings.

IMD Processes

The development of instructional materials in science, mathematics, and technology education is a complex process. Guiding principles frame the processes for the development of NSF sponsored materials. The materials:
- are developed by a collaborative of scientists, mathematicians, teachers, and other educators;

- are based on research in teaching and learn-&G
- align with standards;
- contain appropriate student assessment;
- are field tested in diverse settings; and
- have undergone formative and summative evaluation, which include impact data from field test sites.

Components of this process can be best understood through the diagram shown in Figure 1.

Mathematics Instructional Materials Development

In 1989, the teachers of mathematics in this country through the National Council of Teachers of Mathematics (NCTM) released the *Curriculum and Evaluation Standards for School Mathematics. The* NCTM *Standards* built a framework for what children should know and be able to do at various grade levels of school mathematics. They were a culmination of a grass roots effort by teachers who knew well from experience that the mathematics being taught in classrooms across the country was not working, and it was inadequate in meeting the needs of students, higher education, and the workplace. The membership of *NCTM knew* that *the Standards,* however, needed more than written documents. The country needed new instructional materials to deliver the *Standards,* existence proofs to show that the *Standards* are indeed attainable, and teachers to implement them on the front line.

With the advent of the *Standards, the* NSF saw the need for providing leadership in the development of new mathematics materials, if the U.S. had any hope of being first in the world in mathematics achievement for all students. The *NCTM Standards* provided not only a framework for the development of new materials, but they also brought visibility and an eager constituency for high quality innovative instructional materials. Therefore, in 1989, NSF launched a series of initiatives to fund the development of mathematics materials grades K-5, eighteen months later the development of grades 6-8 materials, and finally in 1993 the development of grades 9-12 materials. Believing that schools and school sys-
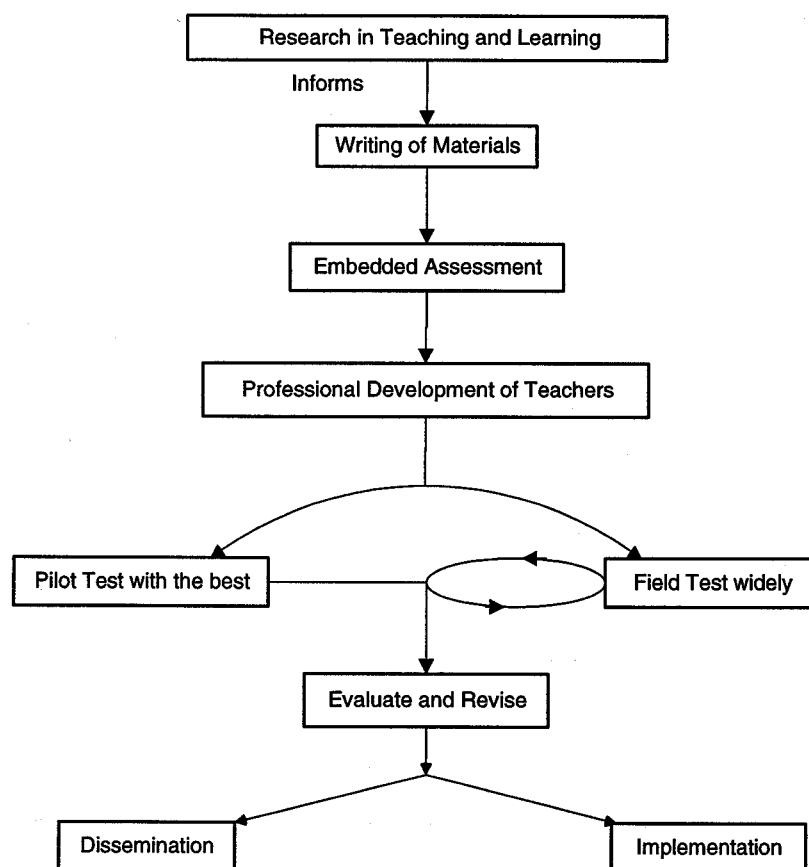
Figure 1. The IMD process.

tems needed a choice of quality materials, a number of curriculum development projects at each of these three levels was funded. Since curriculum materials development is a complex and costly undertaking, each project was funded for five years to complete the development of multiple years of comprehensive materials. In addition, two high school mathematics projects actually began work earlier than the rest with initial funding from other sources and consequently completed development before the rest.

Each of the projects shared the following characteristics:

- a belief that it was possible to significantly improve the teaching of mathematics over what was currently available;
- the materials were to reach as many students as possible;
- the use of the technology of calculators and computers;

- the need to involve students more actively in their own learning;
- that applications of mathematics should be incorporated into the mainstream of experiences of students in classes; and
- the sharp barriers between algebra, geometry, or any other branch of mathematics should be blurred.

Despite these similarities that defined their strengths, the curriculum materials offer unique choices for teachers to use in their own classrooms.

The development of new mathematics instructional materials was not easy, particularly when the NCTM *Standards* called for a paradigm shift in the way teachers would use these materials. These are comprehensive materials, not disjoint units of mathematics, so the development of concepts over time and grade levels is essential and was closely monitored in the pilot and field

testing. At the same time the materials were evaluated, the ability of teachers to use these materials was also evaluated. Both student achievement and teacher achievement were measured. Ongoing professional development for the pilot and field test teachers was essential if the materials were to have a fair test and for the developers to gain maximum feedback. Assessment of student learning was built into the materials. Once the materials were field tested they were revised again, sometimes retested, and then submitted to publishers for publication. At this point the developers thought they would be done, but they found out otherwise. Now, those who eagerly awaited the new materials put even more pressure on the developers and the publishers to get preprint copies of the materials and professional development opportunities to implement the materials.

By the end of 1997, the elementary and middle school mathematics curriculum materials and at least grades 9 and 10 secondary mathematics materials will be published and readily available for use in schools across the nation. By the end of 1998 all of the secondary materials will be published. (See Appendix B for a graph of the development of comprehensive mathematics curriculum materials and a description of these materials.)

Science   Instructional   Materials   Development

Science for All Americans was published in 1989 by Project 2061 of the American Association for the Advancement of Science (AAAS) and contains a set of recommendations on what understandings and habits of mind 'are essential for all citizens in a scientifically literate society. It began the movement toward science standards (even though it dealt with mathematics, social sciences, and technology as well as science) that would eventually describe what children should know and be able to do in the sciences. Project 206 1, in the designers' own words, attempted to establish a conceptual base for reform by defining the knowledge, skills, and attitudes all students should acquire as a consequence of their total school experience, from kindergarten *through* high school. *Benchmarks for Science Lit-*

eracy, published by AAAS in 1993, describes how students should progress toward science literacy, recommending what they should know and be able to do by the time they reach certain grade levels. Building on both *Science for All Americans* and the development of *Benchmarks for Science Literacy,* which was underway at the time, the National Research Council of the National Academy of Sciences brought together diverse groups of scientists, engineers, teachers, and science educators in 1992 and began the development of the *National Science Education Standards (NSES).* The *NSES,* published in December 1995, go beyond the other documents and include not only content standards, but standards for science teaching, for professional development for teachers of science, for assessment, for science education programs, and for science education systems.

Unlike in mathematics, NSF's renewed involvement in the development of instructional materials in science predates issuance of the standards. Starting in 1986, with what were called the TRIAD projects, the NSF IMD program began an extensive process to fund the development of instructional materials for science in elementary schools. Prior to that time, very little science was taught in elementary schools in the nation. In the belief that quality instructional materials would motivate teachers to teach more science and teach science using an inquiry, hands-on approach, NSF began to make a major investment in the development of science materials for school use. Two years later the TRIAD projects were extended to the development of middle school science instructional materials. The TRIAD projects were characterized by development teams of scientists, materials developers, school people, and publishers, under the belief that the materials would have a better chance of dissemination if the publishers and school people were brought in at the outset. Unfortunately, it was too easy for publishers to withdraw support at points along the way, and some projects were never published. However, the TRIAD experiment did give birth to a number of exemplary elementary science projects, for example the *Full Option Science Series (FOSS)* published by the Encyclopaedia Britannica, and *Insights,* origi-

nally published by Optical Data, soon to be transferred to Kendall Hunt. Since early 1992, during the end of the development phase of these projects, the projects were advised to keep close track of the development of the AAAS *Benchmarks for Science Literacy* and *Science for all Americans.* The innovation and quality of these elementary projects set the stage for NRC's *NSES* at the elementary level. Thus, there is considerable alignment of the materials with the *NSES.* In addition, a number of the developers of the TRIAD science materials were participants in the development of the *NSES.*

As the field began to recognize the need for hands-on science materials at all grade levels, developers also began the development of new innovative secondary science instructional materials. The first of these new materials was developed starting in 1986 by the American Chemical Society, called *Chemistry in the Community (ChemCom). ChemCom* is designed to focus on the study of chemical concepts that emphasize chemistry's impact on society. *ChemCom* is designed to teach students about the important role that chemistry plays in their personal and professional lives, how to use knowledge of chemistry to think through and make informed decisions about issues involving science and technology, and the need to develop a lifelong awareness of both the potential and limitations of science and technology. Data indicate that these materials benefit all students, including those intending to major in the sciences in college.

Recently, *the* success of *ChemCom* has motivated the development of comparable materials for physics *(Active Physics),* biology (*BioCom*), and *earth* science (*EarthCom*). In addition, there will soon be the publication of a number of comprehensive materials for secondary science, none of which are filly integrated across all scientific disciplines.

Appendix C gives a graph of the development of science instructional materials.

Middle School Science Study

In an attempt to begin to more fully determine the status of instructional materials for sci-

ence, now that the *NSES* are published, ESIE began a study of instructional materials for middle school science in March 1996. NSF started with middle school science materials because much was already known about the elementary science materials developed through the TRIAD projects and elsewhere. The middle school study provides a review of past and current NSF-supported efforts and identifies areas where more development is necessary. A portion of the middle school study provides brief reviews of noncomprehensive materials, both where NSF has supported the development and otherwise. Preliminary findings from this in-depth study are now available, and a public document will be disseminated in late 1996. The main goals of the study were to answer the following questions:

. What are the characteristics of NSF-supported instructional materials for middle school science?
. How sufficiently do extant materials provide for a comprehensive program for middle school science consistent with national standards for science education?

The study emphasized a review of comprehensive curricula, those that equal a year or more of course material. Secondarily, the study reviewed modules or units that can serve as building blocks for a comprehensive program. The central criteria for reviewing the instructional materials were as follows:

. Is the science content correct?
. How well designed are the materials to provide for conceptual growth in science?
. How well do the materials align with the Science Standards?

Current and Future Plans across IMD: Develop, Evaluate, Disseminate and Implement, and Measure Impact

The mission of the Instructional Materials Development (IMD) Program is to develop, evaluate, and disseminate and implement science, mathematics, and technology education instructional materials and to measure their impact on students and teachers.

Develop-Current and Future

Currently projects are supported in all areas to develop (including pilot and field testing) comprehensive instructional materials in mathematics, science, and technology education. In addition projects are supported that integrate mathematics, science, and technology education in the curriculum in schools. For example, the Integrated Mathematics, Science, and Technology Project (IMAST) is developing integrated materials for seventh- and eighth-grade students around the topics of biotechnology, manufacturing, forecasting, energy transformation, transportation, and others. The materials are designed to be taught by teachers from all three disciplines for at least 120 minutes per day. Although there are separate activities for mathematics, science, and technology, the activities focus on the same key concepts and are coordinated so the students readily see the relationships among the disciplines.

Comprehensive projects often develop a year's worth of materials at a time and then look to completing a sequence useable for multiple years. For example, Chemical Education for Public Understanding (CEPUP) instructional materials were developed in the early 1990s for middle school science students. Two years later, Science Education for Public Understanding (SEPUP) expanded the number of CEPUP units to include units in all of the physical sciences and earth science. Under the name *Issues, Evidence and You,* these units were developed on local issues for middle school students and on global issues for high school students. The development of SEPUP materials contained embedded assessment materials, encouraging the developers to decide early in the development what skills and content the students should learn. NSF has recently funded Life Science Education for Public Understanding to complete the series to a full comprehensive set of instructional science materials for grades 6-10 that cover life, physical, and earth sciences.

It is not always the case that the development of new materials is necessarily required to meet the demand for standards-based up-to-date materials. In some instances, it may be more effective to revise and/or expand existing materials. For example, the popular National Geographic Kids Network consists of technology-based materials that allow students to communicate with other students around the world and to collect and analyze scientific data through a nationwide electronic network (e.g., acid rain). The original materials were designed for students in grades 4-6. The project has now been expanded for students in grades 7-9. A second example is the Middle School Mathematics Through **Applications**— Computers and Design (MMAP), a project that was designed to provide technology-based mathematics units for middle school students. At the time MMAP was funded in the early nineties, it was not clear that enough schools would have computer equipment available for students all day long. However, the demand for the MMAP materials has been so great in both urban and suburban settings, that an extension of MMAP to a comprehensive complete set of mathematics instructional materials for grades 6-8 was funded in 1995.

The advent of new technologies and their availability in schools has created a need for more educational materials that are technology-based. This phenomenon is what led to the expansion of MMAP and the Kids Network materials expansions. Recently other projects have been funded to increase the set of quality technology-based materials, such as the Mapping your City-Geographic Informations Systems (MCGIS). The MCGIS materials engage students in use of the GIS database system to correlate complicated **datasets** on variables such as green space, income, census figures, and ethnicity so they can map their local communities.

Early childhood researchers have been telling the public over the last few years that young children learn much more readily if the teaching methods meet their special needs, that acceleration of standard curriculum to earlier years is not only unproductive but potentially dangerous. The most important ingredients in young children's education are hands-on learning, physical activity, and socialization. Even though the need for hands-on materials in mathematics and science is widely recognized, there are very few **instruc-**

tional materials available for children ages 3 through 5, yet increasing numbers of children are in some kind of school setting during those years. The development of hands-on early childhood mathematics and science instructional materials is a high priority for IMD in the next few years.

Coupled with the need to provide materials for early childhood education is the recognition that instructional materials must be available for parents and other care-givers. A special solicitation that crosses all program areas in the ESIE Division calls for projects that will provide help for parents and care-givers to support their children's science and mathematics endeavors at home and elsewhere, and that will encourage them to be effective advocates for more universally available quality mathematics and science education. The IMD portion of this solicitation is for projects that develop instructional materials for parents.

Instructional materials are needed that will radically change instructional practices and student learning. Particular examples include (a) involving students in research experiences that teach skills related to the scientific process-hypothesis development, data analysis, presentation of results (e.g., providing and understanding geographically distributed data or data retrieved from large image sets); (b) use of advanced instructional technologies to change what can be learned by collaborations of students and teachers; (c) the teaching of the processes of "design under constraint" and or modeling in the context of science, mathematics, and technology education; (d) use of authentic workplace situations that promote learning of disciplinary content, workplace competencies, and career awareness. Such materials must be able to engage students with different learning styles in the study of technical subjects and must provide explicit guidance for teachers as they change their teaching practice.

Increasingly, projects are needed that span the grade levels from 10 through 14. For example, there are many students taking college algebra in colleges or universities, and many students taking calculus in high school. Instructional materials are needed at all these levels that are standards based and that encourage hands-on ac-

tivities and the student as an active learner. Many more students take advanced placement science courses in high school today than previously, yet they go on to college and take courses that have no relationship to what they had in high school. Articulation of materials-secondary and under-graduate-is paramount and a responsibility of IMD in ESIE and the Course and Curriculum Development programs in the Undergraduate Division.

Evaluate-Current and Future

Evaluation of instructional materials during their development is essential. Various paradigms for insuring proper evaluation of materials developed with NSF funding before publication have been in place since early 1993 and will continue to be used in the monitoring process. For example, the developers of the comprehensive mathematics instructional materials projects have met as a group with NSF staff once a year at a Gateways Conference. Developers brought the current version of their materials to Gateways to be critiqued by their fellow developers and the NSF staff. Gateways provides an opportunity for the developers to share their problems, issues, and successes with one another. The Gateways Conference in October 1995 hosted a Public Forum at which time students, teachers, administrators, and others gave public testimony to the success of these materials at the field test sites. The University of Chicago then published the document, The Success of Standards-based Mathematics Curricula for all Students, a Preliminary Report.'

All instructional materials development projects are required to have an external advisory board consisting of scientists, mathematicians, and educators. In addition, because of the limitations on staff travel, some program officers add an external auditor to the project. It is the responsibility of the external auditor to review all materials at each draft and report back to NSF any problems he/she encounters. The external auditor also visits schools where field testing is taking place in order to understand the effectiveness of the materials in classrooms. Often, the second and third year continuation funding for a project

is not made until the materials have been reviewed by an external panel of reviewers appointed by the program **officer.** In addition, the third year funding for a project is not granted unless the project has signed a contract with a publisher to publish the materials. As much as possible, with limited travel funds, program officers make site visits to the development projects and their field test schools.

The Research, Evaluation, and Communication Division is currently performing a programmatic evaluation of the Instructional Materials Development Program. Key aspects to be considered are the dissemination capability of IMD and the quality of materials through random reviews. The tentative completion date for the IMD program review is late 1997.

The Middle School Science Study was designed to provide the staff at NSF, and in turn the field, with a broad understanding of the instructional materials in middle school science, their quality, and their match with the *NSES. The in*strument used to conduct the study was developed by Mark St. John of Inverness Associates to review a specific set of middle school instructional materials. The review panel modified the Inverness instrument, tested it on one set of materials, and then reviewed twenty other sets of materials. The revised instrument is now available to the public to help them choose quality materials for use in **schools.**[2] Each of the sets of middle school science materials rated by the reviewers had been or was currently being developed with NSF funding and is comprehensive in nature. A third-party review of middle school science instructional materials that are either not comprehensive or were not funded by NSF will be completed in late 1996.

For the future, the IMD program will continue to use external auditors, review panels, and others to review draft instructional materials developed with NSF funding prior to publication to assess accuracy and usability of the materials. In addition, meetings similar to Gateways meetings will be initiated for science developers. The first such meeting will take place in late 1996 and bring together all of the current developers of high school science comprehensive materials. After that meeting, NSF may conduct a high

school science study similar to the one conducted for middle school science materials. During FY 1997 NSF will experiment with the use of reverse site visits for all large IMD projects at the midpoint of their grant cycle.

Disseminate and **Implement—Current** *and Future*

Dissemination of instructional materials and information related to these materials is an ever increasing aspect of the work of IMD. In 1993, IMD believed that an electronic database of all instructional materials developed through NSF funding was one solution to the problem of getting usable information out to the field. The NIRL database was created as a searchable database, and one that included developer, publisher, evaluative, and summary information on each project. The database was put on line through the World Wide Web and included a toll free 800 number for more information. NIRL was somewhat successful at getting information about the products quickly to teachers, but without substantial advertising, its potential impact was muted. Updating the database became an impossibly expensive problem, for developers, for NSF, and for the contractor. Consequently, the database is to be brought in-house to NSF, and a system of updates will be undertaken by NSF staff and the grantees through an on-line system. Once the updates are completed and a user-friendly system is established, the database will be advertised and made available over the Web. Meanwhile the Clearinghouse in Ohio sponsored by the Department of Education serves as a repository for NSF-funded materials, as well.

Much of the task of dissemination of the new instructional materials falls on the shoulders of the publishers themselves. The publishers clearly want to see the materials used and they market them through the standard mechanisms at meetings, by telephone calls, with visits, etc. Since the use of these materials requires additional professional development and a cost-effective way of reaching large numbers of teachers, NSF has launched two local systemic change initiatives, one in K-8 for mathematics and science and one in 7-12 for mathematics. These initiatives are

district-based teacher enhancement projects and assume that the districts either have chosen the materials they will use, or will do so during the first year of the award; part of the enhancement is then designed around the selection of new materials. Substantial evaluation paradigms, managed by an outside contractor, were built into the local systemic change project initiatives from the outset. These processes are designed to evaluate changes in teacher practice and effectiveness of instruction and include an analysis of the use of instructional materials. The local systemic evaluation model is being adapted for use in other SIs and in Eisenhower professional development projects. In addition, one component of the yearly accountability portfolios submitted to NSF by the rural, urban, and statewide SI projects includes the use of quality instructional materials at their site. All of these activities provide mechanisms for the dissemination of high quality materials, including those developed with IMD support.

The current dissemination strategies will continue and additional information will be provided to the technical assistance contractors and those managing teacher preparation activities. A number of short monographs describing the new standards-based instructional materials in mathematics, science, and technology in sufficient detail to facilitate choice by schools have been developed, are in production at this time, or will be drafted.[3]

Implementation of quality instructional materials is an ongoing challenge. Since early 1996, the NSF IMD and Teacher Enhancement(TE) Programs have encouraged submission of implementation proposals. These implementation projects can cover a variety of tasks, from the development of videos of classroom usage of sets of the new materials, to guidebooks on how to make good choices of materials, to how to gain parent and community support for the materials, to articulation across grade levels and/or disciplines. The staff hoped that leaving the descriptions open-ended would result in a number of interesting implementation grants. However, in the first round only one award was made. The FY 1997 guidelines are more descriptive and suggestive.

Some of the issues surrounding dissemination and implementation will be discussed in the next and last section of this paper.

Measure Impact---Current and Future

It becomes increasingly clear that NSF must be able to validate claims that students are learning more science, mathematics, and technology, including problem solving, critical thinking, and basic skills, through the use of these new standards-based instructional materials. To help us evaluate these claims, projects are now required to provide student achievement data from their field test sites, and these data must include information from diverse populations. Continuations and new awards are made contingent on the quality and availability of these data. Often, NSF staff visit classrooms to directly assess effectiveness of the materials as well.

IMD has undertaken a few targeted longitudinal studies to measure the impact of the new standards-based instructional mathematics materials on student achievement for a period of up to ten years, during the last years of the project's development and for the first few years that the materials are commercially available, post the end of the grant. One example is Norman Webb's extensive evaluation of the Interactive Mathematics Project secondary materials in six implementation sites across the country. Included in this evaluation is a transcript study that tracks students through the four years of high school and into college.

In addition, the mosaic-of-evidence study is being developed by the RAND Corporation to measure the impact of systemic reform on student achievement in ten SI sites, each of which is using the new mathematics and science instructional materials. It is hoped that this study will also contribute to an evaluation of the use of the materials. Two impact studies were completed in 1993 (Webb & Reynolds) that measured the impact of nine mathematics and ten science sets of instructional materials.[4]

IMD staff are continually looking for new, cost effective ways of measuring the impact of instructional materials projects on students and teachers.

Issues for Instructional Materials Development

Implementation

The motivation for developing innovative, standards-based instructional materials is to improve the teaching and learning of science, mathematics, and technology. But the development is only the first step and will not contribute significantly to improvements in teaching and learning unless the materials are used widely by teachers prepared to use them effectively and who have the understanding and support of school administrators, parents, and others in the community.

The NCTM *Standards* and the NRC *NSES* point to a dramatic departure, in content and pedagogy, from the traditional approach to teaching and learning in mathematics and science. Several serious issues are inherent in the implementation of any standards-based instructional materials:

- Standards-based instructional materials require a significant amount of professional development for teachers in both content and **pedagogy**;
- Publishers are not prepared to provide the needed teacher support activities and often don't realize teachers need more than they did with traditional texts;
- The textbook adoption process is an expensive process that some smaller publishers of innovative materials are not prepared to undertake, yet the process pays big dividends for those who do, for example **Encyclopaedia** Brittanica with *FOSS*;
- Implementation requires support and buy-in from administrators, parents, and the community, and when the support is missing from one group, as initially happened in Palo Alto, CA, and Ames, IA, the whole reform movement can be in jeopardy;
- Assessment of student learning must be linked to the instructional materials, and the design of new assessment tools has not kept up with the development of new materials and the standards;
- Articulation across grade levels and disciplines is essential; and

- Teacher preparation in colleges and universities must be linked with the new materials to facilitate implementation, yet most college and university departments are unaware of either the standards or the new materials.

Until we address these salient issues as a program, and as a directorate, implementation of new innovative materials will be limited.

Publishers and Publishing

Publishing instructional materials in mathematics and science is not only necessary for dissemination and use, but since 1986 has been a requirement of all awards made by NSF for development of materials. Inherent **difficulties** often occur at the outset with securing publishers, but many issues arise after contracts have been signed and the materials are published. Among the issues resident in publishing the materials are the following:

- Many publishers have competing materials that they market and may provide only minimal marketing for the new materials;
- Many small publishers and even some large ones are being bought out by large publishing houses, so that eventually there may be only three or four companies publishing instructional materials;
- Many of the new materials contain substantial teacher materials and very little student materials, making it hard for the publishers to make money in their present form;
- Many of the new materials consist of kits, supplies, videos, not normally produced by textbook publishers;
- Many concerns exist about the intellectual property rights for publishers whose materials are placed on the Internet either by them or by others.
- Internet materials simply are out there, with no indication of quality or use.

The issues of technology and choice of materials and of the publisher's role in the solutions to problems in these areas are enumerated in the next pages.

Materials Selection at Local and State Level

Implementation of quality instructional materials and the dissemination of NSF-developed instructional materials often come together at the site where the selection of school materials takes place. Administrators, teachers, and stakeholders at the state and local levels are forced to make very costly and extremely critical decisions regarding what instructional materials will best serve their students. When faced with the wide array of instructional materials available on the open market, local leaders often become caught in a storm. Spin doctors, money managers, and fast-talking sales people with little understanding of the issues are part of the fracas in which decision makers find themselves. Issues for NSF include the following:

.   Should NSF fund what might be called a seal of approval for instructional materials, those developed through NSF funds and/or those not, together with a published consumer reports listing those that earned the seal, possibly with ratings of the materials?

.   Is it possible to create a guide for school systems to use in selecting materials? This might include evaluative instruments similar to the instrument used for the middle school science study.

.   How can parents and other lay persons, who often participate in curriculum materials selection, be convinced that all students can be educated, while maintaining excellence? What materials should be developed to help groups make better decisions regarding materials?

All of the above issues revolve around getting maximum information concerning instructional materials to a variety of audiences and, at the same time, protecting NSF from being accused of mandating specific materials, particularly in the funded SI sites.

Information Technologies

Educational technologies are changing rapidly. Many school systems, but not all, now have access to many new delivery venues that make possible today what was thought to be science fiction a few years ago. With the rapid, continuous changes taking place in schools and in technology development houses, IMD is faced with an array of issues:

- Should developers design software for state-of-the-art hardware systems or for what is available today for use in schools?
- Should we encourage all IMD projects to include information technologies or leave it to the specialists?
- Is there sufficient research to know how to effectively use educational technologies in classrooms?
- Since it will require massive teacher enhancement to prepare teachers to use both the new hardware systems and the new software available for them, and once a teacher is so trained, in a few years the technology will change once again, how do we build a professional development system for teacher technology education?
- Is it possible to ensure equal access and use of information technologies for all students? It is apparent that poor schools and rich schools have the best chance at providing the technologies, in the first case through donations, in the latter case they can afford it.
- Are parents, students, and business and industry personnel, who expect the use of technology in schools, willing to make the sacrifices necessary to assure quality materials and use?

Summary

Since its inception in 1950, the National Science Foundation has served the Nation by investing in research and education in science, mathematics, and engineering. Over the years NSF's investments in education have included investments in the development of instructional materials for use in classrooms across the country, pre-K-12. NSF believes, and rightly so, that education reform and improved student achievement cannot be accomplished without quality instructional materials in use in classrooms. The IMD program has helped to assure the use of quality materials through development, evaluation, dis-

semination, and implementation activities. These activities will continue and intensify in response to increased calls for improvement through standards-based materials. Data suggest that the materials recently developed and field tested in diverse settings by diverse teachers are making a difference in student achievement, particularly in mathematics.

IMD will continue to develop standards-based quality materials for use by all students and all teachers across the nation. Particular emphasis in the next few years will be placed on completing the portfolio of science materials, expanding the portfolio of early childhood mathematics and science materials, and revising and expanding the materials that incorporate the use of learning technologies in effective ways. Implementation of the materials developed with NSF funding, especially in mathematics, will be increasingly important and receive a considerable amount of attention. Effective ways to measure impact, in particular that students are learning more mathematics, science, and technology, will be sought and used. Materials by themselves, if not used effectively by teachers to improve student achievement, have no value. Therefore it is essential that we determine how

teachers can use the quality materials most effectively to support their students' learning.

As NSF becomes more accountable for its actions, so, too, will the developers and implementers of the portfolio of mathematics, science, and technology materials to be used in the twenty-first century.

Notes

1. *The Success of Standards-based Mathematics Curricula for All Students -A Preliminary Report, Gateways IV Public Forum on the Impact of Mathematics Education Reform,* University of Chicago, Chicago, Illinois, October 1995, available from NSF.

2. *Middle School Science Study Instrument,* available from NSF by contacting any IMD program officer.

3. *Technology Education Instructional Materials, Elementary-High School,* Instructional Materials Development, Division of Elementary, Secondary, and Informal Education, National Science Foundation, NSF 96-6 1, February 1996.

4a. *Impact Study of IO NSF-Supported Precollege Science Instructional Materials Projects,* W.W. Reynolds, I Gawley, F. Pregger, Reynolds & Schaeffer Associates, Inc., October 1993.

4b. *Dissemination of Nine Precollege Mathematics Instructional Materials Projects funded by the National Science Foundation, 1981-91,* N. Webb, H. Shoen, and S. Whitehurst, Wisconsin Center for Education Research, Madison, Wisconsin, April 1993.

Richard Greenberg
University of Arizona

As a developer of science-education programs, I have found myself deeply concerned with structuring programs so that their innovations can actually find their way to being used widely by students and teachers in the real world of schools. In order to achieve that goal, it became clear to me early on that consideration of curriculum cannot be separated from issues of teaching and learning: For practical implementation, the critical link is teacher education. While, for understandable organizational reasons, the program for this conference separates "curriculum" from the topic of "teaching and learning," the distinction seems potentially counterproduc-

tive. The need to address all of these issues in an integrated manner is especially critical in the context of systemic change.

Because this panel is assigned the subject of curriculum, I would like to explore an issue that has been driven home to me during curriculum materials development as part of the Image Processing for Teaching (IPT) project. The specific issue is a dilemma faced in the course of any curriculum innovation, and its resolution illustrates the importance of intimately integrating all components of reform if we are to have successful systemic implementation. Later, I will also offer a cautionary note on how a rigid or bureaucratic

implementation could subvert the process, unless administrators are vigilant in defending the spirit and intent of reform in science education

The use of digital image processing allows a great leap ahead of many classroom experiences, offering the potential for activities using a professional research tool and real, complex image data. It is a technology that can achieve the objective of many curricular innovations: authentic scientific experiences. In the early experimental stages of IPT (beginning in 1990), several teachers with personal subject "hobbies" developed wonderful research-based activities that worked beautifully in their own classrooms, but were too specialized to be widely used by other teachers. It became clear to us that our materials must support curriculum objectives widely recognized and shared among teachers, while at the same time allowing open-ended exploration, discovery, and analysis.

This poses a dilemma that must be faced by any curriculum developers who attempt to incorporate real investigation: You want the activities to be innovative, but you also need to ensure that a substantial number of teachers will recognize such activities as helping them meet their own objectives, and not imposing an arbitrary burden. For traditional teaching objectives, the marketplace has tended to drive curriculum materials toward the lowest common denominator.

The new *National Science Education Standards,* combined with systemic reform, are providing a solution to this dilemma. These standards are motivating teachers to adopt new and desirable objectives. This trend is especially true where systemic change is under way, educating teachers to the advantages and supporting implementation of change. Increasingly, we find that teachers are searching for ways to bring real scientific exploration and discovery into their classrooms. This change is making it possible for developers to provide curriculum materials that encourage the most meaningful learning, while simultaneously addressing widespread demand.

None of this change would be possible without teacher education as a key part of systemic reform. First, teacher education in the context of systemic reform has alerted teachers to the new purpose and approaches of science education reform within their systems. Without such teacher education, the demand for new curriculum support would not be created. Innovations would be local, small-scale, and ephemeral. Second, the specific curriculum innovations that are provided to meet that new demand require teacher education. For example, the IPT project began as a Teacher Enhancement project well before we began serious development of curriculum materials. Thousands of teachers have now learned what digital image processing is, how to do it, and how students can use it for learning. That broad base has been essential for its success. Thus, by providing both general and curriculum-specific teacher education, systemic reform can facilitate the development and implementation of new, enlightened curriculum, and in turn that curriculum is providing an essential part of meaningful reform.

However, I want to raise a concern about ways that organizational structure, which is necessary to coordinate systemic change, can nevertheless subvert the goals of science education reform. We have seen here how the need to structure a conference program begins to separate curriculum from teaching and learning. Here it is not a problem, because the participants are sophisticated enough to avoid being pigeonholed. But imagine what would happen if similar separation occurred in the structure of a systemic initiative. Policymakers and program planners must be careful to design systems to prevent such fragmentation. Even so, operational managers will need to be continually vigilant to ensure that bureaucratic and political tendencies do not separate the components of reform, all of which must be intertwined for systemic change to be beneficial.

Another natural bureaucratic tendency to avoid is applying standards in an overly rigid manner. In general, enlightened developers of curriculum materials do address topics explicitly within the content standards. The practical reason, again, is that we want innovative teachers to recognize our materials as supporting their goals, and to guide all teachers toward meeting the National *Science Education Standards.* However, the Content Standards are deliberately general because they are intended to provide a guiding

philosophy; the Standards explicitly state that these guides are not meant to be proscriptive. This approach means that curriculum can be quite diverse, in order to meet the needs of diverse students in diverse contexts, while still meeting the national Content Standards.

One major theme of the Standards (again reflecting much of the thinking in science education reform, and also confirmed by TIMSS) is that curricula should not try to cover too broad a range of topics, but the topics should be explored in depth. This means that developers can prepare materials with the expectation that the materials will not define curricula themselves. Instead, developers need to provide an array of activities that support a clearly articulated conceptual framework aligned with the Standards. From such an array, teachers can select activities appropriate to their specific context, thus limiting the topical scope while increasing depth of understanding of concepts and processes.

Thus curricula that meet the *National Science Education Standards* may vary widely from school to school, or even classroom to classroom, within any system, and such variation is desirable if it addresses students' diverse needs. We need to encourage such diversity, while at the same time making the fundamental philosophical and pedagogical approaches of science-education reform ubiquitous.

Intellectually, there is no real conflict between the objectives of implementing reform systemically, while promoting internal diversity. However, the challenge in implementing systemic change is to design and monitor administrative and bureaucratic structures that will not pervert those goals. These systems need to be designed to promote diversity, rather than uniformity. Where it is done well, systemic reform with appropriate teacher education can promote the best kinds of curricula, the most meaningful teaching and learning, and the diversity essential to meeting the needs of all students.

Sylvia T. Johnson
CRESPAR
Howard University

Introduction

Systemic educational reform involves coordinated change in the whole education system, particularly in the areas of curriculum, assessment, instruction, standards, and teacher professional development. The process of assessing students is central to the provision of instruction (Cohen & Spence, 1990). Furthermore, assessment impacts the curriculum that is taught as well as the instructional goals, practices, and materials that are used and developed. Presently, nationwide calls for better forms of assessing student achievement raise questions about their relative benefits and drawbacks compared to traditional forms of assessment. Consequently, the diversity of different forms of assessment presents different issues, benefits, and drawbacks. While it has been recognized that traditional assessments have major limitations, the benefits of alternative

forms of assessment are new, and research has not yet had time to study these new forms in depth. Given the intimate, interactive, and sometimes reciprocal relationship between curriculum and assessment, one might ask how assessment is currently being used for curriculum improvement in systemic reform efforts, and what assessments have been generated by systemic reform. Actually, at least three general models for systemic reform are applicable in answering these questions:

1 Assessments are instituted with the goal of changing curriculum and practice.
2 Curricular reforms are instituted, and a need is then seen for more accurately aligned assessments.

3 Curriculum and assessment reform occur in tandem with subsequent modifications of one being rapidly reflected in changes in the other.

This paper reports on examples of each of these models. The problems associated with building and maintaining an effective assessment program are also discussed.

1. An assessment program is instituted with the goal of changing curriculum and practice: The Maryland State Performance Assessment Program (MSPAP).

MSPAP is a statewide performance assessment program piloted in the late 1980s and instituted in 1990. It initially met with reactions of confusion and even hostility among school people and the community. It has steadily made its mark and seems to be a major factor now in how and what teachers teach. Teachers report changes in content emphasis and time spent (Koretz, Mitchell, Barron, & Keith, 1996). Parents and community people are interested and concerned and see the nature of these tasks as important for children to be able to do. While scores are still a good distance from desired levels of proficiency, remarkably steady progress has been made. The state and the school districts have made substantial investments in professional development and community education, and these are important to the current progress.

2. Curricular reforms are instituted, and a need is then seen for more accurately aligned assessments: Curriculum-based assessment.

Curriculum-based assessment (CBA) is defined by Cohen and Spruill (1990) as a broad approach linking assessment to instruction. Idol, Nevin, and Paolucci-Whitcomb (1986) describe CBA as a criterion-referenced test that is teacher-constructed and designed to reflect curriculum content. Blankenship (1985) similarly defines CBA as the "practice of obtaining direct and frequent measurement of student performance on a series of sequentially arranged educational objectives derived from the curriculum

used in the classroom." Clearly, "CBA is the practice of using the material to be learned as the basis of determining the degree to which it has been learned" (Tucker, 1985). It is a procedure for determining the instructional needs of individual students based on ongoing performance in the existing curriculum. According to Tucker, data collection, interpretation, and application are all integral parts of CBA.

Marston and Magnusson (1985) state the primary goal of CBA as aiding the instructional decision-making process. Hargis (1995), on the other hand, declares its primary objective as the success of all students.

Also referred to as frequent measurement, continuous curriculum measurement, therapeutic measurement, and curriculum-based measurement, CBA grew out of work at the Institute for Research on Learning Disabilities at the University of Minnesota (Ysseldyke, Thurlow, Graden, Wesson, Algozzine, & Deno, 1983). The research conducted at the University of Minnesota initially examined the assessment process for students who were disabled. One major finding coming out of the Minnesota Institute was that curriculum-based measurement (CBM), a type of CBA, rather than commercially published norm-referenced tests, was technically adequate and that it was useful for evaluating student progress and improving student performance.

Fuchs (1995) explains that the focus of CBM is long term. She explains that the teacher establishes a broad outcome for students and then uses CBM methods to measure student proficiency by creating a pool of equivalent assessments, each of which samples the key problem types. The student completes one or two assessments each week. Because each assessment is of equal difficulty and incorporates all of the problem types to be learned over the year, the CBM database provides a total score graphed over time to show progress over the year. Fuchs states that CBM satisfies six of the seven criteria for assessment, addresses the three purposes of assessment, and incorporates standardized measurement techniques, providing reliability and validity. She further asserts that it offers detailed information on a student's performance on specific skills and can be used to determine how to im-

prove an instructional program. Since its measurement framework is not tied to any particular model of instruction, a broad range of instructional options can be used. In addition, a teacher can use widely varying methods with the same child to see which method is most beneficial. Moreover, the assessment demands are manageable in the classroom setting, and computer programs have been developed to administer these assessments and manage the data.' It should be recognized, however, that extensive academic background and ongoing professional development and technical support for teachers are necessary for the successful use of CBM. Also, its effectiveness in providing assessment information is probably strongest when used longitudinally for individuals and classrooms rather than for large-scale assessment.

3. *Curriculum and assessment reform occur in tandem, with subsequent modifications of one being rapidly reflected in changes in the other: The Advanced Placement (AP) program and the Pacesetter program.*

Advanced Placement (College Board, 1996a, 1997) is hardly a "new kid on the block" in the secondary education scene, having been in place for more than 40 years. However, its expansion from about 100,000 exams taken by 80,000 students in 1976 to 843,000 exams taken by 537,000 students enrolled in 68,000 AP classes is noteworthy. Curriculum and assessment were developed together, and changes are reflected in both, systemically. The annual involvement of secondary teachers in scoring of examinations provides feedback to teachers through workshops and College Board regional and national meetings on instructional and assessment issues. Much of the program's growth has been in schools that previously had fewer college bound seniors, and the financial involvement of state school systems has been a major factor in this growth. Students get college credit for satisfactory AP assessment performance and develop further confidence in their ability to do college work.

Pacesetter is a much newer program, also from the College Board (1996b), and is designed to provide a demanding, college preparatory curriculum, but with relevance to real-world events. For example, Pacesetter mathematics is a course in Elementary Functions that makes extensive use of graphics calculators. Students derive functional relationships from actual events that are unfolding, strengthening their understanding of abstract conic sections and trigonometric identities. Of course, the same mathematical relations emerge, but they can be understood in context, and they provide meaning to the abstract symbols. We are only beginning to see results from Pacesetter, but students are moving on from these courses to Advanced Placement, and classroom observations show extensive engagement and intensity in the quality and characteristics of student participation (Badger, 1996; Camara, 1996).

Curriculum, Assessment, and Systemic Reform

Systemic education reform is defined as a comprehensive effort to improve education simultaneously from "bottom up" and from "top down" through coordinated state policies that support changes at the local level. States work from "top down" by establishing policies and allocating resources. Local schools work from the "bottom up" by planning and implementing improvements that are consistent with state policies, yet sensitive to local needs.

In addition, systemic reform is based on the assumption that all children can learn challenging content and that curriculum, assessment, instruction, and professional development must be aligned or made consistent with each other to ensure that students achieve high standards.

As a result of the current wave of systemic reform, the nation is undergoing major curricular reform that has fostered the development of content and performance standards. These standards have had a great impact on educational assessment. Consequently, assessment reform often centers on two persistent criticisms of standardized tests: that current standardized tests are external to a particular school system and that standardized test developers do not link tests to specific curricula (Nitko, 1995). Nitko believes that the implication of these criticisms in the context of systemic and curricular reforms is that we

should encourage curriculum-based assessment reform.

What Have We Learned?

CBA has many advantages over traditional methods of assessment (Ysseldyke et al., 1983). These include:

1. *Improved communication of **assessment per-formance**.* The simplicity of data presentation through the use of graphs is easily understood by teachers, parents, and students (Ysseldyke et al., 1983). Progress can be clearly depicted.
2. *Increased sensitivity to and direct impact on the student's curriculum.* CBA directly impacts the student's curriculum by providing the teacher with information about when to alter a student's instructional program (Mirkin, Deno, Tindal, **&** Kuehnie, 1982; Idol et al., 1986).
3. *Peer referencing.* It is possible to norm student performance by sampling regular classroom peers. This shows an individual student's growth in relation to peers in the same curriculum. CBA also enables a school or distinct to establish its own norms (Deno, 1985; **Marston &** Magnusson, 1985).
4. *Bias elimination.* CBA has been shown to reduce substantially the bias in teacher referrals (Mirkin, **Marston, &** Deno, 1982) and the bias inherent in standardized tests (Galagan, 1985).
5. *Streamlined instruction.* Teachers who use CBA to evaluate students' skill areas will **find** that instruction becomes more streamlined; students can be offered instruction for unmastered areas without receiving repetitive instruction in previously mastered areas (Idol et al., 1986).
6. *Increased collaboration.* **CBAs** can be developed collaboratively with other teachers who use the same curriculum. They can be used at the beginning of the school year or in segments throughout the year. They can be used **with groups of individuals or with individu-als.** If multiple forms are developed, they can be used to monitor progress over time (Idol et al., 1986).

7. ***Significant*** *achievement.* Several studies have demonstrated that students achieve at significantly higher levels when CBM is used. The teachers who used CBM were more structured in their instruction and more realistic and responsive to student progress (Fuchs, Deno, **&** Mirkin, 1984). The students were more aware of their educational goals and were more able to accurately predict whether they could meet these goals (Fuchs et al., 1984).
8. *Growth measured over time.* CBM is particularly effective in measuring academic growth over time (Marston, Deno, **&** Mirkin, 198 1). It was demonstrated that the sensitivity of these measures in the early grades suggests that they may be particularly useful for evaluation of the instructional programs of problem learners.

Research Focus

According to Nitko **(1995),** in most of the curriculum-based assessment reform discussion, there is only passing reference to practical **long-**term assessment development procedures. Most of the discussion, according to Nitko, focuses on what is necessary to develop a single assessment, as contrasted with models for sustainable ongoing production processes. He states that, if **CBAs** are to become the primary assessment modes in schools, the process of producing **fresh** annual assessments needs to be articulated. In articulating a production process model, some of the following questions should be answered:

1. How are curriculum-assessment linkages forged?
2. How are curriculum-assessment linkages sustained and validated throughout the development process?
3. What level of technical and financial resources is needed to sustain high quality product production?
4. What types of organizations are capable of sustained production of high quality curriculum-based assessment?

Conclusion

One aspect of attaining the goals of systemic reform is coordinating assessment practices with what is taught. With the push coming from the assessment, the curriculum, or the common development of both, these reforms aim to do just that.

The statewide MSPAP applied consistent pressure and support to alter the way in which school subjects were taught. Advanced Placement has offered a consistent, high level standard for superior performance, and the use of state funds has made it broadly accessible to many schools who could not support it locally. Pacesetter is providing a bridge from elementary algebra and geometry to more advanced mathematics.

The use of curriculum-based assessments can assist in providing a more accurate means of assessing what children have learned. CBA measures the achievement of students in terms of the expected curriculum outcomes of the school (Germann & Tindal, 1985). It is also oriented toward instruction as methodology and can be used to determine the instructional needs of all students based on their performance within existing content concepts (Ginkling & Thompson, 1985). With this process, it is hoped that the instructional decision-making process will be facilitated and that every child will succeed.

Though NSF has a rich lode of reform credits, these programs show that many successful reform activities are state-funded, or supported by private foundations supplemented with local efforts. This climate has probably helped give local systemic activities more time to incubate and take hold, provided adequate attention and staffing to extend scope and to make the idea of systemic change a believable goal in schools and communities.

References

Badger, E. (1996). Finding one's voice: The Pacesetter model for more equitable assessment. *The College Board Review, 179,* 12-19

Blankenship, C. S. (1985). Using curriculum-based assessment data to make instructional decisions. *Exceptional Children, 51,* 232-238.

Camara, W. (1996). *Overview of Pacesetter evaluation efforts* (College Board memorandum, Feb 3, 1996). New York The College Board.

Cohen, L., & Spence, S. (1990). Fundamental considerations of curriculum-based assessment. In L. Cohen & J. Spruill (Eds.), *A practical guide to curriculum-based assessment for special educators* (pp. 3-14). Springfield, IL: Charles C. Thomas.

Cohen, L., & Spruill, J. (1990). *A practical guide to curriculum-based assessment for special educators.* Springfield, IL: Charles C. Thomas.

College Board. (1996a). *Advanced Placement yearbook* (Brochure). New York: Author.

College Board. *(1996b). Pacesetter: A model for districtwide implementation* (Brochure). New York: Author.

College Board. (1997). *A guide to the Advanced Placementprogram* (Brochure). New York: Author Deno, S. (1985). Curriculum-based assessment: The emerging alternative. *Exceptional Children, 52,* 219-232.

Fuchs, L. S. (1995). *Connecting performance assessment to instruction: A comparison of behavioral assessment, mastery learning, curriculum-based measurement, andperformance assessment* (Report No. RR 93002005). Reston, VA: Council for Exceptional Children. (ERIC Document Reproduction Service No. ED 381 984)

Fuchs, L., Deno, S., & Mirkin, P. (1984). The effects of curriculum-based measures and evaluation on student achievement, and student awareness of learning. *American Educational Research Journal, 21, 449-460.*

Galagan, J. E. (1985). Psychoeducational testing: Turn out the lights. The party's over. *Exceptional Children, 52, 288-299.*

Germann, G., & Tindal, G. (1985). An application of curriculum-based assessment: The use of direct and repeated measurement. *Exceptional Children, 52, 224-265.*

Ginkling E. E., & Thompson, V. P. (1985). A personal view of curriculum-based assessment. *Exceptional Children, 52,* 205-218.

Hargis, C. H. (1995). *Curriculum based assessment: A primer.* Springfield, IL: Charles C. Thomas.

Idol, L., Nevin, A., & Paolucci-Whitcomb, P. (1986). *Models of curriculum-based assessment.* Rockville, MD: Aspen.

Koretz, D., Mitchell, K., **Barron**, S., & Keith, S. (1996). *Final report: Perceived effects of the Maryland School Performance Assessment Program* (CSE Report No. 409). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Marston, D., Deno, **S., &** Mirkin, P. (1981). *An analysis of learning trends in simple measures of reading, spelling and written expression: A longitudinal study* (Report No. 49). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities.

Marston, D., & Magnusson, D. (1985). Implementing curriculum-based assessment in special and regular educational settings. *Exceptional Children, 52, 266-272.*

Mirkin, P., Deno, S., Tindal, G., & Kuehnie, K. (1982). Frequency of measurement and data utilization as a factor in behavioral assessment of academic skills. *Journal of Behavioral Science, 4, 36* 1-37 1.

Mirkin, P., **Marston,** D., & Deno, **S.** (1982). *Direct and frequent measurement of academic skills: An alternative to traditional screening, referral, and identification of learning disabled students (Re-*port No. 75). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. ED 224 191)

Nitko, A. J. (1995). Is the curriculum a reasonable basis for assessment reform? *Educational Measurement: Issues and Practice, 14*(3), 5-10,35.

Tucker, J. A. (1985). Curriculum-based assessment: An introduction. *Exceptional Children, 52,* 199-**204.**

Ysseldyke, J. E., Thurlow, M., **Graden,** J., Wesson, C., Algozzine, B., **&** Deno, S. (1983). Generalizations from five years of research on assessment and decision making. *Exceptional Education Quarterly, 14, 75-92.*

Thomas A. Romberg
Sears Roebuck Foundation-Buxom Professor in Education
Director, **National Center for Improving Student Learning and Achievement**
    in Mathematics and Science
University *of Wisconsin-Madison*

There is consensus that throughout the world societies are evolving from the Industrial Age into the Information Age. The most important features of the Information Age are that it represents first a profound switch from physical energy to brain power as its driving force and, second, a change from concrete products to abstractions as its primary products. Information is the new capital and the new raw material; the ability to communicate is the new means of production; and the communication network provides the means of relaying that production. The assumption is that our citizens need a better understanding of mathematics, science, and technology if our society is to prosper in the emerging Information Age.

As a consequence, for over a decade the public has demanded action (e.g., National Commission on Excellence in Education, 1983; National Science Board Commission on Precollege Education in Mathematics, Science, and Technology, 1983). Because of these demands, the National Council of Teachers of Mathematics **(NCTM)** in 1986 formed a Commission on Standards for School Mathematics with the mission to present a vision of the mathematics curriculum, methods of instruction, techniques of assessing student performance, and procedures for evaluating programs that would be needed.' Here I only address issues related to the school mathematics **curriculum.**

The task of developing a new school mathematics curriculum must be viewed as a design task. The emphasis on design implies that current materials are inadequate. Thus, simple alterations of existing programs do not **suffice.** Instead, the fundamental way in which mathematics programs are organized and developed must be changed so that a radically new program is created. In fact, curriculum development is seen as more than a change in content and method; it is an effort to change the instructional culture of schools. By curriculum, it also should be understood that what is to be developed is a total in-

structional package, not just a curriculum guide or a basal textbook series. The message I want understood is that the vision NCTM presented is for a mathematics program, that is, for a mathematics curriculum that expects all students to have an opportunity to learn more and somewhat different mathematics **than** in the past. Such a program is to be significantly different from current practice, and the emphasis in the vision is that mathematics "make sense." It should make sense to students, and students should use mathematics to help them make sense of the world.

Systemic Reform Strategy

The strategy that underlies the curriculum reforms was based on the notion that, because we live in a supply-and-demand economy, if the mathematics community wanted different texts and tests, a demand would have to be created. To respond to this challenge, the mathematical sciences community has followed a seven-step iterative strategy (see Figure 1).

The steps and the relationships between them are as follows:

1 Before any plan can proceed, a need for change must be established.
2 Vision is a key factor. To create a new program we must consider values, goals, and standards. The NCTM standards documents were designed to fulfill this vision.
3 Planning includes involving everybody in a system or school in arriving at consensus about the details of long- and short-range plans (with timetables) for change. It is at this step that demand is created.
4 The next step involves **identifying** specific elements of the system to be targeted for change (curriculum materials, instructional methods, examinations, teachers, technology, etc.) and setting priorities.
5 Any system depends on suppliers. Schools must demand that textbook publishers, testing companies, staff developers, teacher education programs, and others contribute the ingredients necessary for the desired changes in curriculum elements.
6 Then it is time to make the new materials, procedures, and programs operational. Draft

materials and procedures need to be tried out, feedback from this trial phase matched with the vision and the plan, and revisions made.
7 Finally, a product (a curriculum, an instructional procedure, assessment materials) is developed. Quality should then be judged in terms of what students are able to do (the effect of the curriculum, procedure, materials) and whether this meets society's needs.

Design Principles

The following five principles are a consequence of my experience from directing the development of **Mathematics in Context** (MiC; National Center for Research in Mathematical Sciences Education & Freudenthal Institute, in press) a new curriculum for the middle school (Grades **5–8**).[2]

Principle 1. Conceptual domains should be specified.

The mathematical domains that we expect students to engage must be identified, and a curriculum then built around those conceptual domains. The domains should be selected because of their generality and ability to subsume more specialized components of the curriculum deemed desirable for the development of problem-solving ability and quantitative reasoning. These domains should not be considered independent of each other. Although it is true that each domain has some unique properties (signs, symbols, rules), I would rather think of them as the roots of a tree whose trunk involves problem solving, communication, and reasoning.

Principle 2. Curriculum units, each of which takes two to four weeks to teach and each of which tells a story, need to be constructed.

Each unit should provide students with an opportunity to investigate increasingly complex problem situations within, and often across, domains. Students should be expected to construct meanings, interrelate concepts and skills, and use those meanings in a variety of problem situations. Each unit should be similar to a chapter in a Dickens' novel. It should introduce or reintroduce the characters to the reader, and there
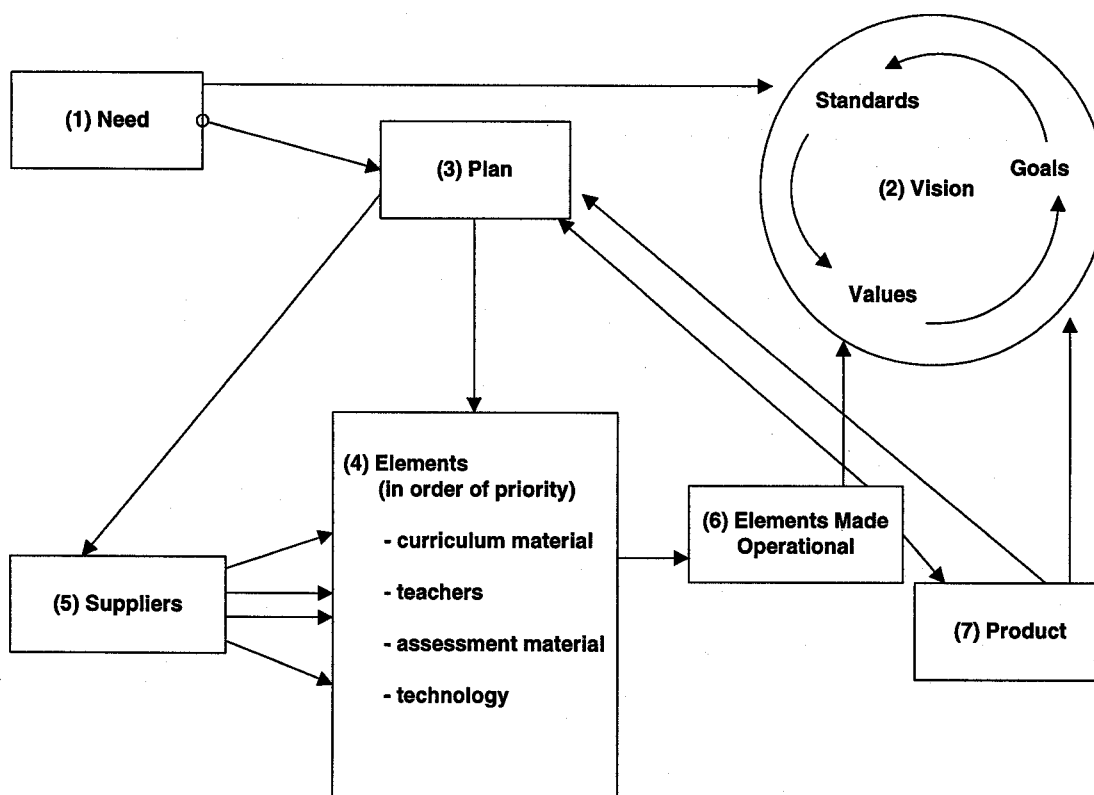
Figure 1. Reform strategy steps

should be a new problem situation to be explored that involves conflict, suspense, crisis, and resolution.

Principle 3. Students should be exposed to the major features of a mathematical domain as they arise naturally in problem situations.

Ideas are best introduced when students see a need or a reason for their use. Promoting the development of understanding requires an integrated curriculum. Problem situations may include the historic reasons for the development of a mathematical domain, the relationship of that domain to other domains, or the uses of ideas from that domain in other fields.

Principle 4. The activities within each unit should be related to how students come to understand a domain.

The activities embedded in these units should include having students (a) construct relationships, (b) extend and apply mathematical knowl-

edge, (c) reflect about their mathematical experiences, and (d) communicate what they know to others. Research shows that learning with understanding can be accomplished if the activities focus on the following interrelated themes:

1 *Modeling*, or representing phenomena in the world by means of a system of theoretically specified objects and relations (often mathematical relations). In classrooms, we find it fruitful to consider modeling as a cycle including model construction, model exploration, and model revision.

2 *Argument* and standards of evidence, with an emphasis on promoting students' skills for generalization in mathematics.

3 *Big ideas*, reflecting that student work needs to be about important mathematical ideas.

4 *Equity*, consistent with contemporary views of mathematics that stress that this discipline is developed in social contexts rather than the isolated pursuit of "truth." This view suggests that the very nature of mathematics is

defined communally, making participation by all not only a fundamental civil right, but also important to the continued vitality of mathematics and science to the nation.

5   *Assessment,* reflecting that classrooms, schools, and parents need ways of ensuring that classroom practices that foster understanding are, in fact, improving student achievement.

Furthermore, each unit should both provide review of prior concepts and skills and lay foundations for concepts and skills to be learned later. Activities used to teach algorithms should differ from those used to teach problem solving. Additionally, the method of instruction is likely to differ. Students might be addressed as a large group when being exposed to new information and work in small groups when conjecturing, proving, or applying. Some activities may require exercises requiring repeated practice, whereas others may involve a dissimilar array of problem situations involving varying cognitive structures. A higher degree of teacher-imposed structure and control may be desirable for lower-level cognitive outcomes, whereas a greater degree of group autonomy may aid higher-level cognitive outcomes.

Principle 5. Curriculum units should always be considered adaptable.

All curriculum sequences need to be adapted and modified in light of what knowledge the students bring to the unit and the context in which instruction takes place. The difference between the intended and the actual curriculum should be apparent. What actually occurs will differ across classrooms. A unit cannot be "teacher proof." Instead, the program should assist each teacher in making reasonable adaptations so that the prior knowledge and interests of the students are taken into account in instruction.

Summary

The design of new mathematics curricula should be seen as a critical ingredient in the current systemic reform movement. New curriculum materials are a necessary, but not sufficient, component in changing schooling practices.

Notes

1   The three standards documents published by the National Council of Teachers of Mathematics are *Curriculum and Evaluation Standards for School Mathematics* (1989), *Professional Standards for Teaching Mathematics* (1991), and *Assessment Standards for School Mathematics* (1995).

2   The development was funded by the National Science Foundation and created with the help of staff from the Freudenthal Institute at the University of Utrecht, The Netherlands.

References

National Center for Research in Mathematical Sciences Education & Freudenthal Institute (Eds.). *(in* press). *Mathematics in context: A connected curriculum for Grades 5-8.* Chicago: Encyclopaedia Britannica Educational Corporation.

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: Author.

National Council of Teachers of Mathematics. (199 1). *Professional standards for teaching mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics.* Reston, VA: Author.

National Science Board Commission on Precollege Education in Mathematics, Science, and Technology. (1983). *Educating Americans for the 21st century: A plan of action for improving the mathematics, science and technology education for all American elementary and secondary students so that their achievement is the best in the world by 1995.* Washington, DC: National Science Foundation.

The Role of Teaching and Learning in Systemic Reform

Christopher Dede
Senior Program Director
National Science Foundation

To reach the next stage of large-scale reform, research-based answers are needed to the following types of practitioner questions about teaching, learning, and the institutional and societal context of schools.

Pedagogy and Assessment

What evidence supports claims that **innovative** pedagogies such as constructivism and situated cognition are important complements to conventional, assimilation-centered instruction?

How can learners' collaborative performance on authentic tasks be assessed without spending so much time and effort on evaluation that teaching and learning are adversely affected?

What additional types of skills and knowledge do teachers need to implement these alternative approaches to instructional design and assessment? What role does educational technology play?

Content and Curriculum

What evidence supports claims that standards-based curricular content better prepares students for 2 1 st century employment and citizenship? How can information technology expand the spectrum of topics taught in the curriculum and convey complex material to a broader range of learners earlier in their schooling?

How can disparate pieces of innovative content developed by various projects be integrated into a seamless curricular framework that meets emerging national content standards, employers' expectations, and requirements for entrance into higher education? For example, how can new types of SMET content that information technology makes accessible be interwoven into the curriculum?

Information Technology

What types of computing and communications equipment in schools should complement the information infrastructures emerging in workplaces, homes, and communities?

What innovative financing strategies and re-allocations among existing expenditures can educators use to fund technology-related expenditures (i.e., initial infrastructure, software, professional development, maintenance and ongoing costs, depreciation)? What other types of policy initiatives can enhance wise usage of educational technologies?

Enhancing Learning Outside of Classrooms

How can educators encourage all **stakeholders** in high quality schooling (e.g., families, businesses, public social services, communities, the media) to assume shared responsibility for students' learning and to act as partners in educational improvement? What role can communities' information infrastructures play in this process?

Management of Schooling

What types of alterations in organizational functioning (e.g., time schedules; flows of information, responsibility, and authority; resource distribution) are required to support these new models of curriculum, teaching, learning, assessment, professional development, and schooling? How can information technology advance this shift?

What additional types of skills and knowledge do administrators need to effectively manage such an organizational structure? What evidence supports claims that these alternative models of educational management are affordable and sustainable?

Professional Development

Beyond those practitioners fluent in innovation, what types of professional development can motivate and prepare typical teachers and administrators to master these new models of curriculum, teaching, learning, assessment, and schooling? How can information technology aid in these efforts?

How can the recruitment and preparation of preservice educators maximize the quality of human resources entering the profession and enhance their ability to implement these innovative educational approaches?

Equity

Given growing diversity in the learner population and increasing imbalances in the resources students can access outside of classrooms, how can educational opportunity be maximized for every type of learner?

How can potential scientists, mathematicians, engineers, and technologists from historically underrepresented groups be encouraged and supported to enter these professions?

Policy

What policy frameworks at the local, state, and national levels best support these new models of curriculum, teaching, learning, assessment, professional development, and schooling?

When educators provide an excellent curriculum, powerful instructional technologies, exemplary teaching and assessment, and support for learning outside of school, what kinds of learning outcomes should be the goal for student performance at various developmental levels?

What evidence can practitioners give policymakers and the public that providing the financing required for this alternative model of education is a wise investment of America's resources?

Through applying the results of research testbeds exploring innovative strategies that address these issues, practitioners can make informed decisions about educational reform. Of course, many intermediate questions must be addressed by researchers, designers, evaluators, and policy analysts to build the knowledge required to generate responses for practitioners.

For example, innovative ways to measure the outcomes of reform efforts are needed. New methods of data collection and presentation must be developed to capture the changes underway in U.S. school systems. As one illustration, the Third International Mathematics and Science Study (TIMSS) included new measures of curriculum, classroom practices, testing practices, and teacher performance. Through initiatives that adapt and extend the methods applied in this study-and other indicators projects-to the full range of educational innovation efforts in the U.S., researchers can help policymakers reach informed decisions about aiding systemic reform.

Researchers must also go beyond the typical questions educational innovators now ask to frame emerging issues central to reform. Often, practitioners are unaware of opportunities potentially available to them. For example, accountants in the late 1970s did not foresee the advent of spreadsheets, yet this information tool revolutionized the knowledge, skills, and organizational structure underlying financial management. Similar "targets of opportunity" that research should explore today include methods by which learners can engage in self- and peer-assessment, ways to teach complex content earlier in the curriculum to a wider range of students, and strategies for developing curricular standards beyond discipline-based ratification of content and skills.

Continual, rich dialogue with practitioners is important in ensuring that educators are not simply passive recipients of research insights, but also active contributors in formulating, selecting, and implementing these studies. Only through collaborative interaction with practitioners can researchers generate findings that are useful for systemic reform and build the intellectual capacity of this community. Inculcating in educational innovators a sense of "ownership" of research studies is vital in enhancing the dissemination and acceptance of their results.

LeRoy Lee
Executive Director
Wisconsin Academy of Sciences, Arts, and Letters

The National Commission on Teaching and America's Future, the National Foundation for the Improvement of Education, and many other organizations and individuals are calling for changes in professional development. Recommendations range from reinventing teacher preparation and professional development to the establishment of professional development centers and learning networks that transcend typical school boundaries.

The Wisconsin Academy, a nonprofit membership organization, has been involved in staff development for over twenty years. Staff development activities of the Wisconsin Academy are characterized by (1) collaborative involvement of the broad educational community, agencies, and business; (2) emphasis on adult learning principles; (3) emphasis on participants as professionals; and (4) long-term contact with participants that provides continued encouragement to become involved in leadership positions as well as professional activities. This paper describes three current efforts, all of which are supported in part by the National Science Foundation

Teacher/Student Model

Field Investigations: Research by Science Teachers (FIRST)

FIRST, now in its seventh year, involves 35 to 40 teachers and their students each year in field research under the overall guidance of field research scientists. The concept is not new. Early studies indicated that teachers without research experience did not perceive research as practicing scientists did. NSF initiated the Research Participation Program (RPP) in 1958 in an effort to provide teachers with research experience. Evaluations of the RPP projects indicated a high percentage of participants could do quality research and did change their perception about the process of science.

FIRST has also found that active engagement in solving real research questions has made a dif-ference in teachers' content knowledge and their teaching practices as well as their perception about the process of science. It differs from other RPP projects by actively engaging the participant and his or her students as a research learning community. FIRST also differs in that the field scientists involved represent a wide array of agencies and research institutions as well as academic specialties.

While the external evaluations have documented success, including the increased professional involvement of participants, there is no long-term study to determine whether there is sustained change in teaching behavior. Equally important, there is no long-term study to determine the impact on participating students. The new NSF ESIE program, Teacher and Student Development through Research Experience Projects, may'provide some data over time. FIRST, like the newly established NSF program, is a one-teacher-at-a-time model and would be **difficult** to scale to a significant number of teachers or schools.

School Team Model (teachers/principal)

Field Experiences for Science Teachers (FEST)

FEST has three primary goals: increased knowledge of geology and ecology; developing positive attitudes toward science teaching and learning; and implementation of science standards. Teams from six elementary schools are selected. Each team consists of the principal and from four to six teachers self-selected across grade levels. Field geology and ecology concepts are developed through a constructivist approach as teams visit Wisconsin sites that exemplify the concept. Teams are guided through a standards-based review of their current earth and environmental science curriculum and work to supplement curriculum to meet standards not addressed.

External evaluations have documented increased content knowledge. Pre-post results of an

efficacy belief instrument are "nothing less than amazing" according to the evaluator. Teacher evaluations have placed standards development as a major highlight of the project. Evaluations have also shown that it is extremely important to have the principal involved on equal status as a participant with the other team members. It is also important to have a team that represents a critical mass from the school, as well as involvement of more than one grade level.

While successful in causing school change, FEST represents a one-school-at-a-time model that would be **difficult** to scale to a level that would involve a significant number of schools. There is no long-term study in place to determine whether change is sustained. While the efficacy instrument has indicated dramatic change, it is not clear what factors were involved in causing the change.

Large Scale, Bottom-Up Staff Development
Wisconsin Academy Staff Development Initiative (WASDI)

WASDI is a coordinated statewide dissemination of a K-12 staff development program to improve science, mathematics, and technology education. WASDI consists primarily of two components: Establishment of ten academies geographically located throughout Wisconsin, and a Lead Teacher Institute to prepare change agents and facilitators for the summer programs.

The model is based on the successful Cray Academy. The Cray conducts two 1-week sessions each summer, each of which provides participants a choice of thirty l-hour hands-on workshops in science, mathematics, or technology. The Cray Academy developed over seven years based on teachers' needs and concerns: (1) workshops that teachers identify as needed, (2) practical information; (3) peer-based instruction; (4) one-week experiences that still provide summer time for other activities; (5) workshops held in a familiar site, such as a school; and (6) workshops within easy driving distance with good parking so that teachers can return to their families in the evening. External evaluations document that teachers have changed attitudes, gained new in-

sights, and significantly increased the amount of hands-on activities used in their classroom.

Collaboration in program development is an important part of the statewide model. Site directors meet regularly to share experiences and directions. They are also electronically linked. As part of the model, each academy site also develops a local board of directors consisting of representatives from business, education, and agencies to provide guidance and oversight.

In 1996 over 1,800 teachers participated at eight different Wisconsin sites. To conduct the large scale program involved in-kind and cash support from fifteen universities and colleges, over 200 businesses, several foundations, and over ten associations and agencies.

Internal and external evaluations have been positive. Questions, however, remain. While the project knows what will draw teachers, it is not known why most teachers elect not to participate. Answers could lead to better designed opportunities. Business and funding agencies desire data relating to student performance. Since Wisconsin does not have a state curriculum or test, it is difficult to show systemic effect as determined by improved student performance.

The second phase of the WASDI project involves working with 60 kindergarten through twelfth-grade teachers selected each year to become change agents in their school and the state. They also serve as primary instructors for the summer academies. The Lead Teacher Institute provides seven weeks of activities relating to areas such as curricula, national standards, the change process, and presentation skills. The approach is collaborative and reflective with the participants proactively involved in the staff development process and program.

A significant part of the Lead Teacher Institute is the on-line learning community using conferencing software. Nine Lead Teachers were selected to facilitate on-line focused discussions. Primary conference sessions include mathematics, science, and technology education. Current conferences on constructivism, leadership, technology, and politics may change as interests or action research questions develop. A private mailbox is also available for each participant. In

the course of one year participants have been on line more than 25,000 hours.

The external Institute evaluation found that the Lead Teacher experience appears to have been particularly good at helping the leaders crystallize their conception of the role they will play as a leader and to provide the confidence and authorization/recognition necessary to implement that role.

Susan *Loucks-Horsley*
Center for Science, Mathematics, and Engineering Education and *WestEd*

The purpose of systemic reform is to improve student learning, which cannot be accomplished without excellent teaching. It is not a surprise, then, that professional development plays a critical role in the success of systemic reform, as it directly influences the quality of teaching and learning in science and mathematics classrooms. This paper draws on my experiences in designing and conducting evaluations of professional development in the context of systemic initiatives at local and state levels, in providing technical assistance to professional developers, and in capturing the experiences of seasoned professional developers in a book on best practice. In the paper I sketch briefly (1) what I believe we know about the role of professional development in systemic reform and (2) what we still need to learn.

What We Know

1. It is a long distance from the policy level to the student, andprofessional development is on the way. In my new role as Director of Professional Development and Outreach for K-12 at the National Research Council's Center for Science, Mathematics, and Engineering Education, I have the task of overseeing the Center's efforts to "disseminate" the *National Science Education Standards* (National Research Council, 1996). It is a constant source of amazement how many people think that you can literally give the book to teachers and expect them to use the *Standards* in their teaching. These standards are a product of a national consensus; the many sets of standards developed at other levels of the system (e.g., by states and districts) similarly result from broad consensus. Their intention has never been to be "implemented" directly, but to guide a system's design for what educators expect of and how they work with students.

Bybee (1996) describes a schema for system change that applies equally well for mathematics reform as it does for science reform; it includes changes in purpose, policies, programs, and practices. According to this schema, purposes relate to the general agreement on the need for science and mathematics literacy for all; standards are the policies that guide education toward those purposes. But in order to move to students, programs need to influence practice, which is the only way that students will have different and better opportunities to learn. This is where professional development enters the picture. Professional development is one of the critical links in this chain, one that can take purposes and policies and influence student learning through its impact on teaching.

We have learned that there is a great distance between systems and students. Although there are many routes that may be chosen (e.g., through new assessment, curriculum, or instructional programs), professional development is a required stop along the route. For students to reach the goals to which the system aspires, teacher learning and change are essential.

2. Investment in people as the primary agents **of change is critical.** Many proponents of systemic reform concentrate on the need to change policies at the state and local levels. Their vision came in part from the California experience of the 1980s and early **1990s,** when the state began to enact a vision that put into place the critical elements of state frameworks, assessments, curriculum adoption criteria, and professional development (Honig, 1990). As other states enact this policy-level focus, they would do well to **exam-**

ine the California situation carefully, as it has evolved. At this point, many of these critical elements are either lost (i.e., the state assessment) or being threatened (e.g., some of the state frameworks). Policies are as good as the politics that help them get established-they may have a shelf life only as long as a current administration.

What is encouraging in California is that the teachers and other educators who have "grown" this reform, not as much from the grassroots but from the developing infrastructure, are keeping the reform alive and well in many locations. The infrastructures are the statewide professional development networks, two of which have been supported through the NSF Statewide Systemic Initiative (SSI), the Mathematics Renaissance and the California Science Implementation Network, and others as well, such as the California Subject Matter Projects. The investment in people through professional development that has been made by these projects has created a strong fabric that is resistant to change, people whose teaching can never return to prereform practices, and who can articulate what is important and why. In evaluating the California SSI, we have seen what we call "inside-out" systemic reform, i.e., changes in the system that result because people are changing and are influencing the structures, procedures, and, in some cases, the policies that guide teaching and learning (Aquarelli & Mumme, 1996). Of the several hundred schools and thousands of teachers who have been touched by the two SSI networks, we have hundreds of examples of network teachers and administrators taking on new leadership roles within buildings and districts (e.g., teachers becoming principals and curriculum supervisors, principals and teacher leaders becoming assistant superintendents), in their local and state professional associations, and as members of state and local committees whose role it is to make curriculum, assessment, and instructional decisions. We have documented dozens of instances of these mathematics and science initiatives influencing changes in other content areas in schools and districts, the nature of professional development offered by county offices and higher education institutions, and teacher preparation programs, both on campus and in clinical settings.

Most interesting, perhaps, is the statewide influence of these professional development networks on assessment and standards development. For example, when CLAS, the new performance assessment system, was canceled by the governor in 1995, a collaborative of districts facilitated by science professional developers was determined to have the kinds of testing program for students that CLAS had offered. Through their collaboration, the CLAS test was revised for use in districts and schools last fall, and an NSF-funded project began at the same time to develop similar tests for the state and other interested systemic initiatives. Another example is the writing of state-mandated science standards, taken on voluntarily by the coalition of state professional development projects, once again determined not to lose the essence of the reforms for which they had worked so hard.

Fullan (1993) emphasizes the importance of all educators being change agents, that it takes people to make change. In a newer article (1996) he "turns systemic reform on its head," arguing for the very people-driven networks that we are seeing stay the course of reform in California. California serves as a warning to those systemic initiatives who have relied heavily on their policy initiatives and neglected the building of strong networks dedicated to professional learning at the individual and school level. They say that it takes a village to raise a child; it takes the people in it to educate the child. As California may have been' seen early as a prototype for systemic reform, it may also be a proving ground for how to sustain reform when there is turbulence in the system. That people and their traditional strategies last tenaciously through policy changes has been a curse of many reform initiatives. That people, once changed, can in fact remain changed may turn this curse into a blessing. Professional development may sustain systemic reform when change at the system level fails.

3. The professional development needed by systemic reform is not the same kind as supported change initiatives in the past. The new paradigm for professional development that Dennis Sparks first called to our attention in 1994 is not about one-time, one-teacher-at-a-time,

expert-driven workshops or institutes held for teachers far from their schools and classrooms. Professional development for systemic reform is larger in:

- scale (i.e., it serves more people in a wide variety of roles),
- scope (i.e., it pays attention to more elements of the system, e.g., curriculum, assessment), and
- duration (i.e., it is intensive and extends over time).

It has many of the characteristics of **effectiveness** identified through research and in the practice of experienced professional developers, such as collaborative work, expertise derived from research as well as expert practice, an emphasis on content understanding, and continuous evaluation (see a synthesis of the national standards related to professional development by Loucks-Horsley, Stiles, **& Hewson,** 1996). Further, like teaching, professional development is dynamic. Bather than selecting from an established set of models to support professional learning, professional developers who successfully design initiatives in the interest of systemic reform use a decision-making process that involves identifying their goals, understanding their context, and creating a unique combination of specific learning strategies that is tailored to their initiatives. A design model and 16 strategies derived from best practice in professional development design have been articulated by the National Institute for Science Education's Professional Development Project (Loucks-Horsley, **Hewson,** Love, **&** Stiles, 1997).

*4. A strong **infrastructure** and deliberate+ developed capacity for change are needed to support the people and change the paradigm.* For educators in large numbers to learn about, try out, and maintain changes in their practice requires a support system with a shared vision of teaching and learning, such as those visions articulated for mathematics and science in the national standards **(NCTM,** 1989; NRC, **1996),** but with greater attention to creating shared images of what the vision looks like in practice-in the classroom interactions of teachers and students, in instructional materials, in student work and

assessments. The support system is staffed by people whose job it is to introduce, facilitate, and support change in the direction of the vision. These people have demonstrated skill in teaching young people as well as the abilities to address the learning needs of adults and build professional networks, both inside and outside of schools, to support ongoing learning (Lieberman **& McLaughlin, 1992).** They have a keen knowledge of the change process and how to work with people at different stages of change (Hall **&** Hord, 1987); skills in communication, problem solving, decision making, team building, and time and task management **(Fullan,** 199 1); and the ability to use pressure and support appropriately (Louis **&** Miles, 1989).

Effective infrastructures build capacity for ongoing change at the local level through design and use of a variety of professional development strategies that help teachers change their practice, through support of collaborative work inside of schools to support individual change and design and implement programs of study, and through building capacity for leadership in various members of the school and community (Friel **&** Bright, 1997).

*5. Professional development must pay **careful** attention to content knowledge.* With a renewed focus on concept development as a valued outcome of science and mathematics education, teachers are no longer able to "cover" for lack of preparation in the area they teach (which assignment is usually not their choice): Shulman's (1987) work in defining and explicating the term "pedagogical content knowledge" has added a new and critical dimension to professional development. Whereas generic professional development (e.g., learning generic teaching skills such as cooperative learning, effective instruction, and questioning techniques) was a hallmark of the **1980s,** we have learned the keen importance of teachers knowing how to teach particular content-understanding the conceptions students are likely to hold about certain mathematics and science concepts, what students of a certain age are developmentally able to learn, and what examples, analogies, and representations help them learn it. Such knowledge is difficult to learn in

preservice education, and is often the province of the experienced expert teacher (Shulman, 1987). This need for learning from a master teacher underlies the use and success of mentor and advising teacher programs (Shuhnan & Colbert, 1990).

6. Instructional materials can play a critical role in teacher as well as curriculum change. Most educators think of teaching and curriculum as two different components of the system, but we are quickly learning the power of materials to help teachers learn (Loucks-Horsley et al., 1997; Friel & Bright, 1997). Materials developed to teach students important concepts and skills represented in national standards, with teaching strategies that address a constructivist view of learning, help teachers try out new behaviors and experience for themselves what new forms of teaching look and feel like. In particular, teachers can see how these approaches work with students. Two professional development strategies use curriculum materials to support teacher learning (Loucks-Horsley et al., 1997). The first is curriculum implementation, in which a set of instructional materials is selected, teachers learn how to use them, try out the materials, reflect on their experiences, and are supported over time to refine their use. The second is curriculum replacement (Burns, 1995), in which teachers try out a unit that embodies new teaching perspectives and strategies and document and discuss their experiences in order to "try on" new ways of helping students learn. Both strategies promise to influence both how teachers teach and the materials they use to do so.

7. Professional development and organizational development must be inseparable. The largest professional association devoted to staff development, the National Staff Development Council, defines professional development as involving both individual and organizational development, because we know that individuals are unlikely to sustain what they learn when their organization does not support them to do so. It is one reason why the "last wave of reform" in science education, which provided opportunities for individual teachers to attend summer institutes away from their schools and districts, fell far short of its potential to change teaching and learning in substantial ways. For teachers to change what they do with their students, the organizations within which they work must change, in two ways. First, their schools and districts must support teachers' changes (e.g., provide materials support and time for collaborative planning and reflection; focus teacher evaluations on the changes). Second, the organizations must themselves become learning organizations, valuing experimentation and collaboration, encouraging deep examination and analysis of teaching and learning, and creating opportunities for extending and enhancing practice (Senge, 1990; Shanker, 1990). Such schools, described by Rosenholtz (1989) as "learning enriched" are characterized by high levels of student as well as adult learning. Without organization development, individual teachers are unlikely to sustain their learning; with it, not only do teachers learn, but their students do so as well.

What We Need to Know

Professional development is a field in which "definitive research" on what is effective does not exist (Frechling et al., 1995). Like teaching, it is too complex to understand by asking simple questions; it is highly influenced by factors out of control of either the professional developer or the researcher; and its success depends greatly upon the goals and context, which are idiosyncratic to a given situation. The ideas discussed above capture what I believe we know; they have come from a combination of research, literature, and the "wisdom of practice." In each case, we have some evidence, but we need closer study, some more existence proofs (i.e., examples of where and how these things work) to increase our certainty. As works-in-progress, professional development efforts lend themselves to examination. While much can be learned from them to further the education community's understanding of how different factors interact, including the people, the context, and the passage of time, they themselves can benefit from ongoing reflection and feedback. Such examination holds great promise for increasing our understanding of the

role of professional development in systemic reform.

As we examine current initiatives, here are some questions I think are important to ask.

1. How can we move from understanding how individual teachers learn and how to help them, to how to support the growth of millions of teachers? Mathematics educators, in particular, have become very expert at understanding how teachers learn and what can help them (Ball, 1996). Science educators, on the other hand, have increased our understanding about what system components are needed to improve the potential of success for change (St. John et al., 1994). The issue of scaling up, however, is still perplexing, as articulated well by Elmore (1996). We need to learn from the many systemic efforts currently underway what mechanisms, strategies, and system elements make learning possible for such magnitude as all teachers in the nation.

2. What are some ways of using scarce resources well, so that teachers have equitable access and opportunity to learn? It is widely acknowledged that, for teachers to make the changes envisioned in national and state standards, many hours, and also resources, must be devoted to their learning. Yet by any metric, there are not enough resources available to provide every teacher in this country the opportunities they need. Professional development initiatives could benefit from understanding the effects and trade-offs involved in selecting different strategies, such as teacher leadership cadres, demonstration sites, and regional professional development centers. What resources actually go to professional development and in what various ways have they been focused? What are some examples of leveraging resources and how might they work in different settings? What are the relative advantages and disadvantages of large-scale, less intense strategies, and those that go deep with fewer people? How can leadership development, assessments, and instructional ma-

terials broaden the reach and impact of professional development?

3. How do professional developers select among different strategies, what combinations seem to work in what situations, and are particular strategies more useful for particular purposes? In our current book, we have identified 16 strategies and suggested that they can serve different purposes (Loucks-Horsley et al., 1997). Are there guides to selecting and combining various professional learning strategies?

4. What outcomes can be expected to result from professional development programs, and how can they best be assessed? This relatively straightforward question is fraught with pitfalls and subject to a multitude of responses. The demand on educators for accountability dictates that professional development must have something to show for itself beyond participant satisfaction. Yet there are many well-regarded arguments for why professional development cannot and should not be examined for its impact on some critical outcomes, e.g., student learning (Hein, 1997). Is this a political question, or can researchers shed some light on the plausibility of drawing relationships between a professional development opportunity and such variables as student learning or teacher behavior change?

5. How can professional development contribute to greater coherence in the educational system? The recent and ongoing releases of data from the Third International Mathematics and Science Study point to the critical importance of coherence in our approaches and support for teaching and learning. With either no helm or too many, teachers are forced to teach too many things superficially, with minimal time for reflection and improvement of their approaches to help students think and learn more deeply. How can professional development help not only teachers, but educators with broader decision-making responsibilities, focus and make critical choices that will ultimately benefit students?

References

Acquarelli, K., & Mumme, J. (March 1996). A renaissance in mathematics education reform. *Phi Delta Kappan, 77,* 478-484.

Ball, D. L. (March 1996). Teacher learning and the mathematics reforms: What we think we know and what we need to learn. *Phi Delta Kappan, 77,* 500-508.

Burns, M. (1995). *Replacement units: A direction for changing math instruction.* Sausalito, CA: Math Solutions Publications.

Bybee, R. W. (Ed.). (1996). *National standards and the science curriculum: Challenges, opportunities, and recommendations.* Dubuque, IA: Kendall/Hunt.

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

Frechling, J.A., Sharp, L., Carey, N., & Vaden-Kieman, N. (1995). *Teacher enhancement programs.. A perspective on the last four decades.* Arlington, VA: National Science Foundation.

Friel, S. N., & Bright, G. W. (Eds.). (1997). *Reflecting on our work: NSF teacher enhancement in K-6 mathematics.* Lanham, MD: University Press of America.

Fullan, M. G. (1991). *The new meaning of educational change.* New York: Teachers College Press.

Fullan, M. G (1993). *Change forces: Probing the depths of educational reform.* New York: Fahner Press.

Fullan, M. G. (1996). Turning systemic reform on its head. *Phi Delta Kappan.*

Hall, G E., & Hord, S. M. (1987). *Changing schools: Facilitating theprocess.* Albany: State University of New York Press.

Hein, G (1997). The logic of program evaluation: What should we evaluate in teacher enhancement projects? In S. N. Friel & G. W. Bright (Eds.), *Reflecting on our work: NSF teacher enhancement in K-6 mathematics.* Lanham, MD: University Press of America.

Honig, B. (June 1994). How can Horace best be helped? *Phi Delta Kappan, 75,* 790-796.

Liebennan, A., & McLaughlin, M. W. (1992). Networks for educational change: Powerful and problematic. *Phi Delta Kappan, 73, 673-677.*

Little, J. W. (1982). Norms of collegiality and experimentation: Workplace conditions of school success. *American Educational Research Journal, 19, 325-40.*

Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educa-tional Evaluation and Policy Analysis, 15,* 129-151.

Loucks-Horsley, S. (1997). Teacher change, staff development, and systemic change: Reflections from the eye of a paradigm shift. In S. N. Friel & G W. Bright (Eds.), *Reflecting on our work: NSF Teacher Enhancement in K-6 mathematics.* Lanham, MD: University Press of America.

Loucks-Horsley, S., Hewson, P. W., Love, N., & Stiles, K. E. (1997). *Designing professional development for teachers of science and mathematics education.* Thousand Oaks, CA: Corwin.

Loucks-Horsley, S., Stiles, K. E., & Hewson, P. W. (1996). Principles of effective professional development for mathematics and science education: A synthesis of standards. *NISE Brief Z(l),* 1-6.

Louis, K. S., & Miles, M. B. (1990). *Improving the urban high school: what works and why?* New York: Teachers College Press.

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

National Research Council. (1996). *National science education standards.* Washington, DC: National Academy Press.

Nelson, B. S. (Ed.). (1995). *Inquiry and the development of teaching: Issues in the transformation of mathematics teaching* (Center for the Development of Teaching Paper Series). Newton, MA: Education Development Center.

Rosenholtz, S. (1989). *Teachers' workplace: The social organization of schools. New* York: Longman.

Schifter, D. (1995). Teachers' changing conceptions of the nature of mathematics: Enactment in the classroom. *In B. S. Nelson (Ed.), Inquiry and the development of teaching: Issues in the transformation of mathematics teaching* (Center for the Development of Teaching Paper Series). Newton, MA: Education Development Center

Senge, P. M. (1990). The *fifth discipline: The art and practice of the learning organization. New* York: Doubleday.

Shanker, A. (1990). Staff development and the restructured school. In B. Joyce (Ed.), *Changing school culture through staff development.* Alexandria, VA: Association for Supervision and Curriculum Development.

Shuhnan, J. H., & Colbert, J. A. (Eds.). (1987). *The mentor teacher casebook.* San Francisco: Far West Laboratory.

Shulman, L. (1987). Knowledge and teaching: Foundations of *the* new reform. *Harvard Educational Review, 57,* 1-22.

Sparks, D. (1994, March 16). A paradigm shift in staff development. *Education Week.*

St. John, M., Century, J. R., Tibbits, F., & Heenan, B. (1994). *Reforming elementary science education urban districts: Reflections on a conference.* Inverness, CA: Inverness Research Associates.

Shirley Malcom
Director, Education and Human Resources Directorate
American Association for the Advancement of Science

I was trained as an ecologist. Jerry Bell always accuses me of wanting to put everything together with everything else, and he's right. I've recently encountered some other people who also want to put everything together with everything else. At a presentation to the AAAS meeting I heard Lee Hood talking about this notion of wanting to understand mind and consciousness. Well, you could look at DNA, and individual neurons, and the connections of axons and dendrites. But at some point the brain itself takes on characteristics that are very different from an individual neuron. For understanding at some level you have to basically let go of the kind of reductionist approach that made us all comfortable.

What I want to do is to put on my ecologist's hat, put the pieces back together, seek interactions, look for points of high leverage within complex systems, worry about key systems components that are being ignored, and search out unintended consequences. I was very lucky when I came into the education arena because I came into it as someone who was trying to focus on the "all," on the equity side. What that did was force me to look at the system, because everything that we tried to do was constantly being thwarted by the system. We ended up creating entities outside the system, because we knew we couldn't affect the system. That experience helped us to understand what we'd have to change if we ever got control. Spending a lot of time thinking about everything that you'd have to fix is very valuable.

So when I reflect on issues of teaching and learning in systemic reform, I need to look from issues of teaching and learning to see how they fit in systemic reform. That is, I'm going to stand on a platform called "professional development," and I'm going to look out at the whole system. And then I will really try to come back in on particular elements.

The first thing I note is that, since teacher professional development is acknowledged as such a critical need, its place within systemic reform is often magnified without clear interconnections to the other key components of the system. For many people, teacher professional development *is* systemic reform. But it is not. It is a necessary, but not sufficient, condition of systemic reform. The connection between teaching and learning is often lost, with the former swapping the focus on the latter. The closer the teaching is to the learning that is expected, the more successful we are likely to be. The closer the teacher's learning is to the form of the student's learning, the more successful we are likely to be. And once we come to that kind of a conclusion and start to look back at the entire system, we have to seek leverage points in other kinds of places.

What other leadership skills and roles in systemic reform besides direct instruction do we need teachers to play-as brokers, as connectors to resources, to tools, to community, to parents? What kinds of things do we need them to learn other than the pedagogy and the content? Where do we think they're going to get these things?

What about the developmental level of children in general? What example is appropriate to a particular age? What kinds of expectations and levels of instruction ought I to

do if I'm trying to get there? What kinds of things are in between?

What kinds of elements of culture, language, interest, etc., do I need to know about a particular child to provide for quality instruction? to understand that people who come to me have a certain history? While curriculum may be something that I do for a school, I have to realize entry points for any particular child, which may be different for one child than for another.

I want to give you a summary of my short list of concerns.

Professional development is necessary but not sufficient to systemic reform. It is often equated to systemic reform because the problem is so large. Prevention is not yet an element of a professional development strategy. It's true. Professional development is forever, but we cannot wait forever. There is a question about needing to affect what we do, as well as how we think about what we do.

There is also an issue of whether there is **some kind of an activation energy effect-that** essentially you won't see a reaction until some dosage has been reached--or whether we ought to expect continuous improvement. After anything has happened, does the opportunity to actually use that show some difference in behavior? I don't know. We do not draw on the models from other professions about professional development. Most programs are incomplete in their focus on content or pedagogy or even both, because they tend not necessarily to reflect on the interaction of these with each other and with other elements of the system. That is, our views tend to be much more like studying the neurons, as opposed to understanding that we have created a system that is different in its entirety and in **toto.**

I looked at the list of things that a teacher would really need to know, and I was boggled by it because it's an incredible list.

- There is the specific content itself, and how it relates. The concepts-how I think about inquiry.

- There are the skills: instructional skills, class management skills, and resource management skills.
- There is the craft of teaching that includes things like diagnosis. If someone does not learn something, how do I then determine what is operating? And how then do I intervene? How do I draw down those strategies that I may have picked up in order to test and really see?
- Questions about attitudes and values. On the knowledge side, we have to expect that we start off with the science and math classes that are provided in the university, but they can't stop there. Somehow professional interaction, journals, meetings, readings, research, etc., continue to keep someone current in the knowledge base. We have to expect that. I haven't been in school in a long time, and if I'm still working on the knowledge I got in school, I'm in trouble.
- Skills: How then do we get them?

The bottom line is that we can't get there without partnerships. The partnerships must be of several kinds, and they must tend to solve multiple problems at the same time.

- Teacher/scientist partnerships to try to help affect higher education at the same time that we are affecting the quality of instruction, by giving teachers a much more authentic sense of what research is.
- Teacher/researcher partnerships by giving teachers a much greater sense of what classroom research is: How do I monitor student learning? And how do I become enabled to figure out whether I am on course and at the right perspective?
- Teacher/parent or other adult in the community partnerships to get me to the point where I can understand the particulars of a child, of the sets of interest that that child happens to bring to the learning environment.

Lauren B. Resnick
Director, *Learning Research and Development Center*
University of Pittsburgh

Systemic reform is about creating teaching and learning environments that can change whole systems. I want to talk about how we might treat our educational systems as learning organizations.

Let me first **define** a "learning organization." This label is hot in business circles, but not so much in our educational circles. I have a very commonsensical definition, in fact, one that's borrowed from some of the leading school districts that I've been working with in the past several years. A learning organization is, quite simply, one in which it is assumed that there will be regular improvement against clear performance criteria for everybody in the system. Not just students but also teachers, professional developers, principals, supervisory personnel, and school superintendents-everyone must continually improve their performance.

That's number one for a learning organization: Regular improvement against clear performance criteria. That means there must be very clear expectations for everyone in the system, and very credible evaluations against those criteria.

Second is the assumption that everyone can improve; everybody can learn. All children can learn, all teachers can learn, school principals can learn, school superintendents can learn. You can go up the ladder, and horizontally as **well**— to include professional developers and others who work in the system. The idea that everyone can, and must, learn is key to the general shift that I have been arguing for-away from our hundred-year-old education system based on assumptions of aptitude and toward one based on assumptions about effort.

Aptitude: we all live with it because the education system we have inherited assumes it. We use IQ tests or their surrogates. We assume that there are gifted and talented students and some other "ordinary" ones who can't learn so well. We assume that there are those with learning disabilities, meaning that somehow they're not learning up to their capacity-presumably some inherited   capacity.

There is an alternative to this aptitude-based way of thinking. It's an alternative that is the underlying assumption in some of our subcultures and also in several **other** countries of the world. The alternative assumption is that directed effort by learners and teachers can actually create talent. I call that an effort-oriented system. A turn to effort instead of aptitude means working on the assumption that everyone in the system actually can get smarter at what they're doing.

Finally, a learning organization is one that provides continuing high quality opportunity to learn. That means providing the best possible **instruction—enough** of it so that everyone can meet the expectations even though some might take longer than others. Consider, for example, what could happen if all of our Title I money were spent on Saturday and summer opportunities instead of pullouts. Think of how much extra learning opportunity some of our children might have before they reach the age of ten or eleven. On a conservative estimate, they could have an extra year of school, without being held back, by just adding an hour a day during the week and half of Saturday.

We must think in terms of learning organizations, hence educational systems. You can, if you like, define your system as a single school-as in the case of charter schools. Even then, you must think organizationally, for nobody is talking about every individual child opposite Mark Hopkins on the end of a log. Just as organizations are made up of all the individual people in them, so people all work within some kind of social structure-a system. I think of a system of education as consisting of its expectations, its forms of assessment, its ways of recognizing accomplishment and its learning opportunities-and, of course, its people.

We can ask four questions about any layer of **the** system. First of all, who is the student? Second, who is the "teacher"'-the provider of learning resources, the organizer of the learning **envi**

ronment as well as, perhaps, a dispenser of specific pedagogy and knowledge? Third, what are the expectations and evaluation instruments? I don't mean tests necessarily, but we do need to know how expectations are conveyed, and how people are assessed against those expectations. Finally, what are the external constraints and enablers?

The answers to these questions are all pretty clear when we think of the classroom as the unit. In that situation the kids are the students, the teacher is the provider and organizer of the learning environment. The teacher probably does some direct instruction, but he or she also organizes all kinds of resources for learning: instructional materials, laboratories, external resource teachers, outside contributors such as companies, community. We've got the students and the teachers clear at the classroom level.

Today we even have a pretty good sense of the expectations for classrooms, in the form of standards, especially if they're expressed as performance standards-describing precisely the kind of work students must do and what a "good enough" piece of work looks like. Expectations become even clearer if they're expressed'in terms of examinations or other assessments that clearly embody the standards so that you can teach to them. So far, the picture for classrooms looks pretty clear-cut. We know who is involved and what they need to do for maximum learning.

But complexities arise when we think about all the external constraints and enablers that impinge upon classrooms. Lo and behold, they turn out to mostly feel like constraints. There's the principal in the school building. Is he or she often viewed as an enabler? There are the union reps in the school building. Are they often viewed as enablers? What about the other teachers-apart from the few that went with you to the special summer academy? What about the school climate as a whole? What about the parent community or the local school board? We tend to view all of those as constraints on the good work we want to do.

Let's see if we can't change that picture. Let's think now of the school as the organizational unit. There is lots of evidence that working only with individual teachers, including lead

teachers, doesn't stick. That is, it is only when whole schools are involved in a reform process that real change, in terms of improved student achievement over time, begins to show through.

Take my four questions: Who are the students if we take the school as the learning organization? They're the teachers; the teachers are the students. Who, then, is the teacher? If we take my definition of a few moments ago of teacher as the person who organizes a whole environment for learning, it's the school principal. The school principal can set up or hinder a learning organization in a whole school-using training programs, mentoring systems, professional developers, local universities, and other resources. But compare the paucity of materials and help available to the school principal in playing that role with the richness of materials for kids.

What about expectations and evaluations for school learning communities? Standards are now being put into place and becoming more and more clear for kids. But for teachers there's basically nothing between basic licensing criteria and the National Board for Professional Teaching Standards. That doesn't add up to a powerful set of tools for a learning organization. We're missing a whole layer of tools for communicating expectations about teacher performance and assessing that performance in ways that support continued learning and improvement.

What about the external constraints and enablers on school organizations? Again, we find mostly constraints. There are the same ones I just mentioned for classrooms: district and state policies, unions, etc. But the biggest constraint is that there are too many things other than creating a learning organization that the school principal is responsible for. If you interview most school principals, follow them around, you find that the learning and teaching performance of their staff is not their primary daily concern. They do not spend their days actively thinking about teacher learning. That is not what they were trained for; it is not what they were selected for; and it is not what they (mostly) do.

Now push up one more level to the school district. Since we've just seen that the principal is critical, where is the principal going to learn? What is the learning organization for the school

principal? It ought to be district leadership or some kind of regional leadership opportunities. Very few of these exist, and almost none of them focus on teaching and learning, especially in math and science. What are the expectations and evaluation tools for school principals as instructional leaders, that is, as teachers? Virtually absent.

Why is this so? What are the constraints on the districts becoming such leaders? Very interesting. There are surprisingly few constraints,

and these are mostly in the heads of district leadership. In virtually every state, almost every bureaucratic rule can be waived. Yet there are hardly any requests for waivers coming into most states. The same is true for federal regulations: hardly any requests. In other words, we aren't taking advantage of the opportunity to create teaching and learning organizations all the way down, and all the way up.

That's something we should work on. It is something we have the tools for now. Let's begin to use them.

The Role of Evaluation in Systemic Reform

Thomas B. Corcoran
Consortium for Policy Research in Education
University of Pennsylvania

What Is Systemic Reform?

Systemic reform as defined by scholars, policymakers, and funding agencies is an approach to school reform that views the policies and actions of government-school districts, states, the federal government-as the critical levers for improving the performance of public education. The central argument is that, if governments set standards for student performance and adopt aligned policies for curriculum, assessment, accountability, and governance, educators will alter their practices, and performance will improve. This reasoning distinguishes systemic reform from other reform strategies that rely on the creation of markets, the professionalization of teaching, or the work of volunteer networks of schools or educators as the critical forces for reform.

Proponents of systemic reform believe the performance problems of public education are primarily the result of low standards, incoherent and fragmented policies, and poor use of resources. Hence, they advocate higher standards and more robust and more coherent policies that send clearer and more compelling signals to educators, students, and parents about what is expected. The core policies addressed by systemic reform are often referred to as drivers or some-

times as the instructional guidance system. Typically they include content and performance standards, aligned assessments and curricular frameworks, an accountability system with rewards and sanctions, changes in teacher development that enable teachers to prepare students to meet the new standards, and decentralization. In some versions of this approach, schools performing well will be rewarded, and those that fail to make progress will be subject to a variety of sanctions. Advocates also propose reducing the regulations, bureaucracy, and policy incoherence that might impede reform and decentralizing decisions about how to best meet the standard (Cohen, 1995). In short, systemic reform seeks to make the system more rational, more coherent, more focused, more **efficient.** The approach is also sometimes called standards-based reform.

Systemic reform has obvious appeal. Its focus, comprehensiveness, and specificity promise a more powerful strategy than the **school-by-** school approaches of the networks and a fairer, more orderly one than the markets envisioned by the choice advocates. Its logic is so compelling that its adherents often act as though it were a proven formula that could guarantee success. The only question of interest to the most committed

enthusiasts often seems to be how to persuade reluctant state and local policymakers or educators to comply with the postulates of their theory.

In fact, systemic reform is a persuasive, but unproven, general theory of action. It is a general theory because it provides only a broad and rather abstract framework for policy design. It must be adapted to tit particular political contexts. As a consequence, the specifics of enacted theories of systemic reform vary from one setting to another. For example, consider the variations in the standards set by states and by localities. Or the variations in the content and form of the assessments used. These variations raise questions of interest to an evaluator. What are their consequences for practice and performance? To what degree can systemic reform be customized to a particular political setting and still be systemic?

I refer to the theory of systemic reform as unproven because as yet we have little empirical evidence to support the efficacy claims made by some of its advocates. We can draw some limited inferences from effects of the basic skills movement in the 1970s and 1980s and from the impact of similar approaches in other countries. And the early data from Kentucky are promising. But the empirical support for systemic reform remains thin. Some advocates seem to confuse evidence documenting the problems they feel are central (e.g., lack of standards, fragmentation, etc.) with evidence of effectiveness. They are not the same thing. And little is known about the importance and relative efficacy of the various components of systemic reform. Are they equally essential? Is high stakes assessment needed to create incentives for improved performance? Are incentives needed for students? Is devolution of authority necessary to achieve the desired results? These are also questions of interest to evaluators.

The effects of systemic reform under varying conditions also need examination. Is fiscal equity a prerequisite condition for success? What are the educational and political effects of implementing standards and high stakes in situations with inequitable resource distributions?

What are the consequences for schools with varying capacity to design and implement changes? What degree of tit is needed between teacher knowledge and skill and the standards?

I could go on. There are many unanswered questions.

I was asked to reflect on the role of evaluation in systemic reform. Taking advantage of the broad scope of this charge, my comments will focus on the assumptions underlying systemic reform and on their implications for evaluation, on some other critical problems faced by those attempting to evaluate systemic reform, and on some of the questions (in addition to those I have already raised) that such evaluation efforts ought to be addressing.

The Premises or Assumptions of Systemic Reform

The theory of systemic reform rests on some assumptions that should be carefully examined and tested. First, systemic reform seeks greater coherence, an alignment of policies, but the education system itself is fragmented by design— fifty states, fifteen thousand districts, countless other agencies impacting the schools-and this fragmentation is intended to permit variation. The agencies of government responsible for the schools are divided from each other by the federal structure and by the separation of powers. They are further divided by powerful traditions of local control and parental rights. On top of that, within any given jurisdiction there are a variety of stakeholders, each with their own views about standards, assessment, locus of authority, etc. What does coherence mean in this environment?

Systemic reform assumes that some consensus can be developed around standards and that the resulting set of coherent policies will focus this fragmented system and improve its performance. From an evaluator's perspective these assumptions must be viewed as problematic. Previous research indicates that context is a dominant influence on the implementation and effectiveness of reforms. And the contexts in the states and school districts vary widely; state politics vary, their histories vary, their resources vary, and so on. Yet some contend that the key components of systemic reform must be essentially the same in every setting, including systemwide standards and assessments, and that these compo-

nents will operate with similar effects under varying conditions. However, states with strong traditions of local control might find it difficult to adopt the recommended policies. Is it not possible to design a systemic approach that respects these traditions of local control? Could a systemic reform strategy that lacked one or more of these critical drivers be effective?

Second, advocates of systemic reform aim to change teaching. They speak of more coherent policies driving instruction. They assume that there is a strong linkage between policy and practice. However, research has found little evidence of such a linkage (Cohen & Spillane, 1992; Elmore, 1995). Moreover, while some advocates of systemic reform argue that they simply want to define the outcomes and that the means of achieving them might vary (hence the emphasis on devolution), others package specific notions of best practice together with their instructional guidance system so that the reform defines both the means and the ends. It is hardly surprising that this approach produces some resistance in a complex decentralized political system.

David Cohen (1994) has suggested that evaluators should be comparing the content and coherence of instructional guidance systems with what teachers, students, and parents comprehend to be its message. His point is that, given the fragmented nature of the system along with the many competing messages about priorities coming from government, professional groups, reformers, and local communities, it would not be surprising if the messages were distorted. Certainly school districts and schools respond differentially to the same policy messages (Spillane, 1996). Evaluators should be examining the efficacy of the links between policy and practice.

Third, there is the question of the variation in schools' capacity to respond (Slavin, 1995). Systemic reformers aspire to raise the performance of all students and close gaps in achievement between advantaged and disadvantaged groups. Yet schools' capacity to respond to the policy signals or use the support that is provided varies. How does systemic reform affect schools at different levels of readiness? How do we engage teachers who resist changing their practice or participat-

ing in professional development? How effective are the intervention techniques being employed?

Fourth, the theory focuses on the system rather than students and seems to treat students as passive vessels with little or no responsibility for their own success. The system is seen as the problem, and if the standards and drivers are put in place, advocates contend the resulting concentration of energy will produce a more powerful pedagogy, higher student motivation, and improved performance. Shouldn't the response of students to these changes in policy and practice be viewed as problematic? What do we know about the link between educational policy and student behavior? How do different incentives affect the behavior of different groups of students? These are questions of importance to an evaluator of systemic reform.

Fifth, there is the problem of the standards. They are sometimes treated as a purely political question in systemic reform: What can we all agree to about what students should know and be able to do? The major concern seems to be that they are high enough. But, as TIMSS has dramatically illustrated, the developmental sequence of the content standards and their breadth and depth matters a great deal and shapes the outcomes. So there are empirical and technical questions to ask about the standards. Is more specificity better? Are interdisciplinary approaches effective? Under what conditions? How many standards can be crowded into the curriculum? Who should set them? Does the locus of power matter?

How Should We Evaluate Systemic Reform?

To design evaluations of systemic reform initiatives, we must know something about the success criteria held by various stakeholders. Are the success criteria to be successful enactment of the policy instruments (e.g., the instructional guidance system), changes in classroom practice, changes in student performance, or all of the above? And with regard to performance, are we looking for improvement, getting all students to , high standards, reducing gaps among significant groups, or all of the above? The answers to these questions seem to vary across audiences.

When federal policymakers talk about evaluating systemic reform, they often seem to be asking whether the government in question-state or district-has aligned its policies, put the drivers in place, provided the necessary supports, and is realizing the desired changes in practice and performance. The underlying assumption here is that systemic reform is a proven strategy and that we know how to do it, and therefore the only important question is, Are they doing it right? If the policy configuration is not as desired or the results are not as expected, the conclusion often is that they simply didn't get it. Such evaluations seldom ask questions about the robustness or costs of the strategy itself, whether some iterations of the strategy might be more effective than others, or how well the strategy works in different political and institutional contexts.

In my judgment this checklist approach to evaluation is a mistake. We should approach variations in strategies with an open mind and examine their design, viability, costs, and impacts. Rather than defining systemic reform rigidly, we should respect the decentralized nature of our system and encourage and evaluate strategies that share a set of core ideas: challenging academic standards, coherence and focus, and systemwide change. We might learn that different policy mixes or sequences of action or distributions of authority work better than others under specific   conditions.

Other Problems of Evaluation

Getting the data. There are serious data problems that compound the problems of evaluating systemic reform. Data on classroom practice are typically collected through surveys, because observations are too labor intensive to collect an adequate sample over space and time. Since the point of systemic reform is to send clearer signals, teachers soon learn what the desired practices are, or at least what the code words are, and in high stakes environments may feel pressured to indicate compliance. Surveys may exaggerate the extent of the changes in practice.

There are worse problems with student performance data. Many states and districts do not assess all areas in which we seek to set standards or do not do so adequately. New measures are under development but are not yet trouble free. Performance measures are expensive and plagued with unresolved technical problems, as are portfolios. The results of more conventional paper-and-pencil measures are not trusted by reformers who argue that such measures are not sensitive to the important outcomes, and furthermore, they undermine efforts to alter practice. So what is an evaluator to do? We may be in the midst of a great transition in assessment, but evaluators need valid, reliable, and stable measures to examine changes over time.

Timelines. What is a reasonable time frame in which to expect results? Policymakers and grantmakers have trouble waiting two or three years, but we know it takes more time than that to put the policies and supports in place and achieve broad scale changes in practice. Five years seems far too short to expect the broad scale changes reformers believe are needed to produce significant gains in achievement. Kentucky s policymakers have said it will take twenty years. Connecticut has a plan for a generation. Philadelphia's leaders talk about a decade. What is a reasonable time frame? How do capacity and previous experience affect time frames?

Attribution. Policymakers and funding agencies want to know that their strategy paid off. They want to take credit for results. How do we deal with attribution? When we get improved results, how can we be sure it was the standards or the best practices that were responsible? Maybe it was just the tests. Maybe it was the extra resources. Maybe it was simply a general cultural shift. How will we know what caused the improvements? Or when we do not get results, has systemic reform failed, or were there inadequate resources, poorly prepared teachers, too many external problems for schools to overcome? What interventions mattered? How will we know? We need well-designed, longitudinal, cross-jurisdictional studies to answer these questions.

What Do We Need to Know about Systemic Reform?

There is a great deal that we do not know, and unfortunately we are not getting answers from current evaluation efforts. Consider for a moment the following questions:

Have the new instructional guidance systems created greater coherence for teachers, students, and parents and reduced the incoherence of a politically fragmented system?

How do policy-driven reforms, particularly those including centralized assessments and high-stakes accountability affect professional norms and practice? To what degrees are they complementary to, or in conflict with, other efforts to set standards of good practice?

What kinds of supports and technical assistance are most effective under different conditions? Do schools that adopt highly developed whole-school designs such as those offered by New American Schools make more rapid progress? Do schools that adopt externally developed curricula make more rapid progress? What do the answers imply for the level of investment needed and the distribution of those resources?

What are the costs? Are some approaches more cost-effective than others? How do the political traditions and values, specific political context, and governance arrangements influence the approach to systemic reform? How do variations in strategy or policy mix affect the costs and the results?

What level of investment does it take, and for how long, to provide teachers of particular backgrounds with the knowledge and skills needed to prepare students to achieve the more challenging standards?

What kinds of institutional arrangements and learning opportunities seem to be most effective and efficient at meeting the needs of teachers?

What are the student benefits? Does it lead to improved student performance? Are the improvements greater than those associated with other strategies? Who benefits? Does it reduce gaps in performance among significant groups of students?

These are just some of the questions that we should be seeking answers to in order to advance systemic reform from the status of a promising general theory to a empirically grounded strategy for school reform.

References

Cohen, D. *(1994). Evaluating systemic reform.* A paper prepared for the Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Education.

Cohen, D. (1995, December). What is the system in systemic reform? *Educational Researcher, 24,* 11-17.

Cohen, D., & Spillane, J. (1992). Policy and practice: The relations between governance and instruction. *Review of Research in Education, 18, 3-5* 1.

Elmore, R. (1995, December). Structural reform in educational practice. *Educational Researcher 24, 23- 26.*

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

Slavin, R. E. (1995). *Sands, bricks, and seeds: School change strategies and readiness for reform.* Baltimore, MD: Johns Hopkins University, Center for Research on the Education of Students Placed at Risk.

Spillane, J. (1996). School districts matter: Local educational authorities and state instructional policy. *Educational Policy, 10, 63-87.*

Paul G *LeMahieu*
Education Research and Development Center
University of Delaware

Chief Justice Holmes once said, "If you can disentangle that which is inextricably intertwined, then you have the mind of an attorney." Yesterday I was made mindful of some language problems that we have, which I'm going to try to disentangle. I've been asked to think about evaluation and the challenge that systemic reform poses to evaluation. There's so much that could be said. The colloquial expression of the topic might just simply be, "How do we find out what's happening? How do we find out what, if anything, we're accomplishing with all that we're doing?"

I want to focus on that word **systemic** for a moment. I have a colleague who's an engineer. I brought him to our operation from the Office of Naval Research, where he was a director of projects for naval underwater warfare systems. He's always asking questions that bring me up short, making me stop and think about some things. The question that he asked me about a week and a half ago was, "Paul, I have a Master's degree in systems research. I have been the director of naval underwater warfare systems. What are you people talking about when you refer to systemic reform?'

Over the last year and a half I have encountered the word "standards" or "standards-based" as meaning systemic reform. I have encountered "constructivist," or "whole language" as meaning systemic reform. I have encountered "performance-oriented systems" suggesting systemic reform. For some the system in systemic reform is an adjective of scale suggesting all of the teachers, or all of the classrooms, all of the schools, in all of the districts. For others, systemic is an adjective of conceptual scope indicating a concern for all of the people and the various constituencies that make up the system: legislation, regulation, policy informing programs and practice, and the multiple roles and people that inhabit the system. For some, the systemic notion is a hybrid of those two, suggesting a theory of action that motivates the way that we proceed. In many places the folks who use the language of systemic reform also use the language of standards-based reform. The two are not necessarily intellectually synonymous. I want to focus on the last definition, the hybrid of systemic reform suggested above, and the challenge that it poses to efforts to evaluate the consequences of our policies, programs, and practices. All elements of the system are interdependent and therefore all have to change in a systemic manner. The fact of the matter is there are unique circumstances of systemic reform that pose challenges to our thinking as evaluators. The context and the approaches of systemic reform would challenge us even within a traditional framework for evaluation. And then, everything has to be rethought yet again because the traditional goals are insufficient.

The first challenge for evaluators is **definitional**. Evaluators are called on to define the "it." Deborah Ball speaks with some amusement of the way we talk to each other about "this kind of teaching," never stopping to clearly define "it" in the midst of our conspiracy of silent assent about "it." The "it" is "this kind of teaching" that we all seek. This position is unnecessarily extreme. As evaluators we are called on to define this "it." Not only do we **define** the "it," but we implicitly declare how much of "it" we're looking for. What is the right balance of "it"? When should "it" occur? When we conduct evaluations, and when we undertake to declare places successful or not depending upon how much of "it" is achieved, we are as much at risk as evaluators of being pressed to an unreasonably extreme assertion as we are at risk as reformers of being pressed into extreme expectations. Looking for nothing but "it" is-even accepting some definition of "it"'-a very big mistake. We have to be mindful of that. Some of the elements and conditions of "systemic" as I described it at the outset, reform broadly conceived and systemic in scope, are significant because they **qualify** the nature and extent of the evaluation challenge. First of all, there are many different instantiations of systemic reform. Each is unique from the others

even in terms of its policies and goals, although there is a tremendous amount of overlap and shared understandings of some larger vision. In point of fact, the explicit goals, or at least what is in the foreground as opposed to the background, differ from place to place. What is in the foreground even differs over time. Not only that, but the policies and the ways in which they are manifested in programs differ. The ways in which programs seek to shape and support practice also differ from place to place. There's a long history of research and evaluation into compensatory education programs. It documents a very long and very disappointing evaluation history of inability to detect effects even where people are convinced effects do exist. One of the lessons of that long history has to do with what happens when an evaluation imposed ex-machina is insensitive to these issues of local context, programmatic variations, and local needs.

A second element has to do with the sheer complexity of systemic reform. Think about what it means in terms of multiplicity of goals. It's dynamic. Someone referred to systemic reformers as dynamic opportunists, right? We're always looking for the moment at which some leverage point makes itself available to us. A good reformer seizes it, acts on it. It is extremely difficult for an evaluation-particularly one that is traditionally conceived---to be responsive to that kind of dynamic complexity.

Lastly, many of the intended outcomes are difficult, even fragile, in the sense that they are bruised if not destroyed by our methods for examining and measuring them. Hence, there is widespread disaffection with traditional measures as "missing the point," if not outright destroying it, and emphasis on the reform of our measures.

The response to this extremely difficult challenge of systemic reform is to reconceive not our methods, but our approaches to evaluation Methods are the particulars, the means of gathering data, the structures within which data are gathered, organized, analyzed.

I think that the appropriate response lies in a larger structural issue, what some refer to as models of evaluation. Instead of thinking about

evaluations of systemic reform as monoliths, think of them as dynamic programs of comprehensively conceived, coherently coordinated studies, each of which is locally sensitive to context needs; each is methodologically appropriate to its questions, its research context, and evaluation goals; all studies are coordinated and complementary of each other so that our knowledge is reinforced not for having asked a single question in the same way over and over and over again in standardized fashion, but for having asked the same questions in different places in appropriately different ways. There is a methodological perspective that insists that standardization is the epistemological strength of desired knowledge claims. Another way of looking at it is to suggest that complementary methods yielding similar observations across varied instantiations of an innovation ought to deliver a stronger epistemological foundation. There is a sense in which this happens currently, but it is by happenstance. It happens because good thinkers are thinking about this problem of systemic reform, and they are looking for which questions aren't being asked, and what opportunities aren't being fully taken advantage of. That's not the same thing as approaching these evaluations in a coordinated, coherent, dynamic, and interdependent manner.

If evaluation is approached in this fashion, it is positioned to better serve the goals of a learning organization; the dynamism that's present in the evaluation system mirrors the dynamism of reform. It enables us to create the information infrastructures that are essential. I was struck as I listened to Lauren Resnick earlier raise the image of learning organization and give her definition. The one thing missing in it is a capacity for institutional reflection, an institutional ability to look at itself, learn from and about itself, and act on that learning for the ongoing improvement of its performance. The approach to evaluation only hinted at here offers the potential to position evaluation to deliver that capacity in ways that our traditional views of it do not. It is a view of evaluation that is essential to our reform because it's consistent with what we hope for from the reform.

Norman L. Webb
National Institute for Science Education

Systemic evaluation is an evolving concept. Over the past six years, gathering information for judging the merit and worth of systemic initiatives has matured as a field. We have a greater understanding about what the nature of systemic change is and what roles evaluation can serve. For the past two years the Strategies for Evaluating Systemic Reform (SESR) Team of the National Institute for Science Education has endeavored to learn about evaluation of large-scale reform from evaluations done of the Statewide Systemic Initiatives (SSIs).

In thinking about evaluation we have depended heavily on the work of those who are doing evaluations. We reviewed the plans for the evaluation of SSIs of 15 states and Puerto Rico. We consulted the reports from SRI International, Policy Studies Associates, Consortium for Policy Research in Education, and ABT Associates, Inc. Other sources of information include organizations engaged in doing, studying, or providing technical assistance on evaluation of systemic initiatives including Horizon Research, Inc., Inverness Research Associates, Westat, Inc., and the McKenzie Group.

What Are We Learning?

Critical functions of systemic initiatives help define roles for evaluation that are somewhat different from the more traditional roles of purely formative and summative evaluations.

Design. The design process of a Systemic Initiative is fluid and continuous. For many of the Initiatives, what was set out at the beginning had to be changed, modified, and thought about again-not only once, but several times. An important role served by many of the SSI evaluators was as a critical friend to judge and help shape the design of the Initiative's strategy. This molding of the Initiative over time and the ongoing strategic thinking elevated the design from a planning document to an outcome of a learning process. The designs have become as much of a product of the Initiatives as they have served to guide the work of the Initiatives.

Management. Closely related to developing a design for the Initiative is managing and making decisions for the Initiative. Evaluators of the SSIs recognized the importance for the Initiatives' staff and others to have access to information about context, implementation, and outcomes for decision making. The number of those making decisions and the range of decisions are large in SIs. Management teams faced issues of congregating a critical mass of stakeholders, developing effective networks of teachers, and planning for long-term sustainability while under pressure to demonstrate improved student learning. With limited resources, evaluators continuously faced the practical and conceptual issue of either attending to producing information on process and capacity building or attending to documenting student and other outcomes. Some evaluators coped with this issue by exerting efforts to identify interim impacts.

Leverage. Reform on the scale of the SIs engages many people at levels of detail that can obscure the larger picture of reform. Mounting an effort to saturate large, multifaceted education systems requires leveraging resources, aligning components, and strategically thinking about "going to scale." Evaluation of an SI requires some attention to where the Initiative is heading along with what the Initiative is doing. One role evaluators assumed was studying continuous improvement and how this improvement would project to the full system. SSIs' evaluators assumed a proactive role to leverage the systems in concrete ways. Some highlighted best practices that were used to promote early successes statewide. Others guided Initiatives to position themselves to take advantage of unanticipated opportunities.

Verification. Systemic reform is based on a theory that assumes the highest level of education will be achieved if all of the pieces and components within a system are aligned and working in cooperation toward important common goals. One role of systemic evaluation is to verify this theory as applied to the system. This formidable task requires understanding the logic of the sys-

tern, what the components are, how they are linked, and what their collective force is. Evaluators of SSIs attended to one piece of this task by developing and seeking answers for "linking" questions. They sought to establish paths from the intervention to local changes in practices and policy.

Student learning. SSI evaluations varied in the attention they devoted to measuring student outcomes, due, in part, to the lack of instrumentation and availability of the information, the Initiative's focus, and the evaluator's perspective. Some evaluators believed that developing a system's capacity for sustained improvement is critical to launching significant change, a development that requires time. Any attempt to measure student outcomes and attribute them to the Initiative in the first stages of its development was felt to be conceptually wrong and detrimental to the Initiative's longevity. If positive outcomes were not produced, the Initiative may be too fragile to withstand strong criticism, even though significant groundwork had been set for important future changes in student outcomes. Other evaluators did not hold this view and incorporated some assessment information in their evaluation. Some compared scores of students whose teachers had participated in the Initiative's training with scores of a control group. Some used trend data from state assessments over a number of years. One Initiative used student assessment as one of its major parts. A three-pronged approach was designed-measuring long-term gains in student outcomes, developing and using pre- and posttests in classrooms, and training teachers to use alternative forms of assessment. The systemwide assessments provided incentives for teachers to learn more about assessment and encouraged them to attend to student outcomes using the developed classroom assessment instruments.

What Do We Need to Learn?

Much is still to be learned about systemic evaluation and measuring change in large education systems. A few areas are listed here.

Equity. Assessing equity in student learning throughout the system is a critical concern and raises important questions for systemic evaluations. More attention needs to be given to developing ways for measuring a system's progress in achieving equity in student learning. Related questions include:

How can progress toward equity in a system be measured in concert with validating a high quality of mathematics and science learning throughout the system?

How can an evaluation effectively identify how well efforts of a Systemic Initiative support helping every student reach his or her full potential?

What levels of disaggregation of data are necessary to study equity within a system?

Achievement measures. Assessment technology is insufficient to measure all important knowledge of science and mathematics. Valid techniques to apply on a large scale still need to be developed to measure how students are able to reason, to solve complex problems, to build arguments, and to do scientific inquiry. "Habits of mind," metacognition, and dispositions are important qualities for pursuing science and mathematics, but are very difficult to measure. Some questions are:

What methods can be used to judge that a system is educating students to do challenging mathematics and science?

How can evidence be produced of how students are developing habits of mind and their competence to think deeply about mathematics and science?

System saturation. Systemic evaluation, like systemic reform, must keep the entire education system within its view at all times. The most common approach to evaluating SIs is to study the components and parts of the system. Less is known about how to consider the interaction among components, their linkages, and what the magnifying effects are. More must be learned about tracking change attributable to the SI through a system. Some related questions are:

How can an evaluation strike a balance between gathering the immense amount of dam that would be useful to know and the amount of information that can be effectively collected and analyzed at a reasonable cost?

How can an evaluation grasp the full extent of a system to make judgements about the effectiveness of an SI to institute change?

How can evaluation produce information on the amount of resources, concentration of efforts, and strategy for change necessary to reach systemic reform for any given system?

How can the quality of classroom experiences and teachers' interactions with students be incorporated into a systemic evaluation effectively and without great expense?

Time frame. It is unclear how much time must be allowed before various changes in an education system should become observable and sustainable. In judging the value of systemic re-form, systemic evaluation must attend to the institutionalization of structures and functions that will sustain movement toward positive outcomes. Evaluation is frequently called on to produce information and judgements before sufficient time has transpired and enough effort has been expended to fully reach goals. More needs to be understood about how to identify and measure interim attainments and progress. We also need to understand more about what is a reasonable amount of time for a system to make significant changes, but time, of course, will depend on a number of factors. Some related questions are:

What is the appropriate time **frame** for detecting change in student learning across the system that can be attributed to an SI and can be judged as sustainable?

What interim outcomes serve as strong indicators of progress toward systemic change?

Iris R. Weiss
President, Horizon Research, Inc.

In 1990, the National Science Foundation requested proposals from states to reform their science and mathematics education systems. While the solicitations defined the overall goals of systemic reform, the choice of reform strategies was left to the proposers. The openness of the solicitation was partly in recognition of the fact that states differ considerably in their needs and resources and, therefore, would require different approaches to reform. In large part, however, the lack of prescription reflected the reality that the theory underlying systemic reform was exceedingly thin, indicating for example, the importance of shared goals and policy alignment, but providing little guidance on how to achieve either.

If the theory of systemic reform and the NSF solicitation for translating the theory into practice were nonprescriptive about implementation, they were nearly silent on how one might evaluate those efforts. Proposers were told it was important to have a plan for formative and summative evaluation and to have that plan carried out by qualified personnel, but there was little guidance about the kinds of evidence to collect.

And so we embarked together, the states and the evaluators, they on figuring out what those exceedingly broad goals like policy alignment meant, and we on how to tell whether they were accomplishing them; they on how to use the modest monies provided by NSF to impact the system both broadly and deeply, and we on how to use the small amount of money earmarked for evaluation to help document and improve their efforts.

The Statewide Systemic Initiatives were followed in fairly rapid succession by the Urban Systemic Initiatives, the Rural Systemic Initiatives, and the Local Systemic Change Initiatives. My company and I have been involved in one way or another with the evaluation of each of them. Most of what I have learned about the evaluation of systemic reform in the last five or six years seems painfully obvious now, although

it certainly wasn't at the time. It is clear that evaluation can be helpful throughout the systemic reform process in the initial planning, helping to fine-tune the initiative as it unfolds, and documenting evidence of impact. But there are potential problems and numerous trade-offs throughout.

1. Assessing Needs and Documenting Progress

Most systemic reform plans start with a description of the current status of the system. Whether called a needs assessment or collection of baseline data, evaluation early on can help document the status of the system prior to the reform initiative. The challenge in systemic reform is that you cannot stop with documenting teacher knowledge or the alignment of the curriculum with national standards or classroom practice or any other component of the system. Rather, you need to understand and describe both the various parts of the system and their interrelationships. Unfortunately, this task is not nearly as straightforward as it sounds, since we aren't always sure what aspects of the system matter or how to measure many of them. As a result, evaluators are essentially making it up as we go along, both in assessing initial status and in documenting progress. We know we need to look at multiple components of the system, but which ones, at what times, and in what depth are not at all clear.

2. Critiquing the Program Design

Perhaps because the goals of systemic reform are so broad, there are many more possible pathways for reform than are typically available to program developers who have more limited objectives. As a result, evaluators of systemic reform efforts are often asked to use evaluation methodologies to help critique and refine the strategy of the initiative, looking at the probable advantages and disadvantages of alternative courses of action. This process can be very helpful to an initiative in discovering mismatches between the goals and the proposed activities while there is still time to correct them. However, while the role of evaluator as critical friend seems appropriate enough, it is sometimes easy for the

evaluator to slip into design consultation and technical assistance, which may or may not be appropriate. In the extreme, this process may lead to the evaluators evaluating the quality and impact of their own ideas, which is clearly not what the funders had in mind.

3. Challenge of Meeting Diverse Information Needs

The most difficult part of the evaluation of systemic reform is that different stakeholders have different information needs and expectations for the evaluation. The PIs may be primarily interested in assessing the quality of their pilot programs so they can improve them before scaling up. Others may want to go very quickly to looking for evidence of impact, which some may define as changes in classroom practice and others as changes in student performance.

Some stakeholders want the numbers surveys provide; others insist on going beyond self-report dam, even if it means observing only a small number of teachers. Most want both, and lots of them. Similarly, those who want to go directly to student measures differ markedly in what counts, with some looking for changes in whatever standardized tests or proficiency measures the state or district uses to assess students and others insisting on authentic measures.

Whatever the measures you choose, do you assess progress for all teachers/students in the system or only those who have been reached directly by the initiative? The former is more in keeping with the goals of systemic reform, but it seems like a waste of resources to look for impacts that are unlikely to have occurred. It is easy to get stuck jumping back and forth between the various options, and, while getting prior agreement on an evaluation plan will in theory resolve these problems, in practice the pressure to collect more and more data to address emerging information needs can be intense, even irresistible.

4. The Fallacy of It's Out There

As PIs modify their initiatives based on experience and in response to emerging opportunities, evaluators need to keep pace with new

evaluation strategies and instruments. Unfortunately, there is a lack of tools for the evaluation of systemic reform. While you rarely have the budget to do extensive instrument development, the evaluation instruments everyone says are out there are awfully hard to find, whether you are looking for classroom observation instruments or performance tasks that are aligned with reform goals, not to mention ways to assess the extent of policy alignment.

5. Deciding Who Gets What Information When

It does not take long before you are awash in data. Finding the time to analyze and report it is difficult, especially because the initiative is on-going and you and the PIs are reluctant to stop collecting data. Tensions abound. PIs want and need honest feedback in order to improve their programs, but at the same time want you to put the best spin possible on anything that appears in writing. We typically give the PI a draft report to check for inaccuracies and misinterpretations, but sometimes their feedback seems aimed less at verification and more at turning your analytic work into a public relations document. On more than one occasion I've had to remind a project that we were preparing a report, not negotiating a treaty. And on more than one occasion a PI has reminded me that they are engaged in an enterprise that is as much political as it is technical, and that critics can use even the smallest negative finding to undermine a promising initiative.

I can end with platitudes. Less is more. Don't collect more data than you can afford to analyze. In deciding on spin, remember you have to look at yourself in the mirror in the morning. Or I can end with the greater truth that the challenges and complexities involved in evaluating systemic reform have helped me and others deepen our understanding of how to bring about improved science and mathematics education on a broad scale. No one said it would be easy.

Andrew A. Zucker and Patrick M. Shields*
SRI International Co-Directors
National Evaluation of the SSI Program

SRI International and its partners' are in the final year of a five-year evaluation of the National Science Foundation's Statewide Systemic Initiatives (SSI) Program. In preparation for writing a report on state strategies for systemic reform, the evaluation team has spent time examining each SSI's approaches to systemic reform.

Preliminary findings about the effectiveness of different strategies are reported here.

Identifying SSI Strategies

In an earlier report[2] we identified eight strategies that the SSIs have used to carry out reform. In this paper, the strategies are grouped under two headings: strategies that focus principally on teachers, classrooms, and schools and strategies that focus on districts, regions, and states. (See Exhibit 1.) This is a useful heuristic, even though the strategies often share some elements of both focuses. (We note that categorizing the strategies under these two headings is similar to what others have done in referring to "bottom up" and "top down" approaches to reform.)

Exhibit 1
Eight Strategies Used by the SSIs

Strategies focused on teachers, classrooms, and schools

    Supporting teacher professional development
Developing, disseminating, or adopting instructional materials
Supporting model schools

Strategies focused on districts, regions, and states

    Aligning state policy
Creating an **infrastructure** for capacity building
Funding local systemic initiatives
Reforming higher education and the preparation of teachers

    Mobilizing public and professional opinion

Typically, an SSI will rely on one or more *primary* strategies and several *secondary* strategies as its participants undertake systemic reform in mathematics and science education. The evaluation team used a variety of documents about each SSI, as well as knowledge gathered during site visits, to make decisions about what the primary and secondary strategies are in each SSI, and these are shown in Exhibit 2. In some cases, SSI strategies have evolved or may even have changed dramatically, making it more **difficult** to assign strategies with certainty. Still, it seems likely that the SSI Principal Investigators would agree with the great majority of the decisions (nearly 100 of them) that have been made by the evaluation team.

It is important to note that Exhibit 2 *only* reflects strategies that are directly part of the SSI. In other words, they are part of the "value added" by the systemic initiative. Thus, for example, although California clearly invested heavily in policy alignment, that largely occurred prior to the SSI and **was** not supported by SSI funds, and so Aligning State Policy is not shown as an SSI strategy in **California.** (Case studies of 12 **SSIs**

that are being distributed as part of the evaluation will reflect the entire context in which each SSI operates, including other key strategies for education reform being used in those states apart from the SSI.)

Quality Ratings

In order to make judgments about the effectiveness of different reform strategies, the team identified a set of six common criteria to use in considering the implementation of each strategy. Each criterion was defined in detail for each of the eight strategies (see Appendix D). The six common criteria are:

1   *Quality:* how well the strategy has been executed (e.g., the quality of instructional materials);

2   *Access and **inclusion:*** the degree to which the SSI-sponsored efforts have been open to the full state community (e.g., access of rural teachers to professional development);

3   *Scale:* the proportion of the total potential target audience reached and the strength of plans to reach the entire state;

4   *Sustainability: the* strength of the **SSIs** strategy for maintaining the reform effort after NSF funding ends;

5   *Impact:* the impact on the target audience (e.g., the degree to which teachers have improved their practice);

6   *Theory ofchange: the* strength of the theory of change guiding the SSI strategy (e.g., support provided through SSI infrastructure reflects **full** needs of participating districts and schools).

Team members then used these criteria to assign an overall rating of the implementation of a particular strategy by a particular SSI. In other words, for each strategy, the **SSIs** that used that strategy were rated on how well it was implemented. Ratings were made on a 5-point scale: very weak, weak, fair, strong, very strong.

The ratings were difficult to do, for a variety of reasons:

■   First, unlike the case of the SSI midpoint reviews, which involved a fair-sized panel, for each SSI these ratings have been mainly the

Exhibit 2
The SSIs' Strategies for Promoting Systemic Reform

State	Aligning State Policy	Supporting Teacher Professional Development	Creating an Infrastructure for Capacity Building	Developing & Disseminating New Content & Materials	Funding Local Systemic Initiatives	Supporting Model Schools	Reforming Higher Education & the Preparation of Teachers	Mobilizing Public & Professional Opinion
AR	S	P	P				S	S
CA		P	S					
CO			S		P		S	S
CT			P		P		S	S
DE	S	P				P		
FL	S	P				P	S	S
GA	S						P	
KY		P	P	S		P	S	S
LA	S	P	S		S		S	S
MA	S	P	P		P			
ME	S	P	S		P		S	
MI	S		P		P		S	
MT	S	S		P				S
NC		P	S					
NE		P	P	S				
NJ			P		P			
NM	P	P	P	S			S	S
NY						P		
OH		P	P					
PR	S	P	P	P		P	S	
SC	P	P	P					P
SD					P			
TX	P		P					
V A		P						
VT	s	P	S	S	S			

Note: P = Primary. S = Secondary. "Primary" means that a strategy constitutes the core thrust (or one of a few) of the SSI. "Secondary" means that it is a key portion of the core thrust, but not central.

responsibility of a single individual. Although the team did try to "calibrate" the ratings, the reliability of specific ratings would have been significantly increased if there were multiple independent raters.

. Second, the team's knowledge of the "non-case study" states is weaker than its knowledge of the case study states; thus, some of the SSIs were harder to rate than others.

. Third, it was more difficult to rate SSIs on some of the strategies than on others. For example, there is little hard information in most cases about the effectiveness of SSI strategies to mobilize public and professional opinion. As a result, ratings on that SSI strategy are more difficult to make with accuracy than on most of the other strategies.

Given these considerations, readers can have greater confidence in ratings that apply to a number of SSIs (e.g., all the SSIs that use teacher professional development as a strategy for systemic reform) than in ratings that were made about a single SSI. Therefore, this paper focuses on ratings of the SSI *strategies.*

Strategies Focused on Teachers, Classrooms, and Schools

Supporting Teacher Professional Development

Professional development is important if reform is to succeed. It is the most common SSI strategy (a high priority in 18 of 25 SSIs), and in our view the majority of the SSIs using this strategy can be rated as strong or very strong in the way they conduct professional development. Delivering high-quality professional development is something that we as a nation know how to do—at least for those teachers who are interested and when there are sufficient funds. The SSIs used three general approaches to professional development: local human resource development, local system capacity building, and state system capacity building. The SSIs typically used a mixed professional development strategy, and at least 9 employed all three approaches. In spite of the strong efforts made by the SSIs, the professional development *system,* such as it is-including state, district, and school policies related to professional development-is in need of restructuring. As long as professional development systems in the state remain unchanged, the SSIs, and the states, will face difficult trade-offs between working with large numbers of teachers superficially or working with small numbers intensively.[3]

Developing, Disseminating, or Adopting Instructional Materials

Instructional materials are basic to what happens in schools. Yet only six SSIs focus on the instructional materials used in mathematics and science classrooms as a major part of their strategy. Of those, we estimate that four can be rated strong or very strong in their efforts. In many of the SSIs, instructional materials are still a "weak link," especially in certain domains (e.g., high school science). High-quality materials need to be identified or developed and decision makers need to be well informed about them. Whether decisions are made at the state or the local level, we believe it would make sense for more SSIs, and more states, to adopt and/or disseminate information about *existing* high-quality curriculum materials.

Supporting Model Schools

Just five SSIs use a model schools strategy, and we would rate only two as doing a strong or very strong job. The evaluation team continues to believe this is a high-risk strategy. It requires a very careful plan for dissemination and scaling up. If such a plan is not well designed and well implemented, the result is, at best, improvement in a handful of schools statewide. In the two states that we believe have strong model schools strategies (Delaware and Puerto Rico), that effort serves as part of a much broader systemic reform strategy.

Strategies Focused on Districts, Regions, and States

Aligning State Policy

State policy is difficult to change because it is driven by myriad political and resource issues beyond the control of the SSI. Only 3 SSIs target state policy as a primary strategy, although 11 others make it a secondary strategy to align the policy system with SSI goals, and in still other SSI states (e.g., California) policy alignment has been carried out under auspices other than the SSI. The majority of the SSIs that have targeted state policy have done a strong job or better at it. The SSIs that have done particularly well in their efforts to align state policy-we estimate there are 9 of them-typically have multiple connections with state policymakers, political "savvy," ongoing systemic reform efforts with which the SSI is integrated, and the good fortune not to be caught in fast-changing political tides.

Creating an Infrastructure for Capacity Building

Success for the SSIs is defined in part by what they will leave behind after the NSF funds disappear, and a new or improved infrastructure

(e.g., new nonprofit advocacy institutions or better regional education centers) would be one tangible legacy. Of the 18 SSIs emphasizing this strategy, we believe about half can be rated as strong or very strong. Doing a good job usually involves having a viable theory of how the infrastructure will help to change the system and making good connections between existing and new components of the infrastructure.

Funding Local Systemic Initiatives

Nine SSIs, mostly in local control states, chose to support local systemic initiatives. More than half of the nine (five) rated as strong or very strong. The key factors in building strong local initiatives are sufficient support for participants at the local level and some kind of quality control mechanism.

Reforming Higher Education and Teacher Preparation

While 14 of the SSIs have adopted this strategy, none rated as very strong for reform in teacher preparation, and just 4 rated as strong. Moreover, only 2 used this as a primary SSI strategy. Changing teacher preparation has been a challenge in part because of the difficulty of changing institutions of higher education, and in part because of the fact that relatively few of the SSI resources have been used for this purpose.

Lessons on how to effectively change teacher preparation may more likely come from NSF's Collaboratives for Excellence in Teacher Preparation program than from the SSI program.

Mobilizing Public and Professional Opinion

Many surveys and studies provide evidence that public and professional opinion is critically important for education reform. Of the 10 SSIs that place a significant emphasis on mobilizing opinion, the evaluation team believes 2 of them are doing a strong job. Although most of the 10 are doing at least a fair job, the SSIs have had a difficult time developing powerful strategies for mobilizing opinion. One reason may be that it is difficult to find appropriate measures of success for public relations initiatives, so it can be hard to know whether one is being successful or not.

Notes
1. Partners include the Consortium for Policy Research in Education (CPRE), Policy Studies Associates, Woodside Research Consortium, and the Council of Chief State School Officers.
2. Zucker, A., Shields, P., Adehnan, N., & Powell. J. (1995). *Evaluation of NSF 's Statewide Systemic Initiatives (SSI) program: Second year report. Cross-cutting themes.* Menlo Park, CA: SRI International.
3. Our report entitled *The SSIs and Professional Development for Teachers* is currently in draft form.

Forum Highlights and Looking Ahead

Cora B. Marrett*
University of Wisconsin-Madison

When I accepted this assignment, I didn't realize just how daunting it would be to try to bring together a number of the themes and talk about where we go from here, particularly around the Forum's emphasis on research on systemic reform. So I'll try to bring some of the observations that have occurred to me.

Some 70 years ago Mary Parker Follett (1924) wrote movingly about the importance of applying the scientific method to human endeavors. Science, she observed, builds from research to an organized body of knowledge. Research uses methods that are orderly and replicable, and it tests ideas that can be falsified, not just demonstrated. Important for this is her notion that any feature of the human theater can be subjected to scientific inquiry. Before this Forum I wondered whether Follett would have been so optimistic if she had had the task of applying scientific inquiry to systemic reform. Consider a couple of very daunting problems. One of these is the elusiveness of the basic concepts-the concepts of system and of reform. In the comments from the several sessions, as well as the presenters, problems associated with the terms have come up time and again, more often around system than around reform. But reform is an interesting one as well if you take the term to suggest re-form. "What's **the** *form* from which things are being changed? To what?" as one of the tables asked with reference to the system idea.

As most people have observed, from general systems theory comes the idea that a system consists of interrelated parts where changes in one will produce identifiable changes in another. In this arena the question that's come up time and again is, What's the system? Does it consist of different actors, units, organizations? Is the sys-

tern what **Bybee** calls "the constructed relationships among conceptual schemes, procedural strategies, contextual factors" or, as he goes on, "the fundamental concepts of disciplines; the actions and behaviors of teachers, of learners, of various technologies; the context within which the content and processes of science and mathematics may be learned." From one of the discussion groups, people wrestling with this question of system and systemic reform asked a very interesting question, How will we know when we have arrived? Or is systemic reform actually a process, and not a destination? That would have profound implications for the way in which we even try to capture systematically the very notion of systemic reform. This notion, then, the elusiveness of the concepts with which we're grappling, would, in many instances, prove a little disheartening to those who try to look as systematically as possible at the world of human endeavors.

But there's another issue that's come to the fore. We talk about systems in a holistic way, and that's often inconsistent with the reductionist model that's been so appropriately used for much of what we do in scientific inquiry. As others have noted-including Norman Webb this morning-from systems theory comes the assumption that a system is not just the sum of its parts. The system itself can be explained only as a totality. In thinking about systems, what merits attention are the connections and relationships, not just the nodes that are there. Yet capturing relationships-capturing the whole-proves somewhat difficult. That's an understatement, of course. Perhaps the difficulty comes out of the kinds of orientations that we bring. I'm thinking in part of Iris Weiss's comments earlier about the particular principal investigator who decided that having scientists in the classroom was the way to go. That might have sounded like a strange approach. But don't we tend to bring to the arena, too, our own orientations, our own interests, and to see

---

components of a system from the perspective that we happen to have? It's often been said that researchers bring to a problem their own tool kits and attempt to fit the problem to the tools they happen to have. I remember someone who knew one tune, and anything he sang had to have that one tune. Sometimes it was a bit difficult to fit the meter to the tune, but he would try because that's the tune that he knew. The same matter of trying to fit the pieces grows out of our interest, both out of the successes we've had in trying to refine things down to the smallest and most manageable parts and out of the orientation we have to look for some things rather than others. A few years ago, as some people were talking about how to think about the nature of research and higher education, there was an argument about the need for greater movement across disciplinary lines. One of the people said, **"We** have imposed those barriers because, after all, nature knows no disciplinary boundaries." The response immediately was, "Yes, but nature doesn't have to get tenure." What we have in place--our reward structures, our systems for keeping us closely linked to the orientations and **models**— might impede our effort to move forward on this holistic agenda.

Before I came to the Forum, I thought those were the daunting questions on the agenda. As I leave, I don't think it's nearly as daunting. I think that Follett would be encouraged by at least three things that have taken place here. First, there is the knowledge that systemic or systematic inquiry has produced, particularly about forces affecting the process of systemic reform; fruitful inquiries have been taking place. Also this systematic inquiry can relate to systemic reform in more than one way; it isn't as if there is only one place in which inquiry can make a difference. As a third matter, there is an excited community ready to engage its skills and its knowledge to move practice forward. This community shares the belief that there are outcomes that will enhance learning for understanding, outcomes so significant that people are willing to wrestle with difficult questions, What do we mean by "system"? How do we take a holistic approach?

But now what are some of these things that we know? Well we know, if one looks at the pa-

pers submitted through the Consortium for Policy Research in Education that some schools serving large numbers of low-income minority students have created transition courses in mathematics consistent with standards-based **performance.** Students in these transition courses appear to surpass those in the general mathematics courses. Admittedly, the former group still lags behind the students in the college preparatory courses. But the efforts to use standards for reforming education apparently redound to more than the most advantaged students into the entire school setting. That research goes on to tell us that implementation is not easy. Often the problems exist in providing the necessary training. We know from the Ohio experience, with that state systemic reform, that through professional development it's possible to produce a culture shift in participants. It can be done, but it's costly; it's time intensive. We know from the several studies reported that this is a complex process that we have in mind. The kinds of inquiries identify and potentially unravel that complexity.

A number of people expressed a little concern about what seemed to be normative statements. It was difficult to determine whether they were statements of what should be done based on some set of ideals or conclusions that had emerged from a body of research. Unfortunately our sessions were not often long enough for people to delve into the knowledge base, but there is clearly an interest in identifying what we know, and in bringing together these ideas into— what I'll return to in a moment-the knowledge base and the building of the knowledge. But research can enter in any number of ways. Consider the discussion on the importance of validated professional packages for professional development. The stuff, as people have put it, that teachers need is not something that has simply evolved out of somebody's own ideas. There ought to be research undergirding a demonstration of the validity of the ideas that are being advanced. One of the groups said that each step in the process should be research validated while one looks in the long term at the big picture. So it is not a matter of just waiting until everything has come. Some of the inquiries have to be done early on; others have to continue to unfold.

Clearly there is a need for evaluation-that's another place in which the research comes in-for determining progress and achievements. I take my hat off to those who have been bold enough to try to undertake evaluation of something in process, something that isn't always easy to get a handle on.

As I've said a number of times, we can sometimes think of systemic reform as having represented a metaphor, a symbol that has now taken on a life of its own. And that's not an easy thing to engage in systematic inquiry about. There's a place, then, for the research that is evaluative. But there is also the need for systematic knowledge at every level to do midcourse corrections, to follow processes. But coming through in a number of the comments is the need now for systematic inquiry at what I'll call the next level, at the level of the system. One comment was that we need mechanisms for identifying and assessing the effectiveness of models for systemic reform; however, we must make sure that we know what models work, in what context. It's paying close attention to what works under what kinds of conditions. Sorting through those, and avoiding the notion that there is one best way to proceed, is the task that a number of people have identified for the next stage of effort. There is also a need to identify important factors, variables essential for change. This idea was captured time and again as people talked about the levers, about the points of pressure. Not only is it intellectually interesting, but using that knowledge lets us lower the cost of introducing change, to intervene to get the benefits without going through what it would take to scale up from where we now are. Other people offered cautions in thinking of scaling up to think about systems: "Beware of the tendency to oversimplify whether we're thinking about assessment, professional development, curriculum-however you define it -or systemic reform. Don't let our eagerness to fix the system lead us to accept silver bullet solutions." So care and thought need to be a part of the close-grained analysis that we have in mind.

A number of people picked up on outcomes, the kinds of outcomes that might not have been so much on the agenda before this Forum. The outcome that seemed to strike people as particularly significant is the notion of the learning organization. One group after another picked up on the idea that perhaps we need better ways to think about learning organizations and about how one achieves that learning across all levels. Possibly, others propose, if we really care about bringing the knowledge together, there are lessons to be learned from other kinds of places, other systems. Just recently, at a conference on systemic change in business organizations, I picked up a new book and saw a section in there on systemic change. I said, "That's got to be on the Forum." Businesses have been moving toward not so much how you get all the external parts linked, but how in the world you put the parts together within your own organization. How do you make sure that **finance,** marketing, and research are connected in some way? Many have asked, Is it not useful to at least think about systems and see what lessons have been gained from other settings? Someone else asked, "Might it not be useful, if we really want to move to a systems approach, to understand how distributed intelligence, the notion that has recently gained a lot of interest, is amassed to affect system performance?" The idea is that the knowledge can be around in a lot of places, and the challenge becomes bringing that knowledge together to affect the notions of performance.

The next stage of moving toward an understanding of systems would suggest a couple of things. One, it seems essential to take the various experiences and try to extract from them what is common. It's the same question that one of the groups asked, Under what conditions do what factors have what effects? It might not be that one can talk uniformly, for example, about the consequences of unions. It might be that that will matter a lot for conditions that can be specified, and specification becomes the task. But aside from trying to extract from the several experiences, this idea that the system itself merits attention, that the relationships warrant study, does deserve attention and interest. It proposes that, at least for the research community, the next task is what we would call systemic research, research on systems, and research that links researchers in quite different ways. If we are indeed **products—**victims, if I can put it that way-of our own **ori-**

entations, is it likely that we're simply going to move outside of that orientation to try to capture other parts of a system? Perhaps not. I'm struck by **Daryl Chubin's** comment yesterday that the researchers who come through the National Institute for Science Education should not look exactly the way they did when they went in; they should have had attention to other kinds of forces that might have rested outside of the concerns they had brought to the Institute. But that's no simple matter. Again, there are not a lot of rewards, there are not a lot of places that acknowledge anything that moves outside of the boundaries we're often accustomed to. Thus this notion of systemic research is something that is going to take a lot of attention and resources. Coming into the picture all of the time is, What does it take to move things to the next level?

Let me then close by remarking again on Follett's idea. Follett said, "There is an imperative for collecting more exact information, for observing more keenly, for experimenting more widely. But," she went on to say, "there is an im-

perative for the organization of knowledge." Her idea was that you do not advance knowledge merely by gathering all of the bits and pieces. There must be a way to put that together so that you can talk about the relationships, understand the patterns. That then is this task. It's the task that is an engaging one, a daunting one, admittedly. But it's one, I would say, clearly worth the investment. For perhaps there is no reason to continue essentially muddling through. Perhaps there is no reason for each place, for each set of actors to have to try separately to determine what connections exist. Perhaps it is the wealth of experience, knowledge that a Forum such as this would represent. Perhaps here would lie the possibilities for developing that kind of knowledge base that's going to be essential if we are going to be able to talk about reform for the purposes this is all intended. And that is improved understanding.

Reference

Follett, M. P. *(1924). Creative experience. New* York Longmans, Green.

Michael W. Kirst
Stanford University

As Cora did, I will draw on a number of sources from the conference and reflect a little bit on the history of systemic reform to look at where we've come from and suggest some thoughts for the future. One of the things we learned from the papers, and in the discussions, and in the feedback we got is how much this is a trial and error process. Iris Weiss hit it right on the head when she said, "The theory underlying systemic reform was exceedingly thin, indicating, for example, the importance of shared goals and policy alignment, but providing little guidance on how to achieve either." We started off with a sort of general notion and have come a long way.

Before reflecting on how far we've come, I want to consider another paper that was presented, an excerpt from the SRI evaluation of the 25 Statewide Systemic Initiatives (SSIs; Zucker, Shields, & others. In "Exhibit 2: The SSIs' Strat-

egies for Promoting Systemic Reform" the authors chart which of eight strategies are used by each SSI. They indicate the primary means, a strategy that constitutes the core thrust or one of the few of the SSI, and secondary means, a key portion of the activity but not central. What you get from this chart would be what we've been reaping from the SSIs, if we assume that what the states have tried to do is what we're getting. Supporting teacher professional development as part of a comprehensive strategy for creating an infrastructure for local capacity building is a priority and is identified as a primary strategy in 16 and 12 states respectively. Aligning state policy, mobilizing public and professional opinion, reforming higher education and the preparation of teachers, and supporting model schools and new materials are secondary. It's a mixed view for funding local systemic initiatives. So if this is roughly correct, what we've started off with is

significant capacity building and support for teacher professional development. We have much weaker support for aligning state policies, mobilizing public opinion, reforming higher education, and supporting model schools.

Well that leads me to reflect on a question that Bill Clune raised, "When are there enough pieces of a complete reform to produce some significant results at a relatively modest cost?" As a high priority we need to understand what the payoff is from partial change. In the presentations and the sessions people said, "We've got to change everything, and then change it again." Well, I've been around a long time, and I don't know whether I'll live to see that. What we have to ask, I think, is, "What payoff do we get from the initial thrusts, which are teacher development and capacity building? How do we bridge out from those into the other areas-mobilizing public opinion, reforming higher education? And what are we getting now in terms of cost effectiveness from the existing teacher and professional development and capacity building? For many years David Tyack and I have been doing work called "What Reforms Last?" We have looked at a hundred years of education reforms and sorted them into three categories: those that have lasted, been imbedded in the system, been left-in a barnyard sense-as a permanent deposit; those that have failed or have faded, still around but in very few schools; and those that ebb and flow, go into the Bermuda Triangle and come back. Things that have lasted include special education, vocational education, Carnegie units, guidance counselors, and graded classrooms. The things that haven't lasted include team teaching, flexible scheduling, the new math of many years ago, and educational television. The ebb-and-flow reforms tend to be things like multicultural education and bilingual education. The things that have lasted tend to have three characteristics: they are structural or organizational changes; they are easily monitored; and lastly, crucial, they develop a supportive long-term constituency. The characteristics of the reforms that have largely failed or faded are that they have tried to change the culture of the classroom, have tried to change pedagogy drastically, and haven't built a strong, powerful constituency

either among a special categorical group like vocational educators or among teachers. And that's what I worry about. If we don't build out fast from this professional development and provide more local capacity building, we'll be in trouble. So our challenge here is to build a constituency for systemic reform over the long run. History suggests that the National Science Foundation may someday move on. How well equipped are the groups around it to mobilize public and professional opinion behind such a constituency? I think that's very important. In the discussion sections we heard that one of those potential constituencies is higher education. We heard that Advanced Placement, a form that seems to have lasted, increased eight times in about the last decade. That's a big change. So if we're not linked into higher education-admissions, teacher education, financial aid-and we're trying to do professional development and capacity building, we may be in a very **difficult** situation. A lot of people in the discussion groups picked up on Uri Treisman's recommendation to try to work from the middle out in picking up groups. This analysis of where we are and what lasts leads me to another discussion we've had, "Do you capture the flag or capture the system?' Capture the flag, as I understood it, consisted of the Systemic Initiatives capturing pots of categorical moneys and bringing them in to reinforce the SIs. Eisenhower is a classic example of that. Eventually we have to move to capturing the system, or we would be relying on a structural add-on. Structural add-ons, such as vocational education and special education, last, if you have a powerful constituency. But they don't tend to impact the core system, which is what we're trying to do. Look again at the focus on supporting professional development. To move beyond capturing the flag and capture the system, you'd have to do something about that thing that I think Moses handed down, the teacher salary schedule. We've had the same teacher salary schedule since the 1920s when we combined the elementary and secondary schools. Teachers get paid for the number of years of service and the number of credits beyond the BA. You heard Susan Loucks-Horsley, the NISE professional development team leader, say that teachers want their one-credit sessions,

so trainers offer them. We need to change the salary structures somehow to get the kind of purchase that we want. How can we bridge out from where we are now? How can we make it last? How can we make systemic reform capture the underlying system rather than a few flags on the outside?

Another persistent theme is how much context matters. There is no statewide systemic reform or systemic initiative; there are states doing things. What Connecticut is doing is very different from what Texas is doing, and it will be different across states. These SSIs will not add up to one grand thing, because context will drive the differences. Within the discussion of context there was not much talk at all, if any, of one of the hallmarks of conflict around systemic reform, which is the centralization/decentralization conflict. We discuss this very little. But systemic reform early on was interpreted as centralizing what ought to be known-content, and performance standards, and so on-at the state level, or in urban systems at the district office. A lot of people say, "I don't look for reform in the Washington, DC, central office or the Milwaukee central office." Or they say, "State level is too centralized." So how we blend centralization and decentralization warrants some careful thinking. We need to discuss that area more carefully in the future. If systemic reform is seen as a centralizing thing, bringing the system together around a central set of concepts or content, it will be resisted unless there is a way to harmonize it with decentralization. And where the central systems lie in that is important. We need some mix. The equity issue also ties into centralization. The history of education reform has been to create equity through more centralized policies, special education, school desegregation, Title I, and other things of that sort. Can we have equity without extensive and excessive centralization of policy? Who's at the center of the system, and who's at the decentralized area?

Another persistent theme was how much the classroom context varies, so that the impact from systemic reform varies a great deal by teacher. There was a vision of teachers as brokers, hearing different signals coming in. The people who originated systemic reform emphasized getting

the signals straight, sending consistent signals to teachers about what mathematics and science is most worth knowing. Right now the colleges and universities in California say to their new students, "We'll give you a multiple-choice placement exam when you get here." California State University gives nineteen multiple-choice questions developed by the professors to determine who goes into remedial education. On the other hand, we have a grand California mathematics framework based on NCTM standards, but there are no real stakes connected to it. How do we keep the signals straight so that the teachers in their brokering function in the classroom have some consistent systemic things they're reacting to? They must adapt to their own context, obviously. What came out in the discussion groups was to start with what teachers care most about and then mutually adapt the systemic policy to the teaching context. Again we have an adaptation theme within a centralization theme. Within the teacher area and teacher context was, of course, teacher and professional development. The presentations and discussion at this Forum made clear that we know a lot about how to do professional development well. This field has grown enormously; last year's Forum added to that. Ten years ago we didn't have a clue, at least at the policy level, about how to do professional development compared to what we now know. And it's such things as you heard of intensive, sustained, teacher collaboration learning organizations. What I hear from people who are in leadership roles about this is, however, "Yes I know more about how to do good professional development. But how do I bring it in on any kind of a massive scale? These good examples really pump you up, but they're fairly isolated. How do I make them in a broad scale-how do I transform them into a broad scale thing other than just a vignette of how to do it?" So we need to move from how to do good professional development, to-picking up some of my earlier themes-how to build a constituency to spread it throughout the entire system, and how to link it more closely to things like salary schedules and unions. When there was talk of professional development and teacher context, one of the themes in the discussion groups and in the papers was the fear of get-

ting lost in the process of systemic reform and losing the holistic nature of it. Cora talked a lot about that. The tendency is that, if you're working with so many parts and relationships, you tend to get lost in the trees and miss the forest.

Some of the discussion groups talked about what might have been a debate when Mondale was a politician, it goes so far back. "'Where is the beef?" A lot of discussion centered around systemic reform, professional development processes, what knowledge is most worth knowing, and who should decide that. To do professional development you have some idea what knowledge you're trying to develop toward. We came up with and then emphasized the idea of pedagogical content knowledge. There's not only pedagogy there, there's content,, There was not as much discussion, so I think there must be some assumption in the room about what knowledge is most worth knowing; however, as you all know, that's still under heavy debate. In California we cannot settle our mathematics curriculum at the state level. We're a state-driven program in some ways through textbooks. We've been able to settle reading through balance--whatever that means-but mathematics is still a blood bath That surprises me a lot I've been writing a lot about the politics of curriculum, and I would have thought math would have been easier. But the factions are too dug in, and they cannot resolve their differences. We have no ability to move ahead on a mathematics policy that would trigger a broader statewide systemic reform Science, for some reason, has flown under the radar screen of tlhe objections; I don't know why. Sci-

ence gets approved automatically. There was a group talking about what they saw as "'whole language science" in California. There are some articles about the spread of whole language science that could be some of the constructivist techniques we're talking about. One state has an interest group called Phonics for Christ. At some point systemic reform has to confront the beef, identify what knowledge is most worth knowing. And that gets us into fundamental political allocation because the reform can't be driven with that.

In this meeting a lot of people were tossed a concept of systemic reform. The concept was pretty vague, even highly underdeveloped, and did not come from a very explicit theory. The evaluation of it, as we heard, is difficult to pull off given the underlying uncertainties. But we've come a long way. We're making progress. There was a lot of defmition of the issues and a lot of concrete ideas. And we're bringing some reform into the classroom. The key to systemic reform in the 'long run is steady work, keep at it. The theme of a CPRE Brief (1997) is persistence and change. Refform is a long struggle. It doesn't happen overnight, but we're moving forward. With steady work and persistence, I think we will make a major difference.

Reference

Massell, D., Kirst, M., & Hoppe, M. (1997). *Persistence and change: Standards-based systemic reform* (RB-2 I). Philadelphia: University of Pennsylvania, Consortium for Policy Research in Education.

*Larry E. **Suter***
*Deputy Division **director**, Research, Evaluation and Communication*
Education and Human Resources Directorate, National Science Foundation

Thank you very much Mike and Cora for those inspiring comments. They certainly helped me hone my perspective on the many interactions and discussions going on at this conference that I was not able to see or hear.

My role in the National Science Foundation for the last few years has been as director of the program that conceived the idea of a National Institute for Science Education and then made the award to the University of Wisconsin from among a field of strong competitors.

The reason that the Education and Human Resources Directorate decided to create a large institute is that, while we award grants to thousands of individual projects, we do not have sufficient opportunity to think about the big picture. So we decided to look outside NSF for help to develop a new Institute. One of our key goals was to create an entity that would be crossdisciplinary. We realized that just having education researchers doing research about education is not enough. We really need the views of scientists who create the science that is taught in the schools and, indeed, taught throughout higher education. All of the projects that have been created at Wisconsin have these features. The NISE projects involve people throughout the university and beyond. The size and the geographic and professional diversity of this audience today shows also the broad reach of the NISE.

We need to have this many communities thinking about these problems because we do not yet have a common knowledge base. The discussions here yesterday suggest that there is, for example, no widely accepted single theory of systemic reform. All of you in this room are somehow involved in the process of trying to create a theory of systemic reform or apply it in practical situations. There is no single Einstein-like genius behind our thinking. The thinking and understanding occurs during the exchange of ideas in forums such as this. I agree with Daryl **Chubin,** who said that all of us attending should not leave here the same. We should leave here with a new commitment to think about how to create the conditions that will help us understand what to do next.

Rochelle **Gelman,** a psychologist, convinced me that the mind is not a tabula rasa, but is active. Luther Williams, a biologist, described how hard it is for him to get his hands around the education system because, once some part is identified, another weakly understood aspect pops up somewhere else. The education system seems closer to a sponge than to a machine that can be understood in terms of inputs and outputs. I suggest that administrators of programs experience the education system this way because active minds are always creating new options. As soon as someone provides a model for understanding the education system or an approach to action, others who also are involved in the process respond with new theories or alternatives to the original plan. This is the human condition. We are always dealing with creative, active people, both in schools and among ourselves as scientists.

Our experience with systemic reform has shown us that making improvements in the U.S. education system also involves improving the use of the public's money. Some education programs invest substantial amounts of money to implement a single plan, working on the assumption that with sufficient funds we can improve professional development, create a curriculum, or generate a better public understanding of assessment. But I am convinced that more essential for reform to be effective is wise investment in the development of ideas and in the use of information. The mere availability of money will not make a program more effective.

That is why we created a single place, an Institute where we can begin to add up the knowledge. As Cora Marrett put it, we need a place where we can share our understanding, criticize it, come back, revise our theories, and through this process develop a more powerful and

informed model that we can share. We need a place where we can nurture the ideas.

NSF does not have enough money to purchase all the change that is necessary in the education system. We can generate this change only by creating the ideas that can convince others. Making sense out of the experiments that we call systemic reform is needed badly. One purpose of this meeting is to begin the process of creating a common understanding about systemic reform. And the purpose of the Institute is to further consolidate our intellectual underpinnings so that this understanding can grow beyond the participants here.

In closing, I want to consider our priorities for the future. I was a little apprehensive about holding a conference on systemic reform because of its complexity. In fact, I asked Andy not to feature systemic reform in the first Forum and in other meetings because I thought we had too little understanding among ourselves to address it publicly. So this year you did it. Thank you, Andy, for being brave enough to take it on. Having come this far, we should frame the next set of priorities on the basis of this conference. What is worth the effort? What is worth knowing? How can we choose what is worth knowing?

Thank you all very much for participating.

Appendix A

Current Agencies Involved in Curriculum Improvement

AGENCY	94 BUDGET* (million)
National Science Foundation Fostering state-of-the-art mathematics and science	43.9
National Aeronautics and Space Administration Fostering space science	4.4
Environmental Protection Agency Fostering environmental education	3.0
Department of Energy Programs associated with laboratories	2.2
Department of Health and Human Services NIH – Fostering health education	.8
Department of Education	0
Department of Interior	0
Department of Defense	0
Department of Agriculture	0

*FCCSET FY 1994 Budget Summary

Appendix B

Comprehensive Mathematics Curricula & Funded Projects with Contact Information

Development   Field Testing   Publication   Implementation
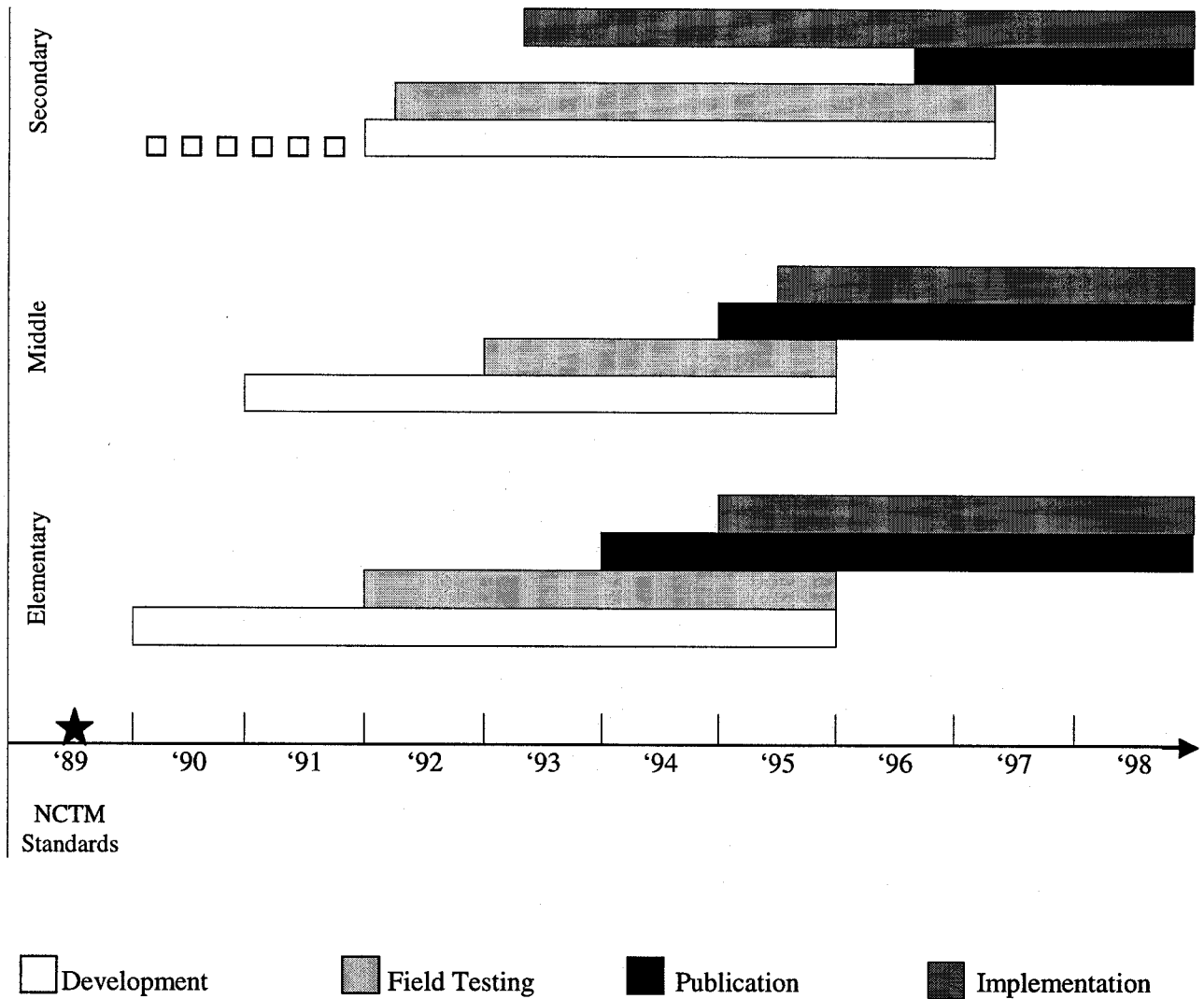
ARISE: Applications/Reform in Secondary Education (H)
    Landy Godbold (404)355-8673
Connected Geometry (H)
    E. Paul Goldenberg (617)969-7100 x2513
The Connected Mathematics Project (M)
    Kathy Burgis (517)432-3635
Cooperative Mathematics Project (E)
    Laurel Robertson (510)533-0213 x247
Core-Plus Mathematics Project (H)
    Christian R. Hirsch (616)387-4526
IMP: Interactive Mathematics Program (H)
    Diana Resek (415)338-2071
Investigations in Number, Data and Space (E)
    Susan Jo Russell (617)547-0430
Math Connections, A Secondary Mathematics Core Curriculum Project (H)
    June G. Ellis (203)244-1942
Mathematics in Context: A Connected Curriculum for Grades 5-8 (M)
    Thomas Romberg (608)263-4285
Middle School Mathematics Through Applications Project (M)
    Shelley Goldman (415)497-7963
Seeing and Thinking Mathematically (M)
    Glen Kleinman (617)969-7100
SIMMS: Systemic Initiatives for Montana Mathematics and Science (H)
    Glenn Allinger (406)994-5351
STEM: Six Through Eight Mathematics (M)
    Rick Billstein (406)243-2603
TIMS: Teaching Integrated Math and Science (E)
    Phil Wagreich (312)413-3019
UCSMP: University of Chicago School Mathematics Project (E, M, H)
    Elementary component: Max Bell (312)702-1563
    Secondary component: Zalman Usiskin (312)702-1560

E Elementary school level
M Middle school level
H high school level

Appendix C

Comprehensive Science Curricula

| Development | Field Testing | Publication | Implementation |
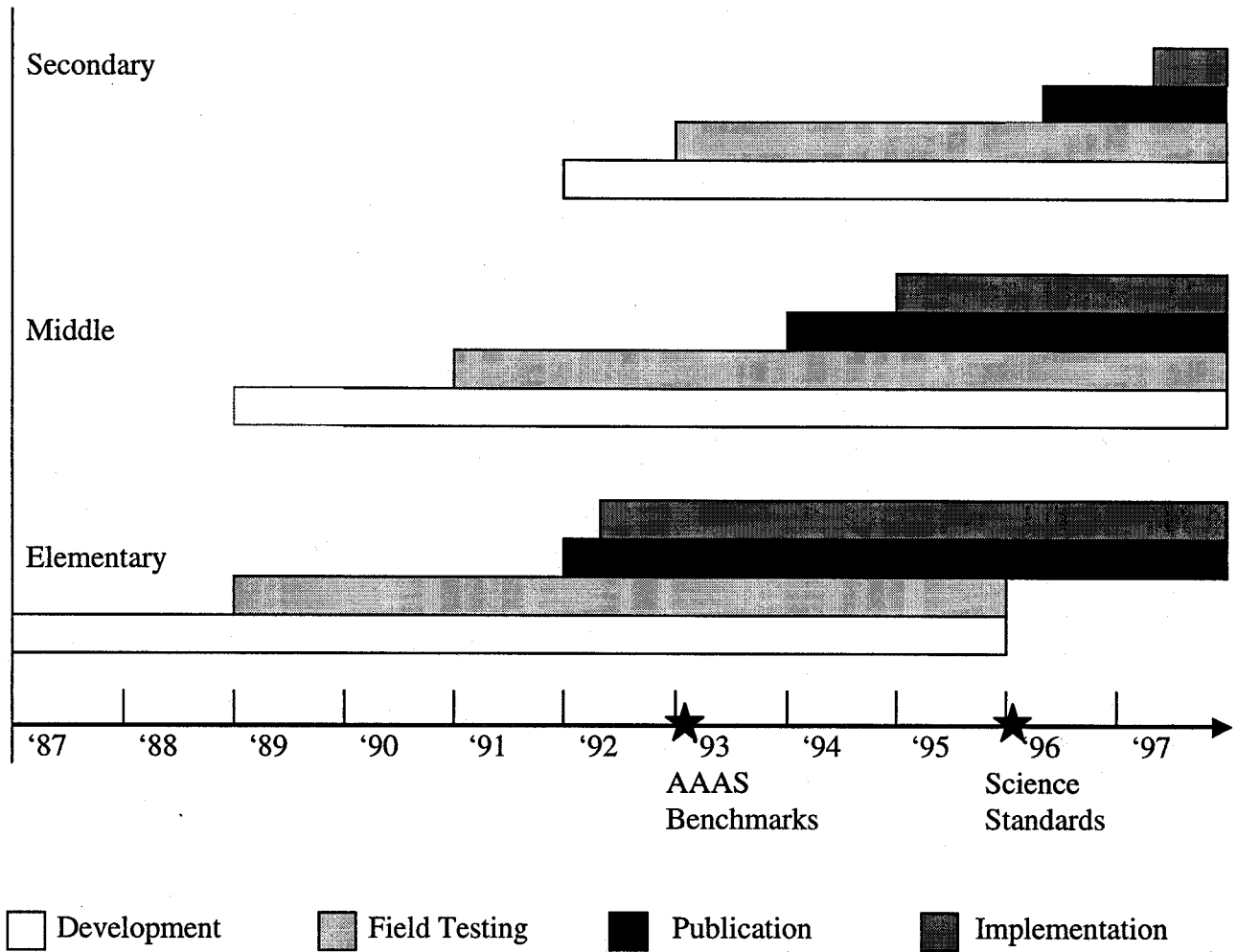
Appendix D

Criteria for Judging SSI Strategies

Criteria	Aligning State Policy	Supporting Teacher Professional Development	Creating an Infrastructure for Capacity Building	Developing, Disseminating, or Adopting Instructional Materials
Quality	Development of ambitious and clear standards for student learning that serve as the basis for alignment of key policies shaping the educational system	Meets subject matter needs; research-based, sustained, adequate time and intensity, models active learning, provides for engagement with ideas and peers; teachers involved in design, designed for varying learners; meets needs of teachers and schools	Qualified personnel, content expertise; sufficient resources, intensity, and duration to have impact on resource users	Instructional materials that reflect national (NCTM, NSES) standards for content and pedagogy
Access/ Inclusion	N/A	Broad array of teachers have access	Access is offered to the full range of teachers and schools in the state	Access to materials is offered to full range of teachers and schools in the state
Scale	Full range of policies (e.g., assessment, professional development, teacher credentialling, school accreditation) are aligned	Serves substantial proportion of teachers with clear strategy for reaching majority of teachers in state	Structure and resources are sufficient to reach all or most of the state	Explicit plan in place for getting materials into all classrooms
Sustainability	Policies are codified or at least formalized; support for policies is broad	System and funding in place to continue to support high quality professional development	Sufficient resources— financial, personnel, material—are available to maintain infrastructure	Explicit and realistic plan (e.g., use of commercial publisher) for maintaining availability of materials
Impact	Clear evidence of substantive policy changes aligned with standards	Evidence that teachers have gained new knowledge, skills, and attitudes	Teachers, schools, and districts are taking advantage of resources available through the infrastructure	Materials are being used in classrooms resulting in improved practice
Theory of Change	Explicit strategy exists, rooted in political context of the state, for policies to impact local practice; incentives for teachers to participate	Professional development includes explicit strategy to support teacher change beyond workshops, including addressing schoolwide change needs and contextual constraints	Support provided through the infrastructure is of sufficient duration and infrastructure, fits the political and resource context of the state	Materials are accompanied by sufficient support for teachers so that they can impact classrooms; incentives to use materials

Criteria	Funding Local Systemic Initiatives	Supporting Model Schools	Reforming Higher Education and Teacher Preparation	Mobilizing Public and Professional Opinion
Quality	Sufficient resources, incentives, and support to ensure that local sites can progress with local systemic reform	Sufficient resources, incentives, and support to ensure schoolwide improvement	Clear standards for high-quality teacher education; sufficient resources, incentives, and support to ensure substantive improvements; focus on both content and pedagogy	Clear and well-articulated vision of high-quality mathematics and science; high-quality public relations material distributed through appropriate channels to target audiences
Access/ Inclusion	Local systemic initiatives represent range of districts in the state; opportunity for participation was open	Model schools represent range of schools in the state; opportunity for participation was open	Full range of IHEs in state have access to SSI activities or to lessons from SSI pilots	Target of public relations to many key groups within the state
Scale	Explicit and realistic plans for transferring lessons from local systemic initiatives to other sites	Explicit and realistic plans for transferring lessons from model schools to other sites	Explicit and realistic plans for reaching all major IHEs	Explicit and realistic plans for reaching audiences within the entire state
Sustainability	Explicit and realistic plans for continuing support of local systemic initiatives	Explicit and realistic plans for continuing support of model schools	Explicit and realistic plans for continuing support of IHEs	Explicit and realistic plans for continuing public relations efforts
Impact	Clear evidence of changed local systems in terms of aligned policies, quality of professional development and curricular materials, and public and professional support	Clear evidence of improved schools in terms of professional culture, explicit planning, and changed classroom practice	Improved undergraduate classes in mathematics, science and/or education; better prepared entering teachers	Evidence of increased support for SSI's vision of high quality mathematics and science among public and/or professionals
Theory of Change	Quantity, quality, scope, and sequencing of support to local initiatives is explicitly tied to a realistic theory of what local systems need to improve mathematics and science for all students	Quantity, quality, scope, and sequencing of support to schools is explicitly tied to a realistic theory of what local systems need to improve mathematics and science for all students	Explicit strategy for addressing the incentive issues associated with changing higher education	Explicit strategy for changing public and professional opinion that takes into account preconceived notion of good mathematics and science

Single copy price is $10.50. To order copies contact:

National Institute for Science Education
University of Wisconsin-Madison
1025 West Johnson Street
Madison, WI 53706

(608) 263-9250
(608) 262-7428 fax
niseinfo @ macc.wisc.edu
http://www.wcer.wisc.edu/nise